



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Ingegneria
ANNO ACCADEMICO OFFERTA	2023/2024
ANNO ACCADEMICO EROGAZIONE	2024/2025
CORSO DILAUREA MAGISTRALE	MANAGEMENT ENGINEERING
INSEGNAMENTO	BIG DATA AND ANALYTICS C.I.
CODICE INSEGNAMENTO	21506
MODULI	Si
NUMERO DI MODULI	2
SETTORI SCIENTIFICO-DISCIPLINARI	ING-INF/05, ING-INF/03
DOCENTE RESPONSABILE	TINNIRELLO ILENIA Professore Ordinario Univ. di PALERMO
ALTRI DOCENTI	LO PRESTI LILIANA Professore Associato Univ. di PALERMO
	TINNIRELLO ILENIA Professore Ordinario Univ. di PALERMO
CFU	6
PROPEDEUTICITA'	
MUTUAZIONI	
ANNO DI CORSO	2
PERIODO DELLE LEZIONI	2° semestre
MODALITA' DI FREQUENZA	Facoltativa
TIPO DI VALUTAZIONE	Voto in trentesimi
ORARIO DI RICEVIMENTO DEGLI STUDENTI	LO PRESTI LILIANA Martedì 16:00 17:00 TINNIRELLO ILENIA Lunedì 9:00 12:00 Ufficio del docente, presso il DEIM, secondo piano.

DOCENTE: Prof.ssa ILENIA TINNIRELLO

PREREQUISITI	Il corso è auto-consistente, anche se è preferibile avere alcune basi di probabilità, algebra lineare, programmazione e basi di dati.
RISULTATI DI APPRENDIMENTO ATTESI	Conoscenza e capacità di comprensione: Al termine del corso lo studente sarà in grado di comprendere e approfondire i problemi relativi alla classificazione e alcune metodologie per ricavare vari tipi di informazioni direttamente dai dati. In particolare sarà in grado di comprendere il funzionamento di alcuni classificatori, come i classificatori bayesiani, lineari e non lineari, classificatori basati su catene di Markov nascoste e su reti neurali profonde, nonché di condurre analisi esplorative sui dati, applicando algoritmi di clustering e estrazione delle features. Avrà conoscenza dei modelli di calcolo, delle architetture e delle infrastrutture necessarie all'elaborazione di Big Data. Avrà infine familiarità con le problematiche legate all'analisi dei Big Data. Conoscenza e capacità di comprensione applicate: Lo studente sarà guidato ad implementare gli algoritmi studiati per applicazioni su dataset reali anche di grandi dimensioni. Sarà inoltre stimolato ad estrapolare gli algoritmi illustrati nel corso dal contesto specifico e ad applicare tali algoritmi (e relative considerazioni) ad altri scenari applicativi e in particolare all'analisi di segnali audio e video. Sarà in grado di fornire opportune soluzioni progettuali per la realizzazione di sistemi in grado di gestire Big Data in maniera efficiente. Autonomia di giudizio Lo studente sarà in grado di condurre diversi tipi di analisi esplorative dei dati in funzione del problema da affrontare. Sarà in grado di generalizzare le tecniche e i concetti acquisiti e stabilirne le relazioni con quelli introdotti nelle discipline a questa correlate. Sarà infine in grado di valutare la bontà delle diverse soluzioni algoritmiche e architetture per la gestione dei Big Data e di valutarne le prestazioni. Abilità comunicative Lo studente dovrà acquisire la capacità di comunicare razionalmente le sue conoscenze sugli argomenti oggetto del corso, con padronanza del lessico specializzato del settore. In particolare, dovrà essere capace di motivare le scelte effettuate nella risoluzione dei problemi di analisi e/o sintesi e di sostenere conversazioni su tematiche relative alla progettazione e gestione di sistemi per la gestione dei Big Data. Capacità di apprendere Lo studente sarà in grado di approfondire autonomamente le tecniche non studiate nel corso e la letteratura scientifica del settore, allo scopo di applicare metodologie più complesse. Sarà inoltre in grado di comprendere e utilizzare nuovi strumenti per la gestione dei Big Data.
VALUTAZIONE DELL'APPRENDIMENTO	MODALITÀ DI ESAME: Una prova scritta obbligatoria e una prova orale facoltativa. La prova orale può essere sostenuta per migliorare la valutazione della prova scritta. Per sostenere la prova orale è comunque necessario superare la prova scritta. La prova scritta è valutata in trentesimi. Il voto minimo per superare la prova scritta è 18/30. La prova orale è valutata con un punteggio aggiuntivo da 0 a 3 punti sul voto in trentesimi. Il voto finale è il voto della prova scritta (nel caso in cui la prova orale non venga sostenuta) o la somma del voto della prova scritta e del punteggio aggiuntivo della prova orale. DESCRIZIONE DELLE PROVE La prova scritta è basata su tre domande aperte e su quattro esercizi, simili a quelli discussi nelle esercitazioni guidate, in cui si richiede di applicare le metodologie presentate nel corso a semplici problemi di analisi di dati, modelli di sistemi e inferenza statistica. La durata complessiva della prova è 2.5 ore. La prova è finalizzata ad accertare: - il grado di conoscenza e padronanza dei concetti e degli algoritmi discussi a lezione; - la capacità di utilizzo dei suddetti concetti e algoritmi per risolvere in modo autonomo problemi di apprendimento; - la capacità di esporre, argomentare e analizzare le scelte effettuate. La prova orale ha una durata di circa 30 minuti e consiste in un colloquio su una discussione di un progetto sviluppato in python per un caso di studio concordato con il docente. La prova è finalizzata ad accertare: - la capacità di applicare i concetti studiati a problemi reali, utilizzando tecniche di programmazione; - la proprietà di linguaggio e la chiarezza espositiva; - le capacità di rielaborazione dei concetti acquisiti e di collegamento tra essi, nell'ambito delle tematiche del corso e/o di tematiche interdisciplinari ad esso correlate.

