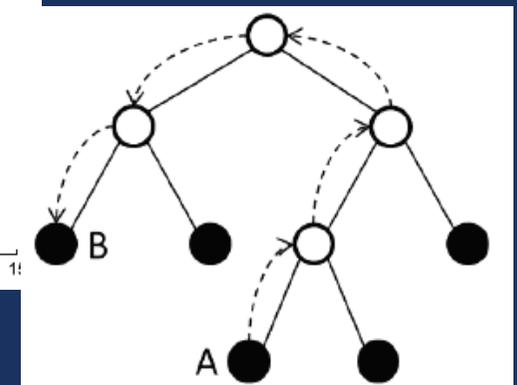
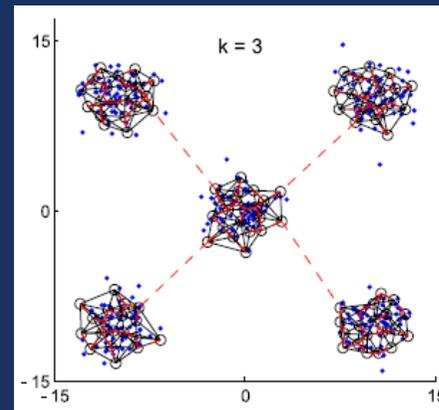
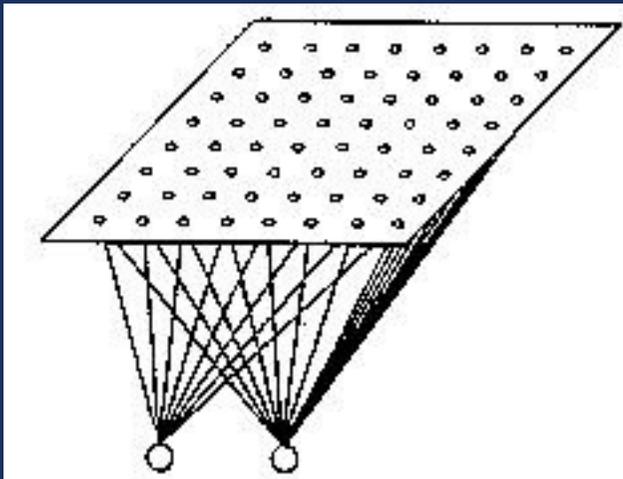


CLUSTERING CON MODELLI SOM DINAMICI

F. ESPOSITO



INTRODUZIONE

Du, K.L., 2010.
Clustering: A neural network approach.
Neural Networks, Elsevier, Vol. 23, 89-107

- Apprendimento Competitivo
- Panoramica metodi clustering
- Tassonomia metodi clustering
- Problemi dei metodi di clustering e possibili soluzioni

Tomenko, V., 2011.
Online dimensionality reduction using competitive learning and Radial Basis Function network.
Neural Networks, Elsevier, Vol. 24, 501-511

- Metodo RBF-NDR
 - Apprendimento dello spazio dei dati
 - Riduzione della dimensionalità
- Esperimenti

SOMMARIO

Prima Parte

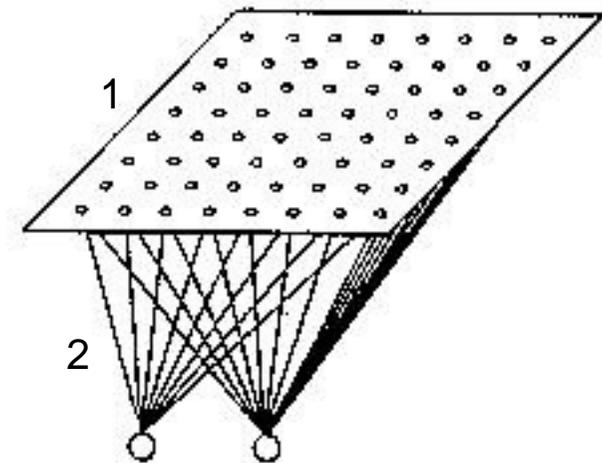
- Regola di base
- Tecniche di clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

REGOLA DI BASE: APPRENDIMENTO COMPETITIVO

- Architettura a strati con neuroni completamente connessi tra loro
- Strategia **Winner Takes All (WTA)**
- La rete così formata è capace di catturare le regolarità statistiche della sequenza di pattern presentati : **Clustering**



1 : Connessioni Inibitorie
2 : Connessioni Eccitatorie

SOMMARIO

Prima Parte

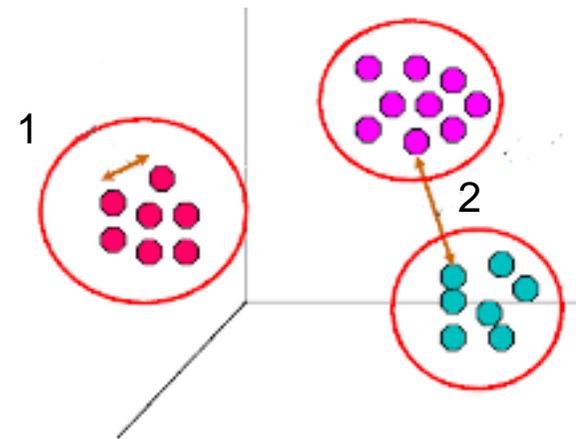
- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

TASSONOMIA CLUSTERING : PARTITIONAL CLUSTERING

- Suddivisione dei dati in partizioni con riassegnazione dinamica dei punti nei cluster
- Complessità $O(N)$
- Scelta preventiva del numero di cluster

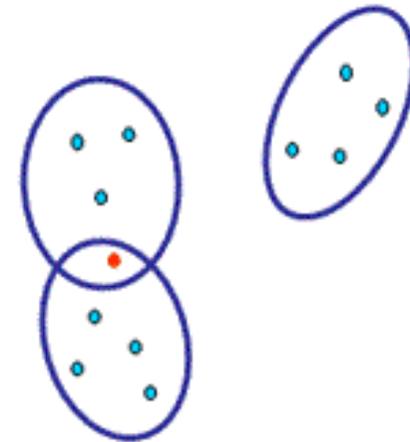


1 : Distanza intra-cluster (Compattezza)
2 : Distanza inter-cluster

HARD E FUZZY CLUSTERING

- **Hard Clustering** : appartenenza totale o nulla
- **Fuzzy Clustering**: Appartenenza di un pattern ad un cluster può variare in maniera continua tra 0 e 1
- Trattazione dei casi limite
- Matrice di Appartenenza $U = \{\mu_{ji}\}$, $\mu_{ji} \in [0, 1]$ (1)
- *Regola* :

$$\sum_{j=1}^K \mu_{ji} = 1, \quad i = 1, \dots, N \quad (2)$$



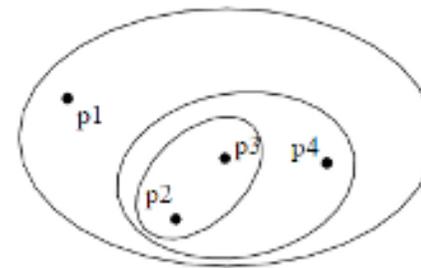
Generazione di cluster sovrapposti

TASSONOMIA CLUSTERING : HIERARCHICAL CLUSTERING

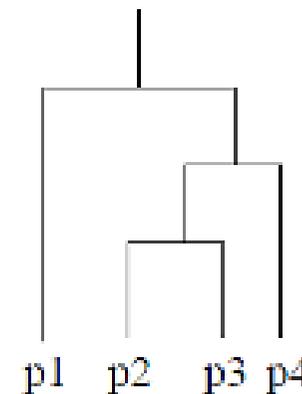
- Sequenza di partizionamenti organizzati in una struttura di tipo gerarchico



- Complessità $O(N^2)$
- Non può incorporare conoscenza a priori di forma e grandezza cluster
- Non è permesso approccio Fuzzy



