

<b>FACOLTÀ</b>	Scienze MM.FF.NN.
<b>ANNO ACCADEMICO</b>	2015/2016
<b>CORSO DI LAUREA</b>	Informatica
<b>INSEGNAMENTO</b>	Algoritmi per l'Analisi dei Dati
<b>TIPO DI ATTIVITÀ</b>	Caratterizzante
<b>AMBITO DISCIPLINARE</b>	INF/01
<b>CODICE INSEGNAMENTO</b>	16479
<b>ARTICOLAZIONE IN MODULI</b>	NO
<b>SETTORI SCIENTIFICO DISCIPLINARI</b>	INF/01
<b>DOCENTE RESPONSABILE</b>	Lo Bosco Giosuè Ricercatore Universitario
<b>CFU</b>	6
<b>NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE</b>	102
<b>NUMERO DI ORE RISERVATE ALLE ATTIVITÀ DIDATTICHE ASSISTITE</b>	48
<b>PROPEDEUTICITÀ</b>	Tutte le materie del I anno, Algoritmi e Strutture Dati, Informatica Teorica, Sistemi Operativi
<b>ANNO DI CORSO</b>	Terzo
<b>SEDE DI SVOLGIMENTO DELLE LEZIONI</b>	Laboratorio informatico - via Ingrassia, Palermo
<b>ORGANIZZAZIONE DELLA DIDATTICA</b>	Lezioni frontali, Esercitazioni in laboratorio
<b>MODALITÀ DI FREQUENZA</b>	Facoltativa
<b>METODI DI VALUTAZIONE</b>	Assegnazione elaborato e Prova Orale
<b>TIPO DI VALUTAZIONE</b>	Voto in trentesimi
<b>PERIODO DELLE LEZIONI</b>	I semestre
<b>CALENDARIO DELLE ATTIVITÀ DIDATTICHE</b>	Come da calendario disponibile presso <a href="http://www.cs.unipa.it">www.cs.unipa.it</a>
<b>ORARIO DI RICEVIMENTO DEGLI STUDENTI</b>	Da concordare

## **RISULTATI DI APPRENDIMENTO ATTESI**

### **Conoscenza e capacità di comprensione**

Il corso fornisce le conoscenze teoriche e pratiche per potere effettuare analisi 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.

<b>ORE FRONTALI</b>	<b>LEZIONI FRONTALI</b>
<b>14</b>	<b>Malab</b> 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; GUI
<b>6</b>	<b>Statistica</b> Metodi di stima, standardizzazione, covarianza, correlazione, analisi dell'componenti principali, intervalli di confidenza, test di ipotesi, regressione, stime dirette di densità.
<b>6</b>	<b>Clustering</b> 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).
<b>4</b>	<b>Classificazione</b> La classificazione come problema di ottimizzazione, classificatore bayeiano, classificatore nearest neighbor, Linear discriminant analysis.
<b>2</b>	<b>Validazione</b> Matrici di confusione, curve ROC, indici di validazione di clustering interni ed esterni.

	LEZIONI DI LABORATORIO ED ESERCITAZIONI
4	ESERCITAZIONI SULL'USO DI MATLAB Vettori e matrici, funzioni di minimo e massimo, funzioni di ricerca, funzioni grafiche, GUI
4	CLUSTERING Utilizzo del kmeans e degli algoritmi gerarchici in Matlab., utilizzo dei dataset piu' comuni inclusi in matlab. Implementazione del calcolo della matrice di adiacenza di una soluzione di clustering, implementazione dell'algoritmo di Minimum spanning tree clustering
4	CLASSIFICAZIONE Implementazione di un classificatore bayesiano, di un classificatore k-vicini, del calcolo della matrice di confusione, della curva ROC
4	VALIDAZIONE Implementazione del calcolo degli indici di Omogeneità, Separazione, Davies–Bouldin, Dunn, Cluster Purity, Rand, Jaccard, Fowlkess, Fowlkess –Mallows.
<b>TESTI CONSIGLIATI</b>	A.K.Jain, R.C.Dubes: <b>Algorithms for clustering data</b> , Prentice Hall.  M.Berthold, D.j.Hand , <b>Intelligent Data Analysis (An introduction)</b> , Springer (ISBN 3-540-65808-4)  Richard O. Duda, Peter E. Hart, David G. Stork, <b>Pattern Classification</b> , Wiley