



Principali informazioni sull'insegnamento

Denominazione dell'insegnamento	<i>Metodi Numerici per la Data Science</i>	
Corso di studio	<i>Laurea Magistrale in Data Science</i>	
Anno Accademico	2024/25	
Crediti formativi universitari (CFU) / European Credit Transfer and Accumulation System (ECTS)	6 CFU	
Settore Scientifico Disciplinare	MATH-05/A	
Lingua di erogazione	Italiano	
Anno di corso	Primo	
Periodo di erogazione	2° semestre	
Obbligo di frequenza	La frequenza è fortemente raccomandata	
Sito web del corso di studio	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea	

Docenti	
Nome e cognome	Francesca Mazzia
Indirizzo mail	Francesca.mazzia@uniba.it
Telefono	0805443291
Sede	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Stanza n.565, 5 ^a piano
Sede virtuale	Piattaforma e-learning UNIBA - https://elearning.uniba.it/
Sito web del docente	https://archimede.uniba.it/~mazzia/
Ricevimento (giorni, orari e modalità, es. su appuntamento)	Venerdì dalle 11:30 alle 13:30, Previo appuntamento per email, per concordare un giorno e/o un orario diverso

Syllabus



Obiettivi formativi	Acquisizione degli strumenti di base della algebra lineare numerica per la modellazione di problemi di apprendimento dei dati e principali tecniche numeriche di fattorizzazione e approssimazione di matrici.
Prerequisiti	Proprietà di base di vettori e matrici. calcolo matriciale, autovalori e autovettori, metodo di eliminazione di Gauss per la risoluzione di sistemi di equazioni lineari, spazi vettoriali e loro proprietà, funzioni lineari e loro proprietà, rappresentazione geometrica di spazi R^2 e R^3 e riferimenti ortogonali.
Contenuti di insegnamento (Programma)	<ul style="list-style-type: none">• Algebra Lineare per la Data Science: Richiami su: numeri di macchina, condizionamento, norme vettoriali e matriciali, sfera unitaria; spazi vettoriali e spazi di matrici; basi di spazi vettoriali e di matrici. Metodi per il calcolo di autovalori e autovettori di una matrice di dati. Il metodo delle potenze. Fattorizzazioni di matrici: Fattorizzazione QR e SVD, Teorema di Eckart/Young/Mirsky, SVD troncata e sue applicazioni in data science mediante lo studio dell'analisi delle componenti principali o PCA (Testi di riferimento C. Meyer, Capitoli 1-3,5,7, Lars Eldèn Capitoli 6,7)• Metodi Numerici per equazioni lineari e non-lineari: Sistemi di equazioni lineari rettangolari. Richiami: metodo di trasformazione in forma di Echelon per righe e forma di Echelon ridotta. Metodo di sostituzioni in avanti e indietro e fattorizzazione LU. Cenni sui metodi iterativi. Sistemi di equazioni non lineari. Metodo di Newton: caso unidimensionale. Metodo di Newton: caso multidimensionale. Cenni di implementazione del metodo di Newton e criteri di arresto.• 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. (Testo di riferimento C. Meyer, Capitolo 4)• Applicazioni in Data science: Applicazioni reali della PCA: il modello Eigenface 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: il modello Pagerank. (Articoli scientifici forniti dai docenti)
Testi di riferimento	<p>Il corso non adotta un testo di riferimento unico; tuttavia, i contenuti didattici possono essere integrati da alcune parti dei seguenti testi:</p> <ul style="list-style-type: none">• C. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2003.• Lars Eldèn, Matrix Methods in Data Mining and Pattern Recognition, SIAM 2007• G. Strang, Linear Algebra and its applications, 4th Edition <p>Gli studenti che lo desiderano possono ottenere i testi in prestito dalla Biblioteca. Può convenire verificarne la disponibilità mediante il Sistema Bibliotecario di Ateneo https://opac.uniba.it/easyweb/w8018/index.php? e contattare la biblioteca per concordare il prestito.</p>
Note ai testi di riferimento	I libri di testo sono integrati con le slide fornite dal docente ed eventuale materiale di approfondimento, resi disponibili sulla piattaforma ADA (vedi sopra "sede virtuale"). Sulla medesima piattaforma sono anche disponibili i codici degli esempi numerici sviluppati a lezione.



