

Insegnamento di: Metodi Numerici in Data Science					
Classe di laurea: L-35 (Matematica)		Corso di Laurea in: Matematica Triennale		Anno accademico: 2020/2021	
Denominazione inglese insegnamento: Numerical Methods in Data Science		Tipo di insegnamento: A scelta		Anno: 3	Semestre: 2
Tipo attività formativa: c- Attività didattiche affini o integrativi di quelli caratterizzanti	Ambito disciplinare:	Settore scientifico-disciplinare: MAT/08		CFU totali: 7 di cui CFU lezioni: 6.5 CFU ese/lab/tutor: 0.5	
Modalità di erogazione, ore di didattica assistita ed ore dedicate allo studio individuale ore di lezione: 52 ore di esercitazione/laboratorio/tutorato: 8 totale ore didattica assistita: 60 totale ore di studio individuale: 113					
Lingua di erogazione: Italiano	Obbligo di frequenza: no				
Docente: Nicoletta Del Buono	Tel: 0805442711 e-mail: nicoletta.delbuono@uniba.it	Ricevimento studenti: Dip. Matematica Piano II, stanza 24	Giorni e ore ricevimento: Lunedì 10-11. In altri giorni e orari solo previo appuntamento.		
Conoscenze preliminari: Le conoscenze che in genere vengono acquisite nella laurea della classe L-35 con riferimento particolare alle discipline di Calcolo Numerico e della Analisi Matematica classica in una e più variabili					
Obiettivi formativi: Acquisizione di alcune tecniche numeriche di base per il trattamento di problemi reali di analisi di dati					
Risultati di apprendimento previsti	Conoscenza e capacità di comprensione: Acquisizione di alcune tecniche matematiche per la risoluzione di problemi di classificazione e clustering. Capacità di utilizzare codici numerici che implementano le tecniche acquisite per l'analisi di dati reali.				
	Conoscenza e capacità di comprensione applicate: Le conoscenze teoriche e pratiche acquisite si utilizzano in vasta parte della matematica applicata e nella costruzione di modelli di apprendimento basato sui dati.				
	Autonomia di giudizio: Capacità di individuare le giuste tecniche numeriche per affrontare e risolvere numericamente problemi di apprendimento basato sui dati.				
	Abilità comunicative: Acquisizione del linguaggio e del formalismo matematico avanzato necessario per la consultazione e comprensione dei testi, l'esposizione delle conoscenze acquisite, la descrizione, l'analisi dei dati di tipo strutturato e la risoluzione dei problemi reali.				
	Capacità di apprendere: Acquisizione di un metodo di studio adeguato, supportato dalla consultazione dei testi e dall'utilizzo di software numerici che implementino le tecniche esposte durante il corso.				
Programma del corso <ul style="list-style-type: none"> - Decomposizione a valori singolari di una matrice di dati e Vector Space Model - Analisi di dati attraverso la fattorizzazione SVD della matrice di covarianza: SVD e Principal Component Analysis. - Clustering e principali algoritmi di clustering: k-means e k-medoids, algoritmi di clustering gerarchico - Fattorizzazioni di matrici (Nonnegative Matrix Factorization) e Clustering. - Algoritmi numerici di classificazione, il metodo di Nearest Neighbors, algoritmo di Learning Vector Quantization, le Support Vector Machine. - Analisi dei metodi HITS e PageRank per il ranking di reti complesse - Algoritmi numerici per problemi su reti e grafi: Algoritmo di Dijkstra (per il problema del cammino di costo minimo), Algoritmo di Ford e Fulkerson (per il problema del massimo flusso). 					

Metodi di insegnamento:

Lezioni in aula ed esercitazioni in laboratorio informatico

Supporti alla didattica:

Dispense disponibili alla pagina <https://sites.google.com/site/nicolettadelbuono/>

Controllo dell'apprendimento e modalità d'esame:

Prova Orale (discussione di un progetto concordato con i docenti)

Testi di riferimento principali:

- V. Comincioli, Metodi numerici e statistici per le scienze applicate, Milano, Ambrosiana, 1992.
- C. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2003.
- I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer, 2002
- A. Cichocki, R. Zdunek, A.H. Phan, S.I Amari, Nonnegative Matrix and Tensor Factorizations, Wiley, 2009
- A. N. Langville, C. D. Meyer: Google's PageRank and beyond. Princeton Univ. Press, 2006.
- T. Hastie, R. Tibshirani J. Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, 2009