

Denominazione	DATA SCIENCE
Moduli componenti	-
Settore scientifico disciplinare	IINF-05/A (ex ING-INF/05)
Anno di corso e semestre di erogazione	3° anno, 2° semestre
Lingua di insegnamento	Italiano
Carico didattico in crediti formativi universitari	6 CFU
Numero di ore di attività didattica frontale	48 (pari a 6 CFU di didattica erogativa)
Docente	Responsabile dell'insegnamento: Prof. Fabio Difonzo (3 CFU) Prof.ssa Felisia Angela Chiarello (3 CFU)
Risultati di apprendimento specifici	Il corso ha la finalità di sviluppare competenze nell'analisi dei dati attraverso lo studio dei metodi di ottimizzazione alla base della Data Science, con particolare riferimento agli algoritmi di discesa del gradiente e alle loro varianti stocastiche, utilizzati per l'estrazione di informazioni da grandi insiemi di dati. Conoscenze e comprensione: Al termine del percorso di studio dell'insegnamento lo studente avrà acquisito conoscenze relative ai fondamenti teorici dell'ottimizzazione, inclusi i concetti di convexità, smoothness, proprietà di convergenza, nonché gli algoritmi di gradient descent, stochastic gradient descent e loro estensioni, anche per funzioni non lisce e problemi compositi. Capacità di applicare conoscenze e comprensione: Lo studente sarà in grado di implementare e applicare algoritmi di ottimizzazione per l'analisi dei dati, includendo metodi deterministici e stocastici, tecniche di subgradiente e approcci prossimali, comprendendone il comportamento in termini di convergenza e complessità computazionale. Autonomia di giudizio e pensiero critico: Al termine delle lezioni lo studente sarà in grado di analizzare un problema in termini di ottimizzazione, valutare le ipotesi strutturali (come convexità e regolarità) e scegliere il metodo algoritmico più appropriato in funzione delle proprietà del problema. Abilità comunicative: Al termine del percorso di studio dell'insegnamento lo studente saprà descrivere in modo chiaro e rigoroso algoritmi e risultati, illustrando le proprietà teoriche e i risultati di convergenza con un linguaggio appropriato. Capacità di apprendimento: Al termine del percorso di studio dell'insegnamento lo studente avrà acquisito strumenti metodologici