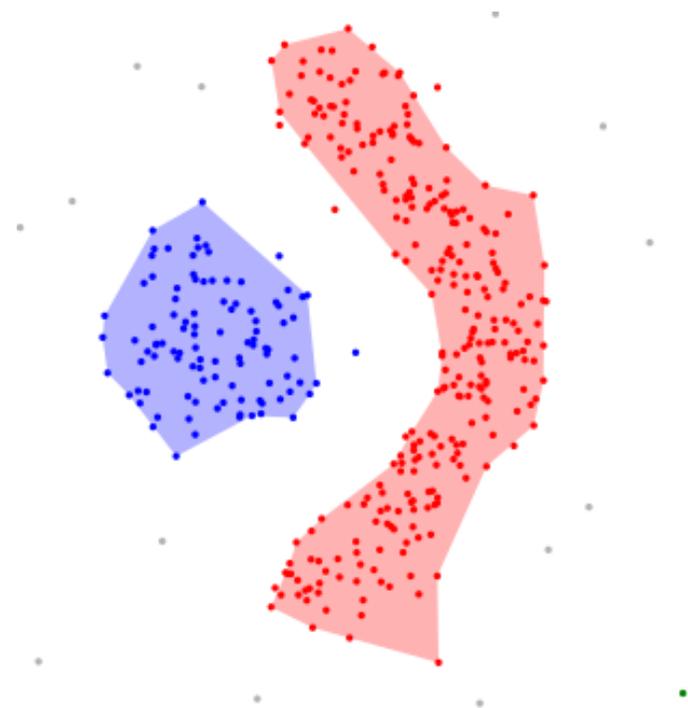
Esempio di clustering gerarchico con approccio top-down



Dendrogramma relativo all'immagine precedente

TASSONOMIA CLUSTERING : DENSITY BASED CLUSTERING

- Raggruppamenti in base a parametri di densità
- Complessità $O(N^2)$



SOMMARIO

Prima Parte

- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

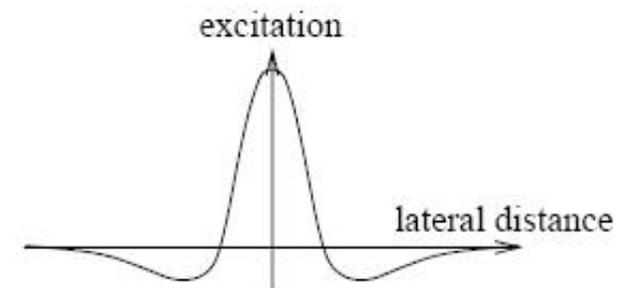
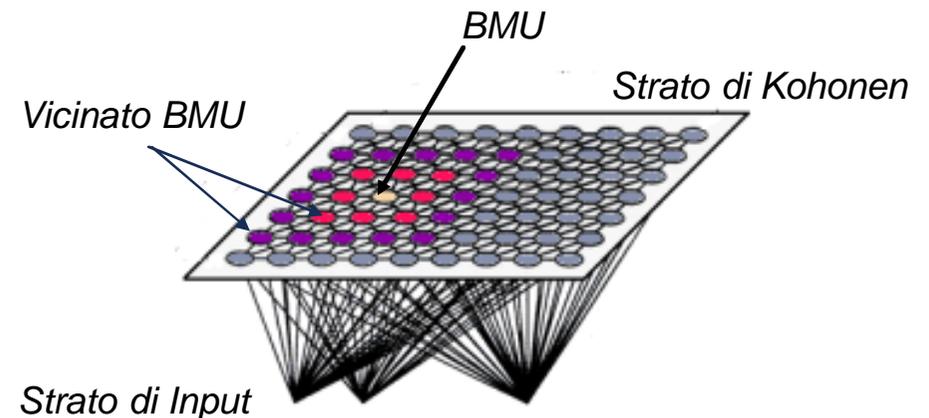
SELF ORGANIZING MAP

KOHONEN, T. (1990). THE SELF-ORGANIZING MAP. PROCEEDINGS OF IEEE, 78, 14641480

- Struttura classica dell' apprendimento competitivo
- Funzione di vicinato determina range aggiornamento pesi
- Attivazione decresce secondo la funzione a Mexican Hat
- Legge di Kohonen per l'aggiornamento dei pesi:

$$w_k^{(t+1)} = w_k^{(t)} + \eta^{(t)} h_{kw}^{(t)} [x^{(t)} - w_k^{(t)}] \quad (1)$$

- k : indice di un neurone
- w : indice del neurone vincente
- x : pattern in input
- η : tasso di apprendimento
- h : funzione di vicinato



Funzione a Mexican Hat

SOM SUPERVISIONATA: LEARNING VECTOR QUANTIZATION (LVQ)

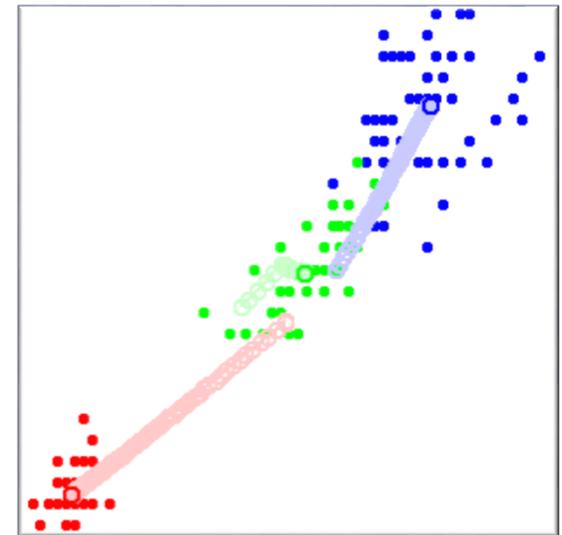
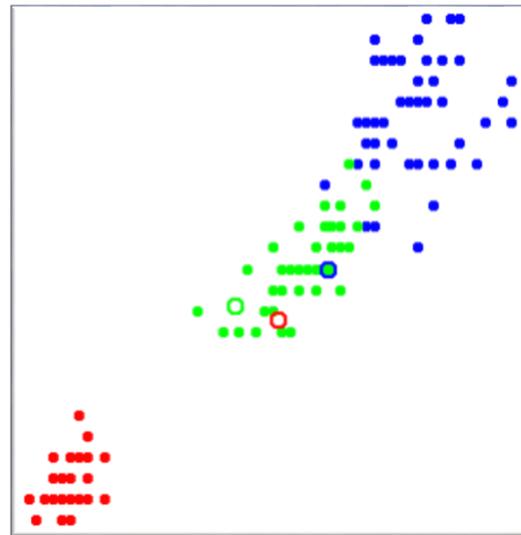
KOHONEN, T. (1990). THE SELF-ORGANIZING MAP. PROCEEDINGS OF IEEE, 78, 14641480

- A differenza della SOM non dà luogo ad una mappa
- Ogni neurone dello strato di output rappresenta un prototipo indipendente dagli altri
- Aggiornamento dei pesi non esteso al vicinato ma limitato al neurone vincente
- Problemi relativi alla scelta del numero di neuroni:
 - Overfitting
 - Pattern diversi nello stesso cluster

$$\mathbf{c}_w(t+1) = \mathbf{c}_w(t) + \eta(k) [\mathbf{x}_t - \mathbf{c}_w(t)], \quad y_{p,w} = 1 \quad (1)$$

$$\mathbf{c}_w(t+1) = \mathbf{c}_w(t) - \eta(t) [\mathbf{x}_t - \mathbf{c}_w(k)], \quad y_{p,w} = 0$$

