

Principali informazioni sull'insegnamento	
Titolo insegnamento	Metodi Numerici per la Data Science
Corso di studio	Laurea Magistrale in Data Science
Crediti formativi	6
Denominazione inglese	Numerical Methods for Data Science
Obbligo di frequenza	No
Lingua di erogazione	Italiano

Docente responsabile	Nome Cognome	Indirizzo Mail
	Nicoletta Del Buono	nicoletta.delbuono@uniba.it

Dettaglio credi formativi	Ambito disciplinare	SSD	Crediti
	Matematico	MAT/08	6 (4 T1+2 T2)

Modalità di erogazione	
Periodo di erogazione	Primo Semestre
Anno di corso	Primo
Modalità di erogazione	Lezioni frontali (4 cfu tipo T1) Esercitazioni (2 cfu tipo T2)

Organizzazione della didattica	
Ore totali	62
Ore di corso	62
Ore di studio individuale	88

Calendario	
Inizio attività didattiche	1 marzo 2021
Fine attività didattiche	4 giugno 2021

Syllabus	
Prerequisiti	<ul style="list-style-type: none"> <li>Proprietà di base su insiemi e gruppi. Scalari, vettori matrici e operazioni su di essi. Sistemi coordinati. Trasformazioni lineari e loro proprietà. Spazi vettoriali, basi e spazi generati. Spazio Range e spazio nullo. Trasformazioni per similarità e autovalori-autovettori di matrici.</li> </ul>
Risultati di apprendimento previsti (declinare rispetto ai Descrittori di Dublino) (si raccomanda che siano coerenti con i risultati di apprendimento del CdS, compreso i risultati di apprendimento trasversali)	<ul style="list-style-type: none"> <li><i>Conoscenza e capacità di comprensione</i> <i>Acquisizione di alcune tecniche numeriche per la risoluzione di problematiche di data science. Capacità di realizzare codici numerici efficienti che implementano le tecniche acquisite.</i></li> <li><i>Conoscenza e capacità di comprensione applicate</i> <i>Le conoscenze teoriche e pratiche acquisite si utilizzano in</i></li> </ul>

	<p>vasta parte della matematica applicata e nella risoluzione di problematiche di data science.</p> <ul style="list-style-type: none"> <li>• <i>Autonomia di giudizio</i> Capacità di individuare le giuste tecniche numeriche per affrontare e risolvere numericamente problemi di data science.</li> <li>• <i>Abilità comunicative</i> 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 e la risoluzione di alcuni problemi in data science.</li> <li>• <i>Capacità di apprendere</i> Acquisizione di un metodo di studio adeguato, supportato dalla consultazione dei testi e dalla implementazione al calcolatore delle tecniche esposte durante il corso.</li> </ul>
<p>Contenuti di insegnamento</p>	<ul style="list-style-type: none"> <li>• Metodi numerici di algebra lineare nella Data Science</li> <li>• Metodi numerici per equazioni lineari e nonlineari</li> <li>• Metodi di low rank approximation per dati strutturati</li> <li>• Analisi delle componenti principali</li> <li>• Approssimazione numerica ai minimi quadrati</li> <li>• Metodi numerici per l'information retrieval e l'analisi di dati su grafi</li> <li>• Applicazioni in Data science: Fattorizzazione QR e Vector space model, SVD e Latent Semantic intexing, PCA e modello Eigenface, NMF per l'analisi di microarray e il topic modelling</li> </ul>

<p><b>Programma</b></p>	<p>Algebra Lineare per la Data Science: Norme vettoriali e matriciali, sfera unitaria. Spazi vettoriali e spazi di matrici. Spazi vettoriali generati. Basi di spazi vettoriali e di matrici. Fattorizzazioni di matrici. Fattorizzazione QR. Metodi per il calcolo di autovalori e autovettori di una matrice di dati. Teorema di Eckart/Young/Mirsky, Decomposizione a Valori Singolari (SVD). Algoritmo per il calcolo della SVD.</p> <p>Metodi Numerici per equazioni lineari e non lineari: Sistemi di equazioni lineari <math>n \times m</math>. Metodo di trasformazione in forma di Echelon per righe. Metodo di sostituzioni in avanti e indietro. Fattorizzazioni. Cenni sui metodi iterativi. Sistemi di equazioni non lineari: esempi per funzioni unidimensionali. Metodo di Newton: caso unidimensionale. Metodo di Newton: caso multidimensionale. Cenni di implementazione del metodo di Newton e criteri di arresto.</p>
-------------------------	---

	<p>Analisi delle componenti principali (PCA): Matrice di covarianza e di correlazione dei dati. Diagonalizzazione di una matrice simmetrica. Autovalori di una forma quadratica definita positiva. Componenti principali di vettori di variabili aleatorie. Combinazioni lineari di massima varianza. Significato statistico e probabilistico delle componenti principali. Analisi delle componenti principali per insiemi di dati multivariati. Euristiche per il calcolo del numero ottimale delle componenti principali. PCA come metodo di feature selection. Matrici di correlazione e di covarianza e PCA. Interpretazione delle componenti principali: il cerchio delle correlazioni, gli z-scores. Applicazioni reali della PCA: il modello Eigenface.</p> <p>Metodi di approssimazione low rank: Approssimazione low rank di matrici di dati. SVD e sue applicazioni in data science, Fattorizzazioni nonnegative di matrici (NMF). Sparse Nonnegative Matrix Factorization per il Clustering.</p> <p>Metodi di approssimazione ai minimi quadrati (ordinary Least squares): Minimi quadrati lineari e polinomiali. Sistema delle equazioni normali. La matrice pseudo-inversa di Penrose. Metodi numerici per OLS: fattorizzazione QR e SVD. Retta di regressione lineare semplice. Retta di regressione multivariata.</p> <p>Applicazioni in Data science: metodi matematici per l'information retrieval e l'analisi di dati su grafi. Fattorizzazioni matriciali per il vector space model. SVD e Latent semantic indexing. Il teorema di Perron-Frobenius e la network centrality. Modelli basati sugli autovettori (HITS e Pagerank)</p>
Testi di riferimento	<ol style="list-style-type: none"> <li>1. V. Comincioli, Metodi numerici e statistici per le scienze applicate, Milano, Ambrosiana, 1992.</li> <li>2. C. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2003.</li> <li>3. I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer, 2002</li> <li>4. A. Cichocki, R. Zdunek, A.H. Phan, S.I Amari, Nonnegative Matrix and Tensor Factorizations, Wiley, 2009</li> </ol>
Note ai testi di riferimento	I testi di riferimento saranno supportati da articoli scientifici e dispense forniti dal docente durante lo svolgimento del corso.
Metodi didattici	Lezioni frontali con slide ed esercitazioni di laboratorio con software open source (Python)
Metodi di valutazione (indicare almeno la tipologia scritto, orale, altro)	Progetto scritto ed esame orale

<p>Criteri di valutazione (per ogni risultato di apprendimento atteso su indicato, descrivere cosa ci si aspetta lo studente conosca o sia in grado di fare e a quale livello al fine di dimostrare che un risultato di apprendimento è stato raggiunto e a quale livello)</p>	<ul style="list-style-type: none"> <li>• Lo studente comprende quali siano i concetti matematici più adatti per trattare problematiche di data science</li> <li>• Lo studente sarà in grado di applicare le conoscenze apprese per proporre soluzioni ad alcuni problemi semplici in data science</li> <li>• Lo studente sarà in grado di organizzare in modo autonomo l'esposizione di una problematica applicativa di data science e di proporre una metodologia numerica di soluzione</li> </ul>
<p>Altro</p>	