

Università degli Studi di Napoli

Federico II

Facoltà di Scienze MM. FF. NN.

Corso di Laurea in Fisica



Ricerca di Nuclei Galattici Attivi in
survey fotometriche multibanda

Tesi di Laurea

Relatori:

Ch.mo Prof. Giuseppe Longo
Dott. Maurizio Paolillo

Candidato:

Stefano Cavuoti
matricola: 60/958

Anno Accademico 2006/2007

Indice

Introduzione	1
1 I Nuclei Galattici Attivi	5
1.1 Il Modello Unificato	8
1.1.1 Lo Zoo dei Nuclei Galattici Attivi	11
1.2 Variabilità, Dimensioni Lineari ed Evoluzione	13
1.3 Le Classificazioni	15
1.3.1 Bontà di una classificazione	17
2 Gli Osservabili	19
2.1 Grandezze dei Campi di Radiazione	20
2.2 Elementi di Fotometria Astronomica	27
2.2.1 Ottico, Infrarosso, Ultravioletto	27
2.3 Emissione nel Continuo	31
2.4 Indici di Colore	35
2.5 Estinzione e Profondità Ottica	37
2.6 Eccesso di Colore	40
2.7 Estinzione Atmosferica	44
2.8 Trasporto Radiativo	45
2.9 Elementi di Spettroscopia	47
2.10 Cenni sulla Teoria Atomica degli Spettri	50
2.10.1 Spettri Atomici	54
2.10.2 Spettri Molecolari	56
2.10.3 Misure Spettrali	58
2.10.4 Formazione delle Righe	60
2.10.5 Allargamento delle Righe	64
2.10.6 Effetto Doppler	67

2.11	Spettri Nebulari	68
2.12	AGN e Rapporti tra Righe Spettrali	70
2.13	Le Survey	72
2.14	Sloan Digital Sky Survey (SDSS)	74
2.15	Il Passaggio da Spettroscopia a Fotometria	75
2.16	I Parametri Usati	76
3	Le Reti Neurali	79
3.1	Reti Neurali Feed-Forward	82
3.2	Storia	84
3.3	Multi-Layers Perceptron	84
3.3.1	L'Algoritmo di Back Propagation	86
3.4	Support Vector Machine	87
3.4.1	Il Kernel RBF	87
3.4.2	Ricerca dei Parametri Migliori	88
3.5	Modalità di Apprendimento	88
3.5.1	Apprendimento Supervisionato	88
3.5.2	Apprendimento Non Supervisionato	89
3.6	Principal Component Analysis	90
3.6.1	Interpretazione Geometrica	90
3.6.2	Fattorizzazione	92
3.6.3	Schema del Lavoro	92
3.7	Probabilistic Principal Surfaces	93
3.7.1	L'Algoritmo delle PPS	94
3.8	Negentropy Clustering	98
3.8.1	L'Algoritmo del NEC	100
4	La Scelta Del Campione	103
4.1	Il Catalogo di G. Sorrentino et al.	104
4.2	Il Catalogo di G. Kauffmann et al.	106
4.3	Il Catalogo di R. D'Abrusco et al.	110
4.4	Il Catalogo Usato	110
5	VO-Neural	113
5.1	Principali Modifiche al Codice delle Reti MLP	114
5.1.1	Aggiunta dei Target Vector	115
5.1.2	Aggiunta della Modalità di Run	115

5.1.3	Rimozione dei Parametri Ridondanti	116
5.1.4	Eliminazione dei Filename Hard-Coded	117
5.1.5	Eliminazione degli Input: hiddenlayers, layer2 e error- func	117
5.1.6	Eliminazione dei File di Debug Prodotti dalla Rete . .	117
5.1.7	Aggiunta di Funzioni che Scrivono un File di Log . . .	117
5.1.8	Aggiunta di un Controllo sulle Dimensioni dei Dataset	118
5.1.9	Aggiornamento dell' Help On-Line	118
5.1.10	Creazione di un Tool Minimale di Validazione	120
5.2	Librerie per le Matrici di Confusione	120
5.2.1	Funzioni per la Creazione delle Matrici	121
5.2.2	Funzioni per l'Incremento delle Matrici	121
5.2.3	Funzioni per il Salvataggio su File delle Matrici	122
5.2.4	Funzioni per Liberare la Memoria Precedentemente Occupata dalle Matrici	122
6	Gli Esperimenti	123
6.1	Features	124
6.2	Selezione dei Candidati AGN Tramite MLP	124
6.2.1	Logistic SCG con 20 Hidden	125
6.2.2	Logistic Conji Grad con 20 Hidden	125
6.2.3	Logistic Quasi Newton con 20 Hidden	126
6.2.4	Linear SCG con 20 Hidden	126
6.2.5	Linear Conji Grad con 20 Hidden	127
6.2.6	Linear Quasi Newton con 20 Hidden	128
6.2.7	Logistic SCG con 52 Hidden	128
6.2.8	Logistic Conji Grad con 52 Hidden	129
6.2.9	Logistic Quasi Newton con 52 Hidden	129
6.2.10	Linear SCG con 52 Hidden	130
6.2.11	Linear Conji Grad con 52 Hidden	130
6.2.12	Linear Quasi Newton con 52 Hidden	131
6.2.13	Riepilogo e Considerazioni	132
6.3	Separazione AGN non AGN Tramite SVM	134
6.4	Separazione Oggetti di Tipo 1 e Tipo 2 Tramite MLP	135
6.4.1	Logistic Linear 20 Hidden	136
6.4.2	Logistic Conji Grad 20 Hidden	136

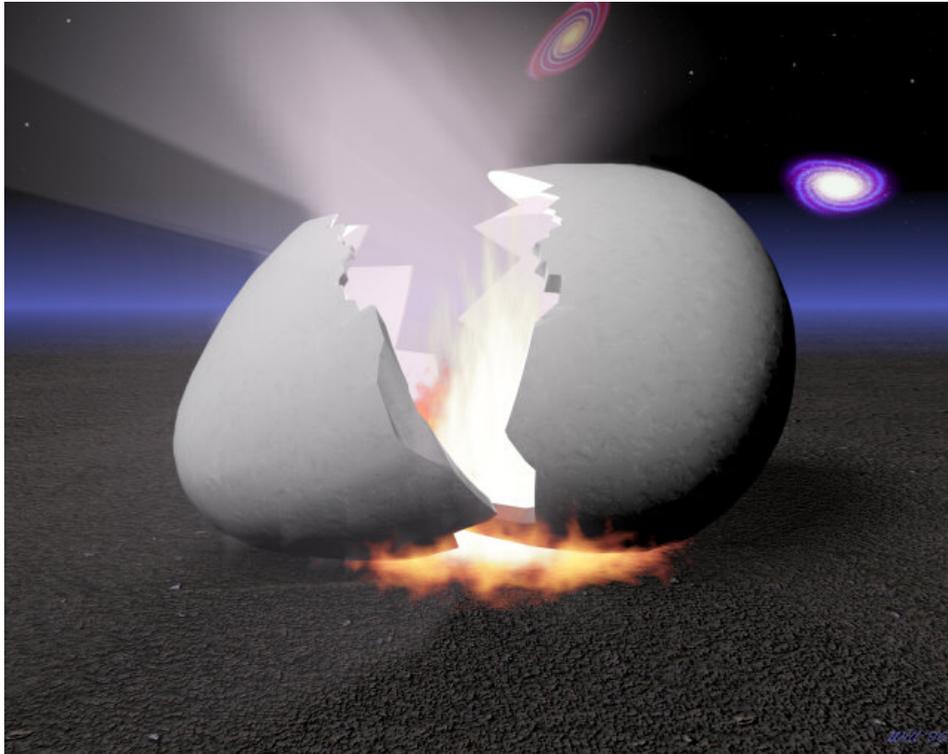
6.4.3	Logistic Quasi Newton 20 Hidden	137
6.4.4	Linear SCG 20 Hidden	137
6.4.5	Linear Conji Grad 20 Hidden	137
6.4.6	Linear Quasi Newton 20 Hidden	138
6.4.7	Logistic SCG 52 Hidden	138
6.4.8	Logistic Conji Grad 52 Hidden	139
6.4.9	Logistic Quasi Newton 52 Hidden	139
6.4.10	Linear SCG 52 Hidden	140
6.4.11	Linear Conji Grad 52 Hidden	140
6.4.12	Linear Quasi Newton 52 Hidden	140
6.4.13	Riepilogo e Considerazioni	141
6.5	Separazione Oggetti di Tipo 1 e Tipo 2 Tramite SVM	142
6.6	Separazione Seyfert LINER Tramite MLP	143
6.6.1	Logistic SCG con 20 hidden	143
6.6.2	Logistic Conji Grad con 20 hidden	143
6.6.3	Logistic Quasi Newton con 20 hidden	144
6.6.4	Linear SCG con 20 hidden	144
6.6.5	Linear Conji Grad con 20 hidden	145
6.6.6	Linear Quasi Newton con 20 hidden	145
6.6.7	Logistic SCG con 52 hidden	145
6.6.8	Logistic Conji Grad con 52 hidden	146
6.6.9	Logistic Quasi Newton con 52 hidden	146
6.6.10	Linear SCG con 52 hidden	147
6.6.11	Linear Conji Grad con 52 hidden	147
6.6.12	Linear Quasi Newton con 52 hidden	147
6.6.13	Riepilogo e Considerazioni	148
6.7	Separazione Seyfert LINER Tramite SVM	148
7	Conclusioni	151
A	Codici Sorgenti	153
A.1	Materiale Relativo alle Reti MLP	154
A.1.1	Codice per l'Aggiunta dei Target Vector	154
A.1.2	Modalità di Run	154
A.1.3	Parametri Ridondanti	155
A.1.4	Tool di Validazione	155

A.1.5	Esempio di Logfile Qualora il Processo Sia Andato a Buon Fine	155
A.1.6	Esempio di Logfile in cui i Dataset Sono di Dimensioni Incongrue	156
A.2	Materiale Relativo alle Matrici di Confusione	156
A.2.1	Codice delle Librerie	157
A.2.2	Esempio di Utilizzo	163
	Bibliografia	165
	Elenco delle figure	171

Introduzione

George Lucas:

“Tanto tempo fa, in una galassia lontana, lontana...”



Questa tesi affronta il problema di come individuare dei potenziali (candidati) nuclei galattici attivi (AGN), nell'universo vicino, utilizzando i soli dati fotometrici.

Un nucleo galattico attivo (o, per estensione, una galassia) si dice attiva quando una frazione significativa dell'energia viene emessa da oggetti diversi dai normali componenti di una galassia: stelle, polveri e gas interstellare.

Questa energia, a seconda del tipo di galassia attiva, può essere emessa lungo quasi tutto lo spettro elettromagnetico.

Alcune di queste regioni attive emettono getti collimati di materia che possono essere molto lunghi, trasportando energia verso strutture estese (come nelle radiogalassie). Ma in tutti i casi è il nucleo, il cosiddetto motore centrale, ad essere la fonte di energia.

Per semplicità basterà per il momento ricordare che il modello utilizzato per descrivere ciò che accade è il seguente: l'energia è generata da materia che cade in un buco nero supermassiccio che ha una massa compresa tra 1 milione e 1 miliardo di volte quella del Sole. Quando la materia cade verso il buco nero, il suo momento angolare la costringe a formare un disco di accrescimento attorno al buco nero. La frizione riscalda la materia e ne cambia lo stato in plasma, e a sua volta produce un forte campo magnetico. Il materiale che si muove all'interno di questo campo magnetico produce grandi quantità sia di radiazione di sincrotrone che di radiazione termica sotto forma di raggi X. Infatti la temperatura vicino al buco nero è di milioni e forse di miliardi di gradi, in quest'ultimo caso migliaia di volte maggiore che nel centro del Sole. Spesso, inoltre vengono osservati getti che si originano dal disco di accrescimento. Tutto questo processo, alimentato dalla gravità del buco nero, è molto efficiente nel trasformare la materia in energia.

Da un punto di vista fenomenologico, le galassie di Seyfert, i quasar e i blazar sono i tipi principali di AGN. I quasar, in particolare, sembrano essere gli oggetti più luminosi dell'universo conosciuto.

Le radiogalassie sono un gruppo eterogeneo di galassie che emettono più onde radio del normale. La maggior parte presentano enormi lobi di intensa emissione radio, disposti in modo simmetrico rispetto al nucleo e da cui viene emessa la gran parte della radiazione. Alcune mostrano uno o due getti (l'esempio più famoso è M87 nell'Ammasso della Vergine) che escono direttamente dal nucleo e si dirigono verso i lobi.

I differenti tipi di galassie attive sono legati dal modello unificato, in cui vengono considerati come lo stesso oggetto visto da differenti angolazioni. L'assorbimento da parte di polveri presenti nella galassia, e gli effetti relativistici del getto rivolto verso l'osservatore sono la causa delle differenze.

Tutto questo verrà illustrato nel capitolo 1

Nel capitolo 2 mi sono occupato dei parametri in gioco, presentando pri-

ma di tutto fotometria e spettroscopia, ho illustrato i modelli che permettono, tramite rapporti tra righe spettrali, di classificare gli AGN, chiarendo che essendo la spettroscopia “time consuming” è cruciale riuscire a indirizzarla sulla base di parametri fotometrici, quindi ho presentato il concetto di Survey, focalizzandomi in particolare sulla SDSS che è la fonte dei dati che ho utilizzato in questo lavoro di tesi, infine ho elencato i parametri fotometrici di cui ho deciso di avvalermi per tentare questa classificazione fotometrica.

Per raggiungere questo obiettivo mi sono servito di alcune metodologie di data mining basate sulle reti neurali. Nel capitolo 3 illustro il concetto di rete neurale, parlandone prima in generale e poi affrontando con maggior dettaglio i modelli che ho ritenuto essere utili per i miei scopi, MLP, SVM, PPS e NEC.

Successivamente, nel capitolo 4, ho spiegato le scelte fatte per costruire la “base di conoscenza” di cui ho avuto bisogno per addestrare le reti neurali; partendo da come erano stati costruiti i cataloghi da cui sono partito.

Quindi ho parlato del progetto VO-Neural; questo perchè, oltre ad utilizzarlo durante questo lavoro di tesi ho collaborato con il progetto, e alcuni dei contributi che ho portato allo sviluppo di nuovi modelli sono elencati nel capitolo 5.

Infine nel capitolo 6 ho riportato nel dettaglio i risultati degli esperimenti principali che ho condotto e che hanno dato risultati degni di nota.

Le conclusioni cui sono giunto si trovano nel capitolo 7.

Nell’appendice A si trovano stralci del codice da me sviluppato nel corso di questo lavoro di tesi.

Capitolo 1

I Nuclei Galattici Attivi

Stephen Hawking:

“Qualsiasi teoria fisica è sempre provvisoria, nel senso che è solo un’ipotesi: una teoria fisica non può cioè mai venire provata.”



I recenti risultati dell’esperimento Auger sulla direzione di provenienza dei raggi cosmici di altissima energia hanno rivitalizzato gli studi finalizzati

all'ottenimento di campioni statisticamente significativi di galassie attive o, come è invalsa l'abitudine di chiamarli, nuclei galattici attivi o AGN.

Da un punto di vista fisico, questi sono oggetti compatti di grande massa presenti al centro delle galassie che tramite processi di accrezione sono in grado di liberare enormi quantità di energia sotto varie forme, radio, ottico, X, γ e raggi cosmici. Fino a pochi anni fa gli AGN erano ritenuti essere oggetti relativamente rari; oggi, anche grazie a misure di tipo dinamico, si ritiene che pressochè ogni galassia ospiti al suo centro un oggetto compatto di grande massa (anche alcuni milioni di masse solari) che, nelle opportune condizioni, può dare luogo al fenomeno degli AGN.

Come si vedrà, le galassie attive emettono una luminosità cento o mille volte maggiore di una galassia normale ed hanno uno spettro diverso da quello delle galassie "normali", sia per quanto riguarda l'intensità della luce emessa nelle diverse bande¹, che per la presenza di righe spettrali in emissione con larghezze equivalenti che implicano gas in movimento con velocità di alcune migliaia di km/s . Le galassie attive sono inoltre spesso sorgenti molto intense proprio nelle bande spettrali più "insolite" per una galassia, cioè la banda radio e la banda X.

Infine la luminosità di molte di queste galassie, almeno in certe bande spettrali, varia con periodi molto brevi, di pochi giorni o anche di ore; questo fatto sarebbe inspiegabile se la sorgente di luce predominante fossero le semplici stelle. Tutto questo fa pensare che la sorgente di tale luminosità non sia stellare, e che essa si trovi in una regione molto piccola della galassia

La maggior parte dell'energia emessa dai nuclei galattici attivi proviene infatti da regioni centrali delle dimensioni di poche ore-luce o giorni-luce. Per avere un'idea di quanto piccolo sia il nucleo emittente rispetto alle dimensioni dell'intera galassia, si pensi che la distanza tra il Sole e Plutone è di 5.2 ore-luce, mentre il diametro di una tipica galassia è di 100.000 anni-luce o più! Questo significa che la regione emittente di un nucleo attivo può avere le dimensioni di un sistema planetario.

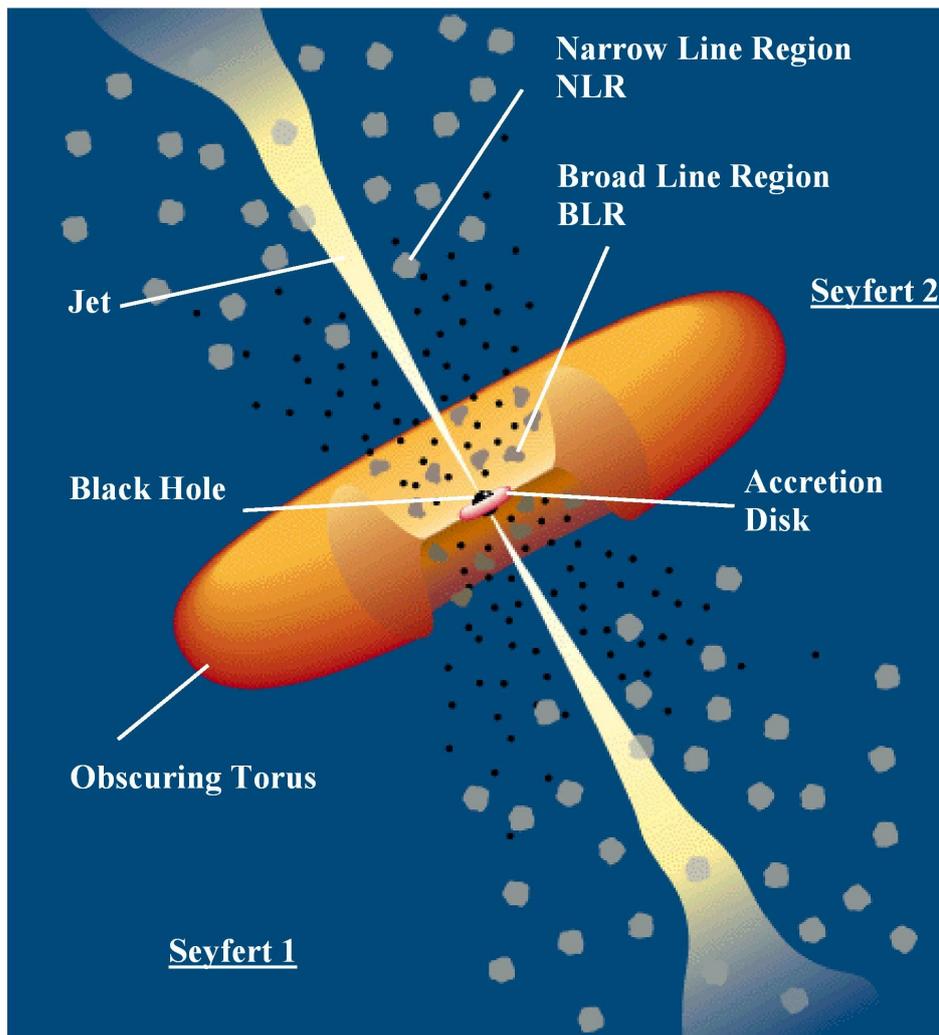


Figura 1.1: Schema di un AGN: il toro oscura una fascia di cielo e a seconda di dove ci si trova rispetto al piano del toro si osserva un AGN di un tipo piuttosto che di un altro

1.1 Il Modello Unificato

Sin dalla loro scoperta (1943) la fenomenologia degli AGN si è andata via via arricchendo e per lungo tempo si è ritenuto che fenomenologie diverse corrispondessero a famiglie di oggetti sostanzialmente diverse tra di loro.

Queste fenomenologie hanno dato luogo alla formulazione di una congerie di teorie e di modelli esplicativi spesso impossibili da conciliare l'uno con l'altro. Senza voler entrare nel dettaglio di tali teorie (che ormai hanno un interesse solo storico) in quanto segue descriverò in un certo dettaglio il cosiddetto modello unificato che, nella forma proposta da Antonucci nel 1993 e rivista da Urry e Padovani nel 1995, fornisce un'interpretazione comune che permette di descrivere le varie categorie di oggetti (che vedremo in seguito nel dettaglio) come diverse fenomenologie di un unico tipo di oggetto visto da diverse orientazioni.

Lo schema previsto di modello unificato (si veda figura 1.1) può essere riassunto come segue.

Al centro dei nuclei galattici attivi c'è un buco nero con una massa compresa tra le 10^6 e le 10^{10} masse solari e la materia circostante (prevalentemente gas e polveri) presente nel nucleo della galassia forma intorno al buco nero uno spesso disco, detto di "accrescimento", di forma toroidale; tale materia alimenta il buco nero cadendovi sopra e nel processo libera grandi quantità di radiazione. La materia viene quindi inghiottita dal buco nero centrale e convertita in energia elettromagnetica con un'efficienza molto alta (circa il 10%). L'accrescimento di una quantità di materia pari alla massa del sole in un anno basterebbe a spiegare l'energia liberata dalla maggior parte degli AGN. Inoltre, se il buco nero ruota intorno al proprio asse, questo motore centrale è in grado di accelerare il gas circostante e di espellerlo dal nucleo sotto forma di getti collimati con l'asse di rotazione del buco nero. Secondo il modello unificato questo toro si troverebbe a ~ 100 anni luce dalla singolarità. tale fenomeno sembra essere stato osservato all'Hubble Space Telescope nella galassia NGC4261 (come mostrato in figura 1.2).

Il toro è otticamente spesso: se esso intercetta la linea di vista la materia che lo compone esso oscura le componenti più interne e nello spettro emer-

¹Il rapporto di flusso in due diverse bande viene definita indice di colore (ad esempio U-B, B-V, ecc.). Il colore misura il contributo relativo di energia nelle varie bande e dà una stima della temperatura di corpo nero.

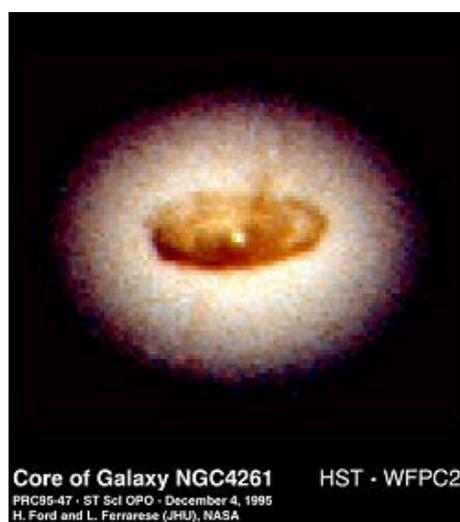


Figura 1.2: Il disco di accrescimento della galassia Seyfert NGC4261 rivelato dall'Hubble Space telescope

gono solo la sua emissione (prevalentemente infrarossa) e quella delle regioni esterne (si veda la figura 1.2). Al di fuori del toro si immagina che siano presenti delle nubi di gas e polveri (zona detta Broad Lines Region o BLR), che sono responsabili dell'emissione della componente allargata delle righe permesse². Si ritiene che le BLR siano causate dalla fotoionizzazione dovuta al disco di accrescimento, molto caldo, attorno al buco nero supermassivo. Si pensa che le BLR si trovino a solo ~ 1 pc dal buco nero.

Alle nubi presenti nella zona detta "Narrow Lines Region" (NLR) è invece dovuta l'emissione delle righe proibite e della componente stretta³ delle righe permesse; si ritiene che le NLR occupino grandi spazi e eccedano lo spessore del toro di polveri e che si trovino tra i 10 e i 1000 pc dal buco nero.

Le emissioni che si misurano dipendono da cosa si interpone tra l'osservatore e il nucleo e, quindi ogni parte del modello è responsabile di un particolare tipo di emissione (come riassunto in figura 1.3). Riassumendo ciò che verrà spiegato in maggior dettaglio in quanto segue:

- i Getti Relativistici (Jet) sono responsabili delle emissioni nei domini Radio, Gamma e X

²Queste righe chiamate larghe (BROAD) hanno ampiezza dell'ordine di migliaia di km/s

³le righe strette (NARROW) hanno ampiezze che non superano 1000 km/s

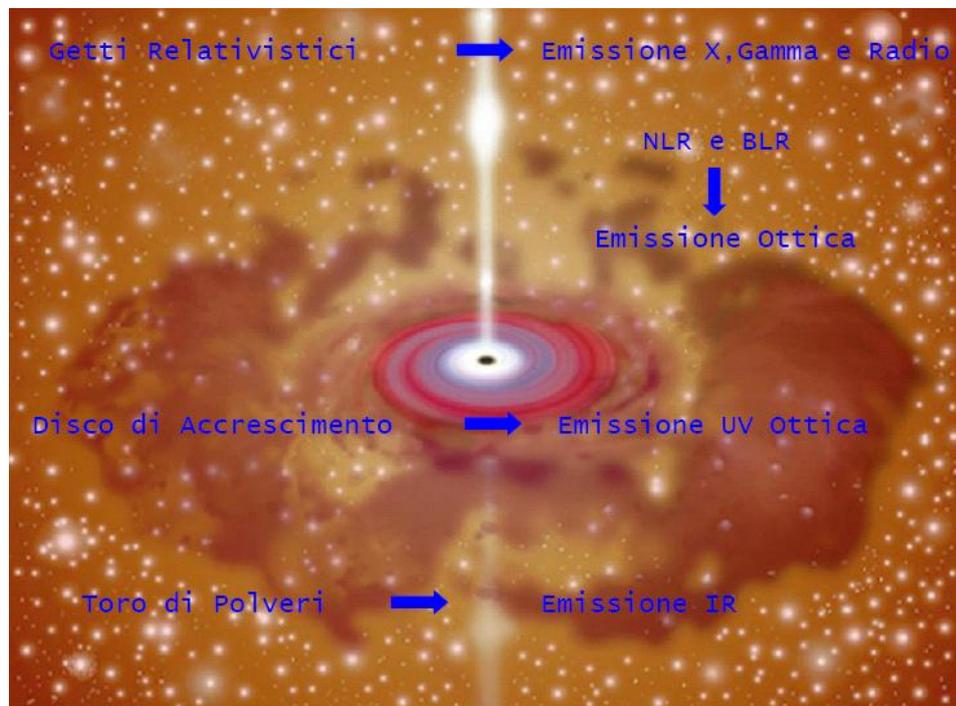


Figura 1.3: Schema delle possibili emissioni di un nucleo galattico attivo in relazione alla porzione di sistema che le genera

- le NLR e BLR originano la quasi totalità dell'emissione ottica
- il disco di accrescimento è responsabile dell'emissione UV e parzialmente di quella ottica
- il toro di polveri circostante produce l'emissione IR

1.1.1 Lo Zoo dei Nuclei Galattici Attivi

Da un punto di vista osservativo ci si può trovare dinanzi a tre casi principali:

Caso 1: linea di vista circa perpendicolare al piano del toro (Blazar):

Gli oggetti che ci appaiono di “faccia” (face-on), in cui il getto punta lungo la linea di vista, sono detti Blazar (da Blaze, vampata). La materia nel jet si muove a velocità relativistica, quindi la radiazione emessa è fortemente collimata (beamed) e può variare con periodi molto brevi (ore o giorni). In alcuni casi la luminosità del nucleo supera quella della galassia e l'oggetto viene chiamato Quasar (quasi stellar) perchè a prima vista appare appunto come una stella ma studiato più approfonditamente si rivela non essere tale e viene a volte indicato come QSO, quasi stellar object. In un universo in espansione, la brillantezza superficiale, in prima approssimazione decresce con $(1+z)^4$ e, per di più, il redshift fa sì che la radiazione raccolta in filtri predeterminati proviene da zone dello spettro sempre più spostate verso l'ultravioletto (dove il rapporto tra la luce delle galassie e quella dei nuclei che esse ospitano è minore) ciò implica che il contrasto tra i nuclei e le immagini delle galassie, e tra queste e il cielo, aumenta all'aumentare di z . Ciò ha fatto sì che siano stati classificati come quasar, anche a redshift non troppo elevati, oggetti che a basso redshift sarebbero stati visti come Seyfert (vedere dopo). I quasar sono stati scoperti inizialmente grazie all'intensa emissione radio e solo in un secondo momento per via ottica. I quasar indicati genericamente come QSO (quasi stellar object) si dividono quindi in RLQ (radio loud quasar) e RQQ (radio quiet quasar) a seconda che siano dotati rispettivamente di una notevole emissione radio o di una emissione radio marginale. Kellerman, nel 1989, per fornire un parametro

per classificare gli RQQ, ha proposto di dividere gli oggetti in queste due categorie secondo un parametro oggettivo e non più tramite classificazioni soggettive; la separazione da lui scelta è

$$R = L_{5GHz}/L_{4400\text{\AA}} < 10 \quad (1.1)$$

All'interno degli RQQ si osserva che il 5 ÷ 10% degli oggetti mostra ampi assorbimenti nella banda blu delle righe di emissione risonanti. A tali oggetti viene dato il nome di Broad Adsorption Lines (BAL). Questo assorbimento è dovuto al materiale espulso dal quasar che si muove verso di noi lungo la linea di vista. I quasar con spettro radio piatto ⁴, oltre che per la compattezza e la frequente variabilità radio, si distinguono per i tempi di variabilità più brevi e per la maggiore variabilità nell'ottico. Tali caratteristiche sono comuni anche alla classe degli Optical Violent Variable (OVV) detti anche blazar, caratterizzati anche da un grado di polarizzazione piuttosto elevato e variabile, sia in intensità che in direzione, del continuo ottico.

La classe dei Blazar comprende sia oggetti che presentano nell'ottico le usuali righe di emissione che oggetti con un continuo praticamente indifferenziato. Questi ultimi, dal loro prototipo, la variabile BL Lacertae, a lungo considerata una stella, vengono denominati BL Lac.

Si ritiene che ciò che rende praticamente invisibili nei BL Lac le righe di emissione che caratterizzano gli altri AGN sia la peculiare ed eventualmente transitoria intensità del continuo. Le righe infatti, quando il continuo diminuisce di intensità, diventano visibili, così come è accaduto nel caso del quasar PKS 0521-36 che presentava nel 1975 le caratteristiche di un BL Lac e che sei anni dopo appariva come una normale Seyfert 1.

I blazar sono anche gli unici oggetti dove di frequente la radiazione γ tra 100 MeV e 3 GeV, o addirittura nel caso del BL Lac Mkr 421 fino a 10^3 GeV, risulta energeticamente dominante.

Caso 2: linea di vista circa giacente sul piano del toro (Seyfert 1):

Quando invece il nucleo è parzialmente oscurato e in parte intercetta la linea di vista, si è in grado di vedere direttamente la regione centrale e dallo

⁴perchè saturato dall'intensità del getto

spettro emergono sia le componenti broad-line che narrow-line e l'emissione da parte del dal disco di accrescimento.

A questo caso appartengono le Seyfert di tipo 1⁵, le broad line radio galaxy e i quasar di tipo 1.

Caso 3: linea di vista in una posizione intermedia tra le due (Seyfert 2):

Se invece si osserva il sistema di taglio (edge-on), la regione centrale, comprendente, il buco nero, il disco di accrescimento e la broad line region, è oscurata. Tutto ciò che si può vedere è il toro di polveri (che emette prevalentemente nell'infrarosso) e le righe di emissione dalla NLR. Dallo spettro emergerà, quindi, la sola componente narrow-line. La luce dal nucleo può essere riflessa nella direzione di vista dal gas caldo che giace al di sopra e sotto il toro, che si comporta come una sorta di specchio.

A questo caso corrispondono: le Seyfert 2, i quasar di tipo 2 e le narrow line radio galaxies.

1.2 Variabilità, Dimensioni Lineari ed Evoluzione

Abbiamo già detto che una delle caratteristiche dei quasar è la variabilità su tempi scala piuttosto brevi del continuo. Prendiamo ad esempio il quasar 3C 273 la cui immagine, con luminosità apparente oscillante tra la 12^a e la 13^a magnitudine, potè essere rintracciata in una serie di lastre che erano state prese fin dagli inizi del secolo. In una particolare occasione, nel 1929, la sua luminosità variò di ben 0.7 magnitudini in soli dieci giorni. Indagini successive hanno mostrato che si tratta di un comportamento diffuso, che coinvolge anche i nuclei delle galassie di Seyfert [1, 3]. L'osservazione nella banda X ha poi permesso di mettere in evidenza tempi caratteristici di variazione ancor più brevi, che per alcuni oggetti sono risultati anche di soli 10² s.

⁵In alcuni casi, galassie che sotto ogni altro punto di vista possono essere considerate normali possiedono nuclei estremamente luminosi e apparentemente puntiformi. Il primo a studiarne gli spettri fu Seyfert nel 1943 che vi trovò righe di emissione sia permesse che proibite. In base a ciò furono quindi denominate Seyfert 1 quelle galassie in cui le righe spettrali permesse risultavano molto allargate nel dominio delle frequenze con sovrapposta una componente stretta (come si vede in figura 1.5), mentre furono denominate Seyfert 2 quelle in cui tutte le righe risultavano tutte egualmente larghe (vedi figura 1.4)

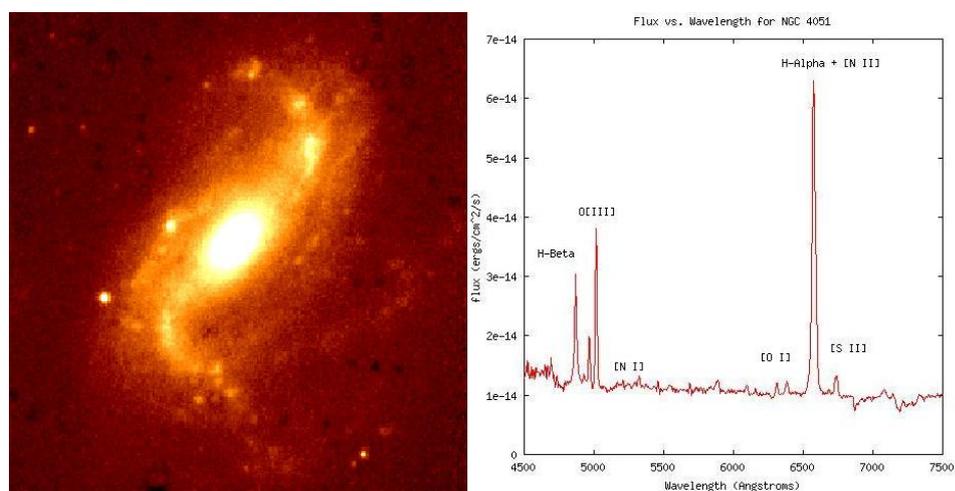


Figura 1.4: Galassia NGC 4051 e il suo spettro

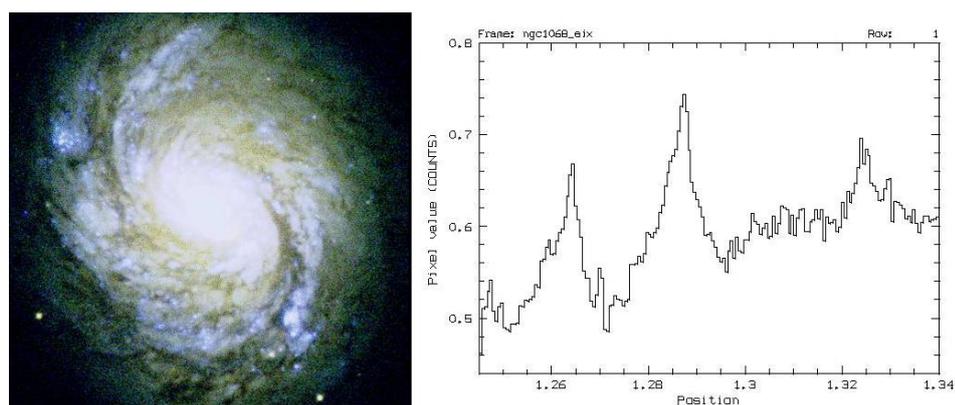


Figura 1.5: Galassia NGC 1068 e il suo spettro

La variabilità mostra immediatamente come le dimensioni della regione dalla quale proviene la luce irraggiata dai nuclei delle Seyfert e dai quasar siano assai ridotte. Infatti, una sorgente di luce estesa non può mostrare ad un osservatore esterno variazioni con tempi scala più brevi del tempo che la luce impiega per percorrere distanze dell'ordine delle sue dimensioni. Ciò perchè un osservatore vede dapprima variare la luce proveniente dalla parte della sorgente più vicina a lui e più tardi quella proveniente dalle zone più lontane. Detti Δl la dimensione della sorgente e Δt il tempo di variazione, si ha:

$$\Delta l \leq c\Delta t \quad (1.2)$$

Nel caso di 3C 273 il variare di 0.7 magnitudini implica dimensioni massime per la zona di emissione dell'ordine di $2.6 \cdot 10^{16} \text{ cm} \cong 0.8 \cdot 10^{-2} \text{ pc}$, pari a circa 30 volte le dimensioni del nostro sistema planetario. Una fotosfera con queste dimensioni per irraggiare termicamente nell'ottico $10^{13} L_{\odot}$ dovrebbe avere una temperatura intorno ai 25000 K: un valore sostanzialmente in accordo con il colore blu osservato nella radiazione emessa.

Il confronto tra i ritardi relativi delle variazioni osservate a più frequenze ci permette di stimare le distanze che intercorrono tra le zone dalle quali provengono le varie emissioni. Normalmente le variazioni nelle bande X e ultravioletta precedono di qualche giorno quelle nel continuo ottico. Le righe larghe e l'infrarosso vicino seguono con ritardi di un ordine di grandezza superiori. Le righe strette risultano essenzialmente stabili [1].

Dal modello unificato si ricava la seguente sequenza evolutiva:

Dal punto di vista morfologico spegnendosi un Quasar molto luminoso diviene o una galassia di tipo Seyfert che nel passo evolutivo successivo diverrà a sua volta una galassia spirale oppure una galassia radio che nel passo evolutivo successivo diverrà a sua volta una galassia ellittica [2].

Tale schema è riassunto nella figura 1.6

1.3 Le Classificazioni

Prima di proseguire ed al fine di inquadrare meglio il presente lavoro di tesi, è opportuno soffermarsi brevemente sul significato metodologico di una classificazione.

Scopo immediato di una classificazione di un insieme omogeneo di oggetti è quello di stabilire una relazione d'ordine⁶ qualitativa $R(p_1, \dots, p_n)$ - mono o pluriparametrica - dove i parametri p_1, \dots, p_n sono predicati o quantità misurabili in modo oggettivo, riproducibile e, se possibile, quantitativo. In quest'ultimo caso, $R = R(p_1, \dots, p_n, \epsilon_1, \dots, \epsilon_n)$.

Scopo ultimo di una classificazione è invece quello di stabilire una tasso-

⁶Quindi una classificazione è innanzitutto una ricerca di regolarità. Linneo, ad esempio, classifica gli animali in uno schema che trova conferma nella teoria dell'evoluzione di Darwin prima e nella differenziazione del DNA poi. Linneo cerca la regolarità nella varietà (principio di uniformità della natura). Questo è un criterio estetico.