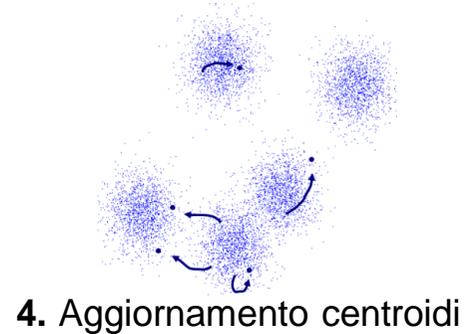
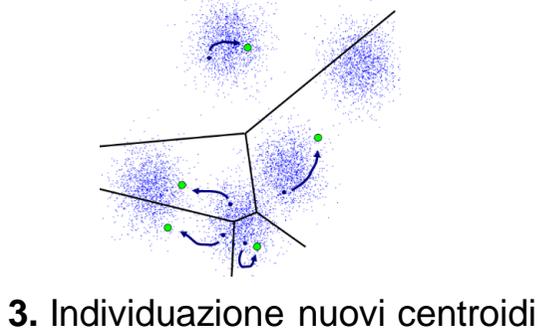
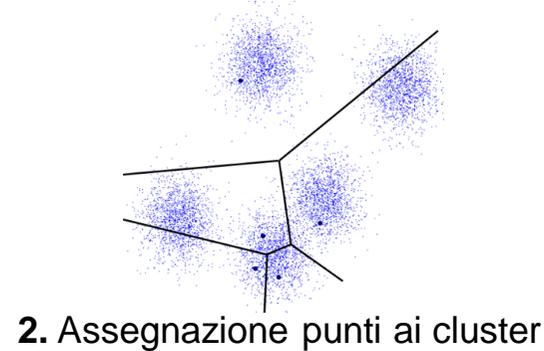
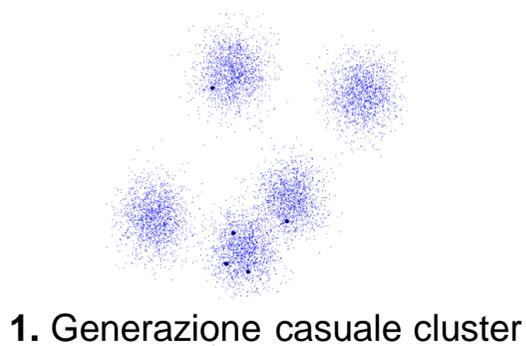
$$\mathbf{c}_i(t+1) = \mathbf{c}_i(t), \quad i \neq w$$



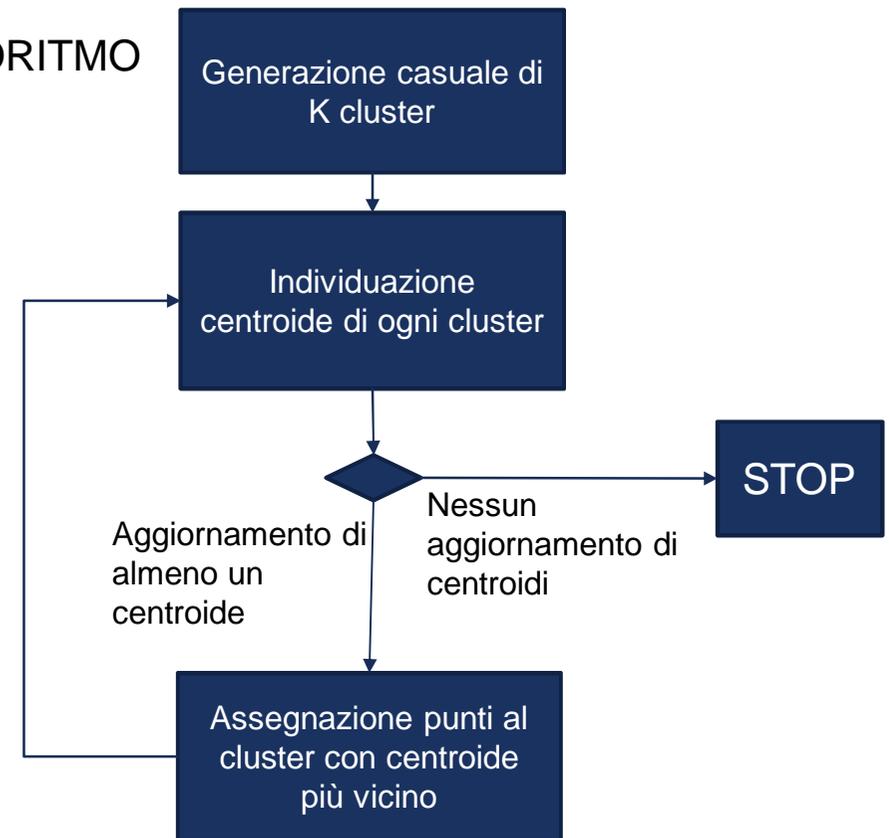
K-MEANS

HARTIGAN, J. A.; WONG, M. A. (1979) "A K-MEANS CLUSTERING ALGORITHM," APPLIED STATISTICS, 28, 100-108

- Tecnica di Partitional Clustering più semplice e nota



ALGORITMO



MOUNTAIN CLUSTERING

YAGER, R. R., & FILEV, D. (1994)

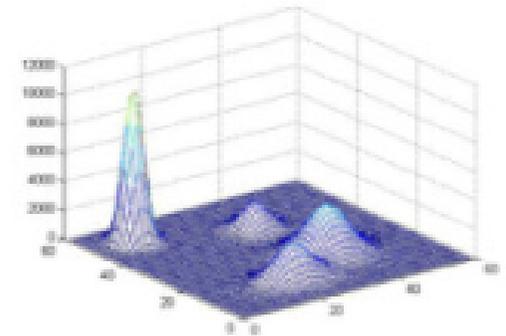
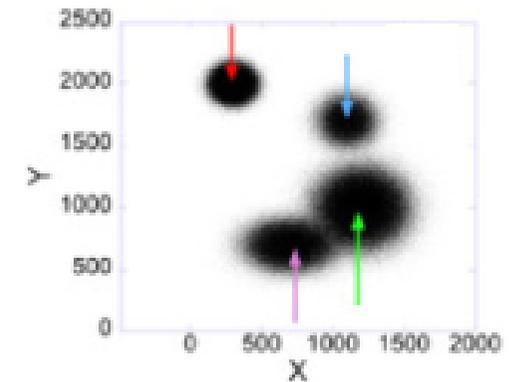
APPROXIMATE CLUSTERING VIA THE MOUNTAIN METHOD. IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS, 24(8), 12791284

- Tecnica per stimare il numero di cluster e locazione iniziale del centro dei cluster
- Primo step di una possibile sequenza di algoritmi
- Sovrapposizione di una griglia sullo spazio dei dati e calcolo del valore potenziale di ogni punto della stessa in base alla densità dei dati circostanti
- Utilizzo di tale struttura causa aumento esponenziale della complessità del problema



SUBTRACTIVE CLUSTERING¹

elimina la griglia utilizzando come potenziali cluster i punti dello spazio stesso

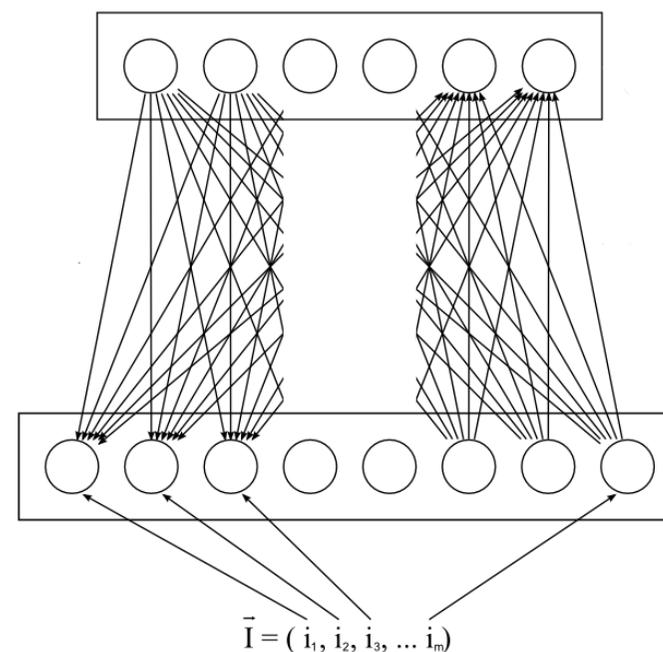


¹ Chiu, S. (1994). Fuzzy model identification based on cluster estimation. Journal Intelligent and Fuzzy Systems, 2(3), 267278

