

FACOLTÀ	Scienze MM.FF.NN.
ANNO ACCADEMICO	2014/2015
CORSO DI LAUREA	Informatica
INSEGNAMENTO	Algoritmi per l'Analisi dei Dati
TIPO DI ATTIVITÀ	Caratterizzante
AMBITO DISCIPLINARE	Discipline Informatiche
CODICE INSEGNAMENTO	16479
ARTICOLAZIONE IN MODULI	NO
SETTORI SCIENTIFICO DISCIPLINARI	INF/01
DOCENTE RESPONSABILE	Lo Bosco Giosuè Ricercatore Universitario Università di Palermo
CFU	6
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	102
NUMERO DI ORE RISERVATE ALLE ATTIVITÀ DIDATTICHE ASSISTITE	48
PROPEDEUTICITÀ	Tutte le materie del I anno, Algoritmi e Strutture Dati, Informatica Teorica, Basi di Dati
ANNO DI CORSO	III
SEDE DI SVOLGIMENTO DELLE LEZIONI	Consultare il sito www.cs.unipa.it
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni frontali/Lezioni laboratorio ed esercitazioni
MODALITÀ DI FREQUENZA	Facoltativa
METODI DI VALUTAZIONE	Prova orale, presentazione di un elaborato
TIPO DI VALUTAZIONE	Voto in trentesimi
PERIODO DELLE LEZIONI	I semestre
CALENDARIO DELLE ATTIVITÀ DIDATTICHE	Come da calendario disponibile presso www.cs.unipa.it
ORARIO DI RICEVIMENTO DEGLI STUDENTI	Lunedì dalle ore 14,00 alle ore 15,00 Venerdì dalle ore 14,00 alle ore 15,00.

RISULTATI DI APPRENDIMENTO ATTESI

Conoscenza e capacità di comprensione

Il corso fornisce le conoscenze teoriche e pratiche per potere effettuare analisi esplorativa di dati sperimentali, con particolare riferimento alla classificazione supervisionata e non supervisionata.

Inoltre fornisce le conoscenze necessarie per l'uso di base e avanzato di Matlab che sarà il linguaggio utilizzato per sviluppare opportuni algoritmi.

Capacità di applicare conoscenza e comprensione

Gli studenti acquisiranno la capacità di utilizzare metodologie per effettuare test di ipotesi, riduzione della dimensione dei dati, analisi cluster e classificazione di dati sperimentali di diversa natura.

Autonomia di giudizio

Gli studenti sono guidati ad apprendere in maniera critica e responsabile tutto ciò che viene spiegato in aula e ad arricchire le proprie capacità di giudizio sia attraverso lo studio del

materiale didattico indicato o fornito dal docente, sia attraverso la realizzazione di elaborati pratici consistenti nell'implementazione di applicazioni per l'analisi esplorativa dei dati.

Abilità comunicative

Attraverso l'interazione durante le lezioni e le attività di laboratorio previste, il corso tenderà a favorire lo sviluppo della capacità di comunicare in modo chiaro ed esaustivo le proprie ragioni ed argomentare le proprie conclusioni. Gli studenti dovranno altresì sviluppare la capacità di lavorare in gruppo, di confrontarsi sulle problematiche proposte rispettando i punti di vista diversi dal proprio e arricchendo le conoscenze acquisite durante il corso con la dialettica e il confronto tra pari.

Capacità d'apprendimento

Attraverso approfondimenti e consultazione dei testi di riferimento, gli studenti saranno stimolati ad una conoscenza più approfondita e critica sui diversi modi in cui i dati possono essere analizzati per estrarne conoscenza.

OBIETTIVI FORMATIVI

Obiettivo del modulo è quello di fornire le metodologie e le tecniche fondamentali per l'analisi e l'elaborazione informatica dei dati sperimentali. Tali metodologie potranno essere applicate a svariati problemi scientifici, non necessariamente inerenti al settore scientifico della materia in oggetto. In particolare verranno trattati gli algoritmi fondamentali per il clustering e la classificazione dei dati, con particolare riferimento ad applicazioni su dati sperimentali reali. Le esercitazioni comprendono l'implementazione di algoritmi in MATLAB. A tale scopo, una parte del corso sarà dedicata a fornire agli studenti la conoscenza di base ed avanzata di tale linguaggio.

ORE FRONTALI	LEZIONI FRONTALI
12	Matlab Comandi base; variabili e matrici; tipi di dato; funzioni built-in; script e funzioni definite dall'utente; funzioni con input e output variabile; strutture di selezione; strutture iterative; funzioni grafiche; funzioni di esportazione di dati ed importazione di dati esterni; funzioni di analisi dei dati.
12	Probabilità, statistica, ottimizzazione Cenni su ottimizzazione, Probabilità, Variabili random continue e discrete, Probabilità condizionata, Teorema di Bayes, Distribuzioni, Campionamento e inferenza statistica, Stima dei parametri, Massima verosimiglianza, statistica descrittiva, test di ipotesi, intervalli di confidenza, regressione, stime di densità.
6	Clustering Il clustering come problema di ottimizzazione, clustering gerarchico agglomerativo (single, complete, average link), clustering gerarchico divisivo (MST clustering), clustering partizionale (k-means, PAM, Isodata, Nearest Neighbor clustering).
4	Classificazione La classificazione come problema di ottimizzazione, classificatore bayesiano, classificatore nearest neighbor, Linear discriminant analysis.
4	Validazione Matrici di confusione, curve ROC, indici di validazione di clustering

	interni ed esterni.
--	---------------------

LEZIONI DI LABORATORIO ED ESERCITAZIONI	
2	Esercitazioni sull'utilizzo di Matlab.
2	Implementazione di metodologie di statistica descrittiva in Matlab.
2	Implementazione di algoritmi di clustering in Matlab e loro utilizzo su dataset di riferimento.
2	Implementazione di algoritmi di classificazione in Matlab e loro utilizzo su dataset di riferimento.
2	Implementazione di indici di validazione e loro utilizzo su problemi di classificazione e clustering.

TESTI CONSIGLIATI	<p>Holly Moore, Matlab per l'ingegneria, Pearson Education Italia (2008) - ISBN: 978-8-871-92447-2</p> <p>M.R. Berthold, D.J. Hand, Intelligent Data Analysis: An Introduction, Springer.</p> <p>A.K.Jain, R.C.Dubes: Algorithms for clustering data, Prentice Hall.</p> <p>Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, Wiley</p>
--------------------------	---