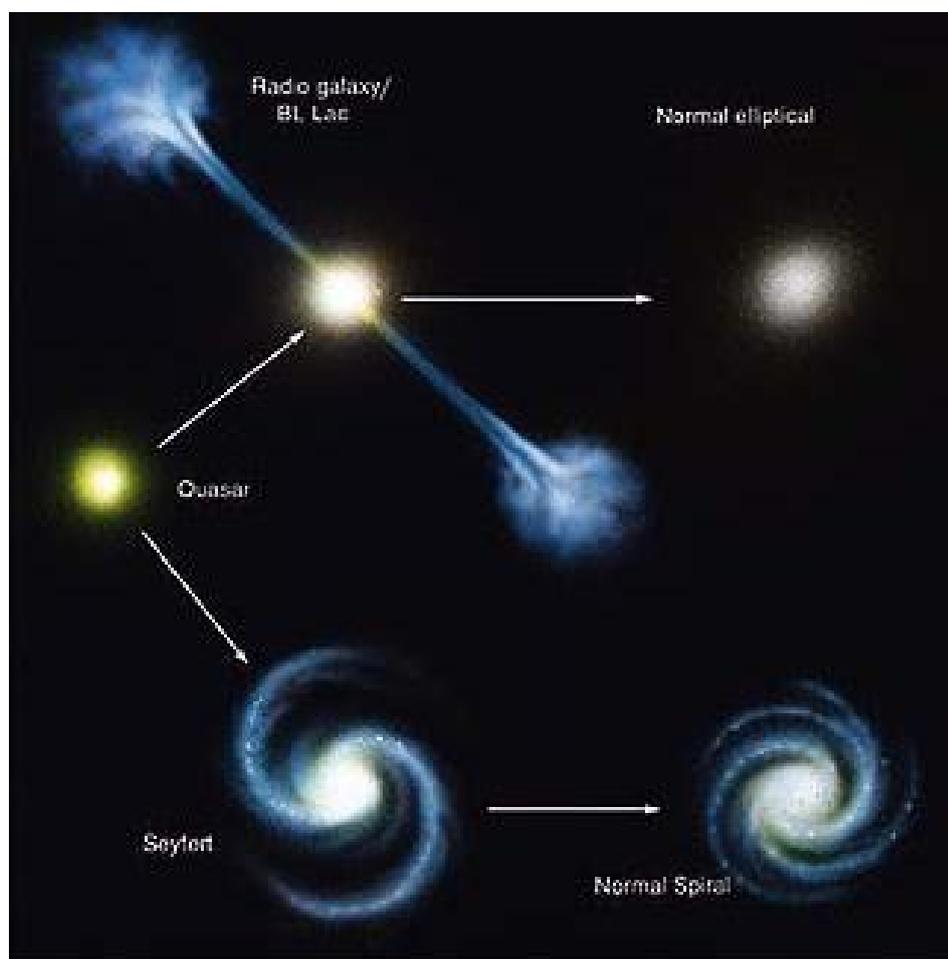


Figura 1.6: Schema di una possibile storia evolutiva dei quasar

nomia fisica: cioè una suddivisione degli oggetti classificati in base a fenomeni fisici (di formazione, evoluzione, struttura fisica, ecc.). In questo caso, R , deve rispecchiare (consapevolmente o inconsapevolmente) una o più leggi di natura L_1, \dots, L_n . In genere, il raggiungimento di una tassonomia comporta l'abbandono dei sistemi di classificazione precedenti. Un buon metodo empirico per riconoscere se una classificazione è o meno una tassonomia, è vedere se essa ha o meno un valore predittivo.

1.3.1 Bontà di una classificazione

Cosa implica la condizione di omogeneità? Innanzitutto, l'omogeneità degli oggetti che dovrebbero essere accomunati da una stessa natura. In secondo luogo, in alcuni casi (ma non in tutti, si pensi ad esempio al problema della classificazione delle forme geometriche o dei cristalli) si deve avere omogeneità del materiale usato per la classificazione.

La classificazione delle galassie introdotto da William Herschel⁷ nel 1816 è, ad esempio, una classificazione che fallisce soprattutto nella non-omogeneità del campione: oggetti galattici ed extragalattici risultano essere accomunati da una comune morfologia pur essendo intrinsecamente molto diversi.

Da un punto di vista matematico il problema della bontà di una classificazione può essere così riassunto:

⁷Herschel introdusse un sistema abbreviato di notazioni morfologiche che, però, non può essere considerato una classificazione vera e propria, ma solo un insieme di codici descrittivi. Il materiale su cui si basò non era omogeneo e la classificazione venne effettuata visualmente (al telescopio).

Herschel tentò un'organizzazione sistematica delle forme ordinandole in base al maggiore o minore grado di concentrazione centrale e di simmetria. Per esempio, una descrizione "vBpLvqmbM" significa very bright, pretty luminous, very gradual, much brighter in the middle.

L'elenco completo dei codici da lui ricavati è quello che segue:

- B Bright
- F Faint
- L Large
- S Small
- G Gradual
- v very
- c considerable
- p Pretty
- e Extremely
- m middle

Questi codici sono stati usati a lungo, ad esempio da Dreyer nel *New General Catalogue* (1888) e da G. de Vaucouleurs (1956), fino a quando non sono stati sostituiti da altri. Il motivo della sostituzione è soprattutto nell'estrema soggettività e nella scarsa riproducibilità della classificazione (forte dipendenza dall'apparato strumentale e dalle condizioni di osservazione). Il principio ispiratore di Herschel era la diffusa convinzione che, nella fisica newtoniana, un sistema di N corpi autogravitante avrebbe dovuto inevitabilmente collassare. Herschel, quindi, diede alla propria sequenza un significato evolutivo. Herschel ha il merito, comunque, di avere applicato per primo il principio di uniformità alle galassie esterne.

John Herschel, sostituì ai codici letterali del padre nuovi codici numerici, ma la situazione non cambiò di molto.

Requisiti	Svantaggi
Devono essere in numero ridotto	Aumentare il numero di parametri riduce l'utilità della classificazione
Facilmente misurabili sui dati a disposizione	La classificazione di un oggetto non deve richiedere più tempo di quanto ne prenderebbe uno studio dettagliato
Non devono portare a risultati contraddittori	Lo stesso oggetto verrebbe ad essere attribuito a classi diverse
Devono, in prima approssimazione riflettere qualche criterio fisico sottostante (anche se al momento sconosciuto)	Si provi a classificare le specie animali in funzione del colore del mantello

Sia $S = \{x_1, \dots, x_n\}$ l'insieme degli oggetti da classificare. Una classificazione è una partizione $S' = \{S_1, \dots, S_m\}$ tale che

$$\bigcup_{j=1}^m S_j = S \quad \text{con } m \ll n \quad \text{e } \forall j S_j \neq \phi \quad (1.3)$$

Per una classificazione ideale dovrebbe anche valere:

$$S_j \cap_{j \neq k} S_k = \phi \quad (1.4)$$

Le classi S_j sono definite dai parametri o predicati $p_{j,l}$ con $l = 1, \dots, L$ definiti in S_j :

$$x_i \in S_j \Leftrightarrow \forall l, p_{j,l}(x_i) = \text{Vero} \quad (1.5)$$

In quanto segue si vedrà come tale formalizzazione possa essere efficacemente usata per trasformare il problema di classificazione in un problema di clustering supervisionato o non supervisionato.

Capitolo 2

Gli Osservabili

Camille Flammarion:

“Cieco chi guarda il cielo senza comprenderlo: è un viaggiatore che attraversa il mondo senza vederlo; è un sordo in mezzo a un concerto.”



In questo Capitolo vengono introdotte le principali grandezze fotometriche utilizzate in astrofisica, ponendo maggiore attenzione sui parametri e

sugli strumenti che saranno utilizzati nel corso di questo lavoro. Per descrivere i campi di radiazione elettromagnetica occorre distinguere due tipi di informazione che si possono ottenere dalla radiazione raccolta:

- un'informazione quantitativa, studiata dalla fotometria, in cui si misura la potenza globale della sorgente. Questa misura è legata alla comprensione della struttura della sorgente e delle fonti di energia che la sostengono.
- Un'informazione qualitativa o specifica, studiata dalla spettroscopia, in cui si misura la distribuzione della radiazione sulle varie frequenze, cioè lo spettro. Queste misure permettono di determinare le caratteristiche fisiche della sorgente, temperatura, densità, pressione, composizione chimica.

Va notato che in relazione al loro irraggiamento le sorgenti astrofisiche si dividono in due categorie:

- sorgenti termiche (stelle normali, mezzo interstellare, mezzo intergalattico) la cui radiazione è essenzialmente di tipo corpo nero, il che indica uno stato di equilibrio termodinamico tra materia e radiazione;
- sorgenti non termiche (stelle attive, pulsar, supernove, nuclei galattici attivi) la cui radiazione indica una situazione di non-equilibrio con presenza di cospicue componenti di particelle sopratermiche.

Le tecniche di analisi fotometrica e spettroscopica si sono sviluppate in relazione alla banda del visibile. La gran parte della materia dell'Universo, stelle, galassie e mezzi diffusi sono in condizioni di equilibrio termodinamico, per cui tali tecniche fanno costante riferimento a sorgenti termiche.

2.1 Grandezze dei Campi di Radiazione

La radiazione emessa da una sorgente dipende dalle caratteristiche fisiche della stessa e la sua misura permette di derivare per le sorgenti termiche, parametri quali densità, temperatura, composizione chimica del plasma emettente, etc. Per le sorgenti non-termiche i parametri ricavabili sono invece il bilancio energetico totale, l'energia specifica delle particelle emettenti, l'intensità e la struttura dei campi magnetici. D'altra parte la radiazione

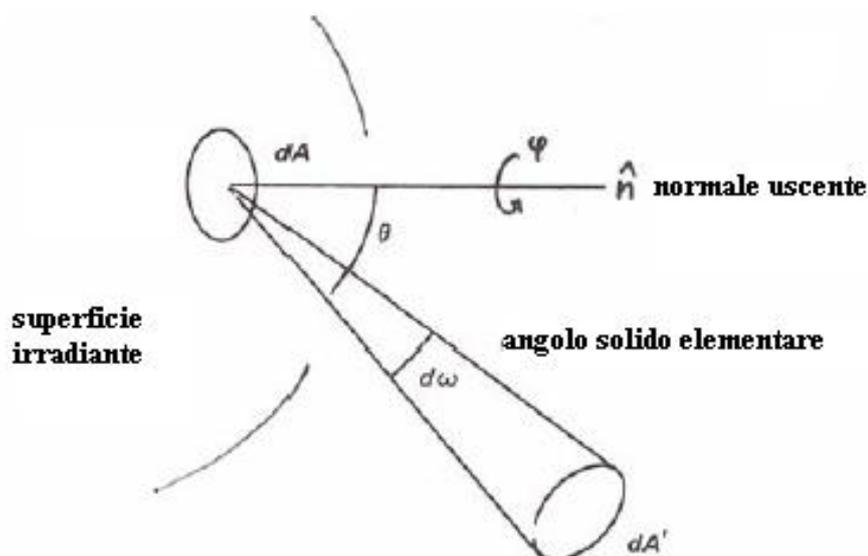


Figura 2.1: Campi di Radiazione

emessa viene modificata nella sua propagazione ed è parzialmente “deformata” all’interno degli strumenti di misura. Occorre quindi ricavare le relazioni principali che legano le caratteristiche dei campi della radiazione osservata a quelle della radiazione emessa. Si consideri la superficie di una sorgente e se ne selezioni un elemento dA con normale uscente \hat{n} (Figura 2.1)

L’energia specifica emessa nell’unità di tempo nella banda $d\nu$ dall’elemento dA entro l’angolo solido $d\omega$ nella direzione ad angolo θ rispetto a \hat{n} è espressa attraverso la formula:

$$dE_\nu = I_\nu(\theta, \varphi) \cos\theta \, dA \, d\omega \, d\nu \, dt \quad (2.1)$$

dove $I_\nu(\theta, \varphi)$ è detta intensità specifica (o radianza); $\cos\theta$ è un fattore geometrico che proietta l’area dA nella direzione θ . Le unità di misura per I_ν sono $\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1} \text{ster}^{-1}$ oppure $\text{W cm}^{-2} \text{Hz}^{-1} \text{ster}^{-1}$

L’intensità integrata su tutte le possibili frequenze è detta intensità totale:

$$I = \int_{\nu=0}^{\infty} I_\nu \, d\nu \quad (2.2)$$

Si dimostra che l’intensità totale (ma la dimostrazione vale anche per

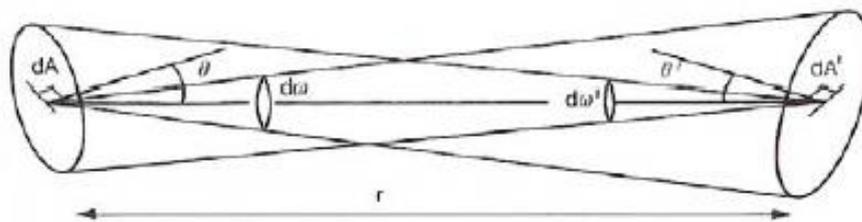


Figura 2.2: Conservazione dell'intensità della radiazione per la propagazione nel vuoto

l'intensità specifica) è una grandezza che si mantiene costante lungo la propagazione della radiazione nel vuoto. Con riferimento alla Figura 2.2 si consideri la radiazione che esce dall'area dA con intensità totale dI nella direzione θ entro l'angolo solido $d\omega$ in un tempo dt :

$$dE = I \cos \theta \, dA \, d\omega \, dt \quad (2.3)$$

e che viene intercettata da un'altra superficie dA' nella direzione θ' . La radiazione può anche essere scritta come intercettata in dA' :

$$dE = I' \cos \theta' \, dA' \, d\omega' \, dt \quad (2.4)$$

D'altra parte la geometria permette di ricavare

$$d\omega = dA' \cos \theta' / r^2 \quad (2.5)$$

$$d\omega' = dA \cos \theta / r^2 \quad (2.6)$$

dove r è la distanza tra le due superfici. E di conseguenza ne risulta:

$$I = I' \quad (2.7)$$

Cioè l'intensità rimane costante nello spazio vuoto. Va peraltro detto che in generale questa grandezza non è misurabile perché gli strumenti osservativi integrano sempre su aree e direzioni. Pertanto occorre definire queste quantità misurabili.

Integrando l'intensità su tutte le direzioni angolari si ottiene la densità di flusso specifico (o emittanza):

$$F_\nu = \int_{\omega} I_\nu(\theta, \varphi) \cos \theta d\omega \quad (2.8)$$

che è la grandezza misurabile per sorgenti non risolte dai telescopi. Infatti in tal caso tutta la radiazione raccolta dal telescopio entro ω_{min} corrisponde a quella dell'intera sorgente emessa su tutte le direzioni di arrivo. Pertanto per sorgenti puntiformi la quantità misurata dai telescopi è la densità di flusso e non l'intensità; se la distanza della sorgente aumenta, la densità di flusso diminuirà $\propto (1/r)^2$. Si definisce densità di flusso totale la quantità:

$$F = \int_{\nu=0}^{\infty} F_\nu d\nu \int_{\omega} I \cos \theta d\omega \quad (2.9)$$

dove l'integrale è esteso a tutte le possibili direzioni. Nel caso di campo di radiazione isotropo, indipendente da θ e φ , va notato che la definizione di densità di flusso totale comporta:

$$F = I \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} \cos \theta \sin \theta d\theta d\varphi = 0 \quad (2.10)$$

perché radiazione entrante e uscente si compensano esattamente. Se si vuole considerare solo la densità di flusso uscente si ha:

$$F = I \int_{\theta=0}^{\pi/2} \int_{\varphi=0}^{2\pi} \cos \theta \sin \theta d\theta d\varphi = \pi I \quad (2.11)$$

E, in generale, per campi non isotropi:

$$F_\nu = 2\pi \int_{\theta=0}^{\pi/2} I_\nu \cos \theta \sin \theta d\theta \quad (2.12)$$

Le unità di misura per F_ν sono $erg s^{-1} cm^{-2} Hz^{-1}$, oppure $W m^{-2} Hz^{-1}$; in radio astronomia si usa il *Jansky*, essendo $1 Jy = 10^{-26} W m^{-2} Hz^{-1}$. La densità di flusso totale (detta anche brillantezza superficiale) si misura invece in $W m^{-2}$. In astronomia ottica si utilizza spesso la densità di flusso distribuita per lunghezze d'onda.

$$F_\lambda = F_\nu \frac{d\nu}{d\lambda} = \frac{c}{\lambda^2} F_\nu \quad (2.13)$$

Le unità di misura sono ovviamente $erg s^{-1} cm^{-3}$ oppure $W m^{-3}$.

Integrando F_ν su tutta la superficie della sorgente si ottiene il flusso totale specifico che in astronomia è detto luminosità specifica:

$$L_\nu = \int_A F_\nu dA \quad (2.14)$$

che nel caso di sorgente a simmetria sferica ed emissione isotropa diventa:

$$L_\nu = 4\pi R^2 F_\nu \quad (2.15)$$

con unità di misura $erg s^{-1} Hz^{-1}$, oppure $W Hz^{-1}$. Infine; la potenza totale della sorgente nella banda $\Delta\nu$, detta flusso totale o luminosità, è data da:

$$L = \int_{\Delta\nu} L_\nu d\nu \quad (2.16)$$

che, per una sorgente sferica e isotropa, assume la forma:

$$L = 4\pi R^2 \int_{\Delta\nu} F_\nu d\nu = 4\pi R^2 F \quad (2.17)$$

che si misura in $erg s^{-1}$ oppure W .

Questa quantità, in assenza di sorgenti o assorbitori, risulta costante al di fuori della sorgente. La radiazione osservata dai telescopi può essere descritta con le stesse grandezze; naturalmente tra le grandezze emesse e quelle osservate esistono vari fattori di “deformazione”, dovuti in parte alla propagazione in parte agli strumenti di misura. La densità di flusso specifico osservato a distanza r dalla sorgente viene ridotta anzitutto per effetto geometrico:

$$f_\nu = \frac{L_\nu}{4\pi r^2} = \left(\frac{R}{r}\right)^2 F_\nu \quad (2.18)$$

Inoltre si deve tener conto degli effetti di assorbimento e strumentali. Si definisce densità di flusso specifico efficace (o raccolto):

$$\ell_\nu = \alpha_\nu P_\nu f_\nu \quad (2.19)$$

dove α_ν è il coefficiente di assorbimento del mezzo (interstellare) interposto tra sorgente e atmosfera terrestre e P_ν è il fattore strumentale, a sua volta composto da tre componenti:

$$P_\nu = A_\nu^{-\sec z} Q_\nu S_\nu \quad (2.20)$$

rispettivamente l'assorbimento atmosferico (z è l'elevazione sull'orizzonte), l'assorbimento all'interno dello strumento (ottiche ed elettronica), la sensibilità strumentale (efficienza quantica). L'assorbimento atmosferico ha fisicamente la stessa origine di quello interstellare; viene tuttavia incluso negli effetti strumentali in quanto è in linea di principio una quantità misurabile in modo diretto.

Si chiama flusso specifico raccolto da un rivelatore di area effettiva Σ (area geometrica effettivamente affacciata alla radiazione incidente):

$$\Phi_\nu = \alpha_\nu P_\nu f_\nu \Sigma \quad (2.21)$$

Integrando sulla banda spettrale del sistema rivelatore - telescopio, si ottengono la densità di flusso efficace:

$$\ell = \int_{\Delta\nu} \alpha_\nu P_\nu f_\nu d\nu = \frac{1}{4\pi r^2} \int_{\Delta\nu} \alpha_\nu P_\nu L_\nu d\nu \quad (2.22)$$

e il flusso efficace raccolto:

$$\Phi = \frac{\Sigma}{4\pi r^2} \int_{\Delta\nu} \alpha_\nu P_\nu L_\nu d\nu \quad (2.23)$$

La stima dei parametri fisici delle sorgenti astrofisiche richiede la valutazione delle caratteristiche dei campi di radiazione alla superficie delle sorgenti astrofisiche, che si devono quindi ricavare partendo dalle grandezze osservate. Sulla base delle definizioni precedenti, il dato di partenza è il flusso efficace φ al rivelatore del telescopio, da cui, essendo nota l'area efficace Σ , si ottengono le densità di flusso ℓ e ℓ_ν .

Se si può misurare o stimare la distanza r della sorgente, si ha una valutazione dell'assorbimento interstellare α_ν (che come vedremo è calcolabile dalla composizione chimica e condizioni fisiche del gas), e con la misura delle caratteristiche strumentali P_ν si giunge a ricavare le luminosità L e L_ν . Qualora infine si abbia anche una stima delle dimensioni R della sorgente, si ricava infine proprio la densità di flusso superficiale F_ν , che è legata alle proprietà fisiche dell'oggetto celeste.

È utile considerare un'ulteriore quantità che viene utilizzata nel caso di sorgenti non puntiformi, cioè risolte dal telescopio; si tratta della brillantezza superficiale, precisamente la densità di flusso per unità di angolo solido. Con riferimento alla Figura 2.3 consideriamo che l'osservatore si trovi all'apice

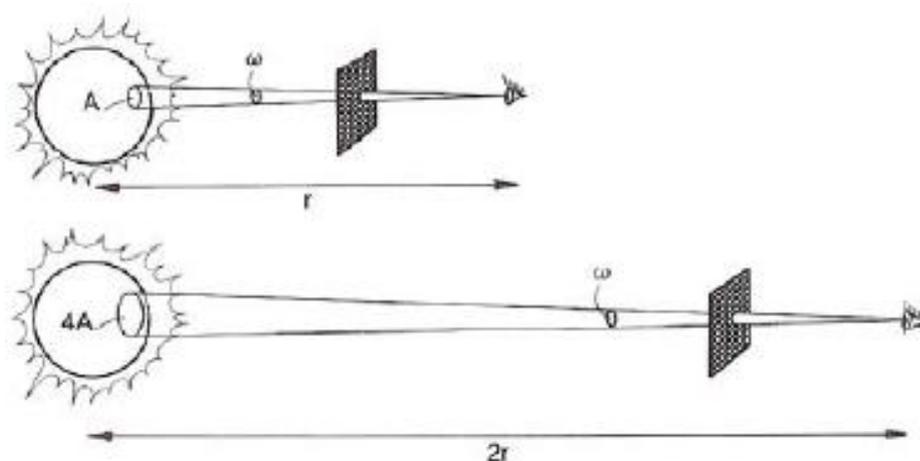


Figura 2.3: Osservazione di una sorgente risolta per un dato angolo solido

dell'angolo solido e che riceva una densità di flusso totale da un'area A della sorgente $f \propto (1/r^2)F$. Tuttavia per un dato angolo solido ω l'area sottesa cresce con r^2 , $A = r^2\omega$ per cui la densità di flusso per angolo solido $B = F/\omega$ di una sorgente non puntiforme uniforme non dipende dalla distanza.

Un campo di radiazione è anche caratterizzato attraverso la sua densità di energia che si misura in $J m^{-3}$. Si consideri una radiazione di intensità I incidente perpendicolarmente alla superficie dA entro un angolo solido $d\omega$.

Entro un tempo dt questa radiazione viaggerà per una distanza $c dt$ riempiendo un volume $dV = dA c dt$ e si potrà quindi definire una densità di energia

$$du = \frac{dE}{dV} = \frac{dAd\omega dt}{dA c dt} = \frac{1}{c} I d\omega \quad (2.24)$$

e integrando su tutte le direzioni di arrivo

$$u = \frac{1}{c} \int_S I d\omega \quad (2.25)$$

Nel caso di radiazione isotropa si ottiene:

$$u = \frac{4\pi}{c} I \quad (2.26)$$

2.2 Elementi di Fotometria Astronomica

Le prime classificazioni delle luminosità delle stelle sono dovute a Ipparco che nel II secolo a.C. introdusse una scala in sei classi di grandezze o magnitudini per le stelle osservate ad occhio nudo, a partire dalle stelle più luminose nella classe di prima magnitudine. Oggi tali classificazioni debbono tener conto delle osservazioni nelle diverse bande elettromagnetiche, per cui in genere si parla di fotometria (in senso lato) dalla banda radio a quella gamma.

La fotometria è la misura dell'energetica globale delle sorgenti astronomiche e della loro variabilità temporale in relazione alla banda di osservazione. Vengono qui indicate alcune formule di riferimento divise per le bande tipiche di osservazione, legate agli specifici strumenti di misura.

2.2.1 Ottico, Infrarosso, Ultravioletto

In fotometria si fa storicamente riferimento alle misure oculari. La sensazione dell'occhio agli stimoli luminosi è regolata dalla legge di Pogson (che nel 1856 formalizzò una legge psico-fisica sperimentata da Weber e Fechner), secondo cui l'occhio è sensibile al logaritmo della densità di flusso efficace ℓ entro la sua banda di sensibilità:

$$m = -2.5 \log \ell + \text{costante} \quad (2.27)$$

dove m viene detta magnitudine apparente e la costante deve essere definita attraverso una scala di riferimento (i logaritmi sono in base 10). I rapporti di luminosità vengono tradotti in differenze di magnitudine:

$$m_2 - m_1 = -2.5 \log(\ell_2/\ell_1) \quad (2.28)$$

La costante viene scelta in modo da riprodurre le scale di magnitudini ottiche definite da Tolomeo; in particolare si fissa la magnitudine della Stella Polare a $m = +2.12$. Si noti che spesso invece che di rapporti di densità di flusso efficace si parla di rapporti di luminosità apparenti intese come quantità di radiazione raccolta dal rivelatore con la sua risposta strumentale.

Si definisce anche una magnitudine assoluta, definita come la magnitudine apparente delle sorgenti qualora fossero portate alla distanza di 10 parsec:

$$M = -2.5 \log \ell_{(10)} + \text{costante} \quad (2.29)$$

Essa è legata alla magnitudine apparente dalla relazione:

$$m - M = 2.5 \log(\ell_{(10)}/\ell) = 2.5 \log(r_{pc}/10)^2 = 5 \log r_{pc} - 5 \quad (2.30)$$

La differenza tra magnitudine apparente e assoluta è chiamata modulo di distanza.

Naturalmente le misure di magnitudine dipendono essenzialmente dallo strumento di misura, cioè dalla funzione P_ν e dalla banda di accettazione $\Delta\nu$ definite precedentemente.

Ciò porta alla definizione di sistemi fotometrici che sono stati progressivamente elaborati sulla base dei telescopi e rivelatori usati:

- sistema visuale, basato sull'occhio umano medio e sull'uso di telescopi rifrattori
- sistema fotografico, basato sulle lastre fotografiche ordinarie e sull'uso di telescopi riflettori a specchi argentati
- sistema fotovisuale, basato su lastre ortocromatiche, più simili all'occhio come banda di accettazione
- sistemi fotoelettrici, basati sull'uso di fotomoltiplicatori e filtri.

Nella figura 2.4 sono dati i diagrammi della funzione P_ν per il sistema fotometrico a molti colori oggi usato: il sistema UBV di Johnson (le iniziali si riferiscono alle bande ultravioletta, blu, visibile), esteso alle bande infrarosse RIJ (Banda rossa, infrarossa, ecc.).

Nella Figura 2.5 sono riportati in dettaglio i valori di calibrazione. I sistemi a molti colori permettono una prima valutazione della distribuzione energetica nelle bande ottiche e/o vicine, fornendo quella che si può chiamare una spettroscopia a banda larga.

Infine si chiama magnitudine bolometrica la magnitudine calcolata non solo nella banda osservata, ma integrata su tutto lo spettro; per sorgenti termiche (in particolare le stelle) si può ottenere la magnitudine bolometrica da quella apparente nell'ottico con un'estrapolazione basata sulla legge di Planck:

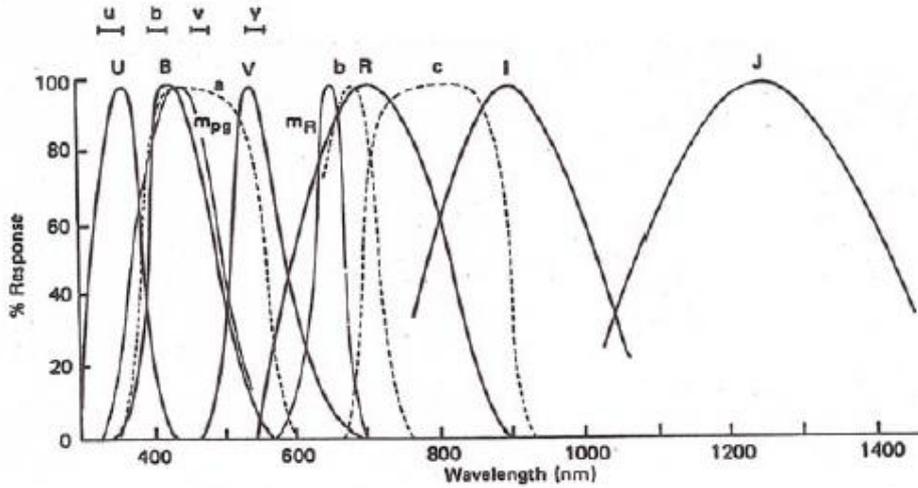


Figura 2.4: Risposta dei rivelatori fotometrici per il sistema fotometrico *UBV* di Johnson esteso alle bande *RIJ* dell'infrarosso. Per confronto sono riportate anche le risposte del sistema fotografico blu m_{pg} e rosso m_r delle lastre di Mount Palomar. Sono inoltre riportate le risposte delle bande del telescopio UK Schmidt (*abc*) e quelle del sistema di Stromgren (*ubvy*).

$$M_{bol} = m_{\nu} - BC \quad (2.31)$$

dove la correzione è fornita da apposite tabelle standard reperibili nei cataloghi stellari.

Le magnitudini bolometriche assolute possono essere espresse in funzione della luminosità. Usando le densità di flusso F per sorgenti alla distanza di 10 pc e riferendosi al Sole per la calibrazione si può scrivere:

$$M_{bol} - M_{bol,\odot} = -2.5 \log \frac{F}{F_{\odot}} = -2.5 \log \frac{L}{L_{\odot}} \quad (2.32)$$

e si può ricavare che la magnitudine bolometrica $M_{bol} = 0$ corrisponde a $L = 3.0 \cdot 10^{35} \text{ erg s}^{-1}$.

Val la pena di citare, infine, i principali cataloghi astrofotometrici ottici : la Sloan Digital Sky Survey o SDSS, la Palomar Sky Survey, lo UK Schmidt Catalog, lo Hubble Space Telescope Guide Star Catalog (costruita con le misure del Telescopio Spaziale in orbita dal 1990), il Carlsberg Astrometric Catalog. Nell'infrarosso il catalogo più completo è quello ottenuto col sa-

Waveband	λ_{eff} (μm)	$\Delta\lambda_{\text{eff}}$ (μm)	ν_0 (Hz)	$S_\nu(0)$ ($\text{W m}^{-2}\text{Hz}^{-1}$)
U	0.365	0.068	8.3×10^{14}	1.88×10^{-23}
B	0.440	0.098	7.0×10^{14}	4.44×10^{-23}
V	0.550	0.089	5.6×10^{14}	3.81×10^{-23}
R	0.700	0.220	4.3×10^{14}	2.88×10^{-23}
I _S	0.800	—	3.7×10^{14}	2.50×10^{-23}
I _J	0.900	0.240	3.3×10^{14}	2.24×10^{-23}
J	1.250	0.380	2.4×10^{14}	1.77×10^{-23}
H	1.650	—	1.80×10^{14}	1.05×10^{-23}
K	2.200	0.480	1.36×10^{14}	6.5×10^{-24}
L	3.400	0.700	8.6×10^{13}	2.95×10^{-24}
M	5.000	—	6.3×10^{13}	1.9×10^{-24}
N	10.200	—	3.0×10^{13}	4.3×10^{-25}
Q	19.500	—	1.55×10^{13}	1.1×10^{-25}

$\log S_\nu = \log S_\nu(0) - 0.4m$
 $\log S_B = \log S_V + 0.07 - 0.4(B - V)$
 $\log S_U = \log S_B - 0.37 - 0.4(U - B)$
 $\log S_K = \log S_V - 0.78 + 0.4(V - K)$
 $\log S_V = -22.42 - 0.4 \times AB$
 $AB = K + 1.92$

Figura 2.5: Calibrazioni assolute di una stella di magnitudine zero in differenti bande passanti

tellite IRAS. Nella banda ultravioletta il catalogo completo del cielo risale alla missione OAO-2, ma verrà presto sostituito dai risultati della missione GALEX.

2.3 Emissione nel Continuo

Le stelle sono sorgenti che emettono radiazione che proviene dal loro interno dove ha raggiunto l'equilibrio termodinamico con la materia. In genere un radiatore in equilibrio termodinamico raggiunge un equilibrio tra radiazione emessa ed assorbita ad ogni frequenza attraverso processi di emissione ed assorbimento per transizioni degli elettroni tra i livelli atomici e per diffusione degli elettroni liberi da parte di ioni (bremsstrahlung), in cui i fotoni vengono distribuiti in modo continuo su tutte le frequenze; la condizione di equilibrio tra emissività j_ν ed assorbimento $k_\nu I_\nu$ per unità di volume, angolo solido e frequenza del radiatore si scrive:

$$j_\nu dV d\omega d\nu = k_\nu I_\nu dV d\omega d\nu \quad (2.33)$$

Ne discende la legge di Kirchhoff (1859) secondo cui il rapporto tra emissività e assorbimento è l'intensità della radiazione emessa ed è costante ad una data temperatura:

$$\frac{j_\nu}{k_\nu} = I_\nu \quad (2.34)$$

In particolare un radiatore ideale che abbia coefficiente di assorbimento pari all'unità, cioè sia in grado di assorbire tutta la radiazione incidente (e quindi sia nero), viene detto corpo nero e la sua emissività dipende solo dalla temperatura assoluta:

$$j_\nu = B_\nu(T) \quad (2.35)$$

Dove la funzione $B_\nu(T)$ è stata ricavata da Planck (1900); la distribuzione di Planck (vedere anche figura 2.6) è:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \quad (2.36)$$

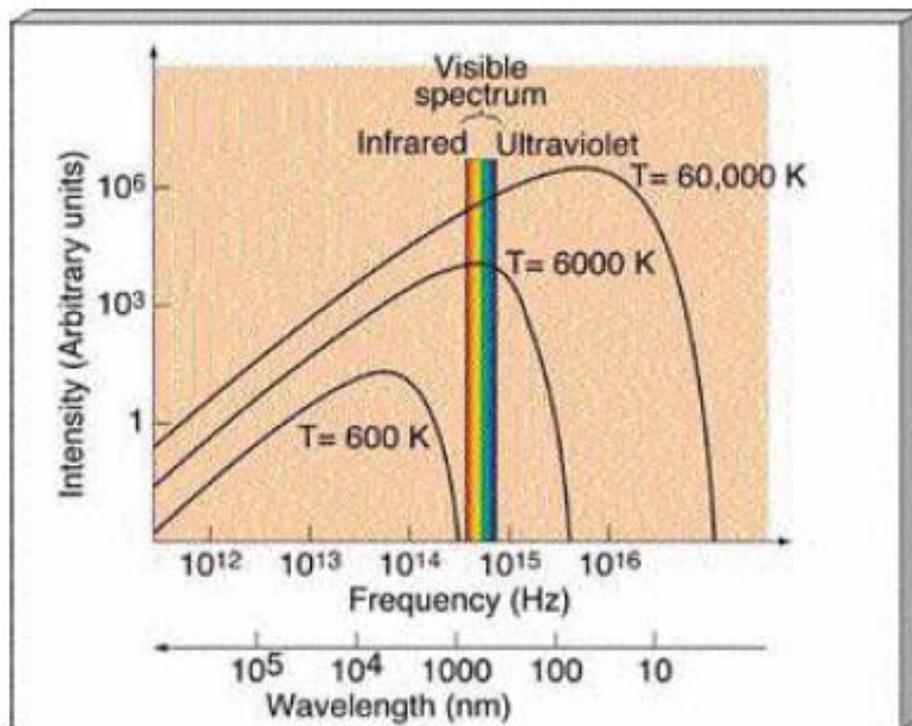


Figura 2.6: Distribuzione di brillantezza della radiazione di corpo nero per diverse temperature

<i>Oggetto</i>	λ_{\max}	T
Nubi fredde, gas, polvere	$10^3 - 10^6$ nm	$10 - 10^3$ K
Stelle, nubi calde	$10^2 - 10^3$	$10^3 - 10^4$
Stelle calde, resti supernova	$10 - 10^2$	$10^4 - 10^6$
Dischi accrescimento, IGM	$10^{-2} - 10$	$10^6 - 10^8$

Figura 2.7: Temperature di plasmi astrofisici stimati con la legge di Wien

con $h = 6.625 \cdot 10^{-27}$ erg s (costante di Planck) e $k = 1.386 \cdot 10^{-16}$ erg K⁻¹ (costante di Boltzmann).

Esistono varie proprietà della legge di Planck che è utile rammentare per futuro riferimento. Anzitutto l'emissione è concentrata in una banda spettrale definita dalla temperatura secondo la legge dello spostamento di Wien:

$$\lambda_{\max} T = 0.290 \text{ cm K} \quad (2.37)$$

Nelle figure 2.7 e 2.8 si riportano alcuni esempi di oggetti astrofisici in cui la legge di Wien può essere utilizzata per valutarne la temperatura.

La densità di flusso totale (integrata su tutte le frequenze) del corpo nero è data dalla legge di Stefan-Boltzmann:

$$F = \sigma T^4 \quad (2.38)$$

con $\sigma = 5.66956 \cdot 10^{-5}$ erg cm⁻² s⁻¹ K⁻⁴ (costante di Stefan).

Tale relazione viene spesso utilizzata per le valutazioni di temperatura di sorgenti termiche: si definisce temperatura effettiva, T_{eff} , la temperatura del corpo nero che produce la stessa densità di flusso totale osservata nelle sorgenti

$$F = \sigma T_{eff}^4 \quad (2.39)$$

Naturalmente tale valutazione è significativa solo per sorgenti simili al corpo nero.

Spesso viene usata un'ulteriore definizione di temperatura legata all'ipotesi che le sorgenti siano del tipo corpo nero: si chiama temperatura di

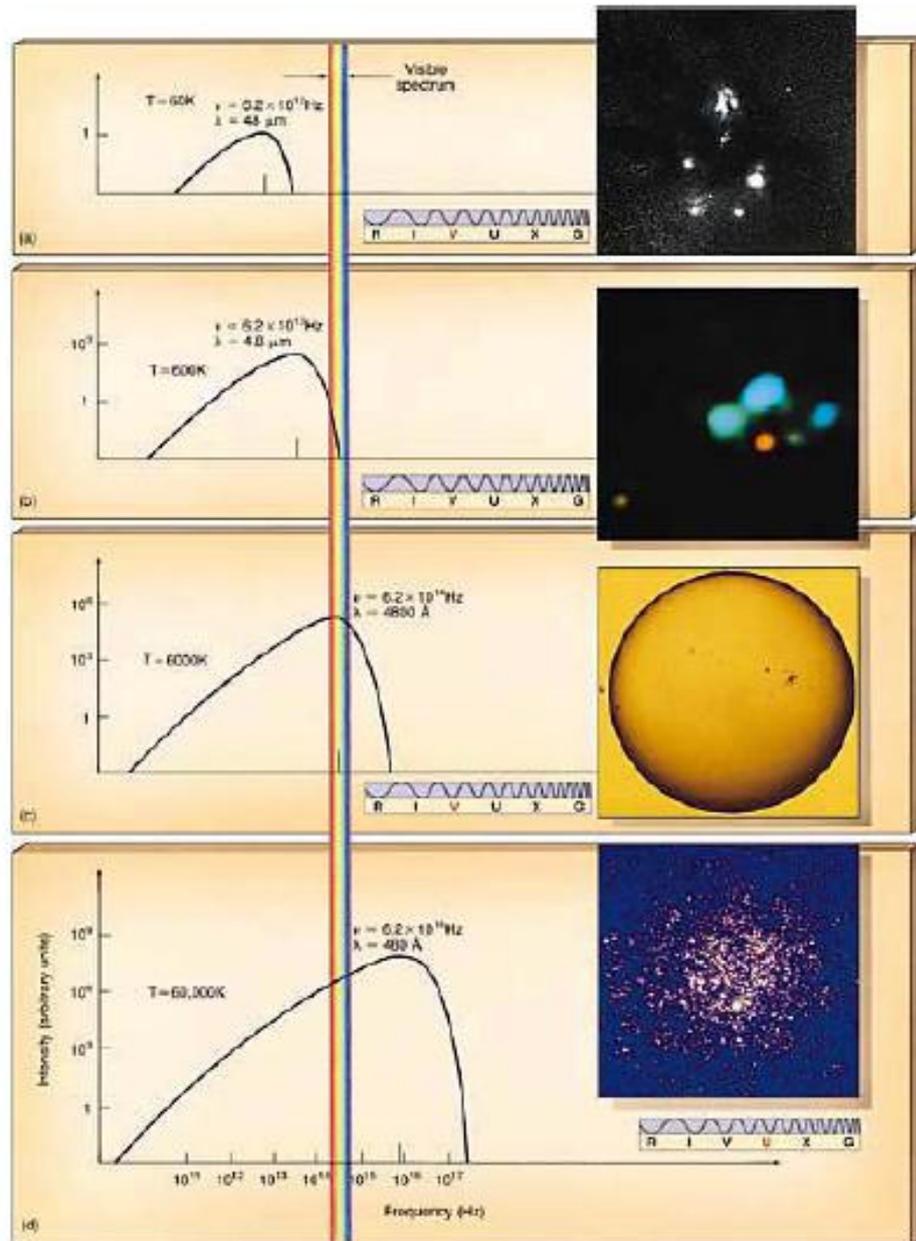


Figura 2.8: Tipiche sorgenti termiche in astrofisica e loro spettri: (a) nubi fredde del mezzo interstellare, (b) nubi di formazione stellare, (c) Sole, (d) ammasso stellare con stelle calde.

brillanza la temperatura che ha un corpo nero capace di produrre la stessa intensità (brillanza) I_ν alla frequenza osservata.