RETI ADAPTIVE RESONANCE THEORY (ART)

GROSSBERG, S. (1976). ADAPTIVE PATTERN CLASSIFICATION AND UNIVERSAL RECORDING: I. PARALLEL DEVELOPMENT AND CODING OF NEURAL FEATURE DETECTORS; II. FEEDBACK, EXPECTATION, OLFACTION, AND ILLUSIONS. BIOLOGICAL CYBERNETICS, 23, 121134. 187202

- Architettura multi-strato con connessioni feed-forward e feed-backward
- **Risonanza** : comunicazione tra strati
- **Adattiva** : aggiunta di un nuovo neurone quando necessario
- Superamento problema stabilità-plasticità



NEURAL GAS

MARTINETZ, T. M., BERKOVICH, S. G., & SCHULTEN, K. J. (1993)

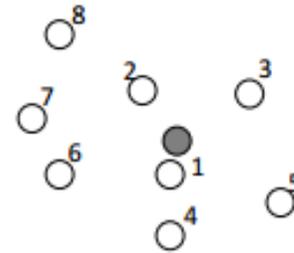
NEURAL-GAS NETWORK FOR VECTOR QUANTIZATION AND ITS APPLICATION TO TIME-SERIES PREDICTIONS. IEEE TRANSACTIONS ON NEURAL NETWORKS, 4(4), 558569

- Neuroni come particelle gassose : non fissi all'interno di una struttura ma liberi e attratti da zone ad alta densità di dati.
- Funzione di vicinato non statica
- Aggiornamento dei pesi in base alla vicinanza al pattern in input

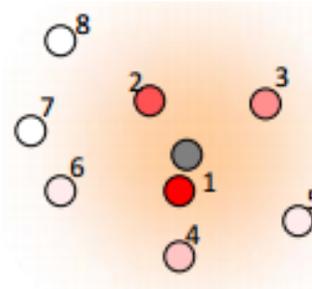
$$\mathbf{c}_k(t + 1) = \mathbf{c}_k(t) + \eta h(r_k(t)) (\mathbf{x}_t - \mathbf{c}_k(t)) \quad (1)$$

$$h(r) = e^{-\frac{r}{\rho(t)}} \quad (2)$$

Intero indicante la posizione nella classifica di vicinanza al pattern in input



I valori numerici indicano la posizione del neurone nella classifica di vicinanza al pattern in input



La gradazione di rosso indica la quantità di aggiornamento, ovvero di avvicinamento dei neuroni al pattern in input

NEURAL GAS – COMPETITIVE HEBBIAN LEARNING (NG-CHL)

MARTINETZ, T. M., & SCHULTEN, K. J. (1994). TOPOLOGY REPRESENTING NETWORKS. NEURAL NETWORKS, 7, 507522

- Combina la due tecniche per la costruzione di **Topology Preserving Maps**: mappatura di un insieme di punti in un grafo che preservi la vicinanza sia nel passaggio dal grafo allo spazio dei dati che viceversa.
- Il **CHL** introduce l'aspetto competitivo nel principio fondamentale per l'aggiornamento dei pesi sinaptici introdotto da Hebb

CHL

Basa la competizione sull'attività di output correlata tra due unità i e j definita come segue:

$$Y_{ij} = y_i * y_j \quad (1)$$

Anziché modificare i pesi in accordo con la regola generale proposta da Hebb

$$\Delta C_{ij} \propto y_i * y_j \quad (2)$$

se ne utilizza una versione competitiva:

$$\Delta C_{ij} \propto \begin{cases} y_i * y_j & \text{if } y_i * y_j \geq y_k * y_l \quad \forall k, l = 1, \dots, N \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

SOMMARIO

Prima Parte

- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- **Problemi Generali**
- Clustering con distanza non euclidea

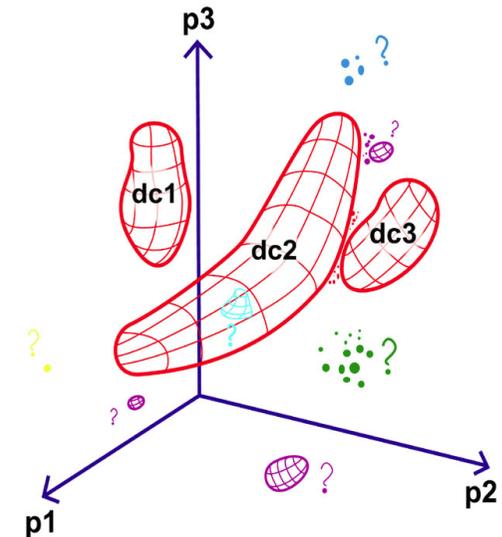
Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

PROBLEMI GENERALI : OUTLIERS

- **Outlier:** anomalia o situazione inattesa che può compromettere il risultato del clustering
- Principali Tecniche per affrontare gli outliers
 - **Noise Clustering**¹
 - **Possibilistic K-means**²
- Metodi di Partitional Clustering suscettibili agli outliers
- SOM abbastanza robusta per sua natura
 - Ogni outlier influenza solo l'unità vincente dello strato di output e il suo vicinato
 - Outliers visualizzabili, in quanto tendono a collocarsi agli angoli della mappa a grande distanza dagli altri Cluster

A Generic Machine-Assisted Discovery Problem:
Data Mapping and a Search for Outliers



¹ Dave, R. N. (1991). *Characterization and detection of noise in clustering*. *Pattern Recognition Letters*, 12, 657664

² Krishnapuram, R., & Keller, J. M. (1993). *A possibilistic approach to clustering*. *IEEE Transactions on Fuzzy Systems*, 1(2), 98110

PROBLEMI GENERALI : MISSING DATA

- **Missing Data** = perdita di informazioni ripristinabile o meno a seconda dell'origine
- Metodo per la manipolazione di dati incompleti : **Expectation – Maximization (EM)**
- SOM poco robusta rispetto a dati incompleti



A causa del degrado introdotto dai missing data, l'unità decretata come vincente potrebbe non essere quella giusta

EM

Algoritmo in grado di stimare parametri secondo il criterio generale della **Maximum Likelihood**.

Il metodo si compone di due passi seguiti iterativamente fino a convergenza:

- **Expectation:** calcolo del valore atteso della massima verosimiglianza
- **Maximization:** determinazione dei valori dei parametri che massimizzano il valore calcolato al passo precedente

PROBLEMI GENERALI : SOTTOUTILIZZO DI UNITÀ

- Alcune unità della rete non vinceranno mai la competizione diventando **unità morte**.
- Soluzioni basate sulla strategia con coscienza:
 - **Frequency Sensitive Competitive Learning (FSCL)** ¹
 - **Rival-Penalized Competitive Learning (RPCL)** ²
- Problema strettamente legato alla regola WTA
- SOM con combinazione di **WTM (Winner Takes Most)** e WTA riesce a far fronte a questo problema

¹Ahalt, S. C., Krishnamurty, A. K., Chen, P., & Melton, D. E. (1990). Competitive learning algorithms for vector quantization. *Neural Networks*, 3(3), 277-290

²Xu, L., Krzyzak, A., & Oja, E. (1993). Rival penalized competitive learning for clustering analysis, RBF net, and curve detection. *IEEE Transactions on Neural Networks*, 4(4), 636-649

PROBLEMI GENERALI : MINIMI LOCALI

- Riguarda il clustering supervisionato
- **Minimi locali:** punti di minimo della funzione che tendiamo a minimizzare durante il processo di addestramento di una rete neurale
- Giungere ad un minimo locale può non essere una buona soluzione del problema
- Algoritmi di Partitional Clustering generalmente suscettibili
- Stocasticità come strumento per sfuggire ai minimi locali