Organizzazione della didattica			
Ore			
Totali	Didattica frontale	Esercitazione	Studio individuale
62 ore	32 ore	30 ore	88 ore
CFU/ETCS			
6 CFU	4 CFU	2 CFU	

Metodi didattici	
	Lezioni frontali ed esercitazioni in aula

Risultati di apprendimento previsti	
Conoscenza e capacità di comprensione	<ul style="list-style-type: none">○ Apprendimento di tecniche numeriche per la risoluzione di problematiche di data science○ Conoscenza di alcune librerie e software disponibili
Conoscenza e capacità di comprensione applicate	<ul style="list-style-type: none">○ Messa in pratica delle conoscenze teoriche○ Implementazione di codici numerici pertinenti ed efficienti
Competenze trasversali	Autonomia di giudizio <ul style="list-style-type: none">○ prendere decisioni relative alla scelta della strategia numerica opportuna per risolvere problemi di data science.



	<p>Abilità comunicative</p> <ul style="list-style-type: none">○ Acquisizione del linguaggio e del formalismo matematico necessario alla descrizione appropriata del problema considerato○ Esposizione delle conoscenze acquisite mediante rigore formale e tecnico. <p>Capacità di apprendere in modo autonomo</p> <ul style="list-style-type: none">○ Comprendere e utilizzare gli strumenti dell'algebra lineare numerica e per modellare e risolvere problemi di data science anche per costruire e risolvere in maniera autonoma nuove soluzioni.
--	--

Valutazione	
Modalità di verifica dell'apprendimento	<p>Durante le lezioni verranno discussi, in modo partecipato, diversi quesiti ed esercizi simili per tipologia a quelli comunemente somministrati durante gli esami e gli esoneri. La finalità è duplice: monitorare in tempo reale lo stato di preparazione degli studenti frequentanti, perfezionandone la preparazione in vista dell'esame o degli esoneri; agevolare lo studio in itinere degli aspetti pratici della disciplina, motivando concretamente i corsisti a sostenere l'esame in tempi brevi, sfruttando possibilmente la modalità degli esoneri. Sono previste due prove parziali scritte per facilitare la preparazione all'esame: la prima durante l'interruzione delle lezioni a metà corso, la seconda a fine corso. Entrambe le date sono concordate, nei limiti consentiti, con gli studenti frequentanti. Di seguito sono elencati ulteriori vantaggi riservati ai frequentanti:</p> <ul style="list-style-type: none">- La partecipazione alle prove parziali consente di avere informazioni sulle parti di programma che vanno approfondite e quindi migliorare la preparazione all'esame finale.- La prova parziale viene valutata complessivamente durante l'esame orale al momento del recupero degli argomenti considerati insufficienti, in modo da migliorare la valutazione complessiva.- Le prove parziali sono prese in considerazione per i tre appelli che seguono la fine delle attività didattiche. <p>Lo studente che non ha sostenuto le prove parziali sosterrà l'esame orale completo in cui dimostrerà la conoscenza degli argomenti teorici e la capacità di risolvere esercizi. Tale conoscenza può essere dimostrata anche risolvendo esercizi per iscritto.</p> <p>Tutti gli studenti all'orale presenteranno e discuteranno in sede di esame la soluzione degli esercizi da svolgere in Python che gli saranno stati assegnati la settimana precedente l'esame orale. Gli esercizi proposti sono simili a quelli svolti a lezione e disponibili sulle piattaforme Teams e/o ADA.</p>
Criteria di valutazione	<ul style="list-style-type: none">● Conoscenza e capacità di comprensione:<ul style="list-style-type: none">○ Dimostrare un'adeguata comprensione e competenza sui contenuti dell'insegnamento.● Conoscenza e capacità di comprensione applicate:



	<ul style="list-style-type: none">○ Conoscenza delle possibili applicazioni dei concetti teorici e adeguata capacità di implementare gli algoritmi proposti.● Autonomia di giudizio:<ul style="list-style-type: none">○ Autonomia nella selezione dei concetti teorici più idonei alla risoluzione di problemi pratici● Abilità comunicative:<ul style="list-style-type: none">○ adeguata capacità espositiva dei contenuti studiati,○ adeguata capacità di analisi e sintesi● Capacità di apprendere:<ul style="list-style-type: none">○ Buona capacità di effettuare collegamenti interdisciplinari
Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	<p>Il voto finale è attribuito in trentesimi. L'esame si intende superato quando il voto finale è maggiore o uguale a 18. Nella valutazione della prova orale e nell'attribuzione del voto finale varrà la seguente scala di valutazione dell'apprendimento:</p> <ul style="list-style-type: none">● Voto insufficiente (<18): Conoscenze frammentarie e superficiali dei contenuti, errori nell'applicare i concetti, esposizione carente● Voto 18-20: Conoscenze dei contenuti sufficienti ma generali, esposizione semplice, incertezze nell'applicazione di concetti teorici● Voto 21-23: Conoscenze dei contenuti appropriate ma non approfondite, capacità di applicare i concetti teorici, capacità di presentare i contenuti in modo semplice.● Voto 24-25: Conoscenze dei contenuti appropriate e ampie, discreta capacità di applicazione delle conoscenze, capacità di presentare i contenuti in modo articolato.● Voto 26-27: Conoscenze dei contenuti precise e complete, buona capacità di applicare le conoscenze, capacità di analisi, esposizione chiara e corretta● Voto 28-29: Conoscenze dei contenuti ampie, complete ed approfondite, buona applicazione dei contenuti, buona capacità di analisi e di sintesi, esposizione sicura e corretto● Voto 30 e 30 e lode: Conoscenze dei contenuti molto ampie, complete ed approfondite, capacità ben consolidata di applicare i contenuti, ottima capacità di analisi, di sintesi e di collegamenti interdisciplinari, padronanza di esposizione
Altro	<p>Si suggerisce agli studenti di affidarsi esclusivamente alle informazioni/comunicazioni fornite sui siti ufficiali del Dipartimento di Informatica, ovvero sui gruppi social solo se costituiti e amministrati esclusivamente dai docenti dei relativi insegnamenti:</p> <ul style="list-style-type: none">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea● https://www.uniba.it/it/ricerca/dipartimenti/informatica● https://elearning.uniba.it/ <p>I programmi degli insegnamenti sono disponibili qui:</p> <ul style="list-style-type: none">● https://elearning.uniba.it/ <p>Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei Regolamenti didattici e manifesti degli studi disponibili nel sito:</p>



- <https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea>

Si suggerisce agli studenti di diffidare delle informazioni e dei materiali circolanti su siti o gruppi social non ufficiali, poiché spesso sono risultati non affidabili, non corretti o incompleti. Per ogni dubbio, chiedere un incontro al docente secondo le modalità previste per il ricevimento.

- *link al corso sulla piattaforma e-learning UNIBA*

<https://elearning.uniba.it/course/view.php?id=2079>

- *link alla pagina Teams del corso:*

<https://teams.microsoft.com/l/team/19%3aS87zz7nwS0rcoaLdA5aa5CCTF1OYk1EQ2Lt-pDHkevg1%40thread.tacv2/conversations?groupId=b0ce0b6d-6b49-4c62-9e75-a70686c4da0e&tenantId=c6328dc3-afdf-40ce-846d-326ead86d49>

codice di accesso:

il3s29c



Main information on the course

Course name	Numerical Methods for Data Science	
Degree	<i>Laurea Magistrale in Data Science</i>	
Academic year	2024/25	
European Credit Transfer and Accumulation System (ECTS), in Italian Crediti Formativi Universitari (CFU)	6 CFU (each CFU corresponds to 25 hours (h) of student's time); CFU are of type T1, T2 or T3 T1 = 8 h lecture + 17 h individual study T2 = 15 h practice + 10 h individual study T3 = 25 h individual study	
Settore Scientifico Disciplinare	MATH-05/A	
Course language	Italian	
Anno di corso	2024/2025	
Periodo di erogazione	2 nd semester	
Obbligo di frequenza	It is highly recommended to attend classes	
Sito web del corso di studio	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea	

Teachers

Name and Surname	Francesca Mazzia
email	francesca.mazzia@uniba.it
phone	0805443291
office	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Room n.565, 5 th floor
e-learning platform	Piattaforma e-learning UNIBA - https://elearning.uniba.it/
Teacher's homepage	https://archimede.uniba.it/~mazzia/
Office hours	Fridays 11:30 am-13:30 pm. The students that require a different time can schedule a meeting via email.