Va ricordato che le curve di corpo nero non si intersecano mai, per cui ad ogni data frequenza solo una data temperatura può produrre l'intensità osservata. Questa definizione è soprattutto usata in radioastronomia, per cui la brillanza del corpo nero può essere scritta nell'approssimazione di Rayleigh-Jeans per basse frequenze $h\nu \ll kT$

$$B_\nu(T) = \frac{2k\nu^2}{c^2}T \quad (2.40)$$

e quindi

$$T_b = \frac{c^2}{2k\nu^2}I_\nu \quad (2.41)$$

Esiste anche un'approssimazione per le alte frequenze $h\nu \gg kT$, l'approssimazione di Wien:

$$B_\nu(T) = \frac{2k\nu^3}{c^2}e^{-h\nu/kT} \quad (2.42)$$

Stelle e galassie attive, per lo più di alta energia, sono invece sorgenti non-termiche perché non in equilibrio termodinamico, e la loro radiazione non segue la distribuzione spettrale del corpo nero. Di conseguenza questi oggetti non sono caratterizzabili in termini di un'energia specifica media o di una temperatura. I meccanismi di emissione nei casi astrofisici di interesse sono radiazione sincrotrone, Compton-inverso, bremsstrahlung non-termico. Il profilo di emissione sulle varie frequenze è generalmente molto meno concentrato su una data frequenza, ma tende ad essere esteso. La banda di emissione ha tuttavia un massimo intorno ad una tipica frequenza $\nu \sim \gamma^2\nu_0$; ove γ è il fattore di Lorentz della popolazione dominante delle particelle emmettenti e ν_0 la frequenza caratteristica del processo: frequenza di sincrotrone e frequenza del gas di fotoni nel Compton inverso.

2.4 Indici di Colore

In pratica non è agevole in fotometria ricostruire la distribuzione continua di una sorgente su tutte le frequenze se non eseguendo molte misure con differenti filtri e differenti strumenti. Tuttavia almeno per le sorgenti ter-

miche di tipo stellare si può caratterizzare l'energia specifica dell'emissione confrontandone le densità di flusso in alcune bande dei sistemi fotometrici. Questo metodo è usato soprattutto nell'astronomia ottica, dove vengono confrontate le magnitudini apparenti nelle bande U , B, V ; va notato che queste tre lettere maiuscole vengono usate per indicare le magnitudini apparenti delle stelle nelle corrispondenti bande.

Si definisce indice di colore (con ovvio riferimento alle osservazioni nel visibile) la differenza di magnitudini in due bande:

$$\text{indice di colore (IC)} = m_2 - m_1 = -2.5 \log \left(\frac{\ell_2}{\ell_1} \right) \quad (2.43)$$

In genere si usano gli indici di colore U - B e B -V , normalizzati con $U = B = V = 0.03$ per stelle di tipo Vega. Nei cataloghi si definiscono le stelle standard primarie con indici di colore di riferimento.

L'indice di colore è soprattutto significativo per sorgenti termiche, come appunto le stelle, in quanto permette una valutazione della temperatura. Infatti il rapporto di densità di flusso in un corpo nero dipende solo dalla temperatura:

$$\frac{B_{\nu 2}}{B_{\nu 1}} = f(T) \quad (2.44)$$

in una sorgente stellare che emette nell'ottico questa relazione può essere scritta nella forma:

$$\frac{\ell_{\nu 2}}{\ell_{\nu 1}} = k + g(T) \quad (2.45)$$

che diventa, in magnitudini:

$$B - V = -0.60 + \frac{7300}{T} \quad (2.46)$$

Per sorgenti termiche questa è la temperatura del corpo nero che darebbe lo stesso colore, ed è quindi la temperatura delle sorgenti con ottima approssimazione. A volte la definizione è applicata anche a sorgenti non-termiche, e la temperatura che se ne ricava viene detta temperatura di colore: indica quale corpo nero sarebbe capace di produrre lo stesso "colore", ma non ha alcuna necessaria relazione con le caratteristiche delle sorgenti (che peraltro non possono essere descritte con una vera e propria temperatura).

In un corpo nero gli indici di colore sono ovviamente tutti correlati in

quanto dipendono solo dalla temperatura. Johnson (1950) mostrò che questo è sostanzialmente vero anche nelle stelle, rinforzando l'idea che abbiano un'emissione molto simile a quella di un corpo nero.

In figura 2.9 è disegnato il diagramma a due colori (U - B, B - V) riferito a stelle a noi relativamente vicine (< 10 pc), in modo che il loro colore non sia sostanzialmente influenzato dalla propagazione attraverso al mezzo interstellare (vedi paragrafo seguente). E' pure data la retta che definisce il diagramma per un corpo nero. Si osserva che in effetti esiste una correlazione molto ben definita, che segue nell'andamento medio quella del corpo nero; vedremo più avanti che l'affossamento della curva intorno a $B - V \approx 0$ è dovuto all'assorbimento del continuo da parte dell'idrogeno a lunghezze d'onda sotto i 3646 Å.

Gli indici di colore, essendo rapporti tra due luminosità dello stesso oggetto, risultano indipendenti dalla distanza, almeno per propagazione nel vuoto. In tal senso il diagramma a due colori dipende solo dalle caratteristiche intrinseche degli oggetti, indipendentemente dalle loro diverse distanze. È questo fatto che rende attraente il suo utilizzo. Come vedremo nel paragrafo seguente, ciò non è più vero se tra sorgente e osservatore intervengono effetti di assorbimento e/o riemissione da parte di un mezzo interposto: ciò accade in effetti a causa della presenza del mezzo interstellare.

2.5 Estinzione e Profondità Ottica

La magnitudine apparente di una stella cresce al crescere della distanza, cioè la sua densità di flusso diminuisce secondo l'equazione 2.30 se lo spazio tra sorgente e osservatore è completamente vuoto. Poiché nella realtà la radiazione delle stelle giunge a Terra attraversando il mezzo interstellare (e interplanetario), la suddetta relazione non è più corretta, in quanto la radiazione può essere assorbita (ed eventualmente riemessa a differente frequenza) e diffusa (riemessa in una differente direzione), il che comporta un'ulteriore effetto di aumento della magnitudine. Tale effetto viene chiamato estinzione, e impedisce, come vedremo più avanti, la rivelazione di stelle che si trovino sul piano galattico a distanze superiori ai 2 kpc. È utile ricavare la legge secondo cui l'estinzione dipende dalla distanza.

Si consideri una stella che emetta un flusso totale specifico $L_{\nu,0}$ nell'angolo solido ω che si propaga entro un mezzo che assorbe e diffonde la

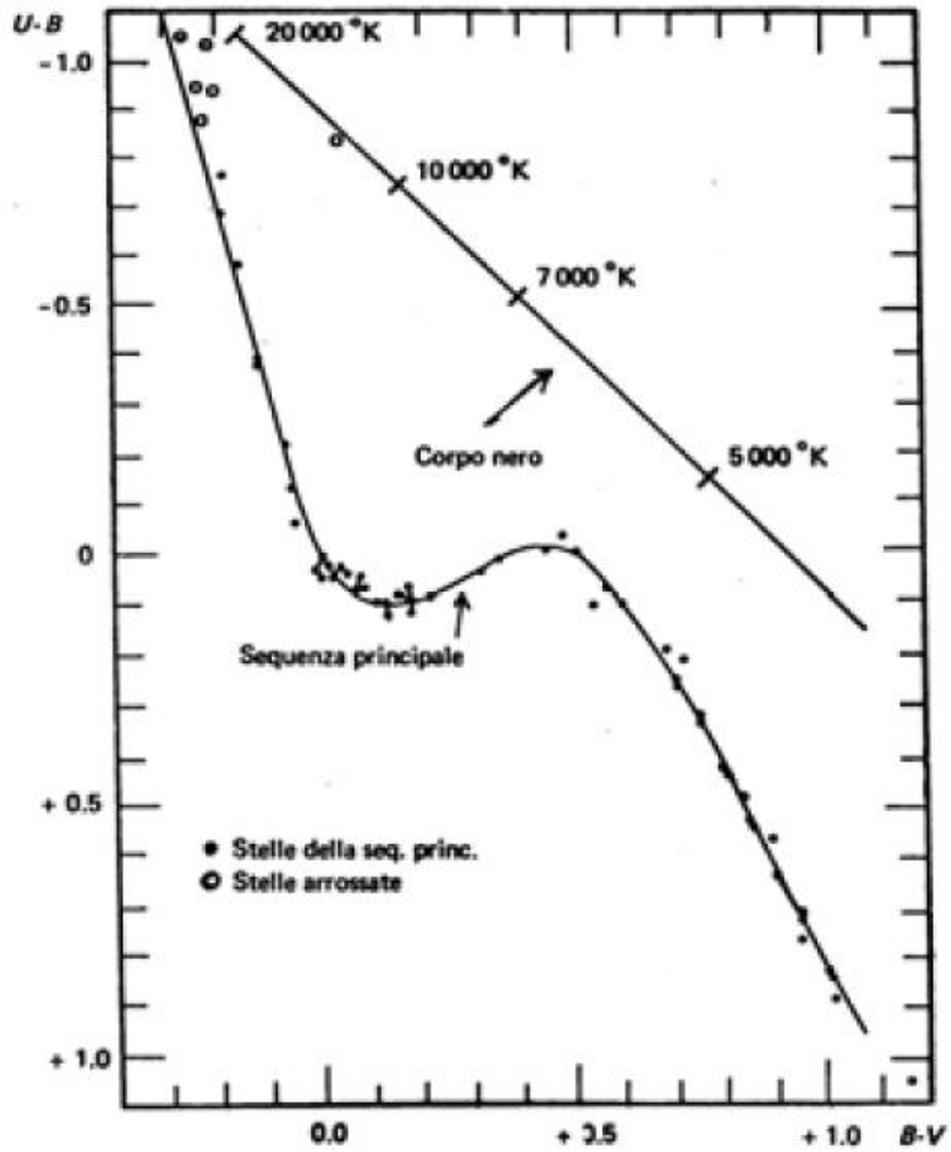


Figura 2.9: Il diagramma a due colori per le stelle

radiazione: al crescere della distanza percorsa il flusso L_ν diminuirà a causa dell'interazione col mezzo. Si può quindi scrivere che in un tratto elementare $(r, r + dr)$ la diminuzione del flusso sarà proporzionale al flusso stesso e alla distanza attraversata:

$$dL_\nu = -\alpha_\nu L_\nu dr \quad (2.47)$$

dove il coefficiente α_ν , detto opacità, dipende dalla fisica dell'interazione fotoni/materia ed è naturalmente funzione della frequenza della radiazione, in particolare è maggiore per frequenze maggiori; le dimensioni del coefficiente α_ν sono $[l^{-1}]$.

Naturalmente l'opacità è nulla nel vuoto e cresce al crescere della densità del mezzo. Discuteremo la fisica del processo nello studio del mezzo interstellare.

In astrofisica si usa spesso una quantità adimensionale detta profondità ottica τ_ν così definita:

$$\tau_\nu = \int_0^{\tau_\nu} d\tau_\nu = \int_0^r \alpha_\nu dr \quad (2.48)$$

con la quale l'equazione 2.47 può essere riscritta e integrata:

$$L_\nu = L_{\nu,0} e^{-\tau_\nu} \quad (2.49)$$

Questa relazione mostra che il flusso della radiazione decresce esponenzialmente con τ_ν , quindi con la distanza e l'opacità del mezzo attraversato: profondità ottica $\tau_\nu < 1$ indica che il mezzo è praticamente trasparente alla radiazione, profondità ottica $\tau_\nu \gg 1$ indica che il mezzo è completamente opaco e impedisce la rivelazione della sorgente.

Vediamo ora di esprimere l'effetto dell'estinzione sulle magnitudini.

Sia $F_{\nu,0}$ la densità di flusso specifica alla superficie di una stella e sia $F_\nu(r)$ il suo valore alla distanza r dopo aver attraversato un mezzo assorbente; la relazione con il flusso totale nell'angolo solido ω sarà:

$$L_\nu(r) = \omega r^2 F_\nu(r) \quad (2.50)$$

$$L_{\nu,0} = \omega R^2 F_{\nu,0} \quad (2.51)$$

dove R è il raggio della stella. Pertanto:

$$F_\nu(r) = F_{\nu,0} \left(\frac{R}{r} \right)^2 \quad (2.52)$$

Le magnitudini apparenti sono legate a tale flusso; invece le magnitudini assolute dipendono dal flusso alla distanza di 10 pc in assenza di estinzione perché a questa distanza il mezzo interstellare risulta trasparente:

$$F_\nu(10) = F_{\nu,0} \left(\frac{R}{10pc} \right)^2 \quad (2.53)$$

Il modulo di distanza diventa dunque:

$$m - M = -2.5 \log \frac{F_\nu(r)}{F_\nu(10)} = 5 \log \frac{r}{10pc} - 2.5 \log e^{-\tau_\nu} = 5 \log \frac{r}{10pc} + A_\nu \quad (2.54)$$

dove

$$A_\nu = (2.5 \log e) \tau_\nu \quad (2.55)$$

è l'estinzione in magnitudini dovuto all'intero mezzo tra la stella e l'osservatore.

Se l'opacità è costante lungo la linea di vista, si può scrivere:

$$\tau_\nu = \alpha_\nu \int_0^r dr = \alpha_\nu r \quad (2.56)$$

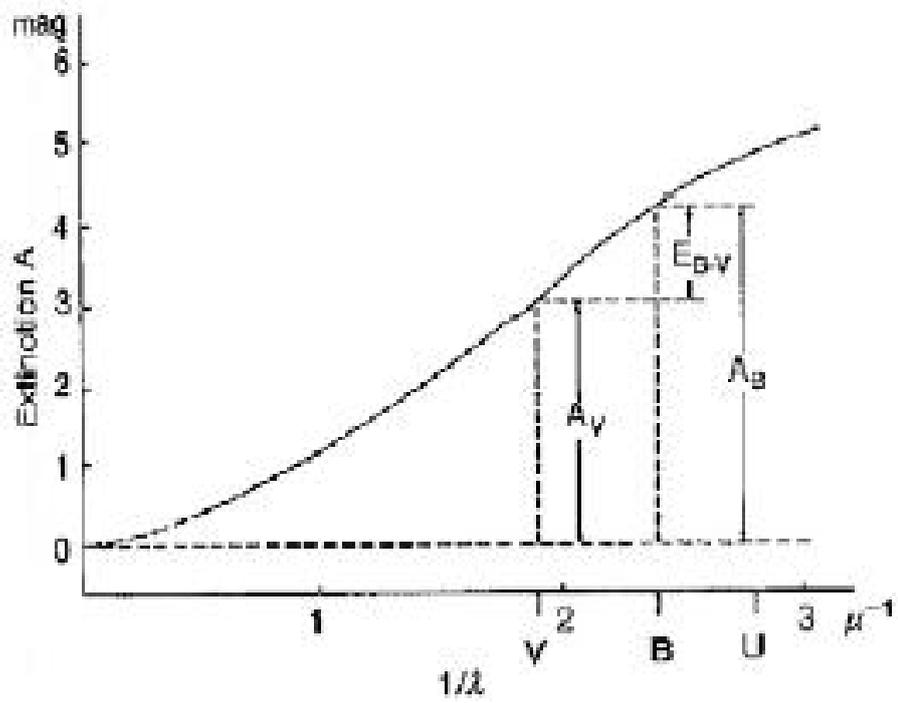
e quindi il modulo di distanza in presenza di un mezzo assorbente diventa:

$$m - M = 5 \log \frac{r}{10pc} - 5ar \quad (2.57)$$

dove $a = (2.5 \log e) \alpha_\nu$ rappresenta l'estinzione in magnitudini per unità di distanza.

2.6 Eccesso di Colore

Come abbiamo detto, l'estinzione dipende dalla frequenza della radiazione, in particolare le frequenze maggiori sono assorbite più di quelle minori: nella banda ottica il blu è più assorbito e diffuso del rosso (vedi Figura 2.10).

Figura 2.10: Attenuazione A_v

Ciò comporta che il colore di una sorgente termica tende ad arrossarsi: in modo più quantitativo possiamo dire che il suo indice di colore $B - V$ cresce. Consideriamo le magnitudini apparenti $V4e4B$ di una stella nel visibile e nel blu secondo l'equazione 2.54:

$$V = M_V + 5 \log \frac{r}{10pc} + A_V \quad (2.58)$$

$$B = M_B + 5 \log \frac{r}{10pc} + A_B \quad (2.59)$$

$$B - V = (M_B - M_V) + (A_B - A_V) = (B - V)_0 + E_{B-V} \quad (2.60)$$

da cui:

dove $(B - V)_0$ è il colore intrinseco della sorgente e $E_{B-V} = (B - V) - (B - V)_0$ è l'eccesso di colore dovuto alla propagazione.

Basandosi sul diagramma a due colori è anche possibile misurare la distribuzione del mezzo interstellare.

Utilizzando l'equazione 2.60 e la sua analoga per l'altro indice di colore si ottiene che l'estinzione determina uno spostamento della posizione di una stella nel diagramma a due colori secondo la relazione:

$$(U - B) - (U - B)_0 = \frac{E_{U-B}}{E_{B-V}} [(B - V) - (B - V)_0] \quad (2.61)$$

In pratica quando una stella si trova fuori dalla posizione della curva di correlazione ottenuta per le stelle vicine (non arrossate), la si riporta sulla curva muovendone il punto rappresentativo parallelamente alla linea di arrossamento, cioè alla retta con coefficiente angolare E_{U-B}/E_{B-V} (vedi figura 2.11).

Tale spostamento dà il valore di $(B - V) - (B - V)_0$ e quindi una misura diretta di E_{B-V} , cioè di A_V , l'assorbimento subito dalla radiazione nel visibile. Quindi, con l'equazione 2.57 e considerando che il coefficiente di assorbimento α_ν è proporzionale alla densità del mezzo, si può stimare la densità di colonna del materiale attraversato:

$$\tau_V = (2.5 \log e)^{-1} A_V = \int_{sorg}^{oss} \alpha_\nu dr \propto \int_{sorg}^{oss} n ds \quad (2.62)$$

Eseguendo la misura di estinzione su stelle di cui si sia già stimata la

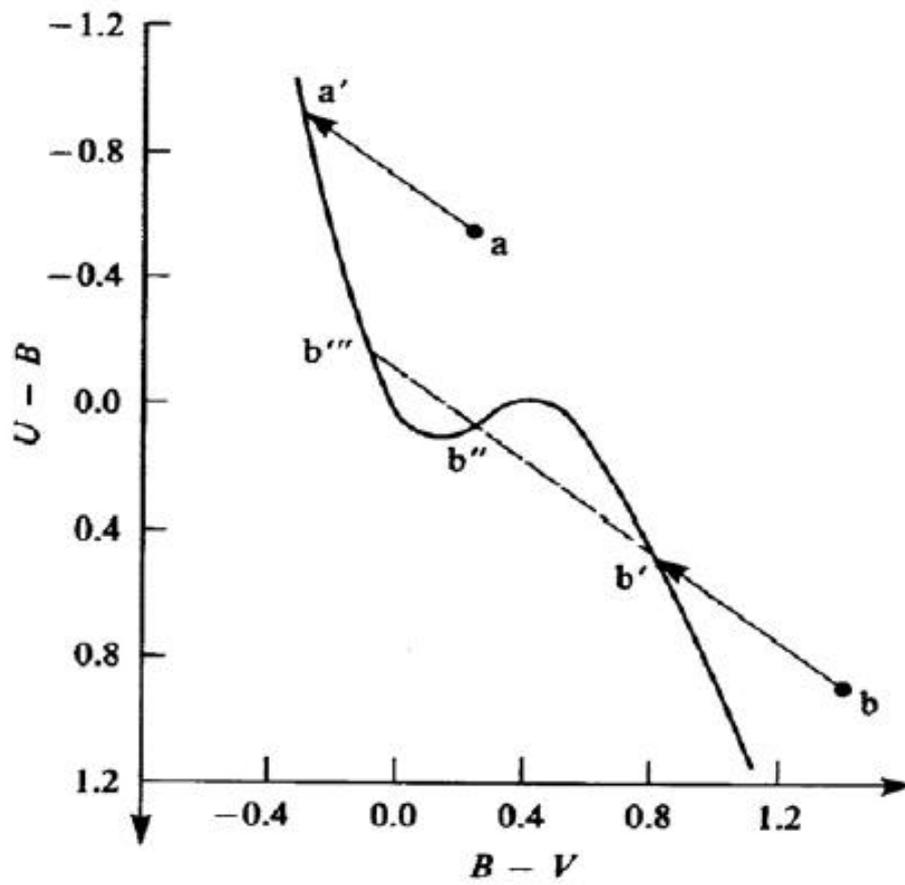


Figura 2.11: Arrossamento degli spettri stellari e misure delle distanze

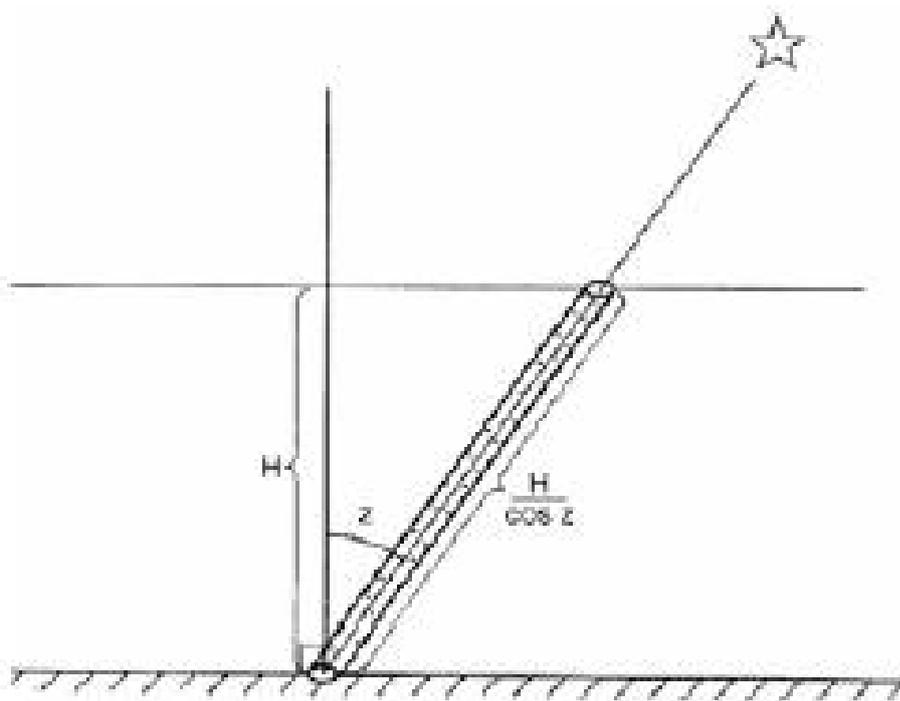


Figura 2.12: Massa atmosferica in funzione della distanza zenitale

distanza, è quindi possibile ricavare la densità media del mezzo interstellare nella loro direzione. In tal modo si ottiene, ad esempio, una mappa della distribuzione di materia intorno al Sole. Alternativamente, determinata la densità di materia in una certa direzione, è possibile ricavare la distanza di altre stelle dalla loro estinzione.

2.7 Estinzione Atmosferica

Possiamo esprimere l'assorbimento atmosferico in modo analogo a quanto visto per l'estinzione del mezzo interstellare. In questo caso però possiamo tenere conto della quantità di materia presente lungo la linea di vista a seconda dell'elevazione della sorgente sull'orizzonte. La minima attenuazione si avrà quando la sorgente si trova allo zenit, ed aumenterà con la distanza zenitale, rappresentata dall'angolo z (vedi figura 2.12).

Nel caso di piccole distanze zenitali, l'atmosfera può essere trattata come uno strato piano per cui la massa di materiale attraversato in funzione di

z sarà proporzionale a $X = 1/\cos z$ e la magnitudine di una stella crescerà secondo la relazione:

$$m = m_0 + kX \quad (2.63)$$

dove k è un coefficiente di estinzione atmosferica che può essere misurato con osservazione di sorgenti di magnitudine nota; va detto che tale coefficiente dipende dalla frequenza, da effetti locali ed è variabile. Inoltre il problema si complica per la curvatura dell'atmosfera quando l'angolo zenitale $z > 70^\circ$.

2.8 Trasporto Radiativo

Per comprendere la fisica dei processi che determinano l'attenuazione della luminosità delle stelle durante la propagazione nel mezzo interstellare, è utile discutere le basi dell'equazione del trasporto radiativo, un argomento fondamentale nell'astrofisica che incontreremo in molte applicazioni.

Si consideri un cilindro di area di base dA e lunghezza dr del mezzo entro cui avviene la propagazione e si assuma che una radiazione di intensità I_ν entri perpendicolarmente ad una delle superfici di base entro l'angolo solido $d\omega$. All'uscita dalla superficie opposta l'intensità sarà variata di dI_ν e corrispondentemente l'energia trasportata sarà variata di

$$dE_\nu = dI_\nu dA d\nu d\omega dt \quad (2.64)$$

a causa dei processi di assorbimento a cui la radiazione è andata incontro per interazione col mezzo e dei processi di emissione con cui il mezzo contribuisce nuovi fotoni. Si possono trascurare processi di diffusione in quanto si assume che il cilindro sia in equilibrio con la materia adiacente, per cui fotoni diffusi uscenti sono bilanciati da quelli entranti.

L'energia assorbita dal mezzo può essere scritta come:

$$dE_{\nu,ass} = \alpha_\nu I_\nu dr dA d\nu d\omega dt \quad (2.65)$$

dove α_ν è l'opacità del mezzo. Se si indica con j_ν il coefficiente di emissione del mezzo, cioè la quantità di energia emessa dal mezzo alla frequenza ν , nell'angolo solido $d\omega$, nell'unità di volume e nell'unità di tempo, l'energia emessa dal cilindro sarà:

$$dE_{\nu,em} = j_{\nu} dr dA d\nu d\omega dt \quad (2.66)$$

Pertanto:

$$dE_{\nu} = dE_{\nu,em} - dE_{\nu,ass} \quad (2.67)$$

da cui

$$dI_{\nu} = -\alpha_{\nu} I_{\nu} dr + j_{\nu} dr \quad (2.68)$$

ovvero, utilizzando la profondità ottica

$$\frac{dI_{\nu}}{d\tau_{\nu}} = -I_{\nu} + S_{\nu} \quad (2.69)$$

dove $S_{\nu} = j_n u / \alpha_n u$ è la cosiddetta funzione sorgente. L'equazione 2.69 trasporto. Nel caso specifico che assorbimento ed emissione si bilancino esattamente si avrà

$$I_{\nu} = \frac{j_n u}{\alpha_{\nu}} = S_{\nu} \quad (2.70)$$

che è la legge di Kirchhoff precedentemente citata e ora esplicitamente derivata. In equilibrio termodinamico (a tutte le frequenze) la radiazione del mezzo è quella di corpo nero, e quindi anche la funzione sorgente è la funzione di Planck. L'equazione del trasporto può essere integrata formalmente:

$$I_{\nu}(\tau_{\nu}) = I_{\nu}(0)e^{(-\tau_{\nu}-x)} S_{\nu}(x) dx \quad (2.71)$$

dove $I_{\nu}(0)$ è l'intensità della radiazione della sorgente che decade esponenzialmente attraversando il mezzo, come già abbiamo visto nell'equazione 2.49.

Il secondo termine invece tiene conto del contributo dovuto all'irraggiamento del mezzo stesso, che pure viene ridotto durante la propagazione.

Per $\tau_{\nu} \gg 1$ il contributo all'intensità osservata è dovuto essenzialmente al mezzo in quanto la sorgente risulta completamente estinta.

La soluzione è formale in quanto la funzione sorgente S_{ν} non è nota e deve essere risolta insieme all'intensità I_{ν} . Come esempio si consideri il caso con S_{ν} costante e nessuna sorgente di fondo: si ottiene

$$I_\nu(\tau_\nu) = S_\nu \int_0^{\tau_\nu} e^{-(\tau_\nu-x)} dx = S_\nu(1 - e^{-\tau_\nu}) \quad (2.72)$$

e per $\tau_\nu \gg 1$ intensità e funzione sorgente si eguagliano

$$I_\nu = S_\nu \quad (2.73)$$

cioè i processi di assorbimento ed emissione nel mezzo si equilibrano.

Una delle applicazioni della teoria del trasporto radiativo è nello studio delle atmosfere stellari e planetarie e nella propagazione della radiazione nel mezzo interplanetario. Nella maggior parte dei casi occorre tener conto della geometria; per lo più è sufficiente lavorare in geometria piana, con le proprietà del mezzo che variano solo nella direzione z perpendicolare agli strati piani.

Pertanto, assumendo z crescente verso l'alto e indicando con r la direzione generica di propagazione dei raggi e con θ l'angolo rispetto alla verticale, si pone:

$$d\tau_\nu = -\alpha_\nu dz = -\alpha_n u dr \cos \theta \quad (2.74)$$

$$\cos \theta \frac{dI_\nu(z, \theta)}{d\tau_\nu} = -I_\nu + S_\nu \quad (2.75)$$

Un'espressione formale per l'energia che emerge ad esempio da un'atmosfera stellare piano-parallela è:

$$I_\nu(0, \theta) = \int_0^\infty S_\nu e^{-(\tau_\nu \sec \theta)} \sec \theta d\tau_\nu = \quad (2.76)$$

2.9 Elementi di Spettroscopia

Lo studio fotometrico nelle diverse bande elettromagnetiche permette di concludere che la radiazione delle stelle e delle galassie normali è distribuita in modo continuo su tutte le lunghezze d'onda e segue da vicino la legge di Planck del corpo nero. Ciò implica che i plasmi emettenti sono in equilibrio termodinamico, sono cioè sorgenti termiche e se ne può valutare la temperatura, definita come temperatura di colore o temperatura effettiva. Il meccanismo di emissione di tali sorgenti è il bremsstrahlung termico basato sulla diffusione di elettroni liberi da parte di ioni positivi, oltre a transizioni

di eccitazione e diseccitazione di atomi e, per plasmi freddi, di associazione o dissociazione di molecole.

Sostanzialmente differente è lo spettro delle nebulose, zone di gas diffuso e caldo nel mezzo interstellare, spesso trasparente in termini di profondità ottica. Queste presentano spettri non continui, ma consistenti di bande più o meno estese, in assenza di un vero e proprio continuo; il loro spettro non segue la distribuzione del corpo nero, per cui se ne conclude che si tratta di plasmi lontani dall'equilibrio termodinamico.

Oltre a queste osservazioni a larga banda della radiazione nel continuo delle stelle e delle nebulose diffuse, è possibile analizzarle disperdendone la radiazione in modo da permetterne un'analisi ad alta risoluzione in lunghezza d'onda; questa è la base della spettroscopia. Ciò consente di ricavare ulteriori caratteristiche sullo stato termodinamico degli atomi o molecole del plasma emettente, in particolare attraverso la rivelazione e lo studio delle righe di assorbimento ed emissione.

Queste righe come abbiamo visto nel capitolo precedente ci permettono di riconoscere un AGN e di individuarne la tipologia.

Vedremo che la spettroscopia fornisce informazioni su:

- energia specifica, stato di eccitazione degli atomi o molecole del plasma emettente
- composizione chimica del plasma emettente
- composizione chimica degli strati assorbenti
- dinamica globale e/o locale.

Lo sviluppo della spettroscopia nella seconda metà del 1800 coincide con la nascita dell'astrofisica, in quanto osservazioni spettroscopiche permisero di indagare la struttura fisica degli oggetti celesti e non solo i loro moti. Lo studio della distribuzione energetica (o fotoni) nelle varie bande di emissione è effettuata con metodi diversi a seconda delle frequenze.

Nell'ottico si ottiene la dispersione della radiazione propagandola attraverso un prisma rifrangente. Nel radio, giacché si lavora con antenne a banda molto stretta, occorre effettuare osservazioni cambiando i ricevitori. Alle alte frequenze si usano filtri alle "finestre" di ingresso della radiazione nei collimatori.

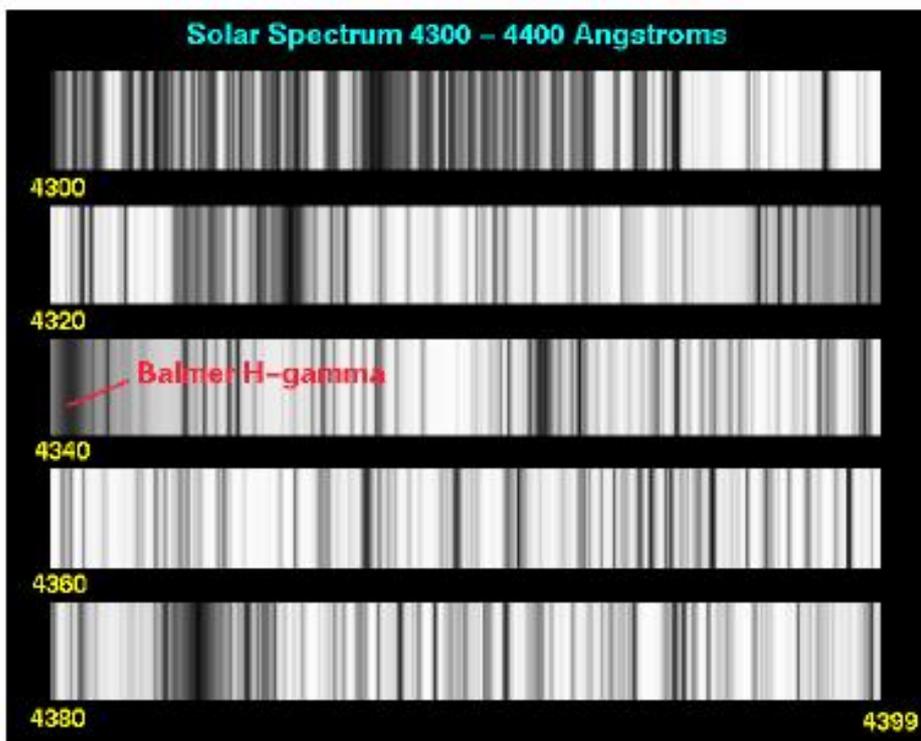


Figura 2.13: Una porzione dello spettro solare

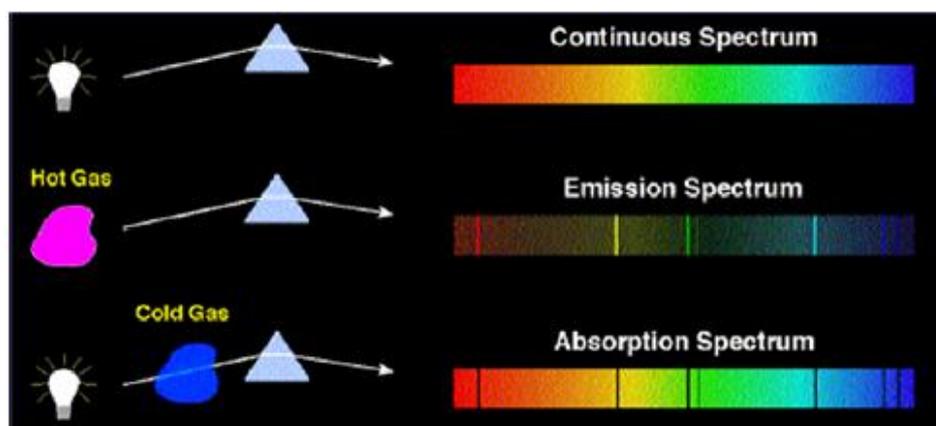


Figura 2.14: Le leggi di Kirchhoff sulla spettroscopia.

Le prime osservazioni di spettri di oggetti astrofisici ad alta risoluzione risalgono a Wollaston, che nel 1802 scoprì le righe di assorbimento nello spettro solare (vedi figura 2.13).

Successivamente Fraunhofer compì analisi sistematiche e nel 1814 propose una classificazione delle righe dello spettro solare (dette appunto righe di Fraunhofer).

2.10 Cenni sulla Teoria Atomica degli Spettri

Nel 1859 Kirchhoff formulò, sulla base di esperimenti in laboratorio, le seguenti leggi della spettroscopia (illustrate schematicamente in Figura 2.14).

1. Solidi e liquidi portati all'incandescenza emettono a tutte le frequenze, cioè hanno uno spettro di emissione continuo, il cui profilo dipende dalla temperatura.
2. Gas rarefatti incandescenti producono uno spettro di righe/bande di emissione, cioè emettono solo a certe frequenze, che risultano essere caratterizzate dalle proprietà fisiche e chimiche del gas.
3. Un gas rarefatto a bassa temperatura interposto tra una sorgente del primo tipo ad alta temperatura e l'osservatore, dà origine a righe di as-

sorbimento sul continuo alle stesse frequenze a cui il gas, se riscaldato, sarebbe capace di emettere.

Balmer nel 1885 trovò una relazione empirica tra le lunghezze d'onda emesse da un gas rarefatto di idrogeno:

$$\lambda_n = \lambda_0 \frac{n^2}{n^2 - 2^2} \quad \lambda_0 = 3646.5\text{\AA} \quad n = 3, 4, 5, \dots \quad (2.77)$$

da cui si individuano specifiche righe osservate: $\lambda_3 = 6563\text{\AA}$ (riga H_α), $\lambda_4 = 4861\text{\AA}$ (riga H_β), $\lambda_5 = 4340\text{\AA}$ (riga H_γ) e così via convergenti verso una lunghezza d'onda minima $\lambda_0 = 3646.5\text{\AA}$ chiamata testa della serie, come effettivamente osservato.

Rydberg trasformò questa relazione in frequenze estendendola ad altri elementi. In particolare i cosiddetti metalli alcalini della tavola periodica (Li, Na, K, Rb, Cs) mostravano più serie di righe, indicate con le lettere P (principal), S (sharp), D (diffuse), F (fundamental), caratterizzabili in modo simile alla serie di Balmer:

$$\nu_n^P = \nu_\infty^P - \frac{R}{(n + P)^2} \quad (2.78)$$

dove $R = 3.29 \cdot 10^{15} s^{-1}$ è la costante di Rydberg eguale per tutti gli elementi, mentre variano da elemento a elemento i valori delle altre costanti (P, S, D, F sono piccoli numeri $\lesssim 1$):

serie P	$\nu_n^P = 1S - nP$	$n = 2, 3, 4, \dots$
serie S	$\nu_n^S = 2P - nP$	$n = 2, 3, 4, \dots$
serie D	$\nu_n^D = 2P - nP$	$n = 3, 4, 5, \dots$
serie F	$\nu_n^F = 3D - nP$	$n = 4, 5, 6, \dots$

Regole simili vennero progressivamente via via elaborate per elementi con spettri sempre più complessi.

Fu la teoria quantistica dell'atomo di Bohr del 1913 che ne permise l'interpretazione fisica. L'atomo di Bohr è costituito da un nucleo elettricamente positivo intorno a cui ruotano elettroni su orbite di differente livello energetico, non distribuite in modo continuo, ma discrete o quantizzate; questi elettroni non emettono nel moto lungo le orbite (come predirebbe la

teoria elettromagnetica classica), ma soltanto quando compiono una transizione tra orbite quantizzate e quindi con emissione o assorbimento di fotoni di ben precisa frequenza $h\nu = \Delta E$ corrispondente al salto di energia ($h = 6.6260755 \cdot 10^{-34} J s$ è la costante di Planck).