SOMMARIO

Prima Parte

- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

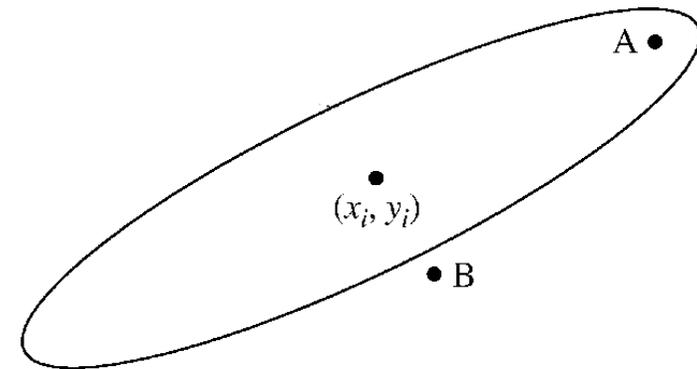
- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

CLUSTERING CON MISURE DI DISTANZA NON EUCLIDEE

- Distanza Euclidea favorisce Cluster circolari
- Distanza di Mahalanobis favorisce Cluster iperellissoidali

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

- Basata su correlazioni tra variabili attraverso le quali differenti pattern possono essere identificati
- Distanza geodetica definita come il percorso di minor lunghezza che unisce due nodi di un grafo



Nell'immagine è mostrata una classica situazione che scaturisce dall'uso della distanza di Mahalanobis. In termini di distanza Euclidea B è nettamente più vicino al punto (x,y) , mentre invece è l'opposto considerando la sopraccitata distanza.

PARTE SECONDA: ONLINE DIMENSIONALITY REDUCTION USING COMPETITIVE HEBBIAN LEARNING AND RADIAL BASIS FUNCTION NETWORK

- Metodo RBF-NDR
 - Apprendimento dello spazio dei dati
 - Riduzione della dimensionalità
- Esperimenti

Online dimensionality reduction using competitive learning and Radial Basis Function network

Vladimir Tomenko*

Environmental Fluid Mechanics, Wessex Institute of Technology, University of Wales, Ashurst Lodge, Southampton, SO40 7AA, UK

ARTICLE INFO

Article history:
Received 1 November 2009
Received in revised form 22 November 2010
Accepted 19 February 2011

Keywords:
Dimensionality reduction
Geodesic distance
Competitive learning
Multidimensional Scaling
Radial Basis Function network

ABSTRACT

The general purpose dimensionality reduction method should preserve data interrelations at all scales. Additional desired features include online projection of new data, processing nonlinearly embedded manifolds and large amounts of data. The proposed method, called RBF-NDR, combines these features.

RBF-NDR is comprised of two modules. The first module learns manifolds by utilizing modified topology representing networks and geodesic distance in data space and approximates sampled or streaming data with a finite set of reference patterns, thus achieving scalability. Using input from the first module, the dimensionality reduction module constructs mappings between observation and target spaces. Introduction of specific loss function and synthesis of the training algorithm for Radial Basis Function network results in global preservation of data structures and online processing of new patterns.

The RBF-NDR was applied for feature extraction and visualization and compared with Principal Component Analysis (PCA), neural network for Sammon's projection (SAMANN) and Isomap. With respect to feature extraction, the method outperformed PCA and yielded increased performance of the model describing wastewater treatment process. As for visualization, RBF-NDR produced superior results compared to PCA and SAMANN and matched Isomap. For the Topic Detection and Tracking corpus, the method successfully separated semantically different topics.

SOMMARIO

Prima Parte

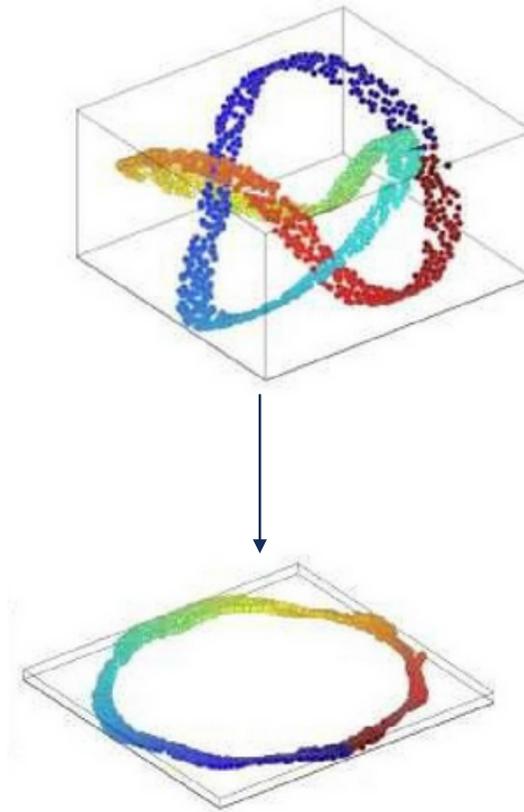
- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

RIDUZIONE DIMENSIONALITÀ

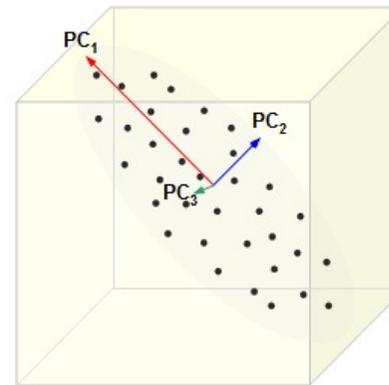
- **Riduzione Dimensionalità:** trovare all' interno dello spazio multidimensionale una rappresentazione significativa ma di ridotta dimensionalità
- **Principal Component Analysis**
- **Multi-Dimensional Scaling**



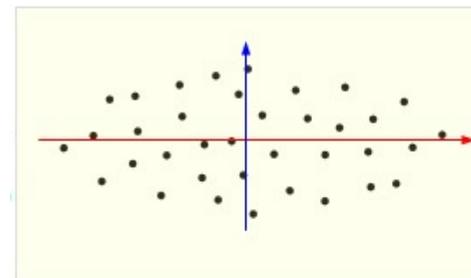
PRINCIPAL COMPONENT ANALYSIS (PCA)

JOLLIFFE, I. T. (1986). PRINCIPAL COMPONENT ANALYSIS. NEW YORK, NY: SPRINGER-VERLAG

- Individuare una Componente Principale per ogni dimensione dello spazio
- Le Componenti Principali devono formare una base ortonormale:
 - Ortogonali tra di loro
 - Di lunghezza unitaria



Le Componenti principali sono ordinate in maniera decrescente in base alla varianza dei dati



La riduzione di dimensionalità avviene eliminando le Componenti Principali con varianza minore

MULTIDIMENSIONAL SCALING (MDS)

TORGERSON, W. S. (1958). THEORY AND METHODS OF SCALING. NEW YORK, NY: JOHN WILEY & SONS

- Basato su criteri di similarità fra coppie di oggetti
- Non riorganizza gli elementi in modo esatto ma approssima il più possibile la similarità osservata nello spazio originale
- Minimizza una funzione di errore chiamata anche di **Stress**

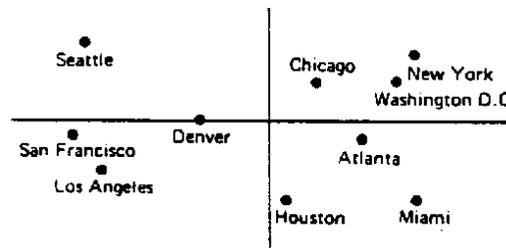
$$STRESS = \sum_{i < j} (d_{ij}^* - d_{ij})^2$$

d_{ij}^* è la distanza calcolata nello spazio multidimensionale

d_{ij} è la distanza calcolata nello spazio a dimensionalità ridotte

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington, DC	
0	587	1212	701	1936	604	748	2139	2182	543	Atlanta
587	0	920	940	1745	1188	713	1858	1737	597	Chicago
1212	920	0	879	831	1726	1631	949	1021	1494	Denver
701	940	879	0	1374	968	1420	1645	1891	1220	Houston
1936	1745	831	1374	0	2339	2451	347	959	2300	Los Angeles
604	1188	1726	968	2339	0	1092	2594	2734	923	Miami
748	713	1631	1420	2451	1092	0	2571	2408	205	New York
2139	1858	949	1645	347	2594	2571	0	678	2442	San Francisco
2182	1737	1021	1891	959	2734	2408	678	0	2329	Seattle
543	597	1494	1220	2300	923	205	2442	2329	0	Washington, DC