Syllabus

Course goals	Acquisition of the basis of numerical linear algebra and of the main techniques to model data-science problems, referencing in particular to matrix approximation and factorizations.
Prerequisites/requirements	Fundamentals of vectors and matrices. Operations with matrices, eigenvalues and eigenvectors computation. Gauss elimination algorithm, vector spaces and their properties; linear functions and their properties; geometric representation of R^2 and R^3 , orthogonal basis.
Course program	<ul style="list-style-type: none">Linear Algebra for Data Science: References: vector and matrix norms; vector spaces and matrix spaces. Eigenvalues and eigenvectors computation. The power method. Matrix Factorizations: QR and SVD, Eckart-Young-Mirsky theorem, Truncated SVD and applications to datascience by the use of principal component analysis (PCA), books: C. Meyer, Chapters 1-3,5,7, Lars Eldèn Chapters 6,7.Numerical methods for linear and non-linear equations:



	<p>Rectangular linear systems. Echelon form. Forward and Backward substitution, Factorizations. Basic concepts on iterative methods. Non linear system: monodimensional examples. Newton method: modimensional and multidimensional cases. Implementation and stopping criteria.</p> <ul style="list-style-type: none"> • Ordinary Least Squares Methods: Linear and polynomial least-squares. Normal equations. Penrose pseudo-inverse matrix. Use of the QR and SVD factorizations Linear regression in the monodimensional case and in the multidimensional case. (Book, C. Meyer, Chapter 4). 		
Books of reference	<p>The topics of the course can be studied in more details by consulting the following:</p> <ul style="list-style-type: none"> • C. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2003. • Lars Eldèn, Matrix Methods in Data Mining and Pattern Recognition, SIAM 2007 • G. Strang, Linear Algebra and its applications, 4th Edition 		
Notes to the books	<p>Textbooks are integrated with the slides provided by the instructor and any supplementary materials, made available on the ADA platform (see above "virtual classroom") or on Teams paltform. The platforms also provide access to the codes of the numerical examples developed during the lessons.</p>		
Organization of the didactic activities			
Hours			
Total	Lectures	Practice sessions	Individual study
62 hours	32 hours	30 hours	88 hours
CFU/ETCS			
6 CFU	3 CFU	3 CFU	

Teaching methods	
	Standard lectures and exercises/lab sessions

Expected learning outcomes	Learning numerical techniques for solving data science problems. Knowledge of some available libraries and software.
Knowledge and understanding	<ul style="list-style-type: none"> ○ Learning numerical techniques for solving data science problems. Knowledge of some available libraries and software.
Applying knowledge and understanding	



	<ul style="list-style-type: none"> ○ Practical application of theoretical knowledge ○ Implementation of relevant and efficient numerical codes
Other skills	<p><i>Making judgements</i></p> <ul style="list-style-type: none"> ○ Making decisions regarding the choice of appropriate numerical strategy to solve data science problems. <p><i>Communication</i></p> <ul style="list-style-type: none"> ○ Acquisition of the language and mathematical formalism necessary for the appropriate description of the considered problem. ○ Exposition of acquired knowledge through formal and technical rigor. <p><i>Learning skills</i></p> <ul style="list-style-type: none"> ○ Understanding and utilizing the tools of numerical linear algebra to model and solve data science problems, including constructing and autonomously solving new solutions.