Bohr mostrò che, quantizzando il momento angolare orbitale degli elettroni trattenuti in orbite circolari intorno al nucleo dalla forza elettrostatica

$$\frac{ma^2\dot{\varphi}^2}{a} = \frac{Ze^2}{a^2} \quad (2.79)$$

$$2\pi ma^2\dot{\varphi} = nh \quad n = 1, 2, 3, \dots \quad (2.80)$$

con a raggio dell'orbita e m massa dell'elettrone, le energie dei livelli discreti in funzione del numero n sono:

$$E_n = -\frac{me^4 Z^2}{2\hbar^2 n^2} eV \quad (2.81)$$

con Z numero atomico e $\hbar = h/2\pi = 1.05457266 \times 10^{-34} J s$.

Un calcolo più corretto che tenga conto della massa M finita del nucleo comporta la sostituzione della massa dell'elettrone con la sua massa ridotta $\mu = mM/(m + M)$. Nel caso dell'atomo di idrogeno:

$$E_n = -13.6 \frac{Z^2}{n^2} eV \quad (2.82)$$

dove per $n = 1$ si ha lo stato più legato o stato fondamentale: l'energia di ionizzazione di un atomo di idrogeno nello stato fondamentale è appunto 13.6 eV. L'intero n è chiamato numero quantico principale. Possiamo ora scrivere la frequenza del fotone emesso quando un elettrone compie una transizione da un livello n_m ad un livello n_n :

$$h\nu = \Delta E = -\frac{\mu e^4 Z^2}{2\hbar^2} \left(\frac{1}{n_m^2} - \frac{1}{n_n^2} \right) \quad (2.83)$$

che riproduce la formula di Rydberg, e dove in effetti la costante avanti la parentesi è proprio la costante di Rydberg espressa in frequenze:

$$R = \frac{\mu e^4}{4\pi\hbar^3} = 3.2898419499 \cdot 10^{15} Hz \quad (2.84)$$

L'equazione 2.83 rappresenta emissione di fotoni quando $n_m > n_n$ e viceversa assorbimento quando $n_m < n_n$, e i due processi corrispondono allo stesso salto di energia se i due numeri quantici sono gli stessi: un atomo ha

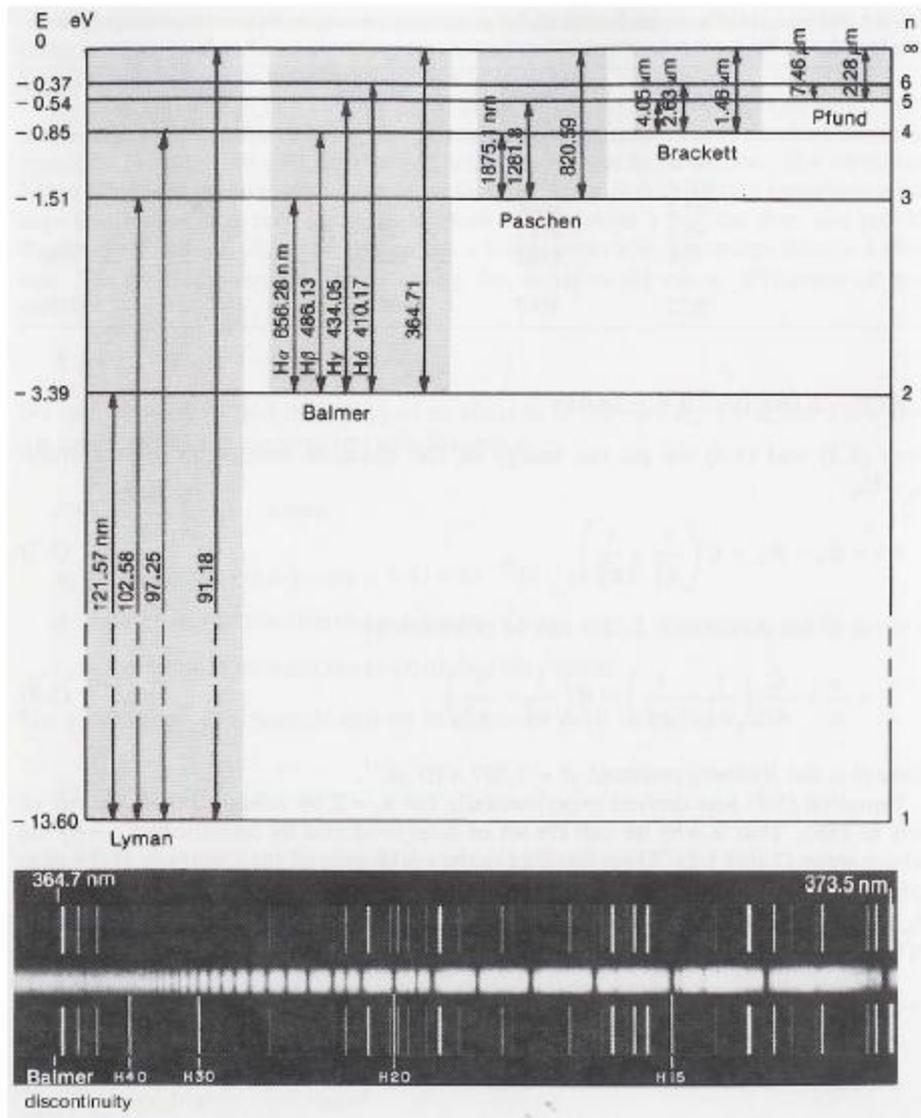


Figura 2.15: Diagramma di Grotrian per le serie delle righe dell'atomo di idrogeno e spettro; le lunghezze d'onda delle transizioni sono date in nanometri e le energie dei livelli in eV.

quindi le stesse righe in assorbimento e in emissione, il che corrisponde a quanto stabilito nelle leggi di Kirchhoff. Una rappresentazione grafica dei processi di emissione e assorbimento in relazione ai livelli energetici delle orbite elettroniche è data dai diagrammi di Grotrian: in Figura 2.15 è riportato il diagramma per l'atomo di idrogeno. Le linee orizzontali rappresentano i livelli energetici, i tratti verticali le possibili transizioni, raggruppate in serie secondo il numero quantico del livello inferiore n con l'altro numero quantico del livello energetico superiore $m = n + 1, n + 2, \dots, \infty$. Le serie sono indicate con il nome dello scienziato che le ha identificate sperimentalmente. Per l'idrogeno la serie che comprende le righe dello spettro visibile è la serie di Balmer. Al crescere di m il salto energetico aumenta e così pure la frequenza corrispondente; esiste un limite superiore per $m \rightarrow \infty$ a cui corrisponde una frequenza che rappresenta la testa della serie:

$$\nu_{\infty} = \frac{\mu e^4 Z^2}{4\pi\hbar^3} \frac{1}{n^2} \quad (2.85)$$

cioè per ogni serie non sono possibili emissioni o assorbimenti per frequenze maggiori tra stati legati dell'elettrone.

Per la serie di Balmer tale limite corrisponde ad una lunghezza d'onda $\lambda_d = 3647.1\text{\AA}$. Fotoni di lunghezza d'onda inferiore che incidano su un gas con popolazione dello stato $n = 2$, che corrisponde appunto ad assorbimenti nella serie di Balmer, sono in grado di fotoionizzare gli atomi: le transizioni atomiche sono del tipo legato-libero cui corrisponde uno spettro di assorbimento continuo in quanto i moti degli elettroni liberi non sono più quantizzati. Ciò corrisponde ad una discontinuità nello spettro con caduta della luminosità al di sotto di tale lunghezza d'onda. Proprio perché l'idrogeno è l'elemento più abbondante nel cosmo risulta per tale ragione molto difficile osservare la radiazione stellare nell'ultravioletto.

2.10.1 Spettri Atomici

Il modello di Bohr fu completato da Sommerfeld nel 1916 con l'introduzione di altri due numeri quantici: il numero quantico azimutale $l = 0, 1, \dots, n - 1$ che tiene conto di orbite elettroniche ellittiche, e il numero quantico magnetico $m = 0, \pm 1, \dots, l$ che tiene conto dell'orientamento tra un campo magnetico esterno e il campo magnetico generato dal moto dell'elettrone inteso come una spira di corrente. Questi numeri quantici

rappresentano una degenerazione (quantizzata) dei livelli corrispondenti ai semplici orbitali circolari. Il numero quantico magnetico corrisponde ad una degenerazione dei livelli in presenza di un campo magnetico esterno: si produce uno splitting delle linee che dà origine al cosiddetto effetto Zeeman con cui si può misurare l'intensità del campo magnetico in cui si trova il plasma emettente.

La teoria semi-classica dell'atomo di Bohr e Sommerfeld è stata poi riscritta all'interno della meccanica quantistica che interpreta la struttura dell'atomo attraverso l'equazione di Schrödinger per la funzione di probabilità degli stati, di cui i livelli energetici sono le autofunzioni. In tale teoria i numeri quantici e le loro relazioni compaiono automaticamente attraverso lo sviluppo della funzione di probabilità nel prodotto di funzioni delle variabili spaziali separate:

$$\psi(r, \theta, \varphi) = R_{nl}(r)\Theta(\theta)\Phi(\varphi) \quad (2.86)$$

La teoria quantistica richiede, tra l'altro, l'introduzione di un ulteriore numero quantico, il numero quantico di spin $s = \pm 1/2$ che tiene conto dell'allineamento tra spin dell'elettrone e momento orbitale (Uhlenbeck e Goudsmit 1925). Nel 1928 da Dirac fu in grado di mostrare come la versione relativistica dell'equazione di Schrödinger permette di spiegare coerentemente il modello atomico a quattro numeri quantici.

La teoria quantistica adotta un ulteriore principio non classico, e cioè che ogni stato energetico, caratterizzato da specifici valori dei quattro numeri quantici (n, l, m, s) , può essere occupato da un solo elettrone: il principio di esclusione di Pauli. A partire dall'atomo di idrogeno con un solo elettrone, si passa agli atomi di numero atomico sempre maggiore aggiungendo nuovi elettroni su livelli energetici sempre più esterni, meno legati: è questa la base del modello a shell degli atomi. I livelli o orbitali sono progressivamente indicati, a partire dai livelli più legati, con le lettere K, L, M, N , etc. corrispondenti ai numeri quantici principali. Quando una shell di dato n viene completamente occupata nelle sue sotto-shell definite dagli altri numeri quantici (esempi sono l'He, F, Ar, Kr, Xe) la struttura atomica corrispondente risulta la più stabile, e ciò rende conto di molte proprietà fisiche e chimiche della tavola degli elementi di Mendeleev. In questi atomi complessi si usano i numeri quantici dati dalla somma, secondo opportune

modalità, di quelli degli elettroni nei livelli più esterni (o di valenza): mentre il numero quantico principale n rimane tale, gli altri numeri quantici totali vengono indicati con lettere maiuscole, L, S, J

Si deve anche notare che un atomo eccitato non può diseccitarsi seguendo indifferentemente una qualunque delle transizioni possibili. Intervengono cioè delle regole di selezione che limitano le combinazioni permesse tra livelli energetici: $\Delta S = 0, \Delta L = 0, \pm 1, \Delta J = 0, \pm 1$, anche se non è possibile la transizione $J = 0 \rightarrow J = 0$. Le transizioni che non soddisfano queste regole sono dette proibite: in realtà queste transizioni sono soltanto poco probabili, nel senso della meccanica quantistica, e sono difficilmente osservabili in esperimenti di laboratorio. Tuttavia righe proibite compaiono negli spettri astrofisici, ad esempio quando il plasma emettente è molto rarefatto; un caso ben noto è quello della transizione dell'idrogeno neutro da spin dell'elettrone da parallelo ad antiparallelo a quello del nucleo, che corrisponde alla riga 21 cm osservata nel radio dal gas interstellare. In questa transizione $\Delta S \neq 0$, ma gli atomi di idrogeno nel gas interstellare sono così numerosi che, pur essendo la probabilità di transizione $\approx 10^{-15} \text{ s}^{-1}$, il processo risulta abbastanza frequente da essere osservabile. Le lunghezze d'onda delle righe dell'idrogeno sono riportate nel diagramma di Grotrian di Figura 2.15

2.10.2 Spettri Molecolari

Nell'astronomia infrarossa e millimetrica è possibile osservare righe emesse da transizioni molecolari, sia di molecole inorganiche ($\text{H}_2, \text{C}_2, \text{FeO}, \text{H}_2\text{O}, \text{CO}, \text{CO}_2$, ecc.) sia organiche sia radicali molecolari ($\text{OH}, \text{CN}, \text{CH}$); il loro interesse sta nella possibilità che offrono di studiare da un lato le componenti fredde dei mezzi diffusi e dall'altro le condizioni di sviluppo di forme di vita nel cosmo. Come nel caso della spettroscopia atomica, anche la spettroscopia molecolare ha rivelato l'esistenza di molecole non conosciute in laboratorio che suggeriscono l'esistenza di processi fisico-chimici peculiari possibili in condizioni di difficile riproduzione sperimentale.

Riportiamo in Figura 2.16 le molecole più importanti dal punto di vista astrofisico e i loro potenziali di dissociazione e ionizzazione.

Gli spettri molecolari presentano un aspetto molto diverso da quelli atomici, in quanto sono essenzialmente bande composte da molte righe. Ciò è dovuto al fatto che le forze elettrostatiche non sono più centrate sul nucleo atomico, ma diventa importante il moto dei nuclei atomici costituenti le

Nome	Potenziale di dissociazione	Potenziale di ionizzazione
H ₂	4.48 eV	15.43 eV
C ₂	6.2	12.0
CH	3.47	10.6
CO	11.09	14.01
CN	7.8	14
O ₂	5.12	12.08
OH	4.39	13.36
MgH	2.3	
CaH	1.5	
TiO	6.8	
FeO	4.4	
H ₂ O	5.11	12.61
N ₂ O	1.68	12.89
CO ₂	5.45	13.77
NH ₃	4.3	10.15
CH ₄	4.4	13.0
HCN	5.6	13.91

Figura 2.16: Molecole di interesse astrofisico

molecole: questi possono oscillare e ruotare rispetto al baricentro. Pertanto i livelli energetici hanno una degenerazione molto più ricca che nel caso atomico.

2.10.3 Misure Spettrali

Importanti informazioni fisiche sullo stato dei plasmi emettenti vengono ricavate dall'esame del profilo delle righe. Si parte dalla fotografia di uno spettro e la si converte in un diagramma di intensità che mostra la densità di flusso in funzione della lunghezza d'onda. Ciò si ottiene utilizzando un microdensitometro che misura l'annerimento delle lastre fotografiche; poiché però l'annerimento non è funzione lineare della radiazione ricevuta occorre operare opportune calibrizioni su lastre campionate su radiazione di intensità nota. Le tecniche digitali più recenti operano la misura del flusso nelle righe direttamente sulla radiazione incidente senza passare attraverso la lastra fotografica. Successivamente lo spettro viene rettificato e normalizzato rispetto al continuo:

$$F_j(\lambda) = F_C(\lambda) - F(\lambda) \quad (2.87)$$

La figura 2.17 riporta la sequenza di queste operazioni per lo spettro di una stella. Le righe di assorbimento appaiono come cadute dell'intensità di varie ampiezze: esistono righe molto chiaramente identificabili, altre righe sono più deboli e la loro identificazione richiede un'analisi del rumore per escludere che si tratti di fluttuazioni dovute alla granularità dell'emulsione. A volte righe molto vicine si possono parzialmente sovrapporre (blending); in tal caso per distinguerle occorre aumentare la dispersione.

Il profilo delle righe dipende dalle proprietà fisiche del plasma emettente e in parte da un allargamento che avviene nello strumento di misura. La quantità $[F_C(\lambda) - F(\lambda)]/F_C(\lambda)$ è chiamata profondità della linea. Una quantità che dipende poco dai dettagli del profilo è la larghezza equivalente, cioè la larghezza di un rettangolo la cui area dia lo stesso assorbimento della riga (figura 2.18

$$W_\lambda = \int_{\lambda_1}^{\lambda_2} \frac{F_C(\lambda) - F(\lambda)}{F_C(\lambda)} d\lambda \quad (2.88)$$

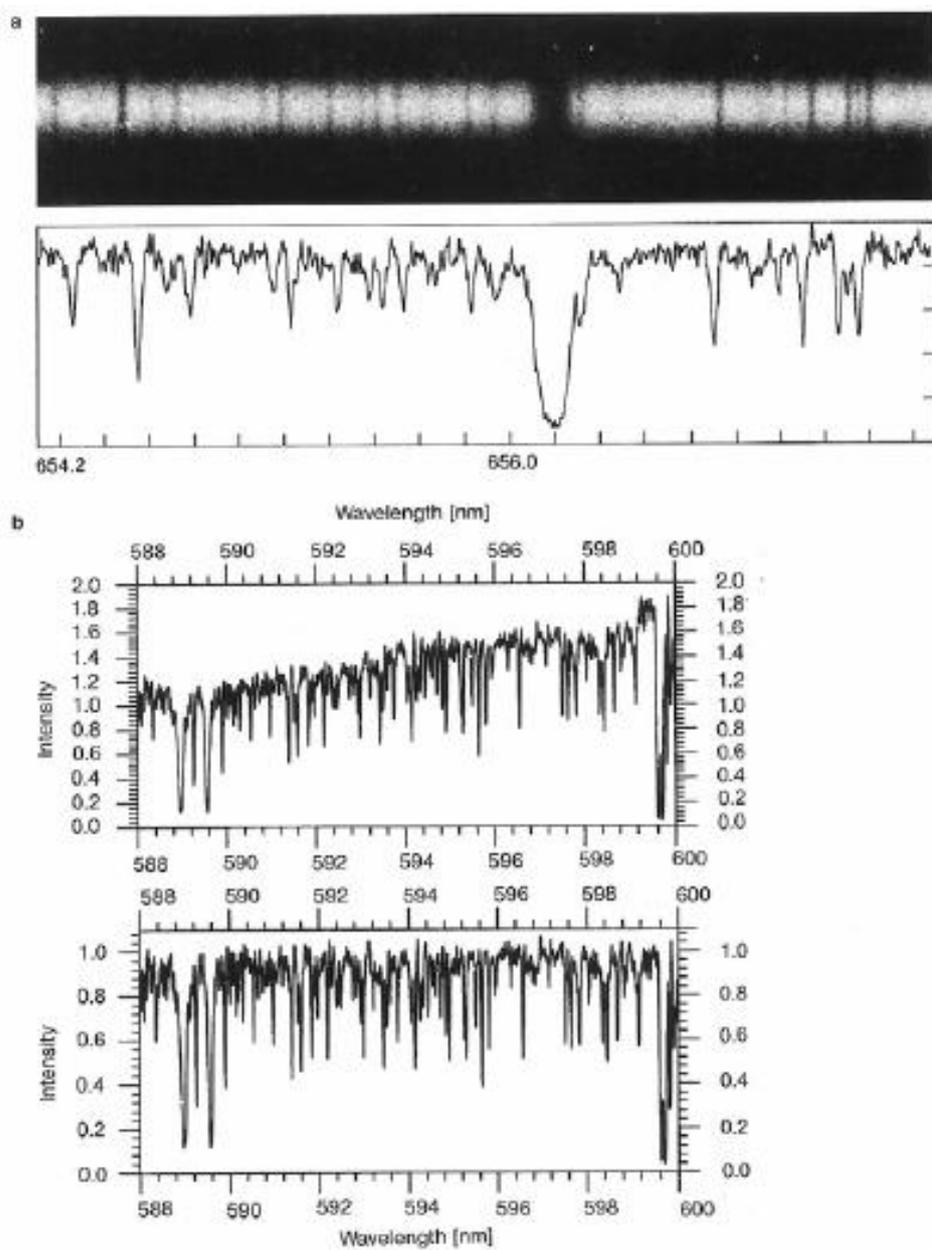


Figura 2.17: Riduzione di uno spettro e normalizzazione al continuo

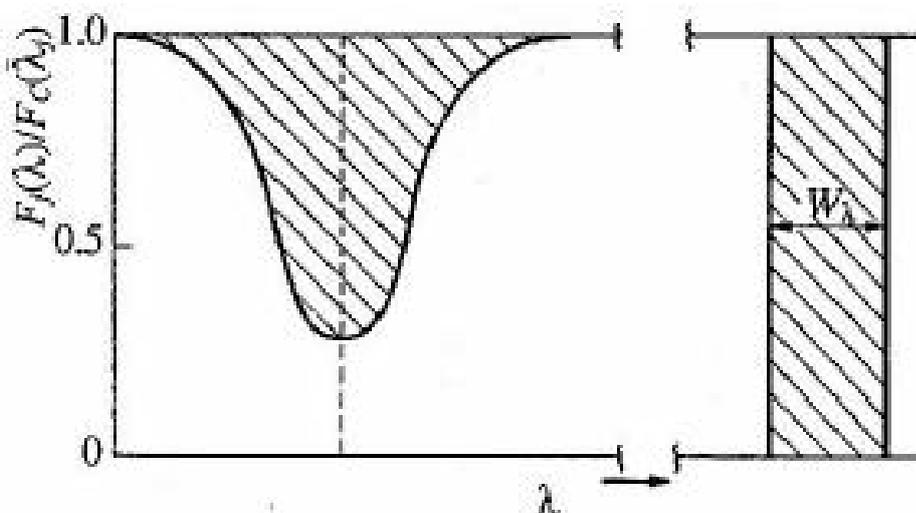


Figura 2.18: Larghezza equivalente delle righe spettrali

L'integrale è esteso all'intervallo $[\lambda_1, \lambda_2]$ di interesse e W_λ è appunto misurato in \AA .

La larghezza dipende dal numero di atomi che sono grado di assorbire la lunghezza d'onda considerata: maggiore il numero di atomi, maggiore la larghezza equivalente della riga e maggiore anche la sua forza o intensità.

Infine l'ampiezza o forza delle linee spettrali è legata all'abbondanza degli atomi responsabili dell'assorbimento e quindi essa consente di determinare la composizione chimica del plasma emettente. Con le equazioni del trasporto per le righe si possono costruire spettri sintetici al computer da confrontare con i dati osservativi.

2.10.4 Formazione delle Righe

In quali condizioni fisiche si formano le righe di assorbimento? Per rispondere a questa domanda occorre valutare in quali orbitali è più probabile trovare gli elettroni e risolve, quindi, un problema di meccanica statistica. In un gas gli atomi guadagnano e perdono energia per collisione; in equilibrio termodinamico, la distribuzione di velocità che agiscono nelle collisioni segue la legge di Maxwell-Boltzmann

$$n(v)dv = n \left(\frac{m}{2\pi kT} \right)^{3/2} e^{-mv^2/2kT} 4\pi v^2 dv \quad (2.89)$$

Gli orbitali di energia più elevata avranno quindi una probabilità minore di essere occupati, perché meno frequenti saranno le collisioni adatte. Supponiamo che s_j sia l'insieme dei numeri quantici che definiscono un livello energetico E_j .

Il rapporto di probabilità di occupazione di due stati s_a e s_b è calcolato dalla meccanica statistica (legge di Boltzmann):

$$\frac{P(s_b)}{P(s_a)} = \frac{e^{-E_b/kT}}{e^{-E_a/kT}} = e^{-(E_b-E_a)/kT} \quad (2.90)$$

dove T è la temperature del sistema. Il fattore esponenziale è detto fattore di Boltzmann. Gli stati energetici possono essere degeneri, con più stati quantici corrispondenti alla stessa energia, e di ciò si tiene conto per mezzo di un fattore chiamato molteplicità o peso statistico. In tal caso il precedente rapporto di probabilità è:

$$\frac{P(s_b)}{P(s_a)} = \frac{g_b e^{-E_b/kT}}{g_a e^{-E_a/kT}} = \frac{g_b}{g_a} e^{-(E_b-E_a)/kT} \quad (2.91)$$

In un sistema di molti atomi il rapporto di probabilità è proporzionale al rapporto dei numeri di atomi negli stati a e b , e quindi dà l'occupazione dei differenti stati eccitati:

$$\frac{N_b}{N_a} = \frac{g_b}{g_a} e^{-(E_b-E_a)/kT} \quad (2.92)$$

Con riferimento ad esempio al caso dell'idrogeno neutro con le energie dei livelli più legati $E_1 = -13.6eV$ ($n = 1$, stato fondamentale, $g_1 = 2$) ed $E_2 = -3.4eV$ ($n = 2$, stato eccitato, $g_2 = 2 \cdot 2^2$) si ricava dall'equazione 2.92 che per avere una buona popolazione nello stato eccitato ($N_1 = N_2$) occorre una temperatura $T = 8.54 \cdot 10^4 K$. Le righe di assorbimento della serie di Balmer sono prodotte appunto da atomi che compiono transizioni a partire dallo stato $n = 2$. Nel prossimo paragrafo vedremo che la temperatura delle stelle che presentano le più intense righe di Balmer è più bassa degli $85.400K$ di un fattore circa 10: anzi le stelle di temperatura superiore ai $10.000K$ non hanno righe dell'idrogeno. Analizziamo brevemente questa apparente incongruenza.

Il fattore di Boltzmann serve anche ad esprimere i numeri relativi di atomi che si trovano in differenti stati eccitati. Tuttavia dobbiamo anche considerare il fatto che per energie di eccitazione superiori al potenziale di ionizzazione gli atomi siano in gran parte ionizzati.

Sia χ_i l'energia di ionizzazione dallo stato fondamentale che porta un atomo a più elettroni dallo stato di ionizzazione i a $i + 1$; naturalmente la ionizzazione può avvenire da uno stato eccitato con minor energia di ionizzazione. Occorre pertanto fare una media sui vari orbitali per tener conto della ripartizione degli elettroni sui possibili stati: questa procedura richiede di calcolare la funzione di partizione Z per gli stati iniziale e finale. La Z è semplicemente la somma pesata del numero di modi in cui un atomo può distribuire i propri elettroni a parità di energia totale. La somma pesata usa il fattore di Boltzmann per indicare che stati più eccitati sono meno probabili, cioè pesano meno. La funzione di partizione diventa:

$$Z = g_1 + \sum_{j=2}^{\infty} g_j e^{-(E_j - E_1)/kT} \quad (2.93)$$

Usando le partizioni per gli stati di ionizzazione Z_i e Z_{i+1} e il potenziale di ionizzazione χ_i dello stato i , si dimostra che vale l'equazione di Saha:

$$\frac{N_{i+1}}{N_i} = \frac{2Z_{i+1}}{n_e Z_i} \left(\frac{2\pi m_e kT}{h^2} \right) e^{-\chi_i/kT} \quad (2.94)$$

dove il fattore 2 viene dalle due possibilità di spin $\pm 1/2$ e la presenza della densità elettronica deriva dal fatto che la ionizzazione produce elettroni liberi; in particolare al crescere di tale densità il rapporto decresce perché sono probabili anche processi di ricombinazione. Utilizzando questa formula si può comprendere perché le righe di Balmer sono intense a temperature più basse di quelle indicate dalla legge di Boltzmann. Infatti l'intensità della riga dipende dal numero totale di atomi N_2/N_{tot} che si trovano nello stato di eccitazione $n = 2$ tenendo conto di quanti siano ionizzati. Questo rapporto si calcola ricordando che per l'idrogeno $N_1 + N_2 \approx N_I$ (atomi non ionizzati) e $N_I + N_{II} = N_{tot}$:

$$\frac{N_2}{N_{tot}} = \left(\frac{N_2/N_1}{1 + N_2/N_1} \right) \left(\frac{1}{1 + N_{II}/N_I} \right) \quad (2.95)$$

che, sostituendo le formule di Saha per N_{II}/N_I e di Boltzmann per

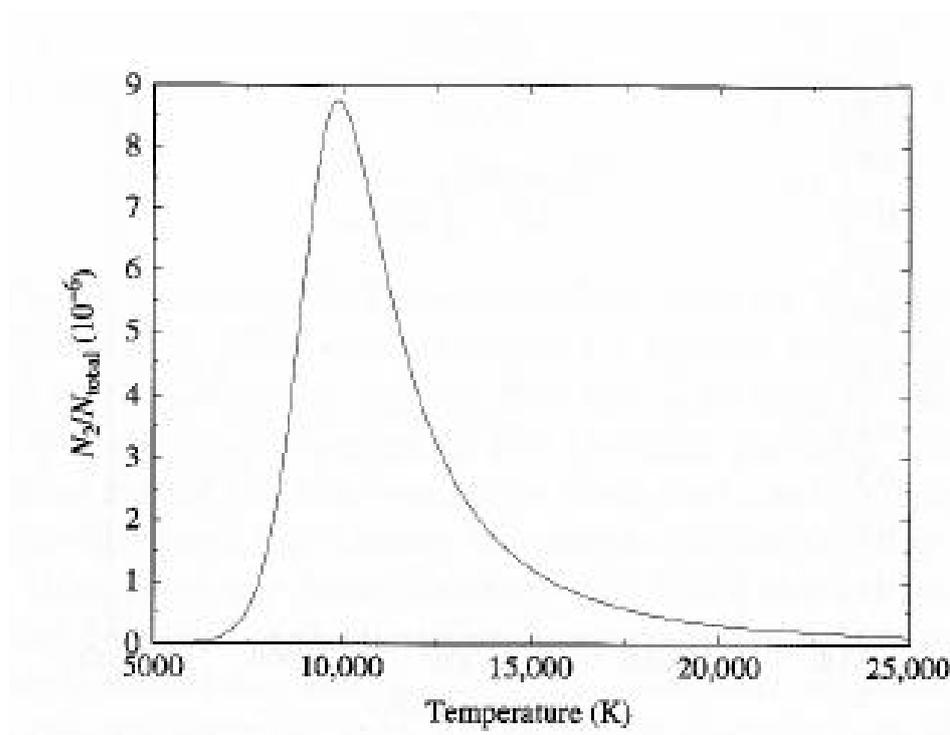


Figura 2.19: Formazione delle righe di Balmer

N_2/N_1 , è rappresentata in figura 2.19 e mostra un picco intorno ai $9.000K$. Il diminuire dell'intensità delle righe di Balmer a temperature oltre i $10.000K$ è dovuto al fatto che l'idrogeno diventa completamente ionizzato.

In conclusione possiamo dire che per formare una riga di assorbimento a una data frequenza occorre che la temperatura sia tale da permettere, secondo l'equazione 2.92, una buona popolazione nello stato (fondamentale o eccitato) da cui il salto energetico può assorbire i fotoni. Allo stesso tempo occorre valutare dall'equazione 2.94, che lo stato non sia ionizzato.

2.10.5 Allargamento delle Righe

La riga rappresentata schematicamente in Figura 2.18 è detta otticamente sottile, perché non esiste nessuna lunghezza d'onda alla quale la radiazione venga completamente bloccata. L'assorbimento è evidentemente più forte alla lunghezza d'onda corrispondente al centro della riga e diminuisce spostandosi alle ali: nel caso delle stelle ciò comporta che l'assorbimento più forte avvenga nelle regioni più esterne e fredde dove il livello energetico che viene eccitato nell'assorbimento è molto popolato, mentre le ali si formano nelle regioni più profonde e più calde dove il livello energetico è già eccitato collisionalmente.

La larghezza delle righe spettrali dipende da vari effetti fisici, ciascuno dei quali genera un particolare tipo di profilo.

1. **Allargamento Naturale.** In base al principio di indeterminazione di Heisenberg una riga non può essere di larghezza nulla neppure se tutti gli atomi che la producono fossero fermi. Poiché un elettrone occupa uno stato eccitato per un tempo Δt , la sua energia è definita entro un intervallo

$$\Delta E \approx \frac{\hbar}{\Delta t} \quad (2.96)$$

e quindi l'energia del fotone emesso nella diseccitazione $E_{fotone} = hc/\lambda$ avrà un'indeterminazione

$$\Delta \lambda \approx \frac{\lambda^2}{2\pi c} \frac{1}{\Delta t} \quad (2.97)$$

che corrisponde all'ordine di grandezza dell'allargamento della riga; nel caso dell' H_α dell'idrogeno, $\lambda = 6563\text{\AA}$ e $\Delta t \approx 10^{-8}s$, si deriva

$\Delta\lambda \approx 3.57 \cdot 10^{-4} \text{ \AA}$, che in effetti rappresenta un valore tipico per tutte le righe.

2. **Allargamento Doppler.** In equilibrio termodinamico gli atomi di un gas si muovono in modo casuale con una distribuzione di velocità maxwelliana, e la velocità più probabile (termica) è $v_p = \sqrt{2kT/m}$. Le lunghezze d'onda assorbite o emesse dai singoli atomi sono quindi spostate per effetto Doppler secondo la relazione $\Delta\lambda/\lambda = \pm|v_r|/c$, per cui le righe prodotte dall'insieme degli atomi sono allargate di

$$\Delta\lambda \approx \frac{2\lambda}{c} \sqrt{\frac{2kt}{m}} \quad (2.98)$$

con una decrescita esponenziale dal centro della riga. Analogamente si può avere un allargamento Doppler per effetto di moti turbolenti del gas; la formula è la stessa, è sufficiente sostituire la velocità media dei moti turbolenti alla velocità termica.

3. **Allargamento Collisionale.** Gli orbitali degli atomi possono essere perturbati per collisioni o per incontri a breve range con i campi elettrici generati da ioni. L'effetto risultante dipende in sostanza dal libero cammino medio tra collisioni o incontri:

$$\Delta t_0 \approx \frac{l}{v} = \frac{1}{n\sigma\sqrt{2kT/m}} \quad (2.99)$$

dove σ è la sezione d'urto delle interazioni. La teoria dell'allargamento collisionale è piuttosto complicata, ma si ottiene un risultato molto simile a quello dell'allargamento naturale, per cui

$$\Delta\lambda = \frac{\lambda^2}{\pi c} \frac{1}{\Delta t_0} \approx \frac{\lambda^2}{\pi c} n\sigma \sqrt{\frac{2kT}{m}} \quad (2.100)$$

dove si nota l'importanza della densità del gas. Spesso ci si riferisce al profilo combinato di questo allargamento e di quello naturale con il nome di profilo di smorzamento.

La combinazione degli allargamenti ora discussi dà origine al cosiddetto profilo di Voigt. Come mostrato in figura 2.20 nelle regioni centrali

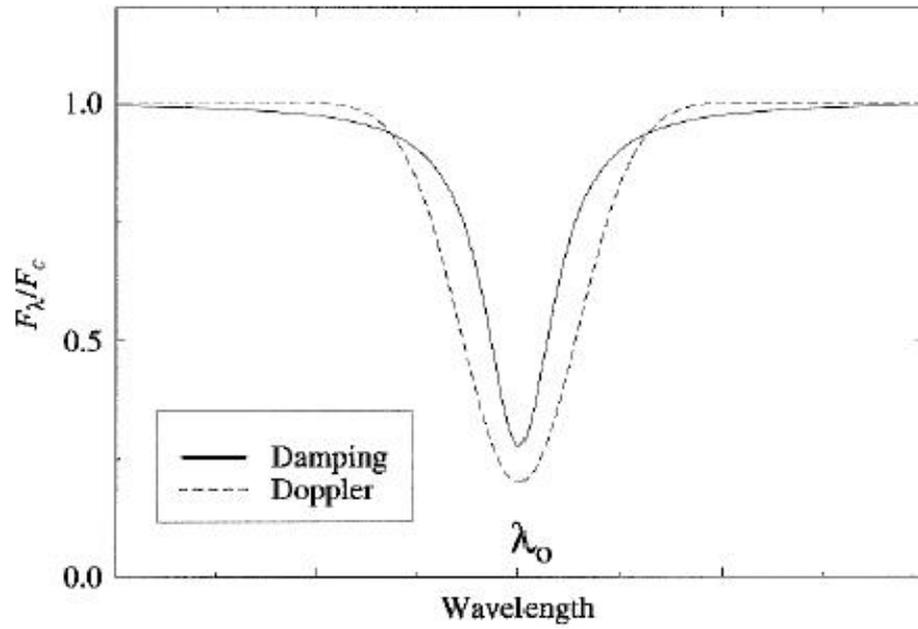


Figura 2.20: Profilo di Voigt

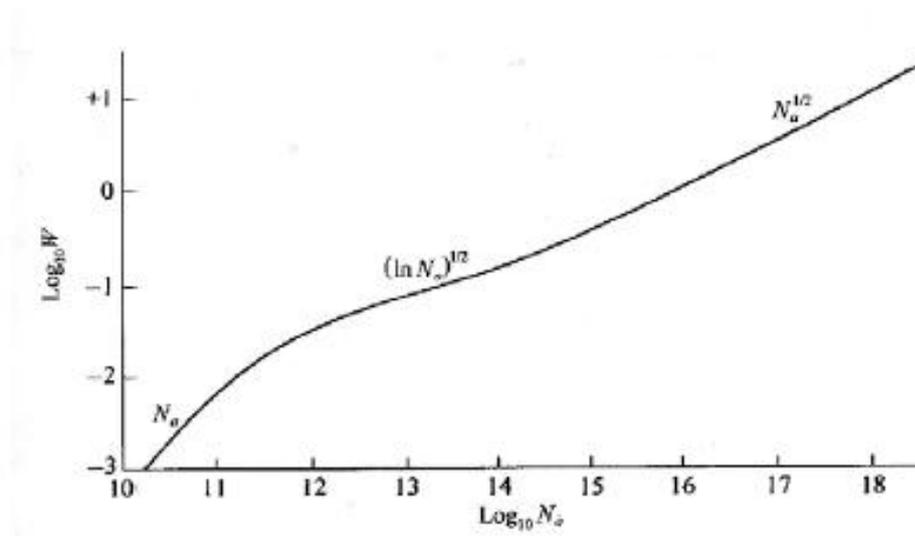


Figura 2.21: Curva di crescita

della riga domina l'allargamento Doppler (nucleo Doppler), mentre lo smorzamento prevale nelle ali (ali di smorzamento). La teoria delle atmosfere stellari è basata sulla costruzione di modelli per ottenere il profilo delle righe teorico da confrontare con quello osservato. Il modello di profilo teorico dipende dalla temperatura, densità e composizione chimica del mezzo dove l'assorbimento ha luogo ed elabora la fisica del processo con i metodi della meccanica statistica quantistica, cioè le equazioni di Boltzmann e Saha. Il risultato è riassunto nella curva di crescita che determina il valore della larghezza equivalente in funzione della densità di atomi assorbenti; studiando più righe non solo si ha una miglior definizione dei parametri fisici, ma si ottiene la composizione chimica del mezzo. Il caso della curva di crescita per la riga K del Ca II è riportato in figura 2.21.

2.10.6 Effetto Doppler

Quando il plasma sia dotato di un moto d'insieme con una velocità radiale relativa all'osservatore v_r le sue linee spettrali sono spostate rispetto alla configurazione di laboratorio per effetto Doppler secondo la già citata formula (nonrelativistica)

$$\frac{\Delta\lambda}{\lambda} = \frac{v_r}{c} \quad (2.101)$$

e lo spostamento è verso le lunghezze d'onda maggiori per moto relativo di allontanamento (redshift) e verso le minori per avvicinamento (blueshift). Lo spettro nel suo insieme risulta quindi deformato, essendo gli spostamenti maggiori per le lunghezze d'onda maggiori. L'osservazione degli spostamenti Doppler è un potente mezzo in astrofisica per lo studio della dinamica dei sistemi in quanto permette una misura diretta di velocità radiali, indipendente dalla distanza, che invece è necessaria per valutare velocità trasverse alla linea di vista con l'osservazione dei moti propri. Con l'effetto Doppler si misurano le velocità di espulsione dei getti nelle regioni di formazione stellare, le velocità del gas in accrescimento su corpi compatti, la dinamica delle stelle doppie, fino alla recessione delle galassie.