La tabella sopra mostra la distanza tra 10 città degli Stati Uniti e può essere usata come matrice di similarità per l'applicazione dell' MDS i cui risultati sono mostrati nella figura in basso



I punti che rappresentano città con un alto grado di similarità sono rappresentati vicini fra loro. Viceversa quelli con bassa similarità sono lontani

SAMMON'S PROJECTION

SAMMON, J. W. (1969)

A NONLINEAR MAPPING FOR DATA STRUCTURE ANALYSIS. IEEE TRANSACTIONS ON COMPUTERS, 18(5), 401–409. DOI:10.1109/T-C.1969.222678

- Estensione dell' MDS
- Utilizza la funzione di Stress proposta da Sammon
- Parte da una configurazione casuale di pattern nello spazio a dimensionalità ridotte e usa il metodo del gradiente discendente per minimizzare la funzione di errore

$$E = \frac{1}{\sum_{i < j} d_{ij}^*} \sum_{i < j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*}$$

Funzione di Stress proposta da Sammon

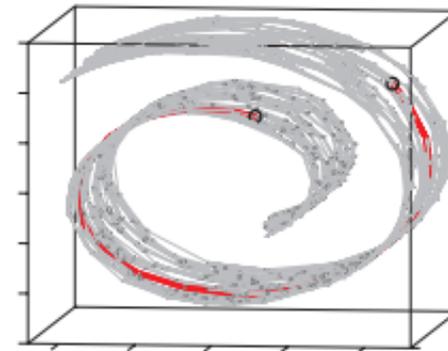
- d_{ij}^* è la distanza calcolata nello spazio multidimensionale
- d_{ij} è la distanza calcolata nello spazio a dimensionalità ridotte

ISOMAP

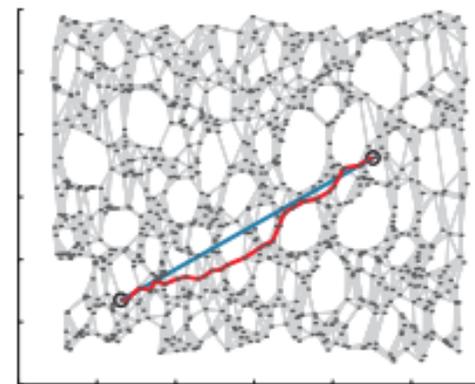
TENENBAUM, J. B., DE SILVA, V., & LANGFORD, J. C. (2000)

A GLOBAL GEOMETRIC FRAMEWORK FOR NONLINEAR DIMENSIONALITY REDUCTION. SCIENCE, 290, 2319–2323

- Estensione dell' MDS
- Utilizza distanza geodetica calcolata su un grafo pesato costruito sullo spazio dei dati
- Vertici del grafo scelti casualmente ed archi creati con **k-rule**:
arco tra x e y se y è uno dei k nodi più vicini ad x



Distanza geodetica tra due punti calcolata sul grafo costruito su una swiss roll



Confronto tra distanza geodetica ed Euclidea dopo la proiezione su uno spazio a dimensionalità ridotta

SOMMARIO

Prima Parte

- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

RADIAL BASIS FUNCTION – NONLINEAR DIMENSIONALITY REDUCTION (RBF-NDR)

TOMENKO, V. (2010). ONLINE DIMENSIONALITY REDUCTION USING COMPETITIVE LEARNING AND RADIAL BASIS FUNCTION NETWORK. UNPUBLISHED DOCTORAL DISSERTATION. UNIVERSITY OF WALES. UK

■ Caratteristiche:

- Preservazione delle correlazioni tra i dati su differenti scale
- Proiezione online di nuovi dati
- Possibilità di processare dati non lineari

■ Si compone di due moduli:

- Pre-clustering dei dati
- Riduzione della dimensionalità

RBF-NDR - Algoritmo

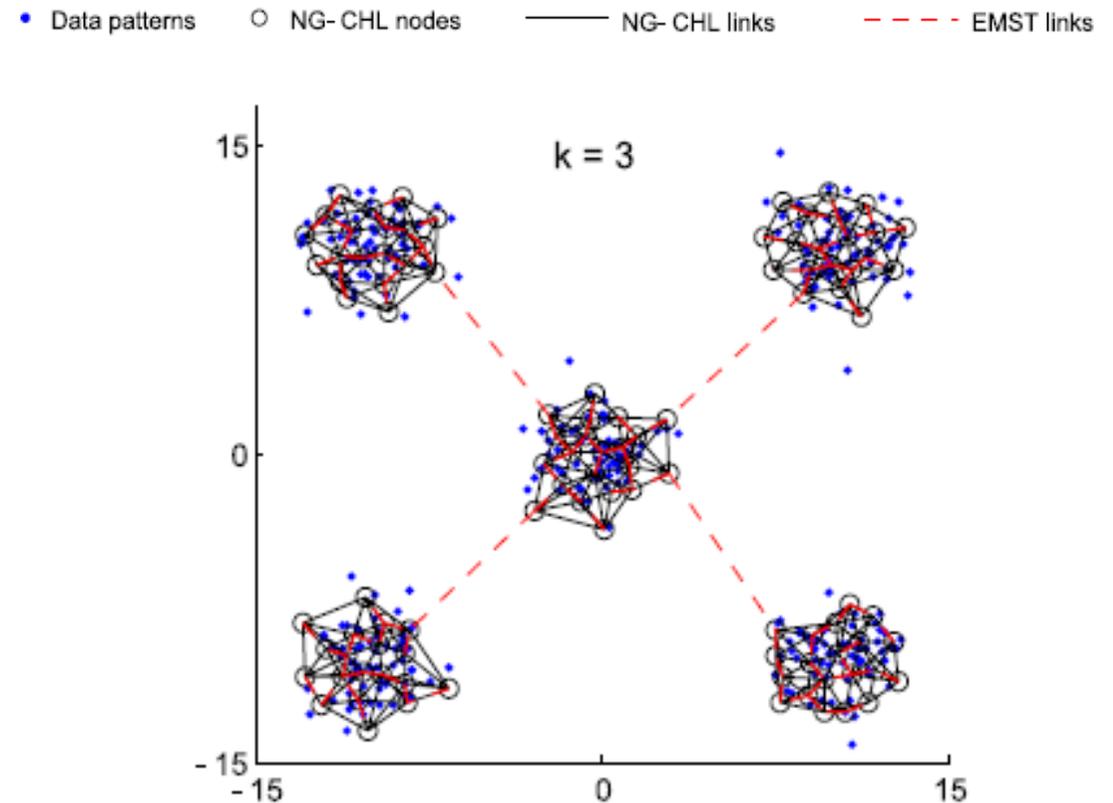
1. Trasforma spazio dei punti iniziale in grafo con vertici (rappresentativi);
2. Connettere topologicamente il grafo precedentemente creato (EMST);
3. Calcola distanze geodetiche tra vertici;
4. Inizializza RBF con η_w , η_c , η_σ , error_threshold, max_iterations, p input, k hidden e m output;
5. Seleziona random coppia di vertici input;
6. Fase forward della RBF;
7. Calcola errore (STRESS);
8. Fase backward della RBF;
9. Se raggiunta max_iterations o error_threshold STOP else goto 5;

RADIAL BASIS FUNCTION – NONLINEAR DIMENSIONALITY REDUCTION (RBF-NDR)