Assessment	
Assessment methods	<p>During the lessons, various questions and exercises similar in type to those commonly administered during exams and exemptions will be discussed in a participatory manner. The aim is twofold: to monitor the real-time preparation status of attending students, enhancing their preparation for the exam or exemptions; to facilitate ongoing study of the practical aspects of the discipline, actively motivating participants to take the exam promptly, possibly utilizing the exemption mode. Two written partial tests are scheduled to facilitate exam preparation: the first during the mid-course break, and the second at the end of the course. Both dates are agreed upon, within permissible limits, with attending students. Below are further advantages reserved for attendees:</p> <ul style="list-style-type: none"> ● Participation in partial tests provides information on the parts of the program that need further study, thus improving preparation for the final exam. ● The partial test is assessed overall during the oral exam when recovering insufficient topics, aiming to enhance the overall evaluation. ● Partial tests are considered for the three exam sessions following the end of didactic activities. <p>Students who have not taken the partial tests will undergo a complete oral exam, demonstrating their knowledge of theoretical topics and their ability to solve exercises. This knowledge can also be demonstrated by solving written exercises. All students during the oral exam will present and discuss the solution of Python exercises assigned to them the week before the oral exam. The proposed exercises</p>



	<p>are similar to those solved during lessons and available on the Teams and/or ADA platforms.</p>
Evaluation criteria	<ul style="list-style-type: none">• Knowledge and understanding:<ul style="list-style-type: none">○ Demonstrate adequate comprehension and competency in the course content.• Applied knowledge and understanding:<ul style="list-style-type: none">○ Understanding possible applications of theoretical concepts and adequate ability to implement proposed algorithms.• Judgement autonomy:<ul style="list-style-type: none">○ Independence in selecting the most suitable theoretical concepts for solving practical problems.• Communicative skills:<ul style="list-style-type: none">○ Adequate presentation skills of the studied contents,○ Adequate ability of analysis and synthesis.• Learning ability:<ul style="list-style-type: none">○ Good ability to make interdisciplinary connections
Measurements and final grade	<p>The final grade is assigned on a scale of thirty. The exam is considered passed when the final grade is greater than or equal to 18. In the evaluation of the oral exam and in the assignment of the final grade, the following learning assessment scale will be used:</p> <ul style="list-style-type: none">• Insufficient (<18): Fragmented and superficial knowledge of the contents, errors in applying concepts, deficient presentation.• 18-20: Sufficient but general knowledge of the contents, simple presentation, uncertainties in applying theoretical concepts.• 21-23: Appropriate but not in-depth knowledge of the contents, ability to apply theoretical concepts, ability to present the contents in a simple manner.• 24-25: Appropriate and extensive knowledge of the contents, fair ability to apply knowledge, ability to present the contents coherently.• 26-27: Precise and comprehensive knowledge of the contents, good application of knowledge, analytical ability, clear and correct presentation.• 28-29: Broad, comprehensive, and in-depth knowledge of the contents, good application of contents, good analytical and synthesis skills, confident and correct presentation.• 30 and "lode" (with praise): Very broad, comprehensive, and in-depth knowledge of the contents, well-established ability to apply the contents, excellent analytical, synthesis, and interdisciplinary connection skills, mastery of presentation.



Further information	<p>It is recommended that students rely exclusively on information/communications provided on the official websites of the Department of Computer Science, or on social groups only if they are established and managed exclusively by the instructors of the respective courses:</p> <ul style="list-style-type: none">• https://www.uniba.it/en/research/departments/computer-science/teaching/degree-programs• https://www.uniba.it/en/research/departments/computer-science• https://elearning.uniba.it/ <p>The course programs are available here:</p> <ul style="list-style-type: none">• https://elearning.uniba.it/ <p>The information that all students should know is written in the Didactic Regulations and study manifests available on the website:</p> <ul style="list-style-type: none">• https://www.uniba.it/en/research/departments/computer-science/teaching/degree-programs/degree-courses <p>Students are advised to be wary of information and materials circulating on unofficial websites or social groups, as they often prove to be unreliable, incorrect, or incomplete. For any doubts, students should request a meeting with the instructor according to the procedures provided for office hours.</p>
	<ul style="list-style-type: none">• Link to the course on the UNIBA e-learning platform: https://elearning.uniba.it/course/view.php?id=2079• Link to the course Teams page: https://teams.microsoft.com/l/team/19%3aS87zz7nwS0rcoaLdA5aa5CCTF1OYk1EQ2Lt-pDHkevg1%40thread.tacv2/conversations?groupId=b0ce0b6d-6b49-4c62-9e75-a70686c4da0e&tenantId=c6328dc3-afdf-40ce-846d-326ead86d49 <p>Access code: il3s29c</p>