Sono interessanti dal punto di vista astrofisico i profili che le righe stesse possono assumere quando le osservazioni combinino i contributi da parti delle sorgenti che abbiano diversa dinamica. In particolare vale la pena citare i cosiddetti profili P Cygni che prendono il nome dalla stella in cui

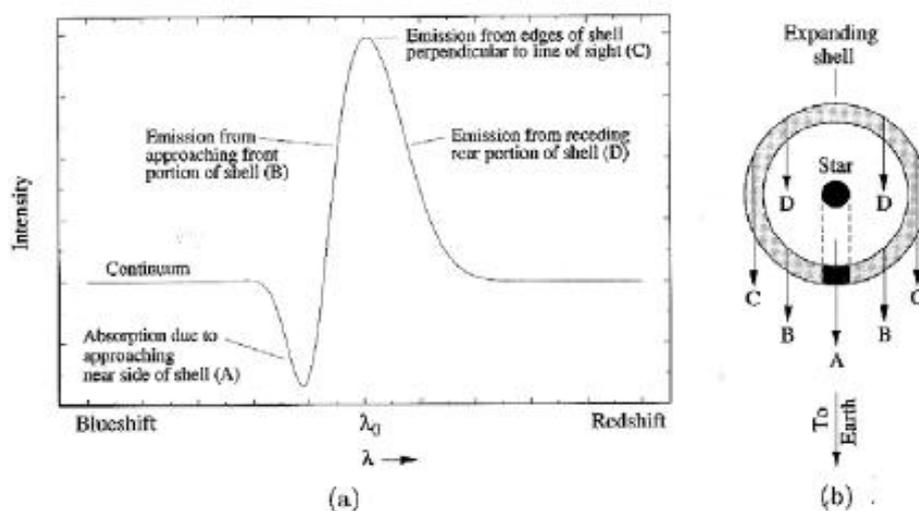


Figura 2.22: Profilo P Cygni: (a) righe osservate, (b) schema della dinamica della perdita di massa

furono osservati per la prima volta. Si tratta di profili che presentano larghe righe di emissione con componenti in assorbimento spostate verso il blu (figura 2.22).

L'origine di questi profili è dovuta alla presenza di forti perdite di massa da questi oggetti. Dalle leggi di Kirchhoff sappiamo che le righe di emissione provengono da gas caldo e rarefatto. Pertanto, come indicato in figura, la massa di gas otticamente sottile espulsa in direzione perpendicolare alla linea di vista dà un contributo in emissione allo spettro continuo dell'oggetto. La linea di assorbimento invece è dovuta al gas che si muove verso l'osservatore lungo la linea di vista e che assorbe la radiazione continua della superficie stellare; perciò il gas assorbente produce una riga di assorbimento spostata verso il blu.

2.11 Spettri Nebulari

Nel 1904 Hartmann osservò che gli spostamenti Doppler di alcune righe di assorbimento negli spettri delle stelle binarie corrispondevano a moti differenti da quelli orbitali. Fu quella la prima indicazione che tra le stelle e la Terra esistesse un gas capace di assorbimento. Le righe più forti nel visibile

sono quella del Na I e del Ca II, nell'ultravioletto quella dell'H I Lyman α a 1216 Å.

Sono anche presenti righe di atomi ionizzati, il che indica processi di ionizzazione da parte della luce ultravioletta stellare e dei raggi cosmici cui non corrispondono rapide ricombinazioni a causa della bassa densità del mezzo.

Le osservazioni delle righe in assorbimento nell'ultravioletto sono un potente mezzo per lo studio dell'idrogeno neutro interstellare; la Ly α negli spettri stellari è prodotta da atomi dell'H I interstellare che compiono un assorbimento dallo stato $n = 1$ a $n = 2$; ciò significa che l'idrogeno neutro nel mezzo interstellare si trova principalmente nello stato fondamentale.

In questo stato gli spin dell'elettrone e del nucleo (protone) dell'idrogeno possono essere paralleli o antiparalleli (struttura iperfine dei livelli atomici), e la differenza energetica tra i due livelli è $\Delta E = 5.87 \cdot 10^{-6} eV$, per cui la transizione dal livello più energetico (spin paralleli) a quello meno energetico (spin antiparalleli) comporta l'emissione di fotoni alla frequenza $\nu = 1420.4 MHz$ ovvero $\lambda = 21 cm$, quindi nella banda radio.

Nel livello iperfine più energetico a causa di processi collisionali e quindi potesse dare origine a righe di emissione fu proposta nel 1944 da van de Hulst che calcolò come le collisioni siano rare (tempo tipico 400 anni) ma molto minori del tempo scala radiativo (11 milioni di anni): osservazioni di Ewen e Purcell nel 1951 confermarono la predizione e aprirono la strada alla possibilità di studiare con la riga 21 cm la struttura e la dinamica (spostamenti Doppler) del mezzo interstellare nella Galassia e nelle galassie esterne.

La riga a 21cm può anche comparire in assorbimento quando la radiazione proveniente da una galassia lontana attraversa nuvole di idrogeno; in effetti la stessa nuvola può essere osservata sia in emissione sia in assorbimento.

Nelle vicinanze di stelle molto calde (stelle O in particolare) l'idrogeno del mezzo interstellare viene fotoionizzato dai fotoni ultravioletti: la ricombinazione porta quindi a righe di emissione dell'H II, e le nuvole che subiscono questo processo sono appunto dette regioni H II. Nell'ottico sono forti le righe della serie di Balmer, in particolare l'H α .

Il numero di processi di ricombinazione è proporzionale al prodotto delle densità di elettroni liberi e atomi ionizzati $n_{ric} \propto n_i n_e$, e nell'ipotesi che la

ionizzazione sia completa $n_{ric} \propto n_e^2$, per cui la brillantezza superficiale di una nuvola permette di valutare la misura di emissione:

$$EM = \int n_e^2 dl \quad (2.102)$$

cioè la densità di colonna degli elettroni lungo la linea di vista.

Le nuvole del mezzo interstellare producono righe molecolari in assorbimento sugli spettri stellari nelle regioni dell'ottico e dell'ultravioletto. Si tratta di righe di molecole diatomiche (CH, CH⁺, CN, H₂, CO) che sono state scoperte dal 1937. Dagli anni 1960 molte altre righe, anche di molecole più complesse (H₂CO, H₂O, NH₃, C₂H₅OH), sono state osservate nella banda radio sia in assorbimento sia in emissione. Le transizioni che originano queste righe sono

- transizioni elettroniche, cambi nella nuvola elettronica che lega la molecola; sono osservabili in ottico e ultravioletto
- transizioni vibrazionali, dovute a cambi nelle proprietà vibrazionali delle molecole; corrispondono a righe nell'infrarosso
- transizioni rotazionali, dovute a cambi nelle proprietà rotazionali delle molecole; corrispondono a righe nel radio.

Alcune nuvole danno origine a compatte sorgenti maser, in cui le righe di emissione della banda radio dell'OH e dell'H₂O appaiono amplificate milioni di volte grazie al processo di emissione stimolata durante la propagazione. Tali molecole vengono portate in uno stato eccitato metastabile da una sorgente esterna, nel caso considerato da una stella; quando vengono investite da una radiazione con energia pari alla differenza tra stato metastabile e stato fondamentale si diseccitano con la stessa fase e direzione di emissione e danno origine a pacchetti coerenti di radiazione.

2.12 AGN e Rapporti tra Righe Spettrali

La selezione tra AGN e Galassie viene usualmente effettuata basandosi su dei rapporti tra particolari righe dello spettro, questi rapporti hanno ovviamente un significato fisico che ora provvedo a illustrare: $[OIII]/H_\beta$ è un indicatore del livello medio di ionizzazione e temperatura $[NII]/H_\alpha$,

$[OI]/H_\alpha$ e $[SII]/H_\alpha$ sono indicatori dell'importanza relativa delle regioni parzialmente ionizzate, regioni che vengono prodotte da fotoionizzazione di alta energia [25].

La separazione tra AGN e Galassie Normali si basa fondamentalmente su due modelli, empirici, dovuti il primo a Kauffman [26] ed il secondo a Kewley [29] che hanno individuato dei criteri atti a dividere lo spazio dei parametri in tre zone, una che identifica con certezza le galassie normali, una seconda che identifica con certezza gli AGN ed una terza dove i due gruppi si sovrappongono.

Il modello di Kewley identifica come AGN ogni oggetto che si trovi al di sopra di una delle curve:

$$\log \frac{[OIII]\lambda 5007}{H_\beta} = \frac{0.61}{\log \frac{[NII]\lambda 6583}{H_\alpha} - 0.47} + 1.19 \quad (2.103)$$

$$\log \frac{[OIII]\lambda 5007}{H_\beta} = \frac{0.72}{\log \frac{[SII]\lambda\lambda 6717, 6731}{H_\alpha} - 0.32} + 1.30 \quad (2.104)$$

$$\log \frac{[OIII]\lambda 5007}{H_\beta} = \frac{0.73}{\log \frac{[OI]\lambda 6300}{H_\alpha} - 0.59} + 1.33 \quad (2.105)$$

Mentre il modello di Kauffman prevede che un oggetto sia sicuramente una galassia normale se si trova al di sotto della seguente curva

$$\log \frac{[OIII]\lambda 5007}{H_\beta} = \frac{0.61}{\log \frac{[NII]\lambda 6583}{H_\alpha} - 0.05} + 1.3 \quad (2.106)$$

A Heckman si deve invece la separazione tra gli oggetti che lui stesso ha chiamato LINER¹ e le Seyfert.

L'equazione con cui si separano è la seguente [32]:

$$\frac{[OIII]\lambda 5007}{H_\beta} = 2.1445 \frac{[NII]\lambda 6583}{H_\alpha} + 0.465 \quad (2.107)$$

Ciò che si trova al di sopra di questa retta è considerato essere una Seyfert, mentre ciò che è un LINER [33, 32].

Inoltre si può distinguere tra Seyfert 1 e 2 tramite il seguente criterio [25]:

Sono definiti Seyfert 1 gli oggetti per cui risulta:

¹Low-Ionization Narrow Emission-Line Regions

$$FWHM(H_\alpha) > 1.5FWHM([OIII]\lambda 5007) \quad (2.108)$$

oppure

$$FWHM(H_\alpha) > 1200Km s^{-1} \quad (2.109)$$

e contestualmente

$$FWHM([OIII]\lambda 5007) < 800Km s^{-1} \quad (2.110)$$

Tutti gli altri vengono classificati come Seyfert 2

Questo insieme di informazioni è stato utilizzato, in maniera diretta o indiretta come vedremo meglio nel capitolo 4, per costruire la nostra base di conoscenza.

2.13 Le Survey

Astronomia, astrofisica e cosmologia hanno c'è una profonda differenza metodologica rispetto alla maggior parte delle branche della fisica: non si può in un laboratorio decidere arbitrariamente di costituire un determinato sistema e di farlo evolvere sotto certe condizioni, questo sia per ovvi problemi di masse e energie in gioco sia per un evidente problema di tempi, che superano non solo la scala di una vita umana, ma addirittura i tempi della storia del genere umano. Quindi ogni problema va affrontato in maniera diversa: raccogliendo tutte le informazioni che è possibile raccogliere e analizzandole statisticamente cercando di desumere, anzichè imporre, quali sono le condizioni iniziali al variare delle quali varia il risultato finale. È quindi cruciale disporre del maggior numero possibile di informazioni sul maggior numero di oggetti possibile.

Questo modo di vedere il problema ha portato alla nascita fin da tempi antichissimi di cataloghi dove era riportato quanto si conosceva del "cielo". Non appena la tecnologia lo ha poi permesso si è passati da un'analisi individuale a un'analisi sistematica. Queste analisi sistematiche, fatte su porzioni di cielo a causa delle limitazioni strumentali, prendono il nome di survey; Uno dei primi esempi è la Shapley-Ames (1932) [4].

L'astronomia galattica ed extragalattica e lo studio della formazione ed evoluzione delle strutture dell'Universo sta attraversando nell'ultimo decen-

nio una fase di rapidissima crescita. Da una parte, l'avvento e lo sviluppo del telescopio spaziale Hubble e dei grandi telescopi terrestri della classe degli 8 e 10 m (VLT, Keck, Subaru) permette di identificare e studiare oggetti di taglia galattica e quasar fino a redshift 5 o 6 (quando cioè l'Universo era oltre 10 volte più giovane di adesso). Con questi strumenti si cerca di ricostruire la storia evolutiva dell'Universo in modo "diretto", osservandolo quando aveva diverse età. D'altro canto, un approccio del tutto complementare, sebbene altrettanto importante, consiste nello studiare nel massimo dettaglio possibile l'Universo "locale", cioè quella parte che si estende entro non più di un paio di miliardi di anni luce da noi (redshift 0.1- 0.2). Per fare questo è necessario compiere un accurato censimento della distribuzione e delle proprietà degli oggetti celesti su grandissime aree di cielo, al fine di ottenere significanza statistica. Sebbene molto meno spettacolare delle osservazioni ad alto redshift con i grandi telescopi, questo tipo di ricerca è fondamentale nel porre in modo quantitativo vincoli osservativi stringenti ai complessi modelli di formazione ed evoluzione delle strutture. Le osservazioni che è necessario raccogliere comprendono principalmente:

- la posizione dell'oggetto in cielo (astrometria);
- la misura del suo flusso o luminosità apparente a varie ampiezze d'onda (fotometria);
- la misura della sua distanza, che può essere nella maggior parte dei casi ben approssimata attraverso una misura dello spostamento verso il rosso (redshift) delle righe spettrali (spettroscopia) o delle più rilevanti strutture del continuo (redshift fotometrico).

Fino a pochi anni fa, le uniche survey ottiche a grande campo disponibili erano basate su lastre fotografiche ottenute con le camere Schmidt ², come quella di Monte Palomar, a cui veniva associato un numero inferiore (di vari ordini di grandezza) di misure spettroscopiche. Solo tra la fine degli anni '90 e il 2000, grazie ai grandi progressi sia nell'ambito dei rivelatori elettronici, sia nell'ambito dell'automazione dei processi di osservazione e riduzione dati, sono divenuti operativi numerosi grandi progetti di survey su grandi aree di cielo, che utilizzano telescopi terrestri di piccole e medie

²camera sviluppata per usi astronomici che permetteva immagini a grande campo con limitate aberrazioni

dimensioni e rivelatori elettronici e stanno fornendo osservazioni in quantità e qualità mai raggiunte prima.

2.14 Sloan Digital Sky Survey (SDSS)

La SDSS rappresenta senza dubbio il progetto più ambizioso fra le survey ottiche su grande scala. La collaborazione, fondata da un gruppo di importanti istituzioni statunitensi comprendenti il Fermilab di Chicago, lo IAS di Princeton e la Johns Hopkins University, riunisce oggi 13 grandi Istituti di USA, Giappone e Germania (Max-Planck- Society). Il progetto, operativo dal 2000, prevede di coprire circa un quarto della volta celeste (10000 gradi quadrati) con immagini in 5 bande dal vicino ultravioletto (banda u) al vicinissimo e vicino infrarosso (i e z), attraverso le bande del visibile, regione blu-verde dello spettro, (g) e rosso (r); le cinque lunghezze d'onda efficaci delle bande utilizzate sono: 3490, 4810, 6230 7640 e 9060 Å [8, 9, 4]. Sulla base di tali immagini, di tutti gli oggetti più brillanti di 17.7 mag in banda r , viene ottenuto uno spettro a media risoluzione al fine di determinarne il redshift e meglio studiarne le proprietà fisiche. A survey completata gli oggetti di cui saranno disponibili posizioni astrometriche e flussi saranno oltre 300 milioni, mentre saranno disponibili i redshift (e quindi le distanze) di oltre un milione di galassie e circa 10^5 quasar. Con tale quantità di dati, lo sforzo tecnologico per l'acquisizione, la riduzione e la distribuzione, tramite successive "Data Release", alla comunità scientifica e al pubblico è stato, è, e sarà notevole [7].

Per le osservazioni della SDSS è stato costruito un apposito telescopio da 2.5m all' Apache Point Observatory (Sunspot, New Mexico, USA). Oltre al telescopio sono stati costruiti tre strumenti per affiancarlo, un telescopio fotometrico, (0.5m), utilizzato come il nome suggerisce per calibrare la fotometria e due strumenti che servono per controllare le condizioni meteorologiche: un seeing monitor e un cloud scanner.

Le immagini sono acquisite da una batteria di 30 CCD (vedi figura 2.23), disposti in 6 colonne e 5 righe (una per banda), con una tecnica detta di drift-scan: il telescopio resta fermo mentre la volta celeste scorre per effetto della rotazione terrestre; la lettura del CCD avviene in modo sincronizzato.

Questo sistema è stato concepito così per eliminare i tempi morti, difatti

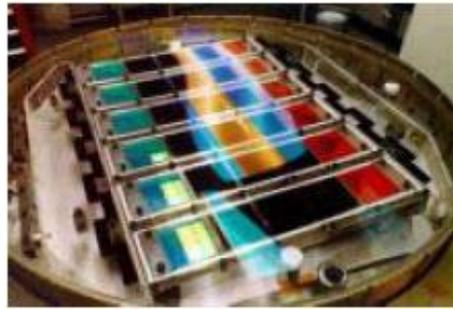


Figura 2.23: I 30 CCD (6 per ogni banda) della camera di acquisizione delle immagini della SDSS

in questo modo non bisogna fermare l’acquisizione ogni volta che occorre leggere il rivelatore, spostare il telescopio e ricominciare l’acquisizione.

Ogni immagine ha un’esposizione complessiva di 54.1 secondi, tempo di transito sul CCD, non molti, ma sufficienti ad assicurare l’identificazione delle sorgenti e a misurarne le proprietà fotometriche, ogni riga è esposta separatamente tramite un filtro diverso, l’intero passaggio dura intorno ai 5.7 minuti. Ogni pixel misura $24 \mu\text{m}$ corrispondente a 0.396 secondi d’arco. Sulla base delle immagini così acquisite vengono ricavate, attraverso un sistema ampiamente automatizzato, le “maschere” per l’acquisizione degli spettri: su piastre forate in corrispondenza delle sorgenti, vengono inserite le fibre ottiche che convogliano la luce allo spettroscopio. Tutta la riduzione dati viene effettuata in una pipeline automatica, in grado di misurare posizioni, flussi, caratteristiche strutturali e le principali informazioni spettrali (tra cui il redshift) degli oggetti osservati. Le osservazioni fotometriche sono state prese solo in notti senza luna mentre le notti classificate come “gray” per via della presenza della luce lunare sono state sfruttate per la spettroscopia [7].

2.15 Il Passaggio da Spettroscopia a Fotometria

Quanto detto finora ci permette di fare alcune considerazioni:

- la fotometria è disponibile per un grande numero di oggetti
- la spettroscopia prende tempo e quindi è disponibile per meno oggetti

- l'analisi dello spettro di una galassia ci permette, tramite i rapporti tra particolari righe, di determinare se questa galassia è o meno un AGN

Sulla base di ciò è evidente che un sistema basato su dati esclusivamente fotometrici che permetta con una certa confidenza di individuare dei potenziali (candidati) AGN è cruciale, perchè permetterebbe di lavorare su cataloghi di AGN (sapendo che alcuni potrebbero non esserlo) molto più ampi e quindi potere lavorare con statistiche significative.

2.16 I Parametri Usati

In questo lavoro di tesi ho utilizzato i seguenti parametri estratti dai cataloghi della SDSS:

1. Raggio Petrosiano al 50 % in Banda r
2. Raggio Petrosiano al 50 % in Banda g
3. Raggio Petrosiano al 50 % in Banda r
4. Raggio Petrosiano al 50 % in Banda i
5. Raggio Petrosiano al 50 % in Banda z
6. Indice di Concentrazione in Banda r
7. Redshift Fotometrico Corretto
8. Magnitudine di Fibra in Banda r
9. Indice di Colore u-g Corretto
10. Indice di Colore g-r Corretto
11. Indice di Colore r-i Corretto
12. Indice di Colore i-z Corretto
13. Magnitudine in Banda r Corretta

I primi cinque parametri sono i cosiddetti raggi petrosiani, uno per banda, in particolare quelli al 50%, ovverossia i raggi che contengono il 50% del flusso petrosiano.

Il flusso petrosiano utilizzato dalla SDSS è definito come [5, 6]:

$$\mathfrak{R}_P(r) = \frac{\int_{0.8r}^{1.25r} dr' 2\pi r' I(r') / [\pi(1.25^2 - 0.8^2)r^2]}{\int_0^r dr' 2\pi r' I(r') / (\pi r^2)} \quad (2.111)$$

Che definisce il “Rapporto Petrosiano ” \mathfrak{R}_P al raggio r dal centro di un oggetto come il rapporto tra la brillantezza superficiale locale in una corona circolare da r alla brillantezza superficiale media entro r e dove $I(r)$ è il profilo di brillantezza superficiale mediato lungo l’azimut.

L’indice di concentrazione, parametro 6, invece è il rapporto tra i raggi delle circonferenze che contengono rispettivamente il 50% ed il 90% del flusso totale proveniente dall’oggetto in approssimazione di sorgente simmetrica ed è un buona misura di quanto la luce sia concentrata nel centro della sorgente.

Il redshift fotometrico corretto, parametro 7, è invece il redshift ricavato dal catalogo di D’Abrusco et al. [27].

La magnitudine di fibra nella banda r, parametro 8, è il flusso contenuto in 3” di diametro d’apertura.

Le magnitudini, come già spiegato, misurano, in una scala logaritmica a base 2.513, pari a 2.5 nella scala logaritmica decimale, i rapporti tra i flussi luminosi F_λ dei singoli oggetti e quello di un oggetto campione.

Seguono i colori, parametri da 9 a 12, e la magnitudine nella banda r, feature 13, tutte queste features sono corrette per l’arrossamento.

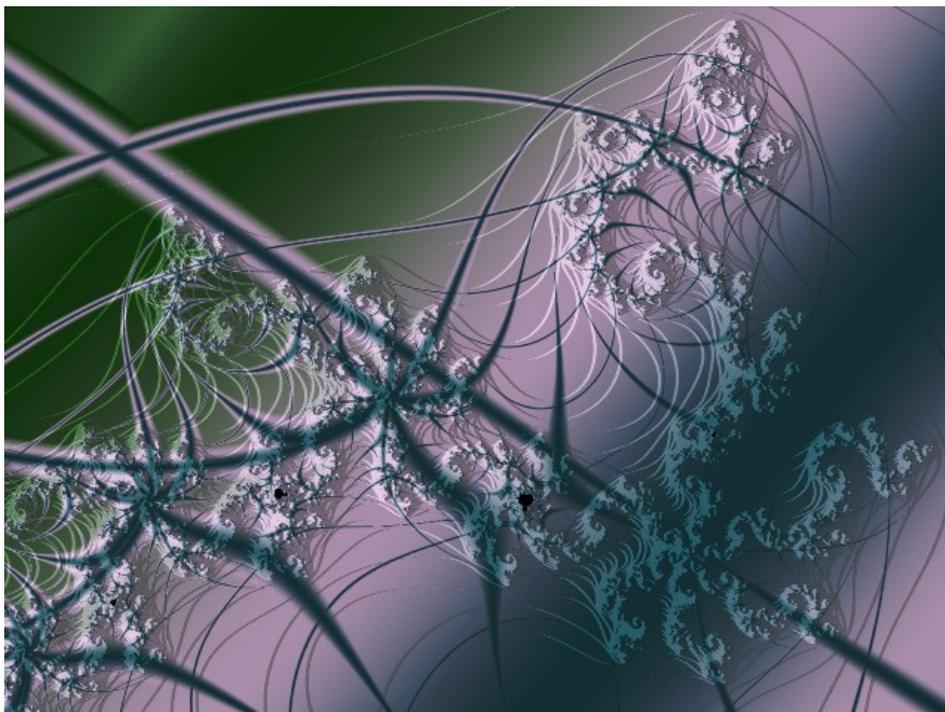
Il colore come già detto è la differenza tra due magnitudini e quindi indica il rapporto tra due flussi.

Capitolo 3

Le Reti Neurali

John von Neuman:

“non c’è alcun modello del cervello più semplice del cervello stesso.”



Una rete neurale è un insieme di neuroni biologici tra loro interconnessi.

Nell'uso moderno si intende però di solito con rete neurale una rete di neuroni artificiali, che cerca di simulare il funzionamento dei neuroni

all'interno di un sistema informatico. Può essere composta sia da programmi che da hardware dedicato.

In quasi tutti gli organismi viventi sono presenti complesse organizzazioni di cellule nervose, con compiti di riconoscimento delle configurazioni assunte dall'ambiente esterno, memorizzazione e reazione agli stimoli provenienti dallo stesso. Il cervello umano rappresenta probabilmente il più mirabile frutto dell'evoluzione per le sue capacità di elaborare informazioni. Al fine di compiere tali operazioni, le reti biologiche si servono di un numero imponente di semplici elementi computazionali (neuroni) fittamente interconnessi in modo da variare la loro configurazione in risposta agli stimoli esterni: in questo senso può parlarsi di apprendimento ed i modelli artificiali cercano di catturare questo tratto distintivo della biologia. Quantifichiamo questa complessità nel cervello umano:

- esistono vari tipi di neuroni
- il cervello è costituito in media di 10 miliardi di neuroni impacchettati con densità dell'ordine di 80000 neuroni per mm^3
- ogni neurone interagisce con 1000-10000 neuroni
- il collegamento sinaptico è possibile anche tra neuroni anche molto distanti
- l'elaborazione risiede in intere regioni ben localizzate
- la struttura cerebrale è sempre in evoluzione
- l'elaborazione dell'informazione che viene diffusa in parallelo tra migliaia di neuroni porta all'emergere dei processi cognitivi

Tipicamente un neurone è costituito di tre parti principali (vedi figura 3.1): il soma (corpo cellulare), l'assone (linea di uscita del neurone unica ma che si dirama in migliaia di rami) e il dendrite (linea di entrata del neurone che riceve segnali in ingresso da altri assoni tramite sinapsi). Il corpo cellulare fa una "somma pesata" (integrazione) dei segnali in ingresso. Se il risultato eccede un certo valore di soglia allora il neurone si attiva ed è prodotto un "potenziale di azione" che è mandato all'assone. Se il risultato non eccede il valore di soglia, il neurone rimane in uno stato di riposo.

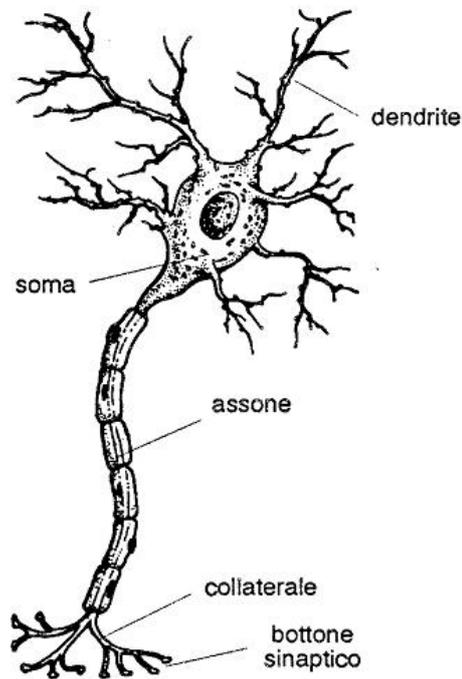


Figura 3.1: Schema di un neurone reale

Una rete neurale artificiale riceve segnali esterni su uno strato di nodi (unità di elaborazione) d'ingresso, ciascuno dei quali è collegato con numerosi nodi interni, organizzati in più livelli. Ogni nodo elabora i segnali ricevuti e trasmette il risultato a nodi successivi.

Le reti neurali, per come sono costruite, lavorano in parallelo e sono quindi in grado di trattare molti dati. Si tratta in sostanza di un sofisticato sistema di tipo statistico dotato di una buona immunità al rumore; se alcune unità del sistema dovessero funzionare male, la rete nel suo complesso avrebbe delle riduzioni di prestazioni ma difficilmente andrebbe incontro ad un blocco del sistema. I software di ultima generazione dedicati alle reti neurali richiedono comunque buone conoscenze statistiche; il grado di apparente utilizzabilità immediata non deve trarre in inganno, pur permettendo all'utente di effettuare da subito previsioni o classificazioni, seppure con i limiti del caso.

Nelle industrie, qualora si dispongano dei dati cosiddetti storici, ovvero delle passate sperimentazioni, le reti neurali risultano interessanti in quanto permettono di estrarre dati e modelli senza effettuare ulteriori prove.

I modelli prodotti dalle reti neurali, anche se molto efficienti, non sono spiegabili in linguaggio simbolico umano: i risultati vanno accettati “così come sono”, da cui anche la definizione inglese delle reti neurali come “black box”.

Come per qualsiasi algoritmo di modellazione, anche le reti neurali sono efficienti solo se le variabili predittive sono scelte con cura.

Non sono in grado di trattare in modo efficiente variabili di tipo categorico (per esempio, il nome della città) con molti valori diversi.

Necessitano di una fase di addestramento del sistema che fissi i pesi dei singoli neuroni e questa fase può richiedere molto tempo, se il numero dei record e delle variabili analizzate è molto grande.

Non esistono teoremi o modelli che permettano di definire la rete ottima, quindi la riuscita di una rete dipende molto dall’esperienza del creatore.

3.1 Reti Neurali Feed-Forward

Le reti neurali si basano principalmente sulla simulazione di neuroni artificiali opportunamente collegati. Il modello rappresentato in figura 3.2 è quello proposto da McCulloch e Pitts.

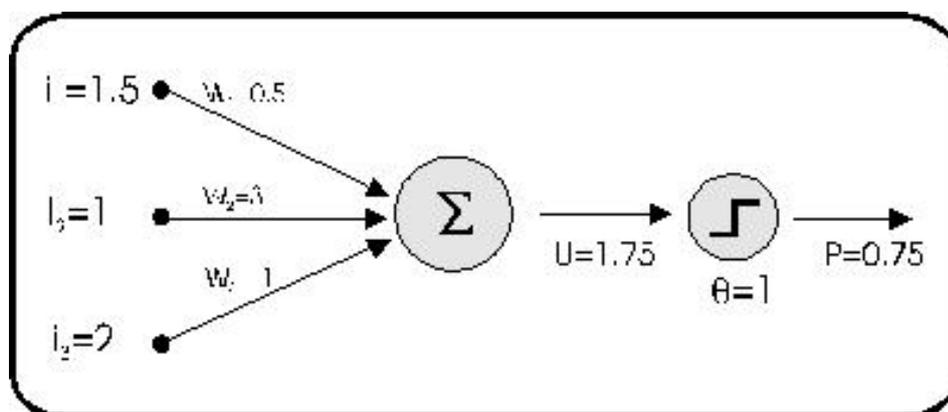


Figura 3.2: Neurone artificiale proposto da McCulloch e Pitts

I suddetti neuroni ricevono in ingresso degli stimoli e li elaborano. L’elaborazione può essere anche molto sofisticata ma in un caso semplice si può pensare che i singoli ingressi vengano moltiplicati per un opportuno valore

detto peso, il risultato delle moltiplicazioni viene sommato e se la somma supera una certa soglia il neurone si attiva attivando la sua uscita.

E' importante capire che il neurone o manda un segnale o non lo manda; non c'è una risposta intermedia!

Il peso indica l'efficacia sinaptica della linea di ingresso e serve a quantificarne l'importanza, un ingresso molto importante avrà un peso elevato, mentre un ingresso poco utile all'elaborazione avrà un peso inferiore. Si può pensare che se due neuroni comunicano fra loro utilizzando maggiormente alcune connessioni allora tali connessioni avranno un peso maggiore.

I singoli neuroni vengono collegati alla schiera di neuroni successivi, in modo da formare una rete di neuroni, come illustrato dalla figura 3.3.

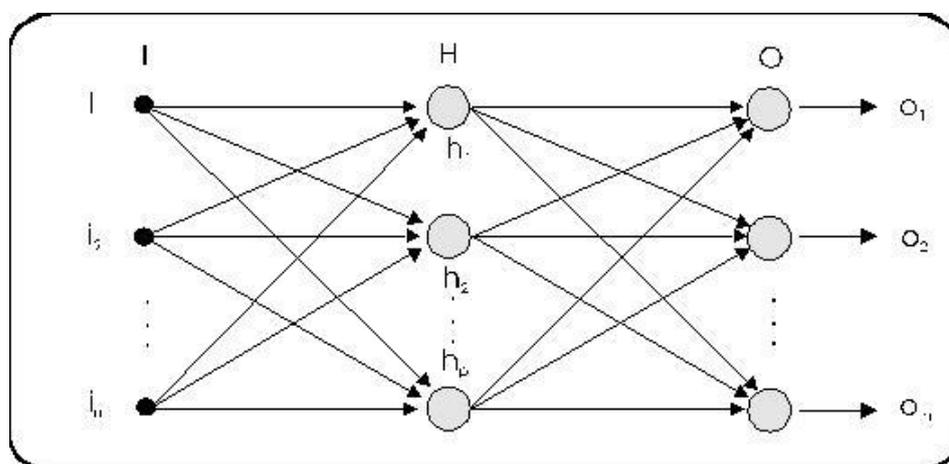


Figura 3.3: Schema di una semplice rete neurale

Normalmente una rete è formata da tre strati.

Nel primo abbiamo gli ingressi (I), questo strato si preoccupa di trattare gli ingressi in modo da adeguarli alle richieste dei neuroni. Se i segnali in ingresso sono già trattati può anche non esserci.

Il secondo strato è quello nascosto (H, hidden), si preoccupa dell'elaborazione vera e propria e può essere composto anche da più colonne di neuroni.

Il terzo strato è quello di uscita (O) e si preoccupa di raccogliere i risultati ed adattarli alle richieste del blocco successivo della rete neurale. Queste reti possono essere anche molto complesse e coinvolgere migliaia di neuroni e decine di migliaia di connessioni[10].

3.2 Storia

L'ampia varietà di modelli non può prescindere dal costituente di base, il neurone artificiale proposto da W.S. McCulloch e W. Pitts in un lavoro del 1943, il quale schematizza un combinatore lineare a soglia, con dati binari multipli in entrata e un singolo dato binario in uscita: un numero opportuno di tali elementi, connessi in modo da formare una rete, è in grado di calcolare semplici funzioni booleane.

Nel 1958 F. Rosenblatt introdusse il primo schema di rete neurale, detto perceptron (perceptrone), antesignano delle attuali reti neurali, per il riconoscimento e la classificazione di forme, allo scopo di fornire un'interpretazione dell'organizzazione generale dei sistemi biologici. Il modello probabilistico di Rosenblatt è quindi mirato all'analisi, in forma matematica, di funzioni quali l'immagazzinamento delle informazioni, e della loro influenza sul riconoscimento dei patterns; esso costituisce un progresso decisivo rispetto al modello binario di McCulloch e Pitts, perché i suoi pesi sinaptici sono variabili e quindi il perceptrone è in grado di apprendere.

L'opera di Rosenblatt stimolò una quantità di studi e ricerche, e suscitò un vivo interesse e notevoli aspettative nella comunità scientifica, destinate tuttavia ad essere notevolmente ridimensionate allorché nel 1969 M. Minsky e S. A. Papert mostrarono i limiti operativi delle semplici reti a due strati basate sui perceptroni, e dimostrarono l'impossibilità di risolvere per questa via molte classi di problemi, ossia tutti quelli non caratterizzati da separabilità lineare delle soluzioni. Questo tipo di rete neurale non è abbastanza potente, infatti non è in grado di calcolare neanche la funzione or esclusivo (XOR). Di conseguenza, a causa di queste limitazioni, ad un periodo di euforia per i primi risultati della cibernetica (come veniva chiamata negli anni '60), seguì un periodo di diffidenza durante il quale tutte le ricerche in questo campo non ricevettero più alcun finanziamento dal governo degli Stati Uniti d'America; le ricerche sulle reti tesero, di fatto, a ristagnare per oltre un decennio, e l'entusiasmo iniziale risultò fortemente ridimensionato.

3.3 Multi-Layers Perceptron

Il contesto matematico per addestrare le reti MLP (Multi-Layers Perceptron, ossia perceptrone multistrato) fu stabilito dal matematico americano

Paul Werbos nella sua tesi di dottorato (Ph.D.) del 1974. Uno dei metodi più noti ed efficaci per l'addestramento di tale classe di reti neurali è il cosiddetto algoritmo di error backpropagation (retropropagazione dell'errore), proposto nel 1986 da David E. Rumelhart, G. Hinton e R. J. Williams, il quale modifica sistematicamente i pesi delle connessioni tra i nodi, così che la risposta della rete si avvicini sempre di più a quella desiderata. L'algoritmo di backpropagation (BP) è una tecnica d'apprendimento tramite esempi, costituente una generalizzazione dell'algoritmo d'apprendimento per il perceptrone sviluppato da Rosenblatt nei primi anni sessanta. Mediante questa tecnica era possibile, come detto, trattare unicamente applicazioni caratterizzabili come funzioni booleane linearmente separabili.

L'algoritmo di apprendimento si basa sul metodo gradient-descent (discesa del gradiente) che permette di trovare un minimo locale di una funzione in uno spazio a N dimensioni. I pesi associati ai collegamenti tra gli strati di neuroni si inizializzano a valori piccoli e casuali e poi si applica la regola di apprendimento presentando alla rete dei pattern di esempio. Queste reti neurali sono capaci di generalizzare in modo appropriato, cioè di dare risposte plausibili per input che non hanno mai visto.

L'addestramento di una rete neurale di tipo BP avviene in due diversi stadi: forward-pass e backward-pass.

Nella prima fase i vettori in input sono applicati ai nodi in ingresso con una propagazione in avanti dei segnali attraverso ciascun livello della rete (forward-pass). Durante questa fase i valori dei pesi sinaptici sono tutti fissati.

Nella seconda fase la risposta della rete viene confrontata con l'uscita desiderata ottenendo il segnale d'errore. L'errore calcolato è propagato nella direzione inversa rispetto a quella delle connessioni sinaptiche.

I pesi sinaptici infine sono modificati in modo da minimizzare la differenza tra l'uscita attuale e l'uscita desiderata (backward-pass).

Tale algoritmo consente di superare le limitazioni del perceptrone e di risolvere il problema della separabilità non lineare (e quindi di calcolare la funzione XOR), segnando il definitivo rilancio delle reti neurali, come testimoniato anche dall'ampia varietà d'applicazioni commerciali: attualmente la BP rappresenta un algoritmo di largo uso in molti campi applicativi.

3.3.1 L'Algoritmo di Back Propagation

Guardiamo ora nel dettaglio il funzionamento dell'algoritmo di Back Propagation:

Quest'algoritmo, come detto, permette di modificare i pesi delle connessioni in modo tale che si minimizzi una certa funzione errore E . Tale funzione dipende dal vettore h-esimo di output $\overline{out}^h = f(\overline{x}^h)$ restituito dalla rete (output vector), dato il vettore h-esimo di ingresso \overline{x}^h e dal vettore h-esimo di output \overline{y}^h (target vector); l'insieme di questi vettori forma il cosiddetto training set¹. Il training set è dunque un insieme di N coppie di vettori $(\overline{x}^h, \overline{y}^h)$, con $h = 1, \dots, N$.

La funzione errore che si deve minimizzare si può scrivere come:

$$E = \frac{1}{2} \sum_h \sum_k (out_k^h - y_k^h)^2 \quad (3.1)$$

dove l'indice k rappresenta il valore corrispondente al k -esimo neurone di output.

E è una funzione dipendente dai pesi (che in generale variano nel corso dell'addestramento). Per minimizzarla si possono usare vari algoritmi, uno dei più diffusi è come detto quello della gradient-descent. L'algoritmo parte da un punto generico \overline{x}^0 e ne calcola il gradiente $\nabla f(\overline{x}^0)$. Il gradiente dà la direzione verso cui muoversi lungo la quale ho il massimo incremento (o decremento se considero $-\nabla$).

Definita la direzione ci si muove di una distanza η predefinita a priori e si trova un nuovo punto \overline{x}^1 sul quale è calcolato nuovamente il gradiente. Si continua iterativamente finché il gradiente non si annulla.

L'algoritmo di backpropagation può essere diviso in due passi:

- Forward pass: l'input dato alla rete è propagato al livello successivo e così via ai livelli successivi (il flusso di informazioni si sposta in avanti, cioè forward). Si calcola dunque $E(w)$, l'errore commesso.
- Backward pass: L'errore fatto dalla rete è propagato all'indietro (backward) e i pesi sono aggiornati in maniera appropriata.

¹eventualmente al training set così formato si può affiancare un validation set che nel corso dell'algoritmo controllerà che la rete non stia andando in overfitting ovvero che mantenga la capacità di generalizzare il problema.

3.4 Support Vector Machine

Dato un training set formato da coppie features-label $(x_i, y_i), i = 1 \dots l$ dove $x_i \in R^n$ e $y_i \in \{1, -1\}^l$.

Le Support Vector Machine (SVM) cercano di risolvere il seguente problema di ottimizzazione [11]:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \quad (3.2)$$

con la condizione:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \quad (3.3)$$

I vettori x_i vengono mappati in un uno spazio a più dimensioni rispetto a quelle che aveva inizialmente dalla funzione ϕ .

Quindi la SVM trova una iperpiano di separazione col maggior margine possibile in questo nuovo spazio.

$C > 0$ è un termine di correzione dell'errore.