TOMENKO, V. (2010). ONLINE DIMENSIONALITY REDUCTION USING COMPETITIVE LEARNING AND RADIAL BASIS FUNCTION NETWORK. UNPUBLISHED DOCTORAL DISSERTATION. UNIVERSITY OF WALES. UK

RBF-NDR - Algoritmo

1. Trasforma spazio dei punti iniziale in grafo con vertici (rappresentativi);
2. Connettere topologicamente il grafo precedentemente creato (EMST);
3. Calcola distanze geodetiche tra vertici;
4. Inizializza RBF con η_w , η_c , η_σ , error_threshold, max_iterations, p input, k hidden e m output;
5. Seleziona random coppia di vertici input;
6. Fase forward della RBF;
7. Calcola errore (STRESS);
8. Fase backward della RBF;
9. Se raggiunta max_iterations o error_threshold STOP else goto 5;



RADIAL BASIS FUNCTION – NONLINEAR DIMENSIONALITY REDUCTION (RBF-NDR)

TOMENKO, V. (2010). ONLINE DIMENSIONALITY REDUCTION USING COMPETITIVE LEARNING AND RADIAL BASIS FUNCTION NETWORK. UNPUBLISHED DOCTORAL DISSERTATION. UNIVERSITY OF WALES. UK

RBF-NDR - Algoritmo

1. Trasforma spazio dei punti iniziale in grafo con vertici (rappresentativi);
2. Connettere topologicamente il grafo precedentemente creato (EMST);
3. Calcola distanze geodetiche tra vertici;
4. Inizializza RBF con η_w , η_c , η_σ , error_threshold, max_iterations, p input, k hidden e m output;
5. Seleziona random coppia di vertici input;
6. Fase forward della RBF;
7. Calcola errore (STRESS);
8. Fase backward della RBF;
9. Se raggiunta max_iterations o error_threshold STOP else goto 5;

$$E(v, f) = \frac{1}{2} [\delta(v, f) - d(y(v), y(f))]^2 \quad (1) \quad \varphi_j(x_i) = \exp\left(-\frac{1}{2} \sum_{k=1}^p (x_{ik} - c_{jk})^2 / \sigma_{jk}^2\right) \quad (2)$$

$$d(y_v, y_f) = \left[\sum_{i=1}^m (y_i(v) - y_i(f))^2 \right]^{1/2} = \left[\sum_{i=1}^m \left(\sum_{j=1}^K w_{ji} (\varphi_j(v) - \varphi_j(f)) \right)^2 \right]^{1/2} \quad (3)$$

$$E(v, f) = \frac{1}{2} \left(\delta(v, f) - \left[\sum_{i=1}^m \left(\sum_{j=1}^K w_{ji} (\varphi_j(v) - \varphi_j(f)) \right)^2 \right]^{1/2} \right)^2 \quad (4)$$

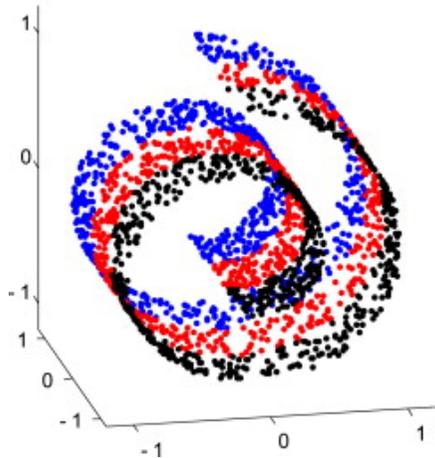
$$\Delta w_{ji} = -\eta_w c(v, f) (y_i(v) - y_i(f)) (\varphi_j(v) - \varphi_j(f)) \quad (5)$$

$$\Delta c_{kj} = -\eta_c c(v, f) \sum_{i=1}^m \left[(y_i(v) - y_i(f)) \frac{w_{ji}}{\sigma_{kj}^2} (\varphi_j(v)(v_k - c_{kj}) - \varphi_j(f)(f_k - c_{kj})) \right] \quad (6)$$

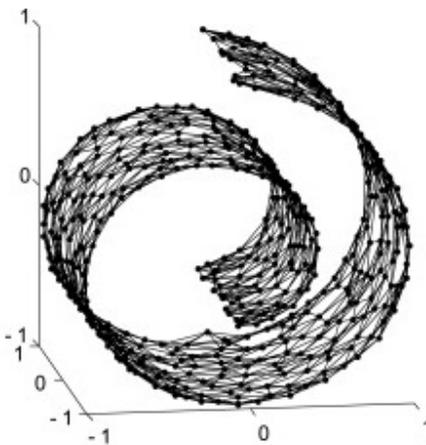
$$\Delta \sigma_{kj} = -\eta_\sigma c(v, f) \sum_{i=1}^m \left[(y_i(v) - y_i(f)) \frac{w_{ji}}{\sigma_{kj}^3} (\varphi_j(v)(v_k - c_{kj})^2 - \varphi_j(f)(f_k - c_{kj})^2) \right] \quad (7)$$

ESPERIMENTO : SWISS ROLL

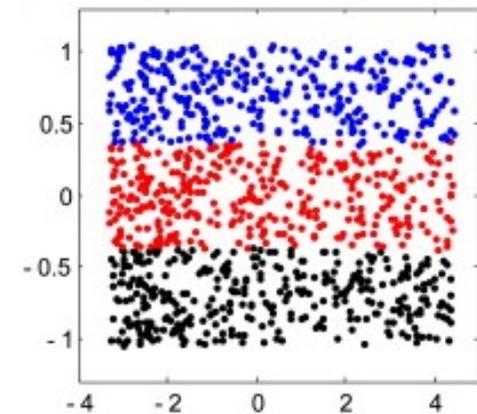
L'algoritmo RBF-NDR viene applicato ad una Swiss Roll



Il primo modulo produce i pattern di riferimento corrispondenti ai vertici del grafo, ovvero la mappa topologica



Proiezione stimata dal secondo modulo



SOMMARIO

Prima Parte

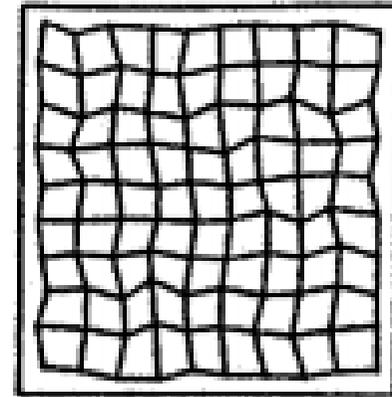
- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - **Ottimizzazioni possibili: ESOM, ETREE**
- Possibili sviluppi tesi

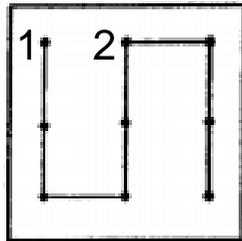
LIMITI SOM

- SOM e Dimensionality Reduction : problema della griglia
- SOM non assicura la costruzione di una Topology Preserving Map



Grafo bi-dimensionale rappresentate lo strato di Kohonen che si distribuisce su uno spazio dei dati di forma quadrata

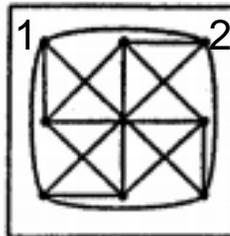
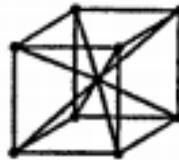
LIMITI SOM (2)



SPAZIO DEI DATI BIDIMENSIONALE
RETICOLO **MONODIMENSIONALE**



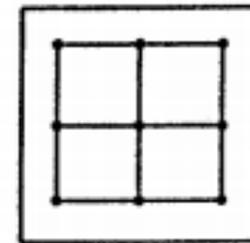
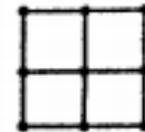
NO TOPOLOGY PRESERVING MAP



SPAZIO DEI DATI BIDIMENSIONALE
RETICOLO **TRIDIMENSIONALE**



NO TOPOLOGY PRESERVING MAP



SPAZIO DEI DATI BIDIMENSIONALE
RETICOLO **BIDIMENSIONALE**



TOPOLOGY PRESERVING MAP

EVOLVING SOM

DENG, D., & KASABOV, N. (2000). ESOM: AN ALGORITHM TO EVOLVE SELF-ORGANIZING MAPS FROM ON-LINE DATA STREAMS. IN PROCEEDINGS OF IEEE-INNS-ENNS INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS. VOL. 6 (PP. 3–8). DOI:10.1109/IJCNN.2000.859364

- Parte da una griglia vuota
- Se l'input corrente non trova corrispondenza adeguata, ovvero la distanza supera una certa soglia ε , viene aggiunto un nuovo nodo alla rete
- Nodo aggiunto uguale al pattern in input: allocazione più semplice che migliora le performance rispetto ad altri algoritmi a struttura incrementale (**GCS,GNG**).
- Per ogni pattern si calcolano i migliori due BMU e se una connessione tra loro esiste, quest'ultima viene aggiornata, altrimenti viene creata

Clustering unsupervised
Legge di Kohonen modificata
Topologia simile alla SOM

$$\Delta \mathbf{w}_i = \begin{cases} \gamma \frac{a_i}{\sum_k a_k} (\mathbf{x} - \mathbf{w}_i) & \text{if } i \in \Omega(b), \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Legge di apprendimento

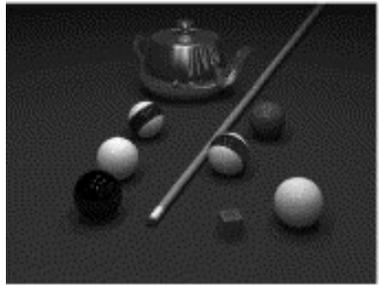
$$a_i = e^{-2\|\mathbf{x} - \mathbf{w}_i\|^2 / \varepsilon^2} \quad (2)$$

Attivazione

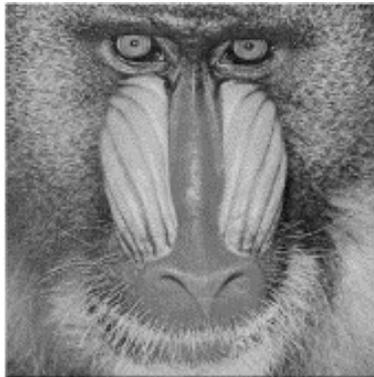
$$s(i, j) = \beta s(i, j) + (1 - \beta) a_i a_j \quad (3)$$

Connessione sinaptica

EVOLVING SOM (2)



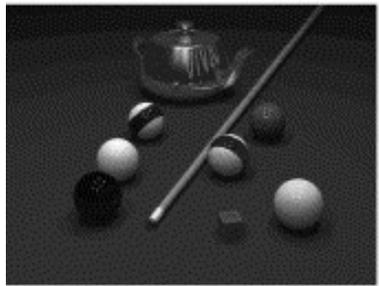
(a)



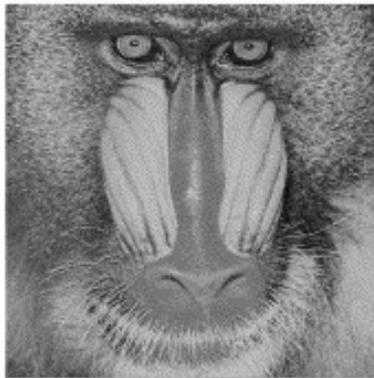
(b)



(c)



(d)



(e)



(f)

Quantisation performances: E_q/σ_q

Methods	Images		
	Pool balls	Mandrill	Lenna
Median-cut	2.58/8.28	11.32/5.59	6.03/3.50
Octree	4.15/3.55	13.17/4.98	7.56/3.83
Wu's	2.22/2.19	9.89/4.56	5.52/2.94
Local K -means	3.54/2.77	11.48/5.42	6.70/2.84
ESOM	2.43/2.56	9.47/3.86	5.28/2.36

EVOLVING TREE

PAKKANEN, J.; LIVARINEN, J.; OJA, E.; 2004. THE EVOLVING TREE—A NOVEL SELF-ORGANIZING NETWORK FOR DATA ANALYSIS. NEURAL PROCESSING LETTERS 20, 199–211, KLUWER ACADEMIC PUBLISHERS

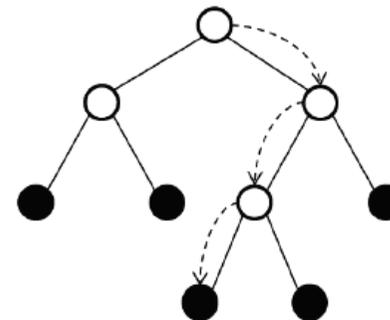
- Evoluzione supervisionata della SOM con struttura incrementale ad albero che supera il problema della griglia statica e aumenta prestazioni
- Inizialmente rete costituita da un singolo nodo che subisce una fase di **splitting** con cui vengono creati i figli
- Ogni nodo ha associata una variabile che conta il numero di volte in cui è stato BMU, quando supera una certa soglia si procede allo splitting del nodo incrementando la struttura dell'albero
- Per aggiornare la posizione delle foglie utilizza la legge di apprendimento di Kohonen

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{w}_i(t)] \quad (1)$$

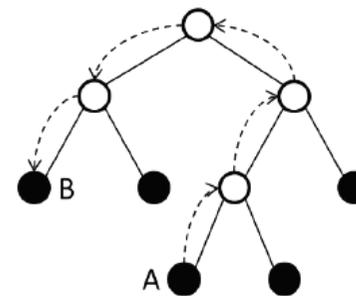
$$h_{ci}(t) = \alpha(t)e^{-\frac{d(r_c, r_i)^2}{2\sigma^2(t)}} \quad (2)$$

$$b_i(t+1) = \gamma b_i(t) \quad (3)$$

Clustering supervised Legge di Kohonen invariata Topologia ad albero



La ricerca del BMU avviene partendo dalla radice e scendono fino ad un nodo foglia



Trattandosi di un albero la distanza tra neuroni è data dal numero di archi che intercorrono tra i due nodi

SOMMARIO

Prima Parte

- Regola di base
- Tecniche di Clustering
 - Partitional Clustering
 - Hierarchical Clustering
 - Density Based Clustering
- Algoritmi di clustering
 - SOM classica e variante supervisionata: LVQ
 - K-Means
 - Mountain Clustering
 - Reti ART
 - Neural Gas
 - NG per Topology Preserving Maps : NG-CHL
- Problemi Generali
- Clustering con distanza non euclidea

Seconda Parte

- Riduzione Dimensionalità
 - Principal Component Analysis
 - Multi Dimensional Scaling
 - Sammon's Projection
 - Isomap
- RBF-NDR
 - Descrizione Metodo
 - Esperimenti
 - Ottimizzazioni possibili: ESOM, ETREE
- Possibili sviluppi tesi

POSSIBILI SVILUPPI TESI

Fasi lavoro completo

- 
1. Implementazione SOM standard
 2. Implementazione ESOM+EMST
 3. Implementazione ETREE
 4. Implementazione NG-CHL+EMST
 5. Implementazione RBF-NDR
 6. Confronto tra NG-CHL, ETREE ed ESOM come metodi alternativi di pre-clustering per RBF-NDR
 7. Test su problemi di astrofisica



1° lavoro di tesi (Francesco)

1. Implementazione SOM standard
2. Implementazione ESOM+EMST
3. Implementazione ETREE
4. Test su problemi di astrofisica e confronto prestazioni

2° lavoro di tesi (TBD)

1. Implementazione NG-CHL+EMST
2. Implementazione RBF-NDR
3. Confronto tra NG-CHL, ETREE ed ESOM come metodi alternativi di pre-clustering per RBF-NDR
4. Test su problemi di astrofisica