	CRITERI DI VALUTAZIONE: Ai fini della valutazione, verra' stimato il livello di raggiungimento dei risultati di apprendimento previsto per il corso, come di seguito elencati. Conoscenze e comprensione: Valutazione del livello di conoscenza e padronanza degli argomenti del corso; Capacita' di applicare le conoscenze: valutazione della capacita' di applicazione delle proprie conoscenze per l'analisi e la soluzione dei problemi proposti, del livello di autonomia, e dell'originalita' delle soluzioni proposte. Autonomia di giudizio: Valutazione delle capacita' logiche e analitiche per orientarsi e formulare giudizi, anche in presenza di informazioni parziali su problematiche/applicazioni riguardanti ambiti disciplinari o interdisciplinari ad essi correlati. Abilita' comunicative e capacita' di apprendimento: Valutazione delle capacita' espositive e di argomentazione, chiarezza e proprieta' di linguaggio. Valutazione delle capacita' di rielaborazione delle conoscenze acquisite e di collegamento multidisciplinare. ESITO DEL VOTO 30-30 e lode: Valutazione eccellente/ottimo. Ottima conoscenza degli argomenti, ottima capacita' analitica anche in nuovi contesti; ottima proprieta' di linguaggio e di apprendimento. 27-29: Valutazione molto buono. Buona padronanza degli argomenti, piena proprieta' di linguaggio, lo studente e' in grado di applicare le conoscenze per risolvere i problemi proposti. 24-26: Valutazione buono. Conoscenza di base dei principali argomenti, discreta proprieta' di linguaggio, con limitata capacita' di applicare autonomamente le conoscenze alla soluzione dei problemi proposti. 21-23: Soddisfacente. Parziale padronanza degli argomenti del corso, soddisfacente proprieta' linguaggio, scarsa capacita' di applicare autonomamente le conoscenze acquisite. 1 8-20: Sufficiente. Minima conoscenza degli argomenti del corso e del linguaggio tecnico, scarsissima o nulla capacita' di applicare autonomamente le conoscenze acquisite. Insufficiente: non possiede una conoscenza accettabile dei contenuti degli argomenti trattati nell'insegnamento.
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni frontali, esercitazioni guidate in aula, esercitazioni in Python.

MODULO MACHINE LEARNING

Prof.ssa ILENIA TINNIRELLO

TESTI CONSIGLIATI

- SergiosTheodoridis, Kostantinos Koutroumbas. Pattern Recognition, Academic Press. eBook ISBN: 9780080949123 Hardcover ISBN: 978159749272.
- Paul J. Deitel, Harvey M. Deitel. Introduzione a Python. Per l'informatica e la data science. Pearson 2021
- Dispense del docente.

TIPO DI ATTIVITA'	C
AMBITO	20929-Attività formative affini o integrative
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	48
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	27

OBIETTIVI FORMATIVI DEL MODULO

Il modulo presenta tecniche di apprendimento automatico basate sui dati per la classificazione basata su approcci statistici e su modelli con memoria.

Più in dettaglio, i principali obiettivi formativi del modulo sono:

- 1) Apprendere tecniche per la stima delle densità di probabilità o distribuzione campionarie delle 'features' usate per la classificazione;
- 2) Applicare tecniche di inferenza Bayesiana anche a processi con memoria;
- 3) Progettare sistemi capaci di apprendere per rinforzo (reinforcement-learning).

PROGRAMMA

ORE	Lezioni
4	Introduzione al corso. Richiami di probabilità e variabili aleatorie.
8	Classificazione: rappresentazione delle ipotesi, regioni di decisione, funzioni di costo. Classificatori Bayesiani. Distribuzioni gaussiane multivariabili. Tecniche di stima delle densità di probabilità delle features: stima parametrica e non parametrica. Stima a massima verosimiglianza.
8	Processi di Markov tempo discreti: matrici di transizioni, condizioni di equilibrio stazionario, distribuzioni di probabilità limite. Esempi di applicazione: l'algoritmo PageRank di Google. L'algoritmo di Viterbi per classificazione in processi con memoria. Processi di decisione di Markov e Reinforcement Learning.
ORE	Esercitazioni
7	Introduzione a Python e alla libreria scikit learn. Applicazioni di tutti i concetti del corso a problemi reali di classificazione ed esempi di implementazione.

**MODULO
DATA ANALYTICS AND STORAGE**

Prof.ssa LILIANA LO PRESTI

TESTI CONSIGLIATI

SergiosTheodoridis, Kostantinos Koutroumbas. Pattern Recognition, Academic Press. eBook ISBN: 9780080949123
Hardcover ISBN: 978159749272
Dispense fornite dal docente / Lecture notes.

TIPO DI ATTIVITA'	C
AMBITO	20929-Attività formative affini o integrative
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	48
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	27

OBIETTIVI FORMATIVI DEL MODULO

Il modulo presenta tecniche di apprendimento automatico basato sui dati. In particolare, i principali obiettivi formativi del modulo sono:

- 1) Apprendere tecniche di classificazione mediante classificatori lineari, non lineari e deep learning e semplici tecniche di clustering.
- 2) Saper implementare semplici tecniche di classificazione, clustering e analisi dei dati in Python.

PROGRAMMA

ORE	Lezioni
4	Classificatori lineari: il percettrone e le tecniche di addestramento per classi separabili. La tecnica del gradiente. Funzioni di costo. Addestramento per classi non separabili. Generalizzazioni ai classificatori multi-classe. Classificatori lineari con funzioni di decisione non a soglia: il regressore logistico. Regressori lineari e dilemma bias/varianza.
4	Classificatori non lineari: reti neurali multi-livello. Il problema dello XOR. L'algoritmo di Backpropagation. Generalizzazioni di classificatori non lineari: trasformazione dello spazio delle features in spazi a più dimensioni. Alberi di decisione. Combinazioni di classificatori non lineari.
6	Introduzione al Deep Learning: CNN, Autoencoder, LSTM, GAN, Graph Neural Networks.
4	Clustering: algoritmi K-Means e fuzzy K-Means. Scelta del numero di cluster.
ORE	Esercitazioni
3	Classificatori lineari e non lineari.
3	Deep learning.
3	Clustering.