Inoltre $K(x_i, x_j) = \phi(x_i)^T(x_j)$ è la cosiddetta funzione di kernel. Esistono vari kernel tra cui:

- linear: $K(x_i, x_j) = x_i^T x_j$
- polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

3.4.1 Il Kernel RBF

In particolare ho concentrato la mia attenzione sul kernel RBF.

Questo particolare kernel fa la mappatura in maniera non lineare, perciò può gestire la situazione in cui la relazione tra features e label non è lineare. Inoltre il kernel lineare è un caso speciale del kernel RBF come dimostrato da Keerthi e Lin [13], inoltre Lin e Lin [14] hanno mostrato come per certi parametri si comporti anche come il kernel sigmoidale.

Inoltre rispetto al kernel polinomiale ha meno iperparametri da gestire e meno difficoltà computazionali.

Questo kernel ha bisogno di due parametri γ e C . Non è noto prima di iniziare a lavorarci quale coppia di parametri vada usata e, quindi, va fatta una ricerca dei parametri migliori, l'obiettivo è ovviamente trovare la coppia che fornisce il risultato migliore. Per prevenire i problemi di overfitting si usa la procedura di cross validation per cui il dataset viene diviso in v parti in sequenza una viene utilizzata per la validazione e le altre $v - 1$ per l'addestramento.

3.4.2 Ricerca dei Parametri Migliori

Per la ricerca dei parametri migliori si è operata una ricerca su una griglia provando l'uno dopo l'altro vari valori di C e γ , utilizzando una sequenza esponenziale: ad esempio $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$, $\gamma = 2^{-15}, 2^{-13}, \dots, 23$. Dopo aver identificato con una prima griglia una regione di massimo si riesegua una griglia più fine nelle regioni dove si sono trovati i risultati migliori.

3.5 Modalità di Apprendimento

Dal punto di vista della modalità d'apprendimento, è necessario operare una distinzione tra almeno due diverse tipologie: supervisionato e non supervisionato

3.5.1 Apprendimento Supervisionato

Un apprendimento si dice supervisionato (supervised learning), qualora si disponga di un insieme di dati per l'addestramento (o training set) comprendente esempi tipici d'ingressi con le relative uscite loro corrispondenti: in tal modo la rete può imparare ad inferire la relazione che li lega. Successivamente, la rete è addestrata mediante un opportuno algoritmo (tipicamente, la backpropagation che è appunto un algoritmo d'apprendimento supervisionato), il quale usa tali dati allo scopo di modificare i pesi ed altri parametri della rete stessa in modo tale da minimizzare l'errore di previsione relativo all'insieme d'addestramento. Se l'addestramento ha successo, la rete impara a riconoscere la relazione incognita che lega le variabili d'ingresso a quelle d'uscita, ed è quindi in grado di fare previsioni anche laddove l'uscita non è nota a priori; in altri termini, l'obiettivo finale dell'apprendimento supervisionato è la previsione del valore dell'uscita per ogni valore

valido dell'ingresso, basandosi soltanto su un numero limitato di esempi di corrispondenza (vale a dire, coppie di valori input-output). Per fare ciò, la rete deve essere infine dotata di un'adeguata capacità di generalizzazione, con riferimento a casi ad essa ignoti. Ciò consente di risolvere problemi di regressione o classificazione.

Per esempio immaginiamo di voler costruire un programma di riconoscimento vocale: addestriamo la nostra rete registrando un numero sufficiente di frasi di cui conosciamo la classe, in questo caso spesso esistono frasi e parole specifiche che ottimizzano questa fase e la rete sarà in grado di riconoscere anche parole non presenti nel campione con cui è stata addestrata. Da questo semplice esempio si può dedurre anche l'importanza della costruzione del campione di addestramento, difatti se l'utente è raffreddato e quindi ha un'alterazione del tono vocale la rete non sarà poi in grado di riconoscere la voce dell'utente, non essendo stata addestrata per quel tono di voce; quindi possiamo o costruire una nuova rete per l'utente raffreddato oppure possiamo costruire una rete in cui come campione mettiamo frasi pronunciate dall'utente sia in condizioni normali sia se raffreddato.

3.5.2 Apprendimento Non Supervisionato

Un apprendimento si dice non supervisionato (unsupervised learning), se è basato su algoritmi d'addestramento che modificano i pesi della rete facendo esclusivamente riferimento ad un insieme di dati che include le sole variabili d'ingresso. Tali algoritmi tentano di raggruppare i dati d'ingresso e di individuare pertanto degli opportuni cluster rappresentativi dei dati stessi, facendo uso tipicamente di metodi topologici o probabilistici. L'apprendimento non supervisionato è anche impiegato per sviluppare tecniche di compressione dei dati.

Vediamo ora anche un esempio di addestramento non supervisionato. Immaginiamo di avere un certo numero di pazienti, soggetti a quella che per noi è una patologia, di ognuno dei quali conosciamo un certo numero di parametri clinici e eventuali sintomi; tramite una rete neurale non supervisionata possiamo raggruppare questi pazienti in un certo numero di sottoclassi aventi secondo la rete caratteristiche comuni e quindi individuare eventualmente pazienti cui è stata diagnosticata erroneamente la patologia suddetta, oppure si potrebbero trovare sottoclassi della patologia con comportamenti diversi.

3.6 Principal Component Analysis

Il metodo delle Principal Component Analysis (ovverossia Analisi delle Componenti Principali, cui ci riferiremo d'ora in avanti semplicemente come PCA) è un metodo di analisi dei dati esplorativo, in quanto volto a cercare una struttura o una configurazione nei dati a differenza delle classiche analisi dati dette confermatorie in cui l'analisi ha il solo scopo di confermare appunto (o rigettare seconda dei casi) una tesi iniziale. Le PCA rientrano nella famiglia dei metodi fattoriali, che sono metodi lineari di analisi dei dati con i seguenti obiettivi:

- descrivere della struttura dei dati, spiegando le correlazioni tra le variabili osservate attraverso dei nuovi fattori o componenti, combinazioni lineari delle variabili di partenza
- ridurre il numero di variabili osservate, trasformando l'agglomerato di osservazioni in una struttura più semplice ma con lo stesso contenuto informativo di quella partenza.

Il termine fattoriale indica che l'analisi si sviluppa attraverso la ricerca di fattori che legano *linearmente* le variabili originali. Il vantaggio che si ottiene con questi metodi è immediato: un set più piccolo di variabili non correlate tra loro e quindi un minor numero di gradi di libertà del sistema. Questa tecnica fu originariamente concepita da Pearson (1901) e indipendentemente da Hotelling (1933)

Ora procederò, come accennato in precedenza, ad illustrare un particolare e specifico metodo fattoriale, quello detto PCA che consiste nell'interpretazione dei dati originali tramite un nuovo sistema di riferimento, costruito appositamente tramite operazioni matriciali, tale che i nuovi assi minimizzino l'inerzia dei punti rispetto ad essi.

3.6.1 Interpretazione Geometrica

Considerato il nostro catalogo di dati è ovvio che tanto maggiore sarà il rapporto tra numero di oggetti presenti e numero di parametri che intendiamo considerare tanto migliore sarà il risultato che ci attendiamo.

Se i parametri fossero soltanto due si potrebbe cercare una relazione anche graficamente come si può vedere in figura 3.4.

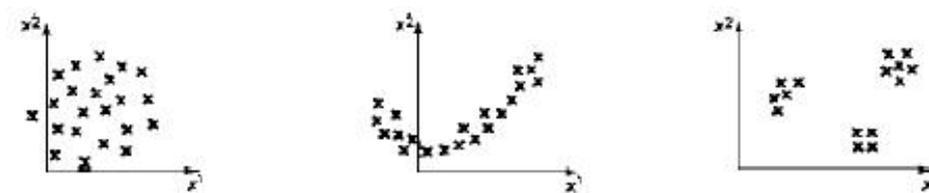


Figura 3.4: Nella figura di sinistra non si vede un legame, in quella centrale si intuisce un forte legame mentre in quella di destra possiamo vedere tre gruppi omogenei

Ovviamente si potrebbe procedere così anche per tre dimensioni ma quando i parametri sono di più è ovvio che ciò non è possibile; l'unica maniera quindi per rendere visibile la forma della nube di punti è la sua proiezione su delle rette o su dei piani, minimizzando la deformazione che è implicita in ogni proiezione.

Quindi nello spazio N -dimensionale (dove N è il numero di parametri che possediamo) si cerca un piano P determinato da due rette su cui proiettare i punti mantenendo al meglio le distanze originali, quindi prendiamo altre rette non correlate alle prime (e quindi a due a due ortogonali a queste) che continuano a rispettare al meglio le distanze e, a questo punto, ricordando che le distanze tra le proiezioni sono sempre minori o uguali delle distanze originali, cerchiamo il massimo delle distanze.

A tale scopo si determina una retta imponendo che le distanze al quadrato tra i punti proiettati su di lei sia massima, la seconda retta si trova imponendo sia l'ortogonalità alla prima sia che le distanze quadre delle proiezioni siano massimizzate, la terza richiedendo l'ortogonalità rispetto alle prime due oltre alla massimizzazione e così via.

Avendo N dimensioni col procedimento precedente troveremo N rette; queste rette sono gli assi principali della nube di punti e tramite le proiezioni si ottengono le nuove coordinate. Geometricamente si è traslato il precedente sistema di riferimento nel baricentro della nube e lo si è ruotato per minimizzare l'inerzia (data dalla somma delle distanze dei punti da un asse) rispetto ai vari assi.

3.6.2 Fattorizzazione

Abbiamo detto che lo scopo delle PCA è ridurre le variabili in gioco in modo da poter visualizzare l'insieme dei punti in maniera ragionevole, perdendo la minore quantità possibile di informazione.

Quanto finora descritto, corrispondendo a un semplice cambiamento del sistema di riferimento, non ci permette di ridurre i dati, ma la cosa importante è che nel nuovo sistema di riferimento la gran parte dell'informazione è presente sui primi assi, e a mano a mano che si procede gli assi trovati hanno un peso sempre minore.

Si può scegliere dove arrestarsi in varie maniere; i metodi più utilizzati sono metodi geometrici, ad esempio il metodo di Pearson (1901), che utilizza i fattori che portano a una percentuale di varianza fissata (usualmente 80-90%); è in pratica il rapporto tra l'inerzia rispetto ai primi h fattori divisa per l'inerzia totale. Questi metodi, come sottolineato da Eastment e Kranowsky (1982), servono per lo più per ottenere un modello descrittivo, se la finalità è predittiva Eastment e Kranowsky suggeriscono invece un metodo detto di cross-validation, il quale metodo prevede l'eliminazione di un punto alla volta e controlla la capacità predittiva della struttura creata in sua assenza, il numero di fattori scelto sarà quello in grado di dare un'adeguata ricostruzione della matrice dei dati da cui si partiva.

3.6.3 Schema del Lavoro

In linea di principio lo schema di lavoro è il seguente: in prima istanza si scelgono le variabili da analizzare, si impone il numero di variabili che devono essere estratte, si fa svolgere l'estrazione a un software apposito, si ruotano opportunamente gli assi e quindi si valuta il risultato, se il risultato non è soddisfacente si ritorna indietro e si riesegue il tutto. Usualmente una prima valutazione si effettua valutando il piano principale, quello determinato dalle prime due rette trovate, su cui si spera di trovare un minimo di percentuale di varianza, usualmente stimato intorno al 70%; se così non fosse è necessario valutare anche il fattore successivo fino ad arrivarci [15].

3.7 Probabilistic Principal Surfaces

Il metodo delle probabilistic principal surfaces (ovvero Superfici Principali Probabilistiche) è per lo più noto con il suo acronimo col quale ci riferiremo a lui d'ora in poi: PPS.

Questo metodo [17, 18, 21, 22] appartiene alla famiglia dei cosiddetti metodi delle variabili latenti [20] e si può considerare un'estensione delle Generative Topographic Mapping (GTM) [19].

Le PCA hanno un limite: la riduzione lineare non è sempre efficace come si può vedere nel semplice esempio di figura 3.5; come abbiamo visto, il loro

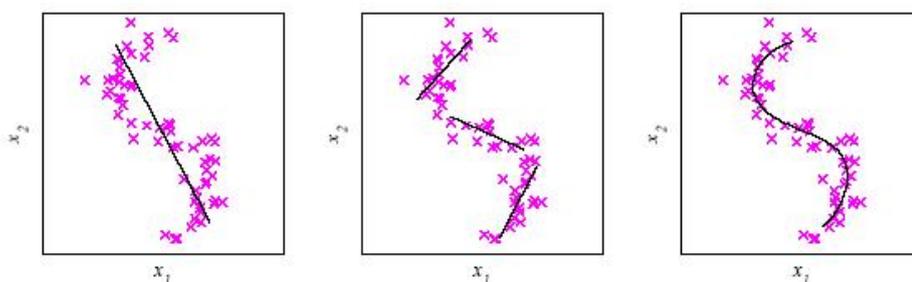


Figura 3.5: Nella figura vediamo come la stessa aggregazione di punti viene vista da una PCA, da varie PCA e con una soluzione non lineare

punto di forza è la semplicità ma possono essere inadeguate nel momento in cui si va a trattare modelli altamente non lineari e il tentativo di aggirare il problema con una polinomiale neanche aderisce sufficientemente alla realtà [18].

In pratica una PPS viene addestrata a riconoscere le migliori funzioni di proiezione dallo spazio N -dimensionale dei parametri ad una superficie sferica in uno spazio tridimensionale; questa superficie è ricoperta da una griglia di variabili latenti, ovvero punti, ognuno dei quali rappresenta il picco di una gaussiana nello spazio N -parametrico. Questo permette di visualizzare il tutto con un grafico tridimensionale indipendentemente dal numero di parametri iniziali e in questo modo l'essere umano può iniziare a controllare l'esistenza o meno di strutture, strutture che così riesce a visualizzare ma che non riuscirebbe neanche ad immaginare senza fare questo lavoro.

3.7.1 L'Algoritmo delle PPS

Affrontiamo ora la descrizione dell'algoritmo che sta alla base delle PPS in maniera più rigorosa.

L'obiettivo di ogni modello basato sulle variabili latenti è quello di esprimere la distribuzione $p(\mathbf{t})$ delle variabili $\mathbf{t} = (t_1, \dots, t_D) \in \mathbb{R}^D$ in termini di un numero variabili latenti minore di quello originario $\mathbf{x} = (x_1, \dots, x_Q) \in \mathbb{R}^Q$ dove $Q < D$. Per raggiungere questo scopo, la distribuzione di congiunzione $p(\mathbf{t}, \mathbf{x})$ viene decomposta nel prodotto della distribuzione di margine $p(\mathbf{x})$ delle variabili latenti e la distribuzione condizionata $p(\mathbf{t}|\mathbf{x})$.

E'conveniente esprimere la distribuzione condizionata come la fattorizzazione sulle variabili originarie, in questo caso la distribuzione di congiunzione diviene:

$$p(\mathbf{t}, \mathbf{x}) = p(\mathbf{x})p(\mathbf{t}|\mathbf{x}) = p(\mathbf{x}) \prod_{d=1}^D p(t_d|\mathbf{x}) \quad (3.4)$$

La distribuzione condizionata $p(\mathbf{t}|\mathbf{x})$ viene quindi espressa in termini di una mappatura dalle variabili latenti alle variabili originarie, cosicchè

$$\mathbf{t} = \mathbf{y}(\mathbf{x}; \mathbf{w}) + \mathbf{u} \quad (3.5)$$

dove $\mathbf{y}(\mathbf{x}; \mathbf{w})$ è una funzione delle variabili latenti \mathbf{x} con parametri \mathbf{w} , e \mathbf{u} è un rumore indipendente dalle \mathbf{x} .

Se le componenti di \mathbf{u} sono scorrelate, la distribuzione condizionata per \mathbf{t} sarà fattorizzabile come in (3.4).

Dal punto di vista geometrico, la funzione $\mathbf{y}(\mathbf{x}; \mathbf{w})$ definisce una varietà nello spazio dei dati dato dall'immagine dello spazio latente. La definizione di modello a variabili latenti necessita per essere completo di specificare la distribuzione $p(\mathbf{u})$, la mappatura $\mathbf{y}(\mathbf{x}; \mathbf{w})$, e la distribuzione di margine $p(\mathbf{x})$. Il tipo di mappatura $\mathbf{y}(\mathbf{x}; \mathbf{w})$ determina quale particolare modello di variabili latenti si utilizza. Il modello desiderato per la distribuzione $p(\mathbf{t})$ dei dati è quindi ottenuta integrando sulle variabili latenti:

$$p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{x})p(\mathbf{x})d\mathbf{x}. \quad (3.6)$$

Questa integrazione non è a priori trattabile analiticamente, è possibile farlo

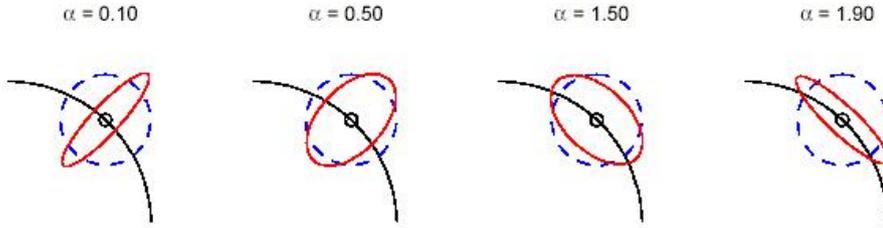


Figura 3.6: Covarianze non orientate $\alpha = 1$ (linea tratteggiata) e covarianze orientate (linea piena) per $\alpha = 0.10, 0.50, 1.50, 1.90$. Il range di validità per α è $0 < \alpha < 2$ per $D = 2$, $Q = 1$ in questo esempio.

solo se le distribuzioni $p(\mathbf{t}|\mathbf{x})$ and $p(\mathbf{x})$ hanno forme particolari.

Le PPS definiscono una mappatura parametrica non lineare $\mathbf{y}(\mathbf{x}; \mathbf{W})$, dove \mathbf{y} è continua e derivabile, che proietta ogni punto nello spazio delle variabili latenti in un punto dello spazio originario. Poichè lo spazio delle variabili latenti è Q -dimensionale, questi punti saranno confinati in una varietà inclusa, non linearmente, nello spazio D -dimensionale delle variabili originarie. Questo implica che i punti proiettati vicino a un nodo della superficie avranno maggior influenza su questo nodo dei punti proiettati lontano da esso (vedi figura 3.6).

Ognuno di questi nodi $\mathbf{y}(\mathbf{x}; \mathbf{w})$, $\mathbf{x} \in \{\mathbf{x}_m\}_{m=1}^M$ ha una covarianza espressa da:

$$\Sigma(\mathbf{x}) = \frac{\alpha}{\beta} \sum_{q=1}^Q \mathbf{e}_q(\mathbf{x}) \mathbf{e}_q^T(\mathbf{x}) + \frac{(D - \alpha Q)}{\beta(D - Q)} \sum_{d=Q+1}^D \mathbf{e}_d(\mathbf{x}) \mathbf{e}_d^T(\mathbf{x}), \quad (3.7)$$

$$0 < \alpha < \frac{D}{Q}$$

dove

- $\{\mathbf{e}_q(\mathbf{x})\}_{q=1}^Q$ è il set di vettori ortonormali tangenti alla varietà in $\mathbf{y}(\mathbf{x}; \mathbf{w})$,
- $\{\mathbf{e}_d(\mathbf{x})\}_{d=Q+1}^D$ è il set di vettori ortonormali ortogonali alla varietà in $\mathbf{y}(\mathbf{x}; \mathbf{w})$.

Il set completo di vettori ortonormali $\{\mathbf{e}_d(\mathbf{x})\}_{d=1}^D$ appartiene a \mathbb{R}^D e il parametro α è un fattore di bloccaggio e determina l'orientamento della

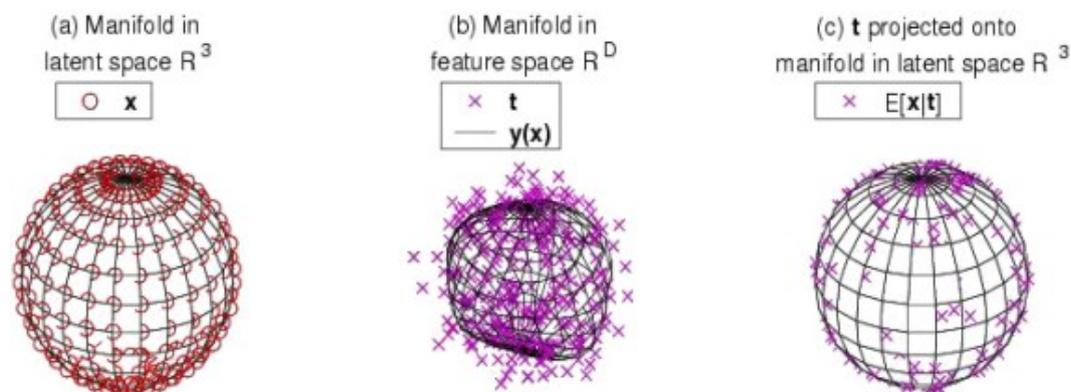


Figura 3.7: (a) Rappresentazione schematica della varietà sferica nello spazio latente tridimensionale R^3 , (b) la stessa varietà distorta nello spazio dei parametri R^D con i punti associati ai dati, (c) la proiezione della distribuzione dei punti sulla superficie della varietà sferica sullo spazio latente R^3 .

matrice di covarianza. Le PPS si riducono alle comuni GTM per $\alpha = 1$ e alle manifold-aligned GTM per $\alpha > 1$:

$$\Sigma(\mathbf{x}) = \begin{cases} 0 < \alpha < 1 & \perp \text{ alla varietà} \\ \alpha = 1 & I_D \text{ o sferico} \\ 1 < \alpha < D/Q & \parallel \text{ alla varietà.} \end{cases}$$

Per stimare i parametri \mathbf{W} e β si usa l'algoritmo Expectation–Maximization (EM) [23], mentre il fattore di bloccaggio è fissato e si assume essere costante durante le iterazioni dell'EM. In uno spazio latente $3D$, allora, una varietà sferica può essere costruita utilizzando una PPS con nodi $\{\mathbf{x}_m\}_{m=1}^M$ disposti regolarmente sulla superficie di una sfera nello spazio latente \mathbb{R}^3 . La motivazione che spinge a utilizzare questo metodo è che le PPS sferiche sono particolarmente adatte a gestire, per dataset particolarmente corposi, dati che si trovano sparsi [19].

Le coordinate della varietà latente $\hat{\mathbf{x}}_n$ di ogni punto \mathbf{t}_n sono calcolate come:

$$\hat{\mathbf{x}}_n \equiv \langle \mathbf{x} | \mathbf{t}_n \rangle = \int \mathbf{x} p(\mathbf{x} | \mathbf{t}) d\mathbf{x} = \sum_{m=1}^M r_{mn} \mathbf{x}_m$$

dove r_{mn} sono le variabili latenti *responsibilities* definite come:

$$\begin{aligned} llr_{mn} = p(\mathbf{x}_m | \mathbf{t}_n) &= \frac{p(\mathbf{t}_n | \mathbf{x}_m) P(\mathbf{x}_m)}{\sum_{m'=1}^M p(\mathbf{t}_n | \mathbf{x}_{m'}) P(\mathbf{x}_{m'})} = & (3.8) \\ &= \frac{p(\mathbf{t}_n | \mathbf{x}_m)}{\sum_{m'=1}^M p(\mathbf{t}_n | \mathbf{x}_{m'})} & (3.9) \end{aligned}$$

Poichè $\|\mathbf{x}_m\| = 1$ and $\sum_m r_{mn} = 1$, for $n = 1, \dots, N$, questi punti giacciono in una sfera unitaria, cioè $\|\hat{\mathbf{x}}_n\| \leq 1$.

Una questione interessante è la stima dell'incidenza di ogni parametro dei dati di ingresso sulle variabili latenti che aiuta a comprendere la relazione tra il parametro e i cluster trovati. L'incidenza dei parametri è calcolata valutando la densità di probabilita delle componenti dei vettori di ingresso rispetto a ogni variabile latente. Più precisamente, sia $\{\mathbf{t}_n\}_{n=1}^N$ il set dei dati di ingresso D -dimensionali, cioè $\mathbf{t}_n = (t_{n1}, \dots, t_{nD}) \in \mathbb{R}^D$, e $\{\mathbf{x}_m\}_{m=1}^M$ sia il set delle variabili latenti con $\mathbf{x}_m \in \mathbb{R}^3$. Per ogni dato $\mathbf{t}_n = (t_{n1}, \dots, t_{nD})$ vogliamo calcolare $p(t_{ni}/t_{n1}, \dots, t_{ni-1}, t_{ni+1}, \dots, t_{nD}, \mathbf{x}_m)$, for $m = 1, \dots, M$ and $i = 1, \dots, D$. In dettaglio:

$$p(t_{ni}/t_{n1}, \dots, t_{ni-1}, t_{ni+1}, \dots, t_{nD}, \mathbf{x}_m) = \quad (3.10)$$

$$= \frac{p(t_{n1}, t_{n2}, \dots, t_{nD}, \mathbf{x}_m)}{p(t_{n1}, \dots, t_{ni-1}, t_{ni+1}, \dots, t_{nD}, \mathbf{x}_m)} = \quad (3.11)$$

$$= \frac{p(t_{n1}, \dots, t_{nD} | \mathbf{x}_m) P(\mathbf{x}_m)}{p(t_{n1}, \dots, t_{ni-1}, t_{ni+1}, \dots, t_{nD} | \mathbf{x}_m) P(\mathbf{x}_m)} = \quad (3.12)$$

$$= \frac{p(t_{n1}, \dots, t_{nD} | \mathbf{x}_m)}{p(t_{n1}, \dots, t_{ni-1}, t_{ni+1}, \dots, t_{nD} | \mathbf{x}_m)}. \quad (3.13)$$

L'ultimo termine si ottiene semplicemente poichè il numeratore è semplicemente il m -esimo termine della gaussiana ricavata dal modello delle PPS centrato su $y(\mathbf{x}_m; \mathbf{W})$ e varianza Σ_m , mentre il denominatore è lo stesso componente Gaussiano in cui l' i -esimo termine manca. Infine il valore della (3.10) sugli N dati di input, per ogni \mathbf{x}_m , è calcolato.

Questo spiega perchè le PPS sferiche possono essere utilizzate come varietà di riferimento per classificare dati a molte dimensioni.

Durante la fase di addestramento viene creata una varietà di riferimento.

Nella fase di test, un dato mai visto dalla rete viene attribuito alla varietà sferica più vicina.

Ovviamente il concetto di più vicino implica il calcolo di una distanza

tra un punto e il nodo dello spazio. Prima di questo calcolo i dati devono essere proiettati sullo spazio. Questo poichè una varietà sferica consiste di zone quadrate o triangolari, ognuna delle quali definita da tre o quattro nodi della varietà, una volta proiettato il dato viene calcolata un'approssimazione della distanza. Nelle PPS esistono tre criteri di approssimazione:

- Nearest Neighbour: trova la minima distanza quadra da tutti i nodi della varietà;
- Grid Projections: trova la più corta distanza di proiezione sulla griglia della varietà;
- Nearest Triangulation: trova la proiezione più vicina alle possibili triangolazioni.

Per lo più dei tre criteri viene utilizzato Nearest Neighbour poichè tale criterio permette di valutare le distanze da ogni dato nello spazio dei parametri a tutti i nodi chiusi sulla varietà sferica; anche se è più pesante in termini di elaborazione dei dati rispetto agli altri due metodi, in pratica fornisce la più affidabile scelta del nodo (o dei nodi, qualora più di uno si trovi alla stessa distanza da un punto).

Per una più dettagliata esposizione sulle PPS e sul loro utilizzo come strumenti nell'ambito del data mining si rimanda alla bibliografia con particolare attenzione a [21, 17, 22]

3.8 Negentropy Clustering

A seconda delle impostazioni si varia il numero di cluster che la PPS creerà; in un primo stadio questo numero conviene venga tenuto alto, essendo ovviamente meglio avere un cluster ridondante, magari composto da uno o due elementi piuttosto che perderne uno fondamentale, il passo successivo è quello di dare il risultato così ottenuto in pasto a quella che viene chiamata usualmente NEC (Clustering Negentropico). Cos'è la Negentropia? E' una quantità che si definisce in teoria dei segnali come la distanza dalla "normalità". Il segnale viene detto normale se è una gaussiana. La Negentropia è nulla se il segnale è normale, altrimenti ha un valore maggiore di zero.

La Negentropia è una quantità invariante per ogni cambio di coordinate lineare.

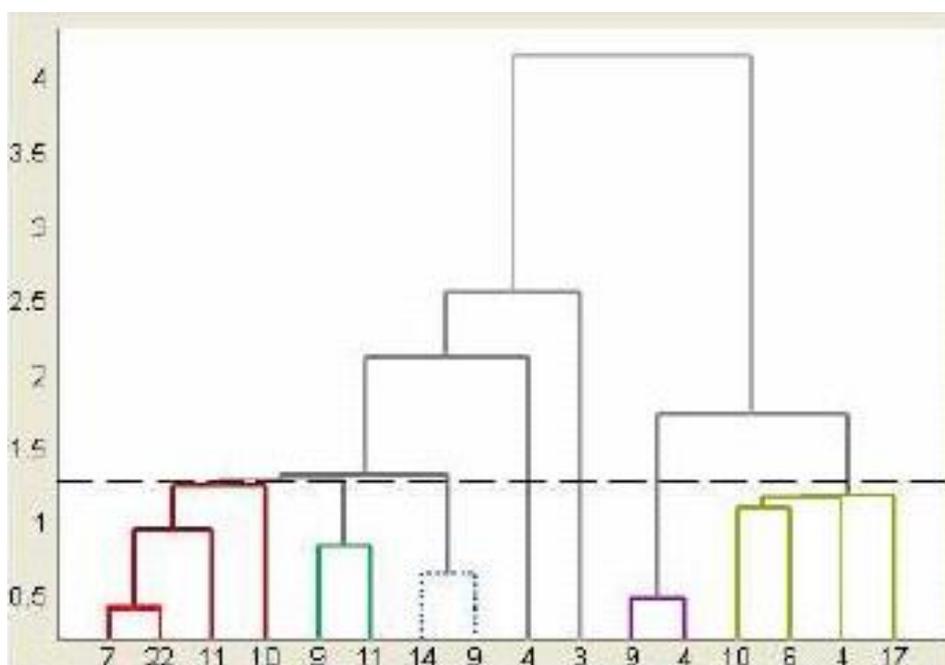


Figura 3.8: Un esempio di dendrogramma, la linea tratteggiata indica la soglia

In pratica una NEC altro non è che una tecnica agglomerativa che serve per passare dalla clusterizzazione fatta con le PPS a quella desiderata. Lo strumento che utilizziamo, facente parte del pacchetto *ASTRONEURAL* produce quindi un dendrogramma (vedi figura 3.8 dove è evidente il significato dell'agglomerazione) in cui ogni strato rappresenta uno dei cluster iniziali che al variare di un valore di soglia vengono mano a mano agglomerati. Può essere utile visualizzarlo in un'altra maniera (vedi figura 3.9) disponendo da sinistra a destra i cluster che si accorpano con valori di soglia diverse (quelli a sinistra si accorperanno con bassi valori della soglia, quelli a destra con alti valori della soglia).

La scelta di quale grado di accorpamento considerare resta all'utente e dipende dallo scopo della sua ricerca. Un esempio di possibile scelta è quello della ricerca dei plateau; graficando il numero di cluster in funzione del valore della soglia, la curva decrescente ottenuta potrebbe avere dei plateau, il cui significato è evidente: in quelle zone i cluster risultano ben separati e quindi è probabile che la classificazione così fatta abbia un forte significato.

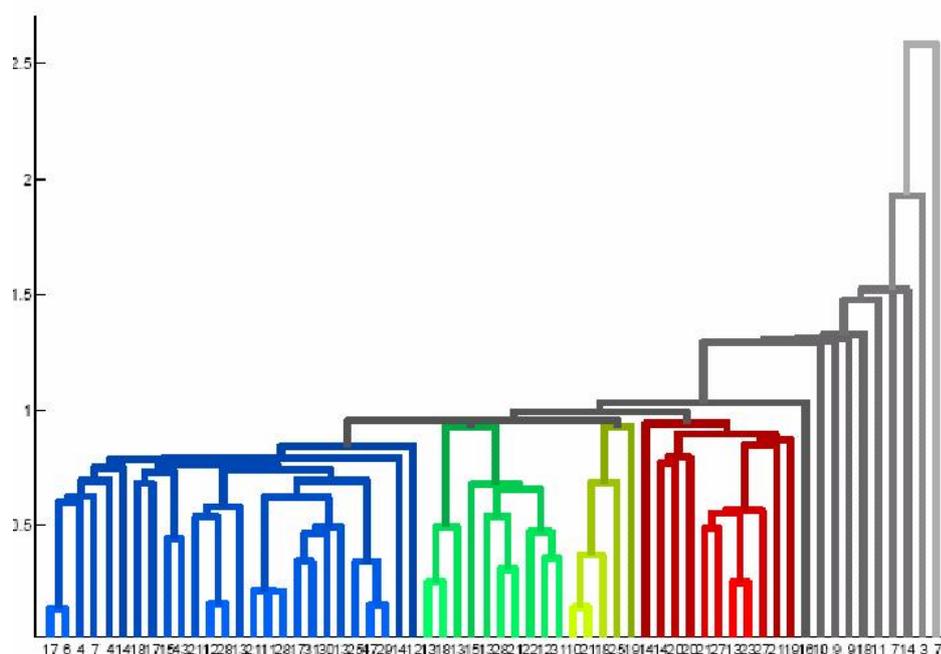


Figura 3.9: Un'altra visualizzazione possibile del dendrogramma

A questo punto abbiamo delle divisioni, però va ancora capito cosa rappresentino e qui entra in gioco la nostra conoscenza del campione che abbiamo utilizzato per addestrare la rete; difatti controllando gli oggetti di cui abbiamo informazioni ricavate per altre vie dove si dispongono possiamo dare un senso fisico ai vari cluster.

3.8.1 L'Algoritmo del NEC

Vediamo ora con maggiore rigore come funziona l'algoritmo detto NEC:

La maggior parte dei metodi non supervisionati, e tra questi le PPS, richiedono che venga, in maniera diretta o indiretta seconda dei casi, fornito *a priori* il numero di cluster, ovviamente questo è un problema quando si sta utilizzando un data set piuttosto complesso dove il numero di cluster può essere molto grande o comunque non predicibile. Un semplice criterio di soglia non è soddisfacente nella maggior parte delle applicazioni astronomiche a causa dell'alta degenerazione e della rumorosità dei dati che possono portare a erronee agglomerazioni dei dati. Si deve quindi stabilire una de-

finizione di distanza e un criterio di accorpamento, grazie ai quali si possa stabilire se due cluster vadano uniti o meno. Ovviamente il processo può andare avanti accorpando i cluster figli del primo accorpamento dando così luogo al dendrogramma di cui ho parlato all'inizio del paragrafo.

La NEC utilizza il discriminante lineare di Fisher che è un metodo di classificazione che prima proietta i dati a molte dimensioni su una retta, e quindi svolge la classificazione sullo spazio lineare risultante [19]. La proiezione si attua massimizzando la distanza tra le medie di due classi e minimizzando la varianza all'interno di ogni classe. Inoltre si definisce l'entropia differenziale H di un vettore casuale

$$\mathbf{y} = (y_1, \dots, y_n)^T$$

con densità $f(\cdot)$ come:

$$H(\mathbf{y}) = \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}$$

cosicchè la negentropia J può essere definita come:

$$J(\mathbf{y}) = J(\mathbf{y}_{Gauss}) - H(\mathbf{y})$$

dove \mathbf{y}_{Gauss} è un vettore casuale Gaussiano con la stessa matrice di covarianza di \mathbf{y} .

La Negentropia può essere interpretata come una misura della non Gaussianità e, poichè è invariante per trasformazioni lineari invertibili, è ovvio che trovare una trasformazione invertibile che minimizza la mutua informazione è praticamente equivalente a trovare la direzione in cui la Negentropia è massimizzata. Il metodo implementato nel pacchetto AstroNeural utilizza una nozione di Negentropia che fornisce un buon compromesso tra le proprietà delle due classiche misure di non Gaussianità date da Kurtosis e la Negentropia canonica.

L'algoritmo del NEC può quindi essere usato per condurre un'agglomerazione non supervisionata dei cluster (detti anche precluster) trovati tramite le PPS. L'unica informazione *a priori* richiesta dal NEC è il valore della soglia di dissimilarità T . Supponiamo di avere n D -dimensionali precluster X_i con $i = 1, \dots, n$ che sono stati determinati dalle PPS; questi cluster sono passati alla NEC che in pratica controlla se ogni coppia di cluster contigui (secondo il discriminante lineare di Fisher) può essere o meno efficientemente

rappresentata da una singola distribuzione Gaussiana multivariata. In altre parole, la NEC determina se due cluster appartenenti a una data coppia possono essere considerati sostanzialmente distinti o parti di un cluster più grande.

Capitolo 4

La Scelta Del Campione

Demostene:

*“L’esito delle azioni umane dipende dagli dei; la scelta dell’azione rivela
l’animo dell’uomo.”*



Scegliere il campione di oggetti da analizzare è un punto cruciale.

Dagli effetti di selezione introdotti e da come si è costruita la base di conoscenza tramite la quale addestrare o interpretare le nostre reti neurali dipende fortemente il risultato che si ottiene.

Il catalogo che è stato utilizzato per i vari esperimenti è il risultato di varie manipolazioni effettuate su tre cataloghi preesistenti ottenuti da:

- G. Sorrentino, M. Radovich e A. Rifatto [25],
- G. Kauffmann et al. [26],
- R. D'Abrusco, A. Staiano, G. Longo, M. Brescia, M. Paolillo, E. De Filippis, R. Tagliaferri [27].

I primi due cataloghi sono stati necessari per costruire una cospicua base di conoscenza, mentre il terzo è stato utilizzato per sfruttare il redshift fotometrico corretto presente in quel catalogo [27].

4.1 Il Catalogo di G. Sorrentino et al.

Questo catalogo contiene dati forniti dalla SDSS data release 4 [28] selezionando un'intervallo di redshift compreso tra 0.05 e 0.095. Il limite inferiore è stato scelto per evitare problemi dovuti all'estensione dell'oggetto, mentre quello superiore altro non è che la distanza a cui il limite della luminosità raggiunge la magnitudine apparente $r = 17.77$, che come detto è il limite di completezza della SDSS. Gli autori selezionano quindi solo gli oggetti più luminosi di $M(r) = -20.00$ che è uguale a $M^+(r) + 1.45$ dove $M^+(r) = -20.8 + 5\log h$.

In totale vengono selezionati quindi 90886 oggetti. Nel 91% dei casi sono oggetti presenti sia nel catalogo fotometrico che in quello spettroscopico. Da questo campione si separano prima di tutto le galassie senza linee di emissione (che verranno da ora in poi chiamate passive), che sono il 18% ovvero 16403. Separate le passive, vengono considerate quelle il cui rapporto tra flusso ed incertezza è maggiore di 2 (che sono il 64% , ovvero 57952), mentre le altre vengono scartate.

A questo punto sono selezionati gli AGN dalle galassie in cui è presente formazione stellare (star forming galaxy, SFN) tramite il modello proposto

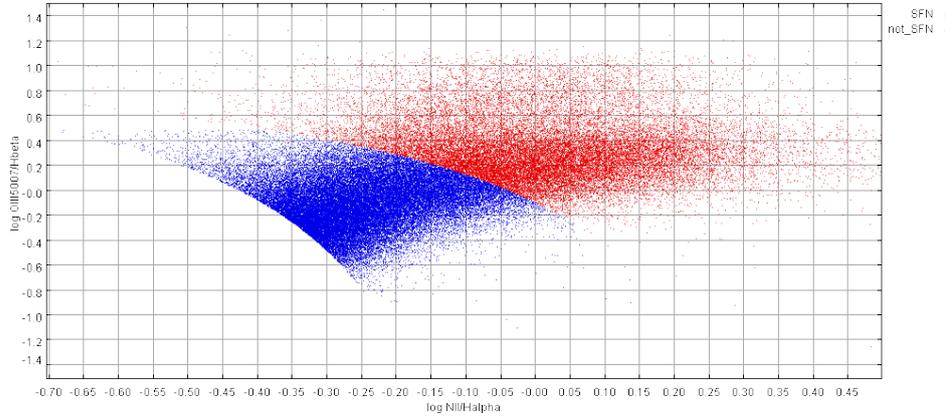


Figura 4.1: Esempio di grafico BPT, NII contro OIII

da Kewley [29]:

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.61}{\log \frac{[NII]\lambda 6583}{H\alpha} - 0.47} + 1.19 \quad (4.1)$$

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.72}{\log \frac{[SII]\lambda\lambda 6717,6731}{H\alpha} - 0.32} + 1.30 \quad (4.2)$$

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.73}{\log \frac{[OI]\lambda 6300}{H\alpha} - 0.59} + 1.33 \quad (4.3)$$

Gli oggetti al di sopra di queste curve sono AGN, quelli al di sotto sono galassie in formazione stellare (in figura 4.1); questo tipo di grafico è detto BPT dalle iniziali di coloro che lo hanno proposto [30].

