

LEZIONE 1 – E-SCIENCE e DATA WAREHOUSE CONCEPTS

1. Quali sono i paradigmi concettuali su cui si basa l'astroinformatica? Qual è il processo alla base di questa disciplina?
2. Qual è il concetto base del Virtual Observatory e quali sono i suoi obiettivi?
3. Quali sono gli obiettivi primari del data mining o data analysis? E come possono essere sfruttati in modo fruttuoso all'interno di una generica disciplina scientifica?
4. Qual'è la differenza tra computational science e computer science?
5. Qual'è l'obiettivo primario della cosiddetta X-Informatics?
6. Cos'è il data warehousing? Che relazioni vi sono con il data mining?
7. Quali sono i cicli virtuosi del data mining?
8. Quali sono le caratteristiche delle tecniche OLTP, OLAP, OLAM e dove si colloca il data mining rispetto a loro?
9. Qual è la migliore soluzione in termini di data mining per minimizzare l'odierna esigenza dell'esplorazione e sfruttamento dei massive data sets?

LEZIONE 2 – DATA MINING e ARTIFICIAL INTELLIGENCE

1. Prendendo spunto dai moderni progetti di data warehouse astrofisiche, quali sono i concetti chiave alla base delle soluzioni adottate o adottabili per l'analisi e processamento dei dati in modo efficiente?
2. Come si relazionano concetti come KDD, Data Mining e Machine Learning?
3. Qual'è la definizione di data mining? Può fare un esempio di cosa è data mining e cosa non lo è?
4. Quali sono i principali task che caratterizzano il data mining?
5. Quali sono i principi del data mining inteso come business intelligence?
6. Quali sono le principali analogie e differenze tra data mining e analisi statistica?
7. Quali sono le principali metodologie impiegate dal data mining?
8. Qual è la definizione di classificazione e qualche esempio pratico in ambito astrofisico?
9. Qual è la definizione di dimensional reduction e su quali tecniche è basato? Qualche esempio pratico in ambito astrofisico?
10. Qual è la definizione di regression analysis e qualche esempio pratico in ambito astrofisico?
11. Qual è la definizione di clustering e qualche esempio pratico in ambito astrofisico?
12. In un sistema di classificazione basato su machine learning come si può incorporare informazione proveniente da fonti esterne?
13. Cos'è un sistema combinato di data mining? Che proprietà deve possedere?
14. Quali sono i due approcci teorici alla base del concetto di Intelligenza artificiale?

LEZIONE 3 – STATISTICA

1. Dato un insieme di punti su un piano cartesiano, qual'è la differenza tra la linea di regressione e quella della media dei punti?
2. Come si fa l'analisi di regressione di un insieme di punti 2D?
3. Come si calcolano i gradi libertà per l'analisi di regressione di un sistema di dati tridimensionali?

4. Qual'è il significato della regola delle 3sigma nell'ambito del post-processing di un sistema di data mining?
5. Quali sono gli estimatori statistici per un sistema di analisi dei risultati basato su metodi di regressione non-lineare?
6. Quali sono gli estimatori statistici più comuni per un sistema di analisi dei risultati basato su metodi di classificazione?
7. Qual'è il significato e ruolo nell'ambito del data mining degli stimatori kurtosis, skewness e modality?
8. Quali test statistici conosce per l'analisi dei dati in un processo di data mining?
9. Come funziona il test di Kolmogorov-Smirnov?
10. Che ruolo ha la statistica della covarianza nei sistemi di analisi dati basati su data mining?
11. Dato un sistema di analisi di regressione su coppie di dati, qual'è la differenza tra la covarianza e la correlazione?
12. Come si può usare il metodo delle curve di ROC all'interno di un sistema basato su data mining?
13. Come si può usare la stima statistica degli intervalli di confidenza in ambito astrofisico per l'analisi dei dati?
14. Che cos'è la Probability Density Function di una distribuzione di dati osservati sperimentalmente?

LEZIONE 4/5 – MACHINE LEARNING SUPERVISIONATO

1. Quali sono i concetti alla base del paradigma supervisionato del Machine Learning?
2. Quali sono i principali domini funzionali di applicazione delle tecniche di Machine Learning?
3. Quali sono le tipologie di apprendimento in sistemi basati sul machine learning?
4. Che cos'è una rete neurale artificiale? E quali sono gli aspetti principali che le caratterizzano rispetto ad altri sistemi di analisi dati?
5. In cosa si differenzia il perceptrone multistrato rispetto al singolo perceptrone?
6. Come si può costruire a livello architetturale una rete neurale ricorrente (time-delay) a partire da un generico perceptrone?
7. In cosa si differenzia la distanza di Mahalanobis rispetto a quella standard euclidea? Per cosa si può usare in ambito astrofisico legato all'analisi dati?
8. Quali sono le principali caratteristiche funzionali delle reti neurali, logica fuzzy e algoritmi genetici?
9. Qual'è la principale prerogativa della tecnica della cross entropy in un modello di machine learning?
10. Quali principali proprietà deve possedere una funzione di attivazione per un neurone artificiale?
11. Quali sono i concetti alla base del meccanismo di learning dei modelli supervisionati? Come si può risolvere o minimizzare il problema dei minimi locali in un problema di ottimizzazione?
12. In cosa differisce una rete MLP da una RBF?
13. Qual'è il ruolo del parametro decay in una rete MLP addestrata con il metodo del Quasi Newton?
14. Quali sono i concetti base di un algoritmo genetico?

15. Quali sono i principali meccanismi di evoluzione e valutazione di una popolazione in un algoritmo genetico?
16. In quali ambiti astrofisici può essere vantaggioso l'uso di una tecnica di data mining basata su algoritmi genetici?
17. Come funziona il modello SVM?
18. Quali sono le principali differenze tra una rete neurale ed una SVM?
19. Quali sono i concetti alla base della logica fuzzy?
20. Qual'è la principale differenza tra inferenza fuzzy e statistica?
21. Come funzionano i decision tree? Come si costruiscono?
22. Come funziona il modello KNN? Quali sono i limiti principali di questo metodo?
23. Quali sono le principali differenze tra un decision tree, un KNN ed una SVM?
24. Quali sono gli aspetti principali alla base del modello Random Forest?
25. Quali sono le principali differenze tra un decision tree, un KNN, RF ed una SVM?
26. Cosa sono i modelli di machine learning basati su logica bayesiana? Come si costruiscono?
27. Qual'è la principale caratteristica dei modelli Naive Bayes?
28. Come funziona un modello Bayesian Belief Network?
29. Cosa sono i gated expert? Quale uso è indicato in ambito astrofisico?

LEZIONE 6 – MACHINE LEARNING NON SUPERVISIONATO

1. Quali sono i concetti alla base del paradigma non supervisionato del Machine Learning?
2. Come funziona la tecnica unsupervised dell'apprendimento competitivo? Quella hebbiana?
3. Quali sono le principali metriche per la valutazione di un esperimento di clustering? Quali aspetti caratterizzano l'una rispetto all'altra?
4. Quali sono le principali metriche per la misura di distanza tra oggetti in un esperimento di clustering? In quale modello è particolarmente indicato l'uso della distanza geodetica?
5. Quali sono le principali caratteristiche di un clustering gerarchico e le differenze con un clustering di tipo partizionale?
6. Come funziona il metodo del K-means?
7. Quali sono i concetti alla base del modello Neural Gas?
8. Quali sono i concetti alla base del modello Self Organizing Map?
9. Quali sono i più noti meccanismi di aggiornamento dei pesi in una SOM?
10. Qual'è un metodo per la visual inspection di un risultato di un esperimento tramite modello SOM? Come funziona?
11. Come si può stimare la qualità di un risultato di un esperimento tramite SOM?
12. Quali sono le principali prerogative degli stimatori di un modello di clustering, errore di quantizzazione, errore topografico, indice di Davies-Bouldin, indici di accuratezza e completezza?
13. Quali sono le principali prerogative di un clustering gerarchico?
14. Quali sono i principali limiti del modello SOM? Come si possono risolvere?
15. Quali sono le principali differenze tra un modello Evolving SOM e Evolving Tree?
16. Come funziona il metodo ibrido CHL (Competitive Hebbian Learning)?
17. Quali sono le tipologie di clustering basate sulla tecnica del Minimum Spanning Tree?

18. Come si può applicare una tecnica di clustering al problema della stima dell'estinzione galattica?
19. Quali sono le differenze principali tra una trasformata di Fourier e una Wavelet?
20. Come si possono affrontare i problemi della segmentazione di immagini astronomiche? E dell'edge detection?
21. Come funzionano le tecniche dell'analisi della Silhouette e del Region Contamination Index?
22. Quali sono le tecniche di ottimizzazione del modello Neural Gas?

LEZIONE 7 – DEEP LEARNING

1. Quali sono i concetti alla base delle tecniche di deep learning?
2. Che cos'è una rete auto-encoder?
3. Quali sono i meccanismi alla base dei modelli di Convolutional Neural Network?
4. Come funziona la tecnica della Softmax? In quale categoria di modelli è particolarmente indicata?
5. In quali casi d'uso astrofisici è potenzialmente utile un modello di deep learning?

LEZIONE 8 – PARALLEL PROGRAMMING

1. Perché è stata recentemente incentivata la programmazione parallela?
2. Come si calcola lo speedup di un sistema computazionale?
3. Cosa implicano le leggi di Moore, Amdahl e Gustaffson?
4. Qual'è la tassonomia base dei sistemi multi-process?
5. Quali sono i principali componenti logici alla base di una scheda GPU?
6. Quali sono le principali evoluzioni recentemente adottate nella tecnologia GPU?
7. Quali sono le principali differenze tra programmazione MPI e GPU?
8. Quali sono i meccanismi di parallelizzazione di un codice seriale pre-esistente?
9. Come si possono parallelizzare modelli di machine learning? Un GA? Una MLP?

LEZIONE 9 – BIG DATA

1. In quanti modi si possono rappresentare i moderni dati astronomici?
2. Qual'è la principale prerogativa dei metadati? Faccia un esempio di metadato in Astrofisica
3. Quali sono le caratteristiche base di una web application e di un web service?
4. Qual'è la tecnica base per ottimizzare un software di cross-matching fra cataloghi astronomici?
5. Quali sono le differenze tra le moderne architetture computazionali? Quale può essere la migliore strategia in ambito astrofisico?
6. Quali sono le principali caratteristiche dei Big data?
7. Su quali concetti si può specializzare un database? Quali soluzioni si possono adottare nel caso di Big Data?
8. Cos'è un Big Data Cycle? E quali sono gli aspetti da analizzare per la costruzione di un sistema basato sui Big Data?

LEZIONE EXTRA – ESEMPI DI DATA MINING IN ASTROFISICA

1. Photometric redshifts;
2. classificazione di oggetti peculiari (AGN, QSO);
3. Time Domain Astronomy;
4. Stima estinzione galattica;
5. Band-merging;
6. Filament detection e edge optimization;
7. galactic clump characterization e star forming evolutionary analysis;
8. Galaxy morphology classification;
9. Strong Lensing classification;
10. Neutrino detectors;
11. Cherenkov radiation detection;
12. Data Quality in astronomical pipelines;
13. Outliers selection and classification;
14. parameter space dimensionality reduction;