I vari rapporti ci danno informazioni di vario tipo: $\frac{[OIII]}{H\beta}$ è un indicatore del livello medio di ionizzazione e temperatura mentre $\frac{[NII]}{H\alpha}$, $\frac{[OI]}{H\alpha}$ e $\frac{[SII]}{H\alpha}$ sono indicatori dell'importanza relativa delle regioni parzialmente ionizzate, regioni che vengono prodotte da fotoionizzazione di alta energia.

Per avere un ulteriore grado di confidenza si decide di trascurare ogni oggetto tale che la barra di errore li rende compatibili con le linee di Kewley

Gli oggetti che superano la riga di Kewley e che come detto non sono compatibili con quest'ultima sono classificati come Seyfert 1 e 2.

Sono definiti Seyfert 1 gli oggetti per cui risulta o

$$FWHM(H\alpha) > 1.5FWHM([OIII]\lambda 5007) \quad (4.4)$$

oppure

$$FWHM(H_\alpha) > 1200 \text{Kms}^{-1} \quad (4.5)$$

e contestualmente

$$FWHM([OIII]\lambda 5007) < 800 \text{Kms}^{-1} \quad (4.6)$$

Tutti gli altri vengono classificati come Seyfert 2

In totale si trovano quindi:

- 1829 AGN [$\sim 2\%$] di cui:
 - 725 Sy 1
 - 1105 Sy 2
- 6061 SFG [$\sim 7\%$]

Non possono invece essere classificate 50062 galassie che rappresentano circa il 55% del campione in esame.

4.2 Il Catalogo di G. Kauffmann et al.

Questo catalogo è disponibile su web all'indirizzo <http://www.mpa-garching.mpg.de/SDSS/DRA/Data/agncatalogue.html> e contiene le linee di emissione di 88178 galassie.

In particolare, per i nostri fini è stato fondamentale che contenesse $\log \frac{[OIII]\lambda 5007}{H_\beta}$ e $\log \frac{[NII]\lambda 6583}{H_\alpha}$ di modo da poterlo sezionare nella maniera più appropriata tramite i grafici BPT[30].

Poichè i dati sono stati aggiornati con la data release 4 della SDSS [28], mentre l'articolo originale era stato scritto prendendo in considerazione i dati della data release 2 [31], nell'effettuare l'aggiornamento sono andati persi i tagli effettuati al momento della stesura dell'articolo [26], quindi il primo lavoro che ho svolto è stato ripristinare i tagli sul redshift ($0.02 < z < 0.3$).

Quindi prima ho suddiviso queste galassie in tre categorie:

- pura formazione stellare (SFN)
- puri nuclei galattici attivi (AGN)

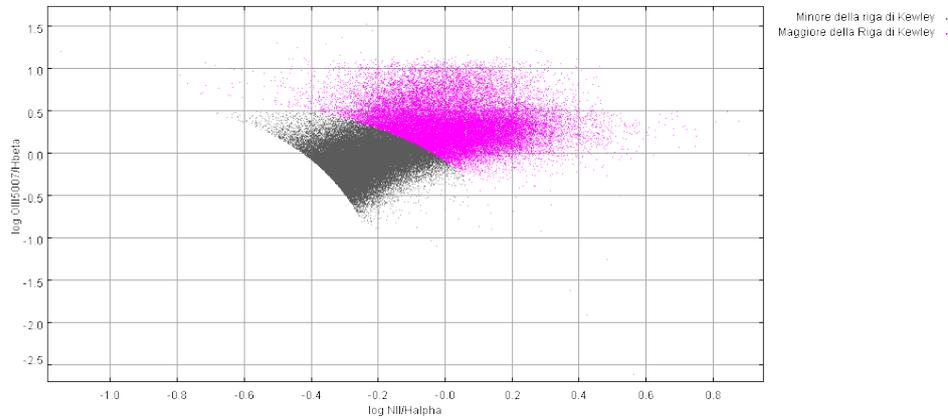


Figura 4.2: Separazione tramite la linea di Kewley

- una zona di mix.

Questa suddivisione è stata effettuata tramite l'equazione 4.1 del modello di Kewley[29] che traccia la linea di demarcazione tra la zona di mix e la zona degli AGN, come si può vedere in figura 4.2, mentre la separazione tra la zona di mix e quella delle SFN è data da:

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.61}{\log \frac{[NII]\lambda 6583}{H\alpha} - 0.05} + 1.3 \quad (4.7)$$

Questa equazione è stata proposta da Kauffman [26, 32].

Il risultato ottenuto è visibile in figura 4.3.

Ho quindi ottenuto tre categorie:

- SFN
- Mix
- Agn

Il risultato ottenuto è visibile in figura 4.4.

Ottenuta questa prima suddivisione, l'ho raffinata ulteriormente suddividendo le galassie presenti nella zona di mix e nella zona degli AGN in Seyfert e LINER ¹.

¹Low-Ionization Narrow Emission-Line Regions, come sono stati definiti da Heckman[33, 32]

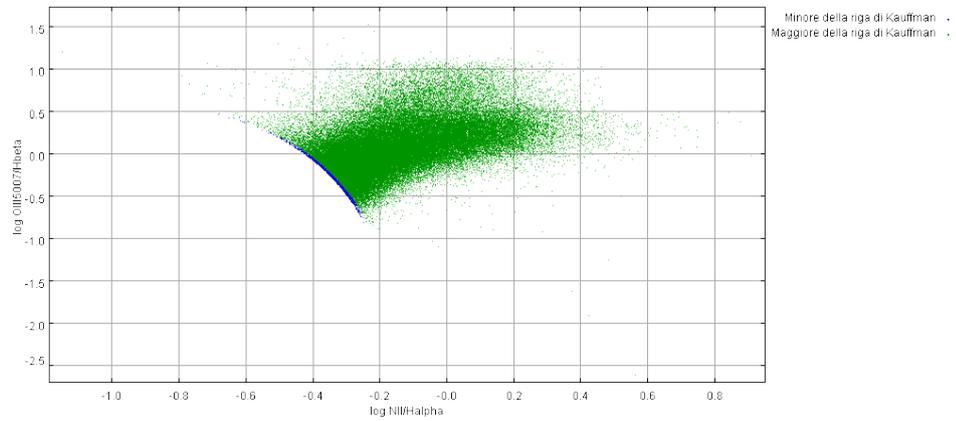


Figura 4.3: Separazione tramite la linea di Kauffman

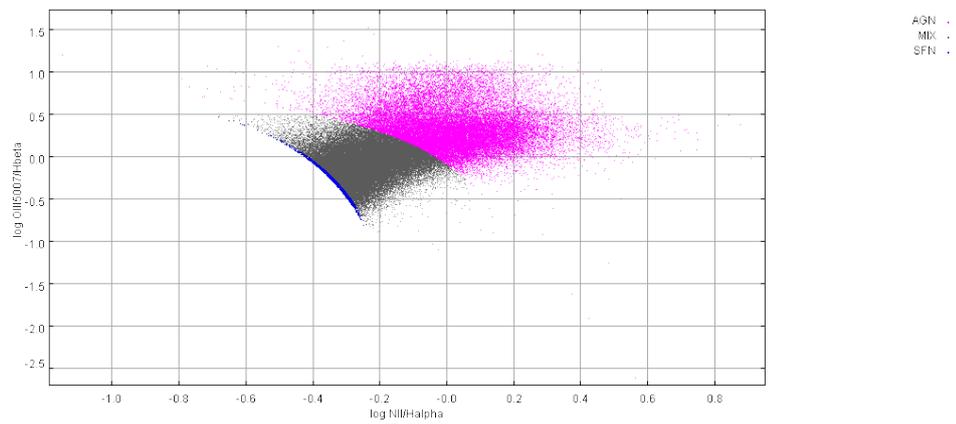


Figura 4.4: Gli oggetti raggruppati nelle tre categorie: SFN, Mix e AGN

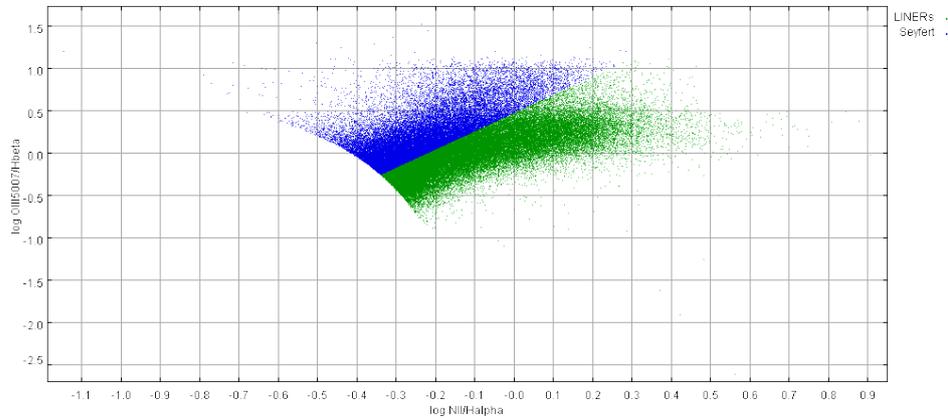


Figura 4.5: Separazione tra LINER e Seyfert

L'equazione utilizzata è la seguente [32]:

$$\frac{[OIII]\lambda 5007}{H\beta} = 2.1445 \frac{[NII]\lambda 6583}{H\alpha} + 0.465 \quad (4.8)$$

Tutto ciò che si trova al di sopra di questa retta è stato classificato come Seyfert, mentre ciò che è al di sotto viene catalogato come LINER come si vede in figura 4.5.

Si ottengono quindi cinque categorie:

- SFN
- Seyfert Mix
- Liner Mix
- Pure Seyfert
- Puri LINER

Questo risultato è riassunto in figura 4.6.

In realtà queste cinque categorie sono state divise a loro volta ognuna in tre sotto categorie in funzione del redshift; per la precisione una categoria comprende gli oggetti con redshift minore del limite inferiore imposto da G. Sorrentino, M. Radovich e A. Rifatto [25], una seconda contiene le galassie con redshift nell'intervallo scelto da questi ultimi e una terza con gli oggetti con redshift superiore.

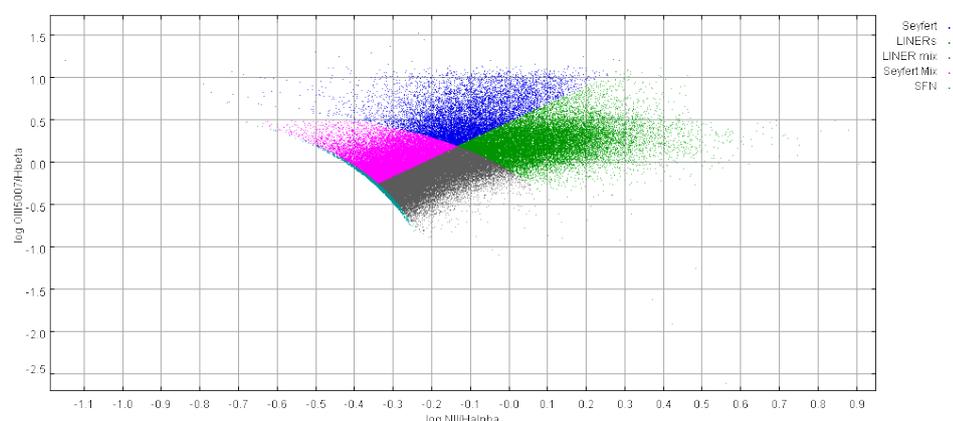


Figura 4.6: Divisione in Cinque Categorie, Galassie Normali, LINER Mix, Seyfert Mix, LINER Puri e Seyfert pure

4.3 Il Catalogo di R. D'Abrusco et al.

Questo catalogo contiene gli oggetti dello spectroscopic subset, SpS (sot-campione spettroscopico) della SDSS data release 4 [28].

Come già accennato l'importanza di questo catalogo è dovuta alla presenza di stime di z fotometrico, con un'accuratezza pari a $\sigma_{rob} = 0.02$, misurata dalla dispersione attorno allo zero della variabile scarto ($z_{phot} - z_{spec}$) [27].

4.4 Il Catalogo Usato

Avendo questi tre cataloghi a disposizione. Il successivo lavoro è stato produrre un catalogo unico che contenesse tutte le informazioni, in termini di conoscenza sulla classificazione dei singoli oggetti, che avevo estratto dai singoli cataloghi. Il processo si è articolato in due fasi: la prima è stata l'unione dei cataloghi di Kauffman e di Sorrentino, utilizzando dove non c'era sovrapposizione l'unica informazione disponibile, ovvero la catalogazione precedentemente ottenuta (quattro possibilità nel caso del catalogo di Sorrentino e quindici per quello di Kauffman), mentre nei casi di sovrapposizione tra i due cataloghi sono state mantenute le etichette ricavate dal catalogo di Kauffman per le passive e per le SFN di Sorrentino che si incrociano con qualunque cosa che non siano le SFN del catalogo di Kauffman. Le passive che si intersecano con le SFN di Kauffman sono state mantenute

come passive. Le Seyfert 1 e le Seyfert 2 di Sorrentino sovrapposte alle Seyfert e Mix di Kauffman) vengono mantenute come Seyfert 1 e Seyfert 2 nel primo caso e vengono indicate come Seyfert 1 mix e Seyfert 2 mix nel secondo caso, Stessa cosa per i LINER che si divideranno in LINER 1 , LINER 2, LINER 1 mix e LINER 2 mix.

Il totale di possibilità è quindi pari a 26 diverse possibili denominazioni.

Dall'unione di questi due cataloghi ho ottenuto un totale di 108162 galassie.

Ho proceduto quindi a incorporare questa "etichettatura" nel catalogo di D'Abrusco [27]; il numero di galassie presenti nell'unione dei due precedenti cataloghi e nel catalogo di D'Abbrusco è pari a 100069.

Capitolo 5

VO-Neural

Albert Einstein:

“I computer sono incredibilmente veloci, accurati e stupidi. Gli uomini sono incredibilmente lenti, inaccurati e intelligenti. Insieme sono una potenza che supera l’immaginazione.”



Parte del mio lavoro di tesi si è svolto nell’ambito del progetto VO-Neural.

VO-Neural rappresenta la naturale evoluzione del pacchetto AstroNeural, progetto nato nel 1994 dalla collaborazione tra il dipartimento di matematica e informatica dell’università di Salerno e il Dipartimento di Scienze Fisiche dell’Università Federico II.

Il pacchetto AstroNeural, terminato nel 2006, comprende un set di strumenti per il data mining astronomico tra cui MLP, PPS e NEC.

Il progetto VO-Neural, fondato all'interno del consorzio europeo chiamato VOTECH, ha come scopo di rendere alcuni degli strumenti sviluppati con AstroNeural, utilizzabili dalla comunità astronomica sotto forma di web service.

Allo stato attuale il progetto ha già rilasciato una prima versione delle reti MLP che è già disponibile per tutti gli utenti di AstroGrid.

La mia collaborazione a questo progetto ha investito diversi aspetti:

- produzione di casi scientifici con cui mostrare l'utilità del software alla comunità
- significativo contributo alla fase progettuale
- modifiche del codice delle reti MLP inizialmente sviluppato da C. Donalek per adattarsi alle nuove esigenze del progetto
- test e debug del codice delle reti MLP
- test e debug dell'intera pipeline del software prodotto prima in versione da shell, successivamente della versione provvisoria XML e poi della versione definitiva
- creazione delle librerie di funzioni in linguaggio C per implementare le matrici di confusione nella prossima release del software
- collaborazione allo sviluppo del sito del progetto
- studio sulla possibilità di importare le SVM sotto VO-Neural

5.1 Principali Modifiche al Codice delle Reti MLP

Questo paragrafo si occupa di esplicitare le modifiche da me apportate, sotto la supervisione di C. Donalek, al codice delle MLP. Di fatto si tratta di piccole modifiche a quello che possiamo considerare il "main" del programma.

Le modifiche apportate sono le seguenti:

- Aggiunta dei target vector nei file di output

- Aggiunta della modalità denominata “run”, dal momento che data la modifica precedente la modalità di test non era più utilizzabile per fare correre la rete addestrata su nuovi dataset
- Eliminazione della richiesta dei parametri ridondanti nelle modalità di test e run
- Eliminazione dei filename hard-coded in modo che tutti i file di uscita fossero dipendenti da un parametro scelto dall’utente.
- Eliminazione degli input hiddenlayers, layer2, errorfunc che si è deciso di lasciare fissati
- Eliminazione dei file di debug prodotti dalla rete
- Aggiunta di funzioni che scrivono lo stato del processo su un file di log
- Aggiunta di un controllo sulle dimensioni dei dataset
- Aggiornamento dell’help del codice.
- Creazione di un tool minimale di validazione

5.1.1 Aggiunta dei Target Vector

Ho provveduto ad aggiungere il target vector nei file di output della nostra MLP. Nelle modalità “train, test e complete” è previsto un file di uscita di tipo ASCII che nella versione originaria conteneva i soli output vector a cui ora sono stati aggiunti i target vector relativi. E’ riportato in appendice A, come esempio, il codice aggiunto per la modalità di “train”.

5.1.2 Aggiunta della Modalità di Run

Dopo la modifica di cui sopra, la modalità di test ha perso per ovvie ragioni la sua doppia funzionalità; non poteva cioè essere più utilizzata, come usualmente si fa, anche per fare correre la rete su nuovi dataset. Quindi si è provveduto ad aggiungere una quarta modalità, che sostanzialmente, riprende il codice della modalità di test senza considerare i target vector.

Il codice che regola questa modalità è riportato in appendice A.

5.1.3 Rimozione dei Parametri Ridondanti

I parametri “nin” e “nout”, nonostante venissero salvati nella rete, non erano da questa prelevati all’atto di eseguire le modalità di test e run. Allo scopo di ridurre al minimo i parametri che l’utente deve dare nel momento in cui lancia la nostra MLP si è provveduto a importarli dal file della rete.

Si noti che alla MLP vanno passati comunque tutti i parametri necessari ma alcuni di essi non sono utilizzati e quindi possono essere forniti in maniera arbitraria.

Questo è importante in quanto questo codice è solo una parte del software integrale, difatti l’utente fornisce esclusivamente i parametri necessari all’utilizzo ed è uno script bash a fornire arbitrariamente quelli “inutili”.

L’elenco dei parametri necessari è riportato in figura 5.1.

TRAIN	TEST	FULL	RUN
input	input	input	input
typeRun	typeRun	typeRun	typeRun
inpNodes		inpNodes	
outNodes		outNodes	
layer1		layer1	
iter		iter	
actFunc		actFunc	
optFunc		optFunc	
errorTol		errorTol	
	inputNet		inputNet
inputDataTr		inputDataTr	
inputDataVa		inputDataVa	
	inputDataTe	inputDataTe	inputDataTe
outputData	outputData	outputData	outputData
outputData	outputData	outputData	outputData
[outputData].tra	[outputData].tes	[OutputData].tra	[outputData].run
[outputData]_netTrain.mat		[outputData]_netTrain.mat	
		[outputData].tes	

Figura 5.1: Schema dei parametri necessari al codice mlp e relative uscite divise in funzione della modalità prescelta

5.1.4 Eliminazione dei Filename Hard-Coded

Nella sua forma originaria il programma prevedeva che alcuni file prodotti dall'MLP avessero un nome fissato; questo comportamento è stato corretto facendo dipendere il nome del file di uscita dal parametro "output-Data" fornito dall'utente all'atto di lanciare il programma. Questa modifica che pareva necessaria in un primo momento è divenuta di utilità marginale nel momento in cui si è deciso di "zippare" i file di uscita.

La modifica in questione non è stata comunque rimossa, visto che non dà luogo a problemi e/o complicazioni di sorta e anzi è una scelta comunque preferibile.

5.1.5 Eliminazione degli Input: hiddenlayers, layer2 e errorfunc

Si è provveduto a rimuovere dal codice la possibilità di impostare i parametri hiddenlayers, layer2 e errorfunc che risultano ora fissati come segue:

- hiddenlayers = 1
- layer2 = 0
- errorfunc = 2 (che corrisponde alla cross-entropy)

5.1.6 Eliminazione dei File di Debug Prodotti dalla Rete

La rete produceva alcuni file nel formato nativo di Matlab (.mat), utili solo ai fini del debug. Questi file contenevano gli output vector ed erano prodotti dalle modalità test, complete e run (nella modalità di run era stato aggiunto per omogeneità).

Tali file sono stati eliminati nella versione definitiva.

5.1.7 Aggiunta di Funzioni che Scrivono un File di Log

Il programma provvede ora alla scrittura su un file di log di alcune informazioni sullo stato del processo. Di seguito sono elencate tale informazioni:

- Data e ora di inizio del job

- Se i dataset sono accettati dal controllo illustrato nel paragrafo successivo scrive che sono stati correttamente caricati e il loro nome, diversamente scrive un messaggio di errore
- Durante ogni fase (train, test e run) scrive data e ora di avvio e di fine
- Data e ora di fine processo
- Qualora il caricamento dei file abbia dato errore conclude il file con la scritta: “error”
- Nel caso si verificano errori li riporta in un file a parte che contiene solo questi ultimi

In appendice A sono presenti due esempi di file di log.

Il file di log viene scritto in modalità “a+” (se il file esiste continua a scriverci senza sovrascriverlo mentre se il file non esiste viene creato).

5.1.8 Aggiunta di un Controllo sulle Dimensioni dei Dataset

Al momento di caricare un file, il programma controlla le dimensioni del dataset e le confronta con nin e nout.

In modalità test, train e complete controlla che il numero delle colonne sia maggiore o uguale della somma di nin e nout.

In modalità di run controlla che il numero delle colonne sia maggiore o uguale del nin e, se questo requisito è verificato, procede nell’elaborazione, diversamente riporta sul file di log il problema, imposta un typeRun diverso da quelli corrispondenti alle quattro modalità possibili e fa proseguire il programma, che non dovendo fare altro, termina e riporta nel file di log la scritta: “error”.

5.1.9 Aggiornamento dell’ Help On-Line

Il programma in intestazione presenta un piccolo help che è stato aggiornato. Questo help è riportato in calce

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% function voneural_MLP
%  Ciro Donalek
% donalek@astro.caltech.edu
```


5.1.10 Creazione di un Tool Minimale di Validazione

Dato che il codice non prevedeva un sistema di validazione, visti e considerati gli stretti tempi per il rilascio della versione, e il fatto che si stava già lavorando a una nuova versione completamente differente si è deciso di non investire troppo tempo nella stesura di un sistema di validazione che doveva essere in breve tempo sostituito.

Il sistema attualmente utilizzato si occupa di calcolare il grado di riconoscimento ottenuto dalla rete sul dataset di train e confrontare questo risultato con quello ottenuto sul dataset di validazione, i risultati, qualora la validazione sia richiesta, vengono riportati nel file di log.

Il codice si trova in appendice A.

5.2 Librerie per le Matrici di Confusione

Questo pacchetto di funzioni serve per produrre dei semplici strumenti di diagnostica che mirano a stimare l'efficienza di una rete neurale.

Il principale strumento prodotto sono le matrici di confusione.

Una matrice di confusione è la rappresentazione discretizzata della relazione tra target e output vector, il suo essere discreta la rende ovviamente inadatta a rappresentare le reti neurali che mirano a ricostruire un parametro continuo ma è perfetta per tutti gli altri casi, in particolare per i tentativi di classificazione.

Per discretizzata intendo che gli elementi della matrice sono il numero di oggetti che ha valore all'interno di un certo bin¹.

In che modo questa matrice rappresenti l'efficienza della rete è presto detto: ogni elemento della matrice rappresenta un possibile "accoppiamento" tra target e output vector e la traccia della matrice corrisponde al numero di oggetti correttamente classificati.

Il pacchetto oltre a calcolare questa matrice permette di calcolare il tasso di successo, ovvero la percentuale di oggetti correttamente classificati (il valore della traccia diviso il numero di oggetti totali); oltre a questo calcola il numero di "bit fault" che sono gli oggetti che cadono fuori dalla traccia e la percentuale di oggetti correttamente classificati per ogni classe.

Ovviamente la percentuale di oggetti correttamente classificati per ogni

¹unitario nel lavoro che segue ma che in linea di principio può essere qualsiasi

classe è rappresentata per ogni riga dall'elemento presente sulla diagonale principale diviso per la somma degli elementi della riga.

Ognuna delle funzioni è sviluppata in due versioni, chiamate normali e avanzate; la differenza tra le due risiede nel formato in cui vengono forniti i vettori.

La versione normale prevede vettori monodimensionali in cui sono presenti un numero di classi determinato dalla variabile "num " mentre nella versione avanzata i vettori sono "num"-dimensionali e su ogni colonna sono possibili due classi (0 e 1).

In questo caso le classi vengono ricavate utilizzando una conversione dal binario (dove ogni colonna rappresenta un bit) al decimale.

5.2.1 Funzioni per la Creazione delle Matrici

alloc_confusion_matrix e *alloc_confusion_matrix_advanced* inizializzano una matrice $N \times N$ dove N è il numero di classi richiesto.

Input:

- Numero di classi o di colonne a seconda della scelta della versione della funzione.

Output:

- Una matrice vuota $N \times N$ dove N è il numero di classi richiesto.

5.2.2 Funzioni per l'Incremento delle Matrici

update_confusion_matrix e *update_confusion_matrix_advanced* prendono un target vector e un output vector e aggiungono 1 all'elemento della matrice corretto.

Input:

- Una Matrice di Confusione
- Un Target Vector
- Un Output Vector

- Numero di classi o di colonne a seconda della scelta della versione della funzione.

Output:

- Una Matrice di Confusione modificata

5.2.3 Funzioni per il Salvataggio su File delle Matrici

print_confusion_matrix e *print_confusion_matrix_advanced* calcolano il tasso di successo, il numero di bit fault e il tasso di successo per ogni classe; dopodichè scrivono tutto ciò sul file di output.

Input:

- Numero di classi o di colonne a seconda della scelta della versione della funzione.
- Una Matrice di Confusione
- Nome del File di Output

Output:

- Un File Output

5.2.4 Funzioni per Liberare la Memoria Precedentemente Occupata dalle Matrici

destroy_conf_mat e *destroy_conf_mat_advanced* liberano le celle di memoria utilizzate dalla matrice di confusione.

Input:

- Una Matrice di Confusione
- Numero di classi o di colonne a seconda della scelta della versione della funzione.

Output:

- Nessuno

Capitolo 6

Gli Esperimenti

Arthur Bloch:

*Legge della futilità: “Nessun esperimento è mai completamente fallito:
può sempre servire da esempio negativo.”*



Prima di giungere a dei risultati utili sono passato per un cospicuo numero di esperimenti falliti che però hanno sempre avuto un ruolo importante

nella progressiva comprensione dell'utilizzo dei parametri in gioco.

In quanto segue presenterò tuttavia soltanto quegli esperimenti che ritengo in qualche misura significativi.

Tutti gli addestramenti sono stati condotti con il pacchetto AstroNeural e con il massimo numero di interazioni permesso da quest'ultimo: 2010.

6.1 Features

Le features utilizzate sono le seguenti, se non viene diversamente indicato gli esperimenti si intendono effettuati utilizzandole tutte:

1. *petroR50_u*
2. *petroR50_g*
3. *petroR50_r*
4. *petroR50_i*
5. *petroR50_z*
6. *concentration_index_r*
7. *z_phot_corr*
8. *fibermag_r*
9. $(u - g)_{dered}$
10. $(g - r)_{dered}$
11. $(r - i)_{dered}$
12. $(i - z)_{dered}$
13. *dered_r*

6.2 Selezione dei Candidati AGN Tramite MLP

Quelli che seguono sono alcuni esperimenti tramite i quali ho ottenuto risultati molto simili tra loro che sono degni di nota.

Le reti sono state addestrate con un target vector pari a 1 per gli oggetti ritenuti sicuri AGN, ovvero gli oggetti superiori alla riga di Kewley e 0 tutti quelli al di sotto di questa linea: zona di mix e galassie normali.

Lo scopo di queste reti è di trovare dei candidati AGN.

Una volta addestrata la rete, si possono identificare futuri candidati AGN con i soli parametri fotometrici.

Si ricorda che il redshift degli oggetti utilizzati è compreso tra 0.02 e 0.3

6.2.1 Logistic SCG con 20 Hidden

La rete ha un tasso di riconoscimento del 75.35% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3444	2535
0(known)	1650	9348

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 57.60%

resto: 84.99%

Quindi solo $\sim 15\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia. Il risultato è 1.20%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 1.32% dei casi le classifica come AGN.

6.2.2 Logistic Conji Grad con 20 Hidden

La rete ha un tasso di riconoscimento del 75.03% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3423	2556
0(known)	1683	9315

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 57.25%

resto: 84.70%

Quindi solo $\sim 15.3\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 1.24%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 1.37% dei casi le classifica come AGN.

6.2.3 Logistic Quasi Newton con 20 Hidden

La rete ha un tasso di riconoscimento del 75.63% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3385	2594
0(known)	1529	9469

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 56.61%

resto: 86.10%

Quindi solo $\sim 14\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 0.95%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 0.98% dei casi le classifica come AGN.

6.2.4 Linear SCG con 20 Hidden

La rete ha un tasso di riconoscimento del 75.46% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3358	2621
0(known)	1545	9453

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 56.16%

resto: 85.95%

Quindi solo $\sim 14\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 0.85%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 0.88% dei casi le classifica come AGN.

6.2.5 Linear Conji Grad con 20 Hidden

La rete ha un tasso di riconoscimento del 75.53% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3395	2584
0(known)	1571	9427

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 56.78%

resto: 85.72%

Quindi solo $\sim 14.3\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 0.98%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 1.03% dei casi le classifica come AGN.

6.2.6 Linear Quasi Newton con 20 Hidden

La rete ha un tasso di riconoscimento del 75.67% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3391	2588
0(known)	1543	9455

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 56.72%

resto: 85.97%

Quindi solo $\sim 14\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 1.00%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 1.03% dei casi le classifica come AGN.

6.2.7 Logistic SCG con 52 Hidden

La rete ha un tasso di riconoscimento del 75.74% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3356	2658
0(known)	1460	9403

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 55.12%

resto: 87.23%

Quindi solo $\sim 12.8\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 1.12%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 1.10% dei casi le classifica come AGN.

6.2.8 Logistic Conji Grad con 52 Hidden

La rete ha un tasso di riconoscimento del 75.54% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3401	2713
0(known)	1440	9423

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 55.63%

resto: 86.74%

Quindi solo $\sim 13.3\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 1.15%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 1.11% dei casi le classifica come AGN.

6.2.9 Logistic Quasi Newton con 52 Hidden

La rete ha un tasso di riconoscimento del 75.96% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3455	2659
0(known)	1422	9441

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 56.51%

resto: 86.91%

Quindi solo $\sim 13.1\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 0.98%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 0.94% dei casi le classifica come AGN.

6.2.10 Linear SCG con 52 Hidden

La rete ha un tasso di riconoscimento del 75.67% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3370	2744
0(known)	1387	9476

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 55.12%

resto: 87.23%

Quindi solo $\sim 13.8\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 1.05%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 0.98% dei casi le classifica come AGN.

6.2.11 Linear Conji Grad con 52 Hidden

La rete ha un tasso di riconoscimento del 75.44% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3336	2778
0(known)	1391	9474

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:

agn : 54.56%

resto: 87.20%

Quindi solo $\sim 12.8\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 1.06%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 0.99% dei casi le classifica come AGN.

6.2.12 Linear Quasi Newton con 52 Hidden

La rete ha un tasso di riconoscimento del 75.99% dando luogo alla seguente matrice di confusione:

	1(net)	0(net)
1(known)	3402	2712
0(known)	1364	9499

Quindi si ottengono le seguenti percentuali di riconoscimento corretto:
percentuali:

agn : 55.64%

resto: 87.44%

Quindi solo $\sim 12.6\%$ degli oggetti di questa incertezza viene classificata come AGN.

A questo punto si può verificare quale percentuale di questi oggetti che viene classificata come AGN pur non essendolo è in realtà una galassia.

Il risultato è 0.89%.

Inoltre passando al nostro classificatore un elenco di galassie sicure questi solo nell' 0.82% dei casi le classifica come AGN.

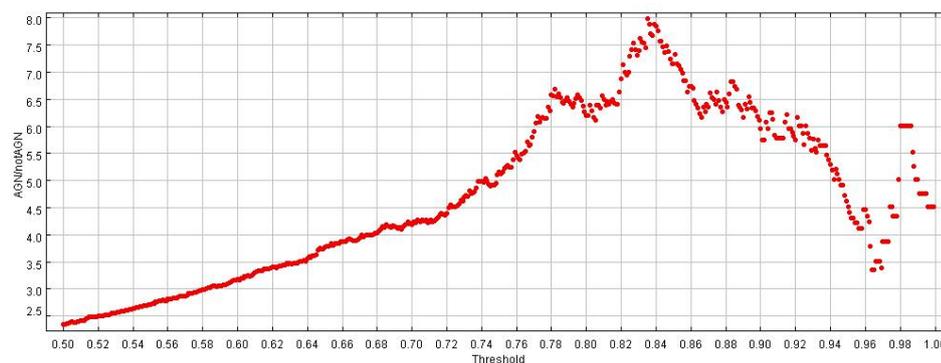


Figura 6.1: Rapporto tra AGN Correttamente Valutati e Non AGN Classificati Male, Prima Metà del TestSet

6.2.13 Riepilogo e Considerazioni

Il riepilogo dei risultati sinora presentati è presente in figura 6.4.

La migliore classificazione è stata ottenuta con funzione di attivazione Linear e di ottimizzazione Quasi Newton, con 52 neuroni nascosti.

Questo risultato va analizzato sottolineando ancora una volta che sotto l'etichetta 0 si trovano sia galassie certe che gli oggetti della zona di MIX per i quali i grafici BPT non permettono di dire molto.

Si è quindi cercata una soluzione che da un lato garantisca una buona purezza e dall'altro non diminuisce troppo l'efficienza.

Come si è visto la contaminazione dovuta a galassie certe è molto piccola, e la contaminazione causata dagli oggetti della zona di MIX potrebbe essere parzialmente attribuibile a AGN non riconosciuti. Ad ogni modo ho tentato di massimizzare il rapporto tra AGN correttamente riconosciuti e oggetti etichettati come zero attribuiti alla categoria 1 dalla rete. Ciò è stato fatto facendo variare il valore della soglia, prima selezionando casualmente metà del TestSet, quindi l'ho rifatto con l'altra metà e infine con l'intero TestSet, questo per evitare di considerare fluttuazioni statistiche, i risultati sono nelle figure 6.1, 6.2 e 6.3.

Si osserva che c'è un massimo intorno ad una soglia di 0.837, considerando però di utilizzare questa soglia la matrice di confusione diverrebbe:

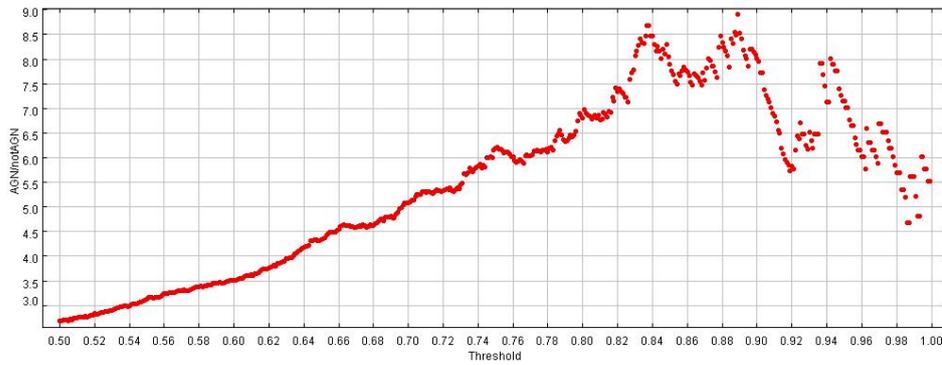


Figura 6.2: Rapporto tra AGN Correttamente Valutati e Non AGN Classificati Male, Seconda Metà del TestSet

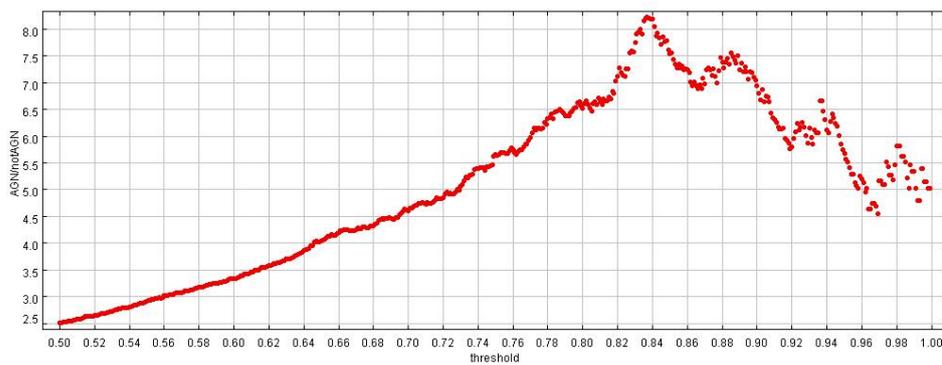


Figura 6.3: Rapporto tra AGN Correttamente Valutati e Non AGN Classificati Male, Analisi Totale del TestSet

1	2	3	4	5	6	7	8
Logistic	SCG	20	75.35%	57.60%	84.99%	1.20%	1.32%
Logistic	Conji Grad	20	75.03%	57.25%	84.70%	1.24%	1.37%
Logistic	Quasi Newton	20	75.63%	56.61%	86.10%	0.95%	0.98%
Linear	SCG	20	75.46%	56.16%	85.95%	0.85%	0.88%
Linear	Conji Grad	20	75.53%	56.78%	85.72%	0.98%	1.03%
Linear	Quasi Newton	20	75.67%	56.72%	85.97%	1.00%	1.03%
Logistic	SCG	52	75.74%	55.12%	87.23%	1.12%	1.10%
Logistic	Conji Grad	52	75.54%	55.63%	86.74%	1.15%	1.11%
Logistic	Quasi Newton	52	75.96%	56.51%	86.91%	0.98%	0.94%
Linear	SCG	52	75.67%	55.12%	87.23%	1.05%	0.98%
Linear	Conji Grad	52	75.44%	54.56%	87.20%	1.06%	0.99%
Linear	Quasi Newton	52	75.99%	55.64%	87.44%	0.89%	0.82%

Figura 6.4: Schema Riassuntivo dei Risultati Ottenuti: nella colonna 1 è presente la funzione di attivazione usata, nella colonna 2 è presente l'algoritmo di ottimizzazione utilizzato, nella colonna 3 è presente il numero di neuroni nascosti usati, nella colonna 4 è riportato il tasso di riconoscimento globale della rete, nella colonna 5 è riportato il tasso di riconoscimento dei soli AGN, nella colonna 6 è riportato il tasso di riconoscimento per il resto, nella colonna 7 viene riportata che percentuale degli oggetti catalogati e mal classificati dalla rete è una galassia sicura, nella colonna 8 viene riportata che percentuale di un catalogo di sole galassie viene classificata dalla rete come AGN.

	1(net)	0(net)
1(known)	558	5556
0(known)	68	10795

Che porterebbe il tasso di riconoscimento degli AGN al 9.13% valore estremamente basso.

6.3 Separazione AGN non AGN Tramite SVM

Il procedimento utilizzato per la ricerca della rete migliore è quello proposto da Chih-Wei Hsu, Chih-Chung Chang, e Chih-Jen Lin [12]; in pratica sono stati eseguiti 110 addestramenti che sono corrispondenti a 110 accoppiamenti tra il parametro C e il parametro gamma, il range di variazione è $C = 2^{-5}, 2^{-3}, \dots, 2^{15}, \gamma = 2^{-15}, 2^{-13}, \dots, 23$; c'è un fattore 4 tra un punto e il successivo, gli addestramenti sono stati fatti utilizzando la cross validation facendo dividere il dataset di train in 5 parti.

Dato il peso computazionale è praticamente impossibile eseguire questi processi in serie su un desktop, e si è quindi deciso di usare la tecnologia

6.4. SEPARAZIONE OGGETTI DI TIPO 1 E TIPO 2 TRAMITE MLP135

GRID, usando le macchine del progetto SCoPE. Ottenendo i risultati in tempi ragionevoli. A tal fine si è usato uno script in python per l'invio e l'analisi dei risultati dei processi.

Il dataset è lo stesso considerato nel caso delle MLP, l'80% dei dati è stato utilizzato per addestrare la rete e il 20% per testarla.

La coppia di parametri risultata migliore è $C = 32768$ e $\gamma = 0.001953125$, il grafico che descrive la ricerca è in figura 6.5, il tasso di riconoscimento ottenuto è stato del 75.76%

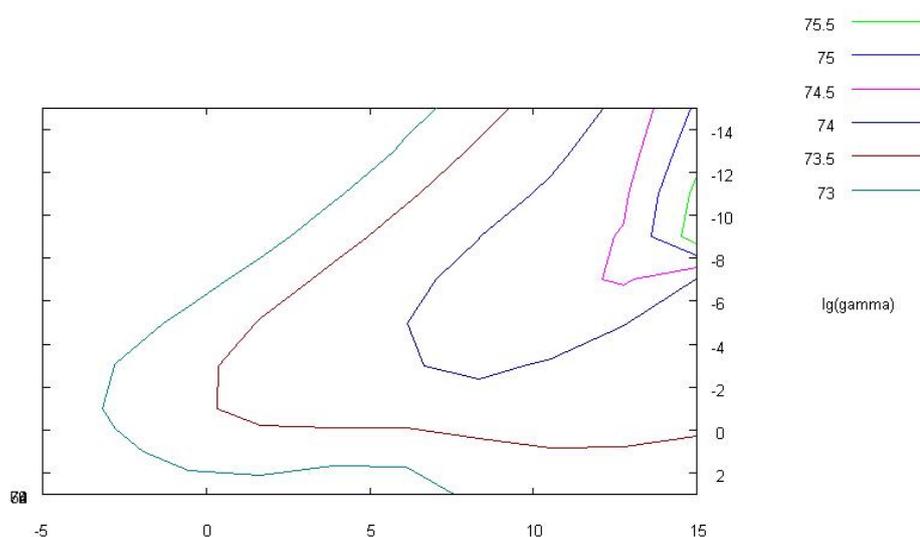


Figura 6.5: Ricerca a Griglia della miglior coppia di parametri, i logaritmi si intendono in base 2

Il risultato è peggiore di quello ottenuto con le MLP

6.4 Separazione Oggetti di Tipo 1 e Tipo 2 Tramite MLP

L'esperimento è stato effettuato selezionando soltanto gli oggetti che nel catalogo di G. Sorrentino et al. venivano indicati come Tipo 1 e Tipo 2, questo indipendentemente da se risultassero o meno appartenere alla categoria Seyfert o LINER e avendo cura di selezionare solo quelli la cui appartenenza agli AGN fosse certa.

Il dataset si componeva quindi di 1570 oggetti: ho indicato come 1 gli oggetti di Tipo 1 e come 0 gli oggetti di Tipo 2.

Va ricordato che gli oggetti utilizzati appartenendo esclusivamente al catalogo di G. Sorrentino et al. hanno redshift compreso tra 0.05 e 0.095

6.4.1 Logistic Linear 20 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	116	10
0(known)	21	167

Il tasso di riconoscimento è quindi pari al 90.1%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 92.1%

Tipo 2 : 88.8%

6.4.2 Logistic Conji Grad 20 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	124	2
0(known)	7	181

Il tasso di riconoscimento è quindi pari al 97.1%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 98.4%

Tipo 2 : 96.3%

6.4. SEPARAZIONE OGGETTI DI TIPO 1 E TIPO 2 TRAMITE MLP137

6.4.3 Logistic Quasi Newton 20 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	126	0
0(known)	2	186

Il tasso di riconoscimento è quindi pari al 99.4%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 100%

Tipo 2 : 98.9%

6.4.4 Linear SCG 20 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	119	7
0(known)	20	168

Il tasso di riconoscimento è quindi pari al 91.4%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 94.4%

Tipo 2 : 89.4%

6.4.5 Linear Conji Grad 20 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	114	12
0(known)	17	171

Il tasso di riconoscimento è quindi pari al 90.8%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 90.5%

Tipo 2 : 91.0%

6.4.6 Linear Quasi Newton 20 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	114	12
0(known)	9	179

Il tasso di riconoscimento è quindi pari al 93.3%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 90.5%

Tipo 2 : 95.2%

6.4.7 Logistic SCG 52 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	122	4
0(known)	9	179

Il tasso di riconoscimento è quindi pari al 95.9%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 96.8%

6.4. SEPARAZIONE OGGETTI DI TIPO 1 E TIPO 2 TRAMITE MLP139

Tipo 2 : 95.2%

6.4.8 Logistic Conji Grad 52 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	124	2
0(known)	7	181

Il tasso di riconoscimento è quindi pari al 97.1%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 98.4%

Tipo 2 : 96.3%

6.4.9 Logistic Quasi Newton 52 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	126	0
0(known)	3	185

Il tasso di riconoscimento è quindi pari al 99.0%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 100%

Tipo 2 : 98.4%

6.4.10 Linear SCG 52 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	92	34
0(known)	21	167

Il tasso di riconoscimento è quindi pari al 82.5%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 88.8%

Tipo 2 : 73.0%

6.4.11 Linear Conji Grad 52 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	116	10
0(known)	20	168

Il tasso di riconoscimento è quindi pari al 86.0%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 92.1%

Tipo 2 : 89.4%

6.4.12 Linear Quasi Newton 52 Hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	122	4
0(known)	8	180

6.4. SEPARAZIONE OGGETTI DI TIPO 1 E TIPO 2 TRAMITE MLP141

Il tasso di riconoscimento è quindi pari al 96.2%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 96.8%

Tipo 2 : 95.7%

6.4.13 Riepilogo e Considerazioni

Il riepilogo dei risultati ottenuti per questa classificazione è presentato in figura 6.6.

Nonostante i pochi punti a disposizione la separazione appare evidente, i tassi di riconoscimento sono estremamente alti, il miglior risultato è stato ottenuto con la modalità Logisti Quasi Newton con 20 neuroni nascosti.

1	2	3	4	5	6
Logistic	SCG	20	90.1%	92.1%	88.8%
Logistic	Conji Grad	20	97.1%	98.4%	96.3%
Logistic	Quasi Newton	20	99.4%	100.0%	98.9%
Linear	SCG	20	91.4%	94.4%	89.4%
Linear	Conji Grad	20	90.8%	90.5%	91.0%
Linear	Quasi Newton	20	93.3%	90.5%	95.2%
Logistic	SCG	52	95.9%	96.8%	95.2%
Logistic	Conji Grad	52	97.1%	98.4%	96.3%
Logistic	Quasi Newton	52	99.0%	100.0%	98.4%
Linear	SCG	52	82.5%	88.8%	73,0%
Linear	Conji Grad	52	86.0%	92.1%	89.4%
Linear	Quasi Newton	52	96.2%	96.8%	95.7%

Figura 6.6: Schema Riassuntivo dei Risultati Ottenuti: nella colonna 1 è presente la funzione di attivazione usata, nella colonna 2 è presente l'algoritmo di ottimizzazione utilizzato, nella colonna 3 è presente il numero di neuroni nascosti usati, nella colonna 4 è riportato il tasso di riconoscimento globale della rete, nella colonna 5 è riportato il tasso di riconoscimento dei soli oggetti di tipo 1, nella colonna 6 è riportato il tasso di riconoscimento dei soli oggetti di tipo 2.

6.5 Separazione Oggetti di Tipo 1 e Tipo 2 Tramite SVM

Il dataset è lo stesso considerato nel caso delle MLP, l'80% dei dati è stato utilizzato per addestrare la rete e il 20% per testarla.

La coppia di parametri risultata migliore è $C = 512.0$ e $\gamma = 0.03125$, il grafico che descrive la ricerca è in figura 6.7

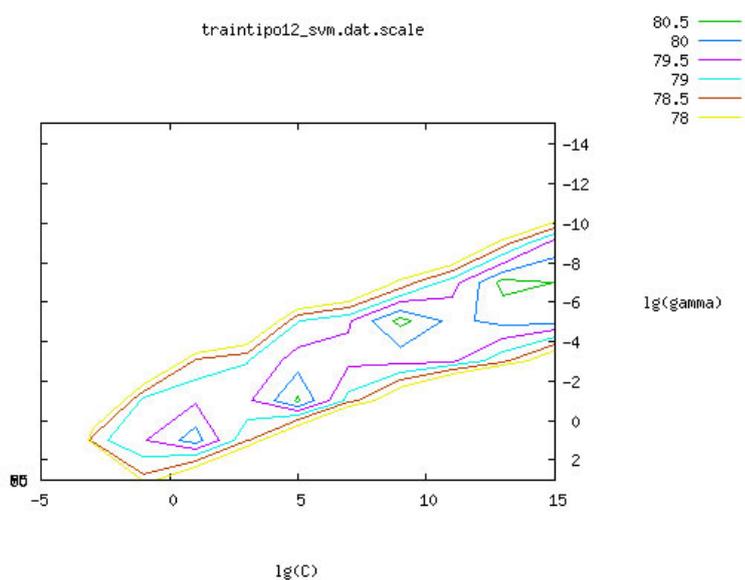


Figura 6.7: Ricerca a Griglia della miglior coppia di parametri, i logaritmi si intendono in base 2

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	87	36
0(known)	24	177

Il tasso di riconoscimento è quindi pari al 81.5%.

La rete, chiamata a distinguere tra Tipo 1 e Tipo 2, risponde riconoscendo le due classi con le seguenti percentuali:

Tipo 1 : 70.7%

Tipo 2 : 88.1%

Il risultato è peggiore di quello ottenuto con le MLP

6.6 Separazione Seyfert LINER Tramite MLP

Il dataset utilizzato è stato costruito selezionando tutti gli oggetti che trovandosi al di sopra della linea di Kewley avendo classificato come Seyfert o come LINER. Il numero di oggetti presenti nel dataset è pari a 30380. Sono stati indicati gli oggetti di tipo Seyfert con 1 e quelli di tipo LINER con 0.

6.6.1 Logistic SCG con 20 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	903	969
0(known)	353	3851

Il tasso di riconoscimento è quindi pari al 78.24%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 48.23%

LINER : 91.60%

6.6.2 Logistic Conji Grad con 20 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	881	991
0(known)	348	3856

Il tasso di riconoscimento è quindi pari al 77.96%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 47.06%

LINER : 91.72%

6.6.3 Logistic Quasi Newton con 20 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	899	973
0(known)	344	3860

Il tasso di riconoscimento è quindi pari al 78.32%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 48.02%

LINER : 91.82%

6.6.4 Linear SCG con 20 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	873	999
0(known)	332	3872

Il tasso di riconoscimento è quindi pari al 78.09%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 46.63%

LINER : 92.10%

6.6.5 Linear Conji Grad con 20 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	863	1009
0(known)	329	3875

Il tasso di riconoscimento è quindi pari al 77.98%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 46.10%

LINER : 92.17%

6.6.6 Linear Quasi Newton con 20 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	904	968
0(known)	352	3852

Il tasso di riconoscimento è quindi pari al 78.28%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 48.29%

LINER : 91.63%

6.6.7 Logistic SCG con 52 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	981	916
0(known)	338	3841

Il tasso di riconoscimento è quindi pari al 79.36%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 51.71%

LINER : 91.69%

6.6.8 Logistic Conji Grad con 52 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	876	996
0(known)	335	3869

Il tasso di riconoscimento è quindi pari al 78.09%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 46.74%

LINER : 92.03%

6.6.9 Logistic Quasi Newton con 52 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	1001	896
0(known)	338	3841

Il tasso di riconoscimento è quindi pari al 79.69%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 52.77%

LINER : 91.69%

6.6.10 Linear SCG con 52 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	855	1045
0(known)	292	3898

Il tasso di riconoscimento è quindi pari al 78.05%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 45.00%

LINER : 93.03%

6.6.11 Linear Conji Grad con 52 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	840	1032
0(known)	314	3890

Il tasso di riconoscimento è quindi pari al 77.85%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 44.87%

LINER : 92.53%

6.6.12 Linear Quasi Newton con 52 hidden

La matrice di confusione risultante è la seguente:

	1(net)	0(net)
1(known)	993	904
0(known)	334	3845

Il tasso di riconoscimento è quindi pari al 79.62%.

La rete chiamata a distinguere tra Seyfert e LINER risponde riconoscendo le due classi con le seguenti percentuali:

Seyfert : 52.35%

LINER : 92.01%

6.6.13 Riepilogo e Considerazioni

Il riepilogo dei risultati ottenuti è presente in figura 6.8.

In questo caso la classificazione non è particolarmente riuscita, probabilmente questo oltre a una eventuale cattiva scelta dei parametri, che potrebbero semplicemente non essere significativi in questo caso, dipende dalla poca differenza fisica tra le due categorie, che anche nei grafici BPT (vedi capitolo 4) è poco caratterizzata (vedi figure 4.5 e 4.6).

Ad ogni modo il risultato migliore è stato ottenuto con la modalità Logistic Quasi Newton utilizzando 52 neuroni nascosti.

6.7 Separazione Seyfert LINER Tramite SVM

Come nel caso della classificazione degli AGN il peso computazionale rende impossibile eseguire questi processi in serie su un desktop, ho inviato ogni singolo processo su GRID.

Il dataset è lo stesso considerato nel caso delle MLP, l'80% dei dati è stato utilizzato per addestrare la rete e il 20% per testarla.

La coppia di parametri risultata migliore è $C = 8192$ e $\gamma = 0.03125$, il grafico che descrive la ricerca è in figura 6.9, il tasso di riconoscimento ottenuto è stato del 78.18%

Il risultato è lievemente peggiore di quello ottenuto con le MLP.

1	2	3	4	5	6
Logistic	SCG	20	78.24%	48.23%	91.60%
Logistic	Conji Grad	20	77.96%	47.06%	91.72%
Logistic	Quasi Newton	20	78.32%	48.02%	91.82%
Linear	SCG	20	78.09%	46.63%	92.10%
Linear	Conji Grad	20	77.98%	46.10%	92.17%
Linear	Quasi Newton	20	78.28%	48.29%	91.63%
Logistic	SCG	52	79.36%	51.71%	91.69%
Logistic	Conji Grad	52	78.09%	46.74%	92.03%
Logistic	Quasi Newton	52	79.69%	52.77%	91.69%
Linear	SCG	52	78.05%	45.00%	93.03%
Linear	Conji Grad	52	77.85%	44.87%	92.53%
Linear	Quasi Newton	52	79.62%	52.35%	92.01%

Figura 6.8: Schema Riassuntivo dei Risultati Ottenuti: nella colonna 1 è presente la funzione di attivazione usata, nella colonna 2 è presente l'algoritmo di ottimizzazione utilizzato, nella colonna 3 è presente il numero di neuroni nascosti usati, nella colonna 4 è riportato il tasso di riconoscimento globale della rete, nella colonna 5 è riportato il tasso di riconoscimento delle sole Seyfert nella colonna 6 è riportato il tasso di riconoscimento dei soli LINER.

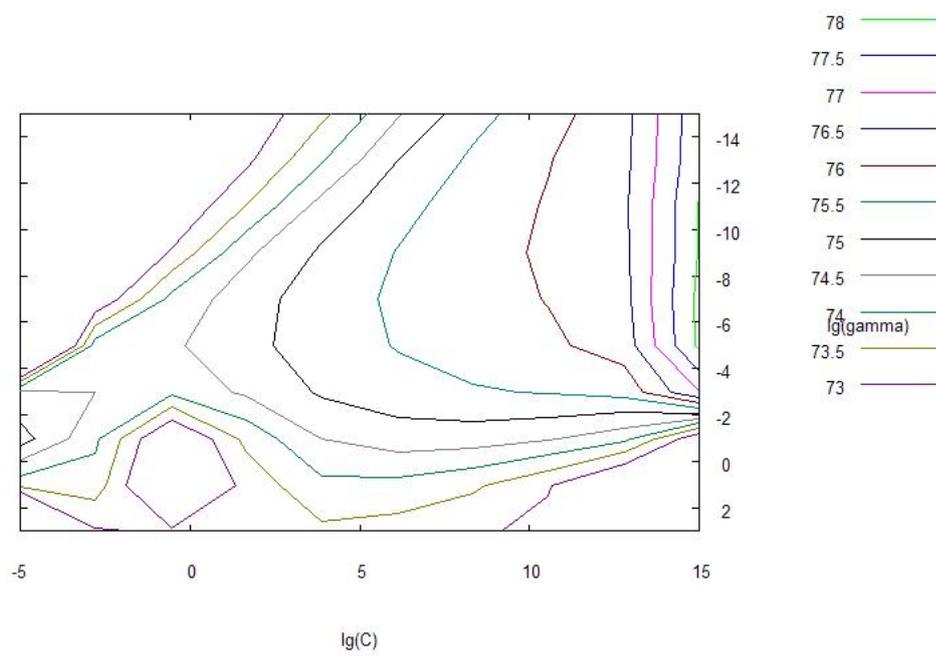


Figura 6.9: Ricerca a Griglia della miglior coppia di parametri, i logaritmi si intendono in base 2

Capitolo 7

Conclusioni

Daniel Pennac:

“A forza di riflettere, si finisce per arrivare a una conclusione. A forza di giungere a una conclusione, succede che si prende una decisione. E una volta presa la decisione, succede che si agisce per davvero.”



L'obiettivo principale di questa tesi è stato raggiunto ottenendo una rete

neurale in grado di selezionare su base puramente fotometrica forti candidati AGN.

La rete è stata pensata non per cercare tutti i candidati ma per avere un'elevatissima confidenza sui candidati selezionati.

Ovviamente servirebbero le misure spettroscopiche di alcuni dei nostri candidati per poterli verificare e validare ulteriormente fermo restando che la procedura di Test effettuata dalla rete altro non è che questo e quindi possiamo già dire che parte di questo lavoro di verifica è stato svolto.

Per quanto concerne la separazione tra oggetti di Tipo 1 e Tipo 2 sono anche stati ottenuti ottimi risultati pur essendo pochi i dati in nostro possesso.

La classificazione è molto probabilmente possibile e non appena si dovessero avere nuovi dati sarebbe interessante ripetere gli esperimenti in modo da poter ottenere un classificatore ancor più efficiente.

La separazione tra LINER e Seyfert invece non è netta e i risultati conseguiti non ci permettono di dire molto. Ciò probabilmente dipende da una differenza tra le due categorie non particolarmente marcata, cosa che in effetti si può osservare nelle figure 4.5 e 4.6, dove è evidente che la separazione anche nei parametri spettroscopici avviene molto al di sopra della riga di Kewley per quanto concerne gli aspetti fenomenologici.

Nell'ambito del progetto VO-Neural si è giunti alla release della prima versione del software che fa funzionare le reti MLP e tale software è già disponibile per tutta la comunità scientifica.

Inoltre sono state poste le basi per la seconda versione delle reti MLP mentre si stanno finendo di sviluppare le SVM.

Ancora una volta va sottolineato che l'importanza principale di questo lavoro prima ancora che nei risultati risiede nella metodologia adottata che risulta assolutamente innovativa rispetto a quanto è sinora apparso in letteratura.

Sia questo lavoro che alcuni di quelli che lo hanno preceduto, e mi riferisco agli articoli di D'Abrusco et al. [22, 27], dimostrano che questo tipo di approccio è percorribile e apre nuove porte all'analisi dell'enorme mole di dati prodotta dalle Survey moderne e si auspica quindi che permetterà nel breve periodo di avere una migliore conoscenza dell'universo e in particolare migliori classificazioni.

A.1 Materiale Relativo alle Reti MLP

I seguenti paragrafi contengono stralci del codice da me modificato per la prima release di VO-Neural MLP.

A.1.1 Codice per l'Aggiunta dei Target Vector

L'esempio che segue è valido per la modalità di train:

```
outputdata_train=[trainTarget NeuralOut_train];
saveascii(outputdata_train,outFileTrain,'%f', 'w');
```

Le righe di codice sono perfettamente analoghe nelle altre modalità.

A.1.2 Modalità di Run

Il codice che regola questa modalità è il seguente:

```
if (typeRun==4)
    c=fix(clock);
    dataclock= [ num2str(c(1)) '/' num2str(c(2)) '/'
    num2str(c(3)) ' ' num2str(c(4)-1) ':' num2str(c(5)) ':'
    num2str(c(6)) ' GMT: run session started'];
    fid = fopen(logfile, 'a+');
    fprintf(fid, '%s\n', dataclock);
    fclose(fid);

    load('-mat', inputNet);
    nin=net.nin;
    runSet=test(:,1:nin);
    NeuralOut_run=mlp fwd(net,runSet(:,:));
    outFileRun=strcat(outputData, '.run');
    saveascii(NeuralOut_run,outFileRun,'%f', 'w');
    %netRun=strcat(outputData, '_netRun.mlp');
    %save(netRun, 'NeuralOut_run');
    c=fix(clock);
    dataclock= [num2str(c(1)) '/' num2str(c(2)) '/'
    num2str(c(3)) ' ' num2str(c(4)-1) ':' num2str(c(5)) ':'
```

```

    num2str(c(6)) ' GMT: run session finished'];
    fid = fopen(logfile, 'a+');
    fprintf(fid, '%s\n', dataclock);
    fclose(fid);
end

```

A.1.3 Parametri Ridondanti

I parametri presenti nella rete eventualmente caricata sono stati importati dalla stessa nella maniera che segue:

```
nin=net.nin;
```

L'esempio valido per il nin, è analogo anche per il nout.

A.1.4 Tool di Validazione

```

if (strcmp(inputDataVa, 'empty')==0)
    [ConfTrain, RateTrain]=confmat(trainTarget,NeuralOut_train);
    % confusion matrix
    validSet=valid(:,1:nin);
    validtarget=valid(:,nin+1:nin+nout);
    NeuralOut_Valid=mlpfwd(net,validSet(:,:));
    [ConfValid, RateValid]=confmat(trainTarget,NeuralOut_train);
    c=fix(clock);
    dataclock= [num2str(c(1)) '/' num2str(c(2)) '/'
    num2str(c(3)) ' ' num2str(c(4)-1) ':' num2str(c(5)) ':'
    num2str(c(6)) ' GMT: training set reaches a success rate
    of: ' num2str(RateTrain(1)) '% while validation set reaches
    a success rate of: ' num2str(RateValid(1)) '%'];
    fid = fopen(logfile, 'a+');
    fprintf(fid, '%s\n', dataclock);
    fclose(fid);
end

```

A.1.5 Esempio di Logfile Qualora il Processo Sia Andato a Buon Fine

```
2007/10/11 11:14:58 GMT: job started
```

```
2007/10/11 11:14:58 GMT: training set correctly loaded:  
inputDataTr_9.tmpgen
```

```
2007/10/11 11:14:58 GMT: validation set correctly loaded:  
inputDataVa_10.tmpgen 2007/10/11 11:14:58 GMT: training session  
started
```

```
2007/10/11 11:15:16 GMT: training set reaches a success rate of:  
100% while validation set reaches a success rate of: 100%
```

```
2007/10/11 11:15:16 GMT: training session finished 2007/10/11  
11:15:16 GMT: job finished
```

A.1.6 Esempio di Logfile in cui i Dataset Sono di Dimensioni Incongrue

```
2007/10/11 11:14:58 GMT: job started 2007/11/11
```

```
2007/10/11 11:14:58 GMT: invalid training set, number of column less  
than requested
```

```
2007/10/11 11:14:58 GMT: invalid validation set, number of column  
less than requested
```

```
2007/10/11 11:14:58 GMT: invalid test set, number of column less  
than requested
```

```
2007/10/11 11:14:59 GMT: job finished error
```

A.2 Materiale Relativo alle Matrici di Confusione

I prossimi due paragrafi contengono il codice delle librerie e un esempio di programma che le utilizza.

A.2.1 Codice delle Librerie

```

#include<stdlib.h> #include<stdio.h> #include<math.h>
#include<string.h> /* *_advanced function are multicolumn
ones */
int ** alloc_confusion_matrix(int num_output) {
    int ** confusion_matrix = calloc( num_output ,
    sizeof(int *));
    int i;
    for(i = 0; i != num_output; i++)
    {
        confusion_matrix[i] = calloc(num_output , sizeof(int));
    }
    return confusion_matrix;
} int update_confusion_matrix(int ** confusion_matrix,
float target,
float output, int num_output) {
    int i, j;
    i=0;
    j=0;
    for(i=0; i != num_output && abs(target-i)>0.5; i++);
    for(j=0; j != num_output && abs(output-j)>0.5; j++);
    confusion_matrix[i][j]++ ;
    return 0;
} int print_confusion_matrix (int num_output, int **
confusion_matrix, char * log_file) {
    int i, j;
    int total = 0;
    int success = 0;
    int * successclass = calloc( num_output , sizeof(int));
    int * totalbit = calloc(num_output , sizeof(int));
    FILE * stream = fopen(log_file, "w");
    for(i = 0; i != (num_output); i++)
    {
        success +=confusion_matrix[i][i];
        successclass[i]=confusion_matrix[i][i];
        totalbit[i]=0;
    }

```

```

for(j = 0; j != num_output; j++)
{
    fprintf(stream, "%lu" , confusion_matrix[i][j]);
    fprintf(stream, "\t");
    total+= confusion_matrix[i][j];
    totalbit[i]+= confusion_matrix[i][j];
}
fprintf(stream, "\n");
}
fprintf(stream, "\nSuccess rate is ");
float succrat = 100*success/total;
fprintf(stream, "%f" ,succrat);
fprintf(stream, "% \n \n");
int bitfault = total - success;
fprintf(stream, "%hu", bitfault);
fprintf(stream, " Bits are fault\n\n");
float * successpercent = calloc(num_output , sizeof(float));
int k = 0;
for(k = 0; k != num_output; k++)
{
    fprintf(stream, "Success rate for class ");
    fprintf(stream, "%hu" , k);
    fprintf(stream, " is: \t");
    if (totalbit[k]!=0)
    {
        successpercent[k]=100*successclass[k]/totalbit[k];
        fprintf(stream, "%f" , successpercent[k]);
        fprintf(stream, "% \n");
    }
    else
    {
        fprintf(stream, "class not present in the
        dataset\n");
    }
}
}
free(successclass);

```

```

        free(totalbit);
        free(successpercent);
        fclose(stream);
    }
int destroy_conf_mat(int ** confusion_matrix, int num_output) {
    int i;
    for(i = 0; i != num_output; i++)
    {
        free(confusion_matrix[i]);
    }
    free(confusion_matrix);
}

int ** alloc_confusion_matrix_advanced(int num_output) {
    int ** confusion_matrix = calloc( (int)pow(2, num_output) ,
    sizeof(int *));
    int i;
    for(i = 0; i != pow(2, num_output); i++)
        confusion_matrix[i] = calloc((int)pow(2, num_output) ,
        sizeof(int));
    if (confusion_matrix == NULL)
    {
        printf("Memory not allocated in matrix");
    }
    else
        return confusion_matrix;
} int update_confusion_matrix_advanced(int ** confusion_matrix,
float * target, float * output, int num_output) {
    int i, indice_i , indice_j;
    i=0;
    indice_i=0;
    indice_j=0;
    while (i != num_output)
    {
        if (target[i]<0.5)
        {

```

```

        target[i]=0;
    }
    else
    {
        target[i]=1;
    }
    indice_i += target[i]*(pow(2 , (num_output-i-1)));
    if (output[i]<0.5)
    {
        output[i]=0;
    }
    else
    {
        output[i]=1;
    }
    indice_j += output[i]*(pow(2 , (num_output-i-1)));
    i++;
}
confusion_matrix[indice_i][indice_j]++ ;
return 0;
} int print_confusion_matrix_advanced (int num_output, int **
confusion_matrix, char * log_file) {
    int i, j;
    int total = 0;
    int success = 0;
    int * successclass = calloc(pow(2, num_output) ,
sizeof(int));
    int * totalbit = calloc(pow(2, num_output) , sizeof(int));
    FILE * stream = fopen(log_file, "w");
    for(i = 0; i != pow(2, num_output); i++)
    {
        success +=confusion_matrix[i][i];
        successclass[i]=confusion_matrix[i][i];
        totalbit[i]=0;
        for(j = 0; j != pow(2, num_output); j++)
        {

```

```

        fprintf(stream, "%lu" , confusion_matrix[i] [j]);
        fprintf(stream, "\t");
        total+= confusion_matrix[i] [j];
        totalbit[i]+= confusion_matrix[i] [j];
    }
    fprintf(stream, "\n");
}

fprintf(stream, "\nSuccess rate is ");
float succrat = success/total;
fprintf(stream, "%f" ,succrat);
fprintf(stream, "\n \n");
int bitfault = total - success;
fprintf(stream, "%hu", bitfault);
fprintf(stream, " Bits are fault\n\n");
float * successpercent = calloc(pow(2, num_output) ,
sizeof(float));
int k = 0;
for(k = 0; k != pow(2, num_output); k++)
{
    printf("\n");
    /* char foo(int class)
    {
        char b = "";
        while (class > 0)
        {
            int    j = i & 1;
            char sid;
            sid=sprintf(sid, "%d" ,j);
            b=strcat( sid , b);
            class >>= 1;
        }
        int z;
        for(z = 0; z != (num_output-strlen(b)); z++)
        {
            b= strcat( "0" , b);

```

```

        }
        return b;
    }
    char classbin = foo(k);
    fprintf(stream, classbin); */
    if (totalbit[k]!=0)
    {
        fprintf(stream, "Success rate for class ");
        fprintf(stream, "%hu" , k);
        fprintf(stream, " is: \t");
        successpercent[k]=successclass[k]/totalbit[k];
        fprintf(stream, "%f" , successpercent[k]);
        fprintf(stream, "% \n");
    }
    else
    {
        fprintf(stream, "class ");
        fprintf(stream, "%hu", k);
        fprintf(stream, " not present in the dataset\n");
    }
}
free(successclass);
free(totalbit);
free(successpercent);
fclose(stream);
} int destroy_conf_mat_advanced(int ** confusion_matrix, int
num_output) {
    int i;
    for(i = 0; i != pow(2, num_output); i++)
    {
        free(confusion_matrix[i]);
    }
    free(confusion_matrix);
}
}

```

A.2.2 Esempio di Utilizzo

```
int main() {
    char * log_file = "/home/cavuoti/Desktop/confusion.txt";
    int num_output = 3;

    int ** confusion_matrix = alloc_confusion_matrix(num_output);

    float target1[5];
    target1[1]=0;
    target1[2]=2;
    target1[3]=1;
    target1[4]=0;
    target1[0]=1;
    float output1[5];
    output1[1]=0;
    output1[2]=2.4;
    output1[3]=1;
    output1[4]=0;
    output1[0]=0;
    float target[5];
    float output[5];
    int k;

    for (k=0; k!=5; k++)
    {
        target[k]=1*target1[k];
        output[k]=1*output1[k];

        update_confusion_matrix(confusion_matrix, target[k],
            output[k], num_output);
    }

    print_confusion_matrix (num_output, confusion_matrix,
        log_file);
    destroy_conf_mat(confusion_matrix, num_output);
}
```


Bibliografia

William Shakespeare:

“La mia biblioteca era per me un ducato grande abbastanza.”



- [1] A. Braccisi: *Dalle Stelle All'Universo, lezioni di astrofisica*, Zanichelli Editore, (2000)

- [2] J. Brau: *ASTRONOMY 123* <http://physics.uoregon.edu/~jim-brau/astr123/> (2001)
- [3] M.A. Ulrich et al, *Ap. J.*, 347, 29, (1989)
- [4] G. Sorrentino: *Galaxy Environment In The Public Surveys 2dFGRS and SDSS*, Ph.D. Thesis, (2007)
- [5] M. R. Blanton et al. *The Luminosity Function of Galaxies in SDSS Commissioning Data*, (2001)
- [6] Yasuda et al. *Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data*, (2001)
- [7] Gruppo Astronomico Tradatese: *Lettera N.98*, <http://www.gruppoastronomicotradatese.it/lettere/lett98.pdf>, (2004)
- [8] M. Fukugita et al.: *AJ*, 111, 1748, (1996)
- [9] J.E. Gunn et al.: *AJ*, 116 ,3040 (1998)
- [10] E. Pessa: *Statistica con le reti neurali*, Di Renzo Editore, (2004)
- [11] B. E. Boser, I. Guyon, e V. Vapnik . *A training algorithm for optimal margin classifiers*. Nei Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp. 144-152. ACM Press. (1992)
- [12] Chih-Wei Hsu, Chih-Chung Chang e Chih-Jen Lin: *A Practical Guide to Support Vector Classification*(2007)
- [13] S. S. Keerthi e C.-J. Lin: *Asymptotic behaviors of support vector machines with Gaussian kernel*. *Neural Computation* 15 (7), 1667-1689.(2003)
- [14] H.-T. Lin e C.-J. Lin: *A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods*. Technical report, Department of Computer Science, National Taiwan University. (2003)
- [15] S. Camardella: *PRINCIPAL COMPONENT ANALYSIS: Un Nuovo Approccio Metodologico alla Classificazione dei Corpi Idrici*, Ph.D. Thesis, (2001)

- [16] Kyu-yu Chang & J. Ghosh: *A Unified Model for Probabilistic Principal Surfaces*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol.23, Iss.1, Jan 2001 Pages:22-41 (1999)
- [17] Kyu-yu Chang & J. Ghosh: *Three-Dimensional Model-Based Object Recognition and Pose Estimation Using Probabilistic Principal Surfaces*, in SPIE: Applications of Artificial Neural Networks in Image, 192-203 (2000)
- [18] Kyu-yu Chang & J. Ghosh: *A unified Model for Probabilistic Principal Surfaces* in IEE Transactions on Pattern Analysis and Machine intelligence, 23, 22-41 (2001)
- [19] C. M. Bishop, M. Svensen & C. K. I. Williams: *Neural Computation*, 215-234 (1998)
- [20] C. M. Bishop: *Latent variable models*, in M. I. Jordan (Ed.), Learning in Graphical Models, MIT Press, (1999)
- [21] Staiano A.: *Unsupervised Neural Networks for the Extraction of Scientific Information from Astronomical Data*, PhD thesis, University of Salerno (2003)
- [22] R. D'Abrusco, G Longo & G. Walton: *Quasar candidates selection in the Virtual Observatory era* (2007)
- [23] A.P. Dempster, N.M. Laird, D.B. Rubin, *Maximum-Likelihood from Incomplete Data Via the EM Algorithm*, J. Royal Statistical Soc., Vol. 39, NO. 1, (1977)
- [24] Staiano A., De Vinco L., Ciaramella A., Raiconi G., Tagliaferri R., Longo G., Miele G., Amato R., Del Mondo C., Donalek C., Mangano G., Di Bernardo D.: *Probabilistic principal surfaces for yeast gene microarray data-mining*, in ICDM'04 - Fourth IEEE International Conference on Data Mining, pp. 202-209, (2004)
- [25] G. Sorrentino, M. Radovich, and A. Rifatto: *The environment of active galaxies in the SDSS-DR4*, A&A 451, 809-816, (2006)
- [26] G. Kauffman et al.: *The host galaxies of active galactic nuclei*, MNRAS, 346, 1055, (2003)

- [27] R. D'Abrusco, A. Staiano, G. Longo, M. Brescia, M. Paolillo, E. De Filippis, R. Tagliaferri: *Mining the SDSS archive. I. Photometric redshifts in the nearby universe.*, arXiv:astro-ph/0703108v2 9 Mar, (2007)
- [28] SDSS: *data release 4*, <http://www.sdss.org/dr4>
- [29] L.J. Kewley et al.: ApJ, 556, 121 (2001)
- [30] J.A. Baldwin, M.M. Phillips & R. Terlevich: PASP, 93, 5 (1981)
- [31] SDSS: *data release 4*, <http://www.sdss.org/dr2>
- [32] L.J. Kewley et al.: *The host galaxies and classification of active galactic nuclei*, MNRAS, 372 (3), 961-976, (2006)
- [33] T.M. Heckman: A&A, 87, 182 (H80) (1980)

Elenco delle figure

1.1	Schema di un AGN	7
1.2	Il disco di accrescimento della galassia Seyfert NGC4261	9
1.3	Schema delle possibili emissioni di un nucleo galattico attivo in relazione alla porzione di sistema che le genera	10
1.4	Galassia NGC 4051 e il suo spettro	14
1.5	Galassia NGC 1068 e il suo spettro	14
1.6	Schema di una possibile storia evolutiva dei quasar	16
2.1	Campi di Radiazione	21
2.2	Conservazione dell'intensità della radiazione per la propaga- zione nel vuoto	22
2.3	Osservazione di una sorgente risolta per un dato angolo solido	26
2.4	Risposta dei rivelatori fotometrici	29
2.5	Calibrazioni assolute di una stella di magnitudine zero in differenti bande passanti	30
2.6	Distribuzione di brillantezza della radiazione di corpo nero per diverse temperature	32
2.7	Temperature di plasmi astrofisici stimati con la legge di Wien	33
2.8	Tipiche sorgenti termiche in astrofisica e loro spettri	34
2.9	Il diagramma a due colori per le stelle	38
2.10	Attenuazione A_ν	41
2.11	Arrossamento degli spettri stellari e misure delle distanze	43
2.12	Massa atmosferica in funzione della distanza zenitale	44
2.13	Una porzione dello spettro solare	49
2.14	Le leggi di Kirchhoff sulla spettroscopia.	50
2.15	Diagramma di Grotrian per le serie delle righe dell'atomo di idrogeno e spettro	53

2.16	Molecole di interesse astrofisico	57
2.17	Riduzione di uno spettro e normalizzazione al continuo	59
2.18	Larghezza equivalente delle righe spettrali	60
2.19	Formazione delle righe di Balmer	63
2.20	Profilo di Voigt	66
2.21	Curva di crescita	66
2.22	Profilo P Cygni	68
2.23	I 30 CCD (6 per ogni banda) della camera di acquisizione delle immagini della SDSS	75
3.1	Schema di un neurone reale	81
3.2	Neurone artificiale proposto da McCulloch e Pitts	82
3.3	Schema di una semplice rete neurale	83
3.4	Rappresentazione di possibili relazioni con due parametri	91
3.5	Raffronto singola PCA, multi PCA e soluzione non lineare	93
3.6	Covarianze non orientate	95
3.7	Rappresentazione schematica della varietà sferica	96
3.8	Un esempio di dendogramma	99
3.9	Un'altra visualizzazione possibile del dendogramma	100
4.1	Esempio di grafico BPT, NII contro OIII	105
4.2	Separazione tramite la linea di Kewley	107
4.3	Separazione tramite la linea di Kauffman	108
4.4	Gli oggetti raggruppati nelle tre categorie: SFN, Mix e AGN	108
4.5	Separazione tra LINER e Seyfert	109
4.6	Divisione in Cinque Categorie, Galassie Normali, LINER Mix, Seyfert Mix, LINER Puri e Seyfert pure	110
5.1	Schema dei parametri necessari al codice mlp e relative uscite divise in funzione della modalità prescelta	116
6.1	Rapporto tra AGN Correttamente Valutati e Non AGN Clas- sificati Male, Prima Metà del TestSet	132
6.2	Rapporto tra AGN Correttamente Valutati e Non AGN Clas- sificati Male, Seconda Metà del TestSet	133
6.3	Rapporto tra AGN Correttamente Valutati e Non AGN Clas- sificati Male, Analisi Totale del TestSet	133
6.4	Schema Riassuntivo dei Risultati Ottenuti	134

6.5	Ricerca a Griglia della miglior coppia di parametri, i logaritmi si intendono in base 2	135
6.6	Schema Riassuntivo dei Risultati Ottenuti	141
6.7	Ricerca a Griglia della miglior coppia di parametri, i logaritmi si intendono in base 2	142
6.8	Schema Riassuntivo dei Risultati Ottenuti	149
6.9	Ricerca a Griglia della miglior coppia di parametri, i logaritmi si intendono in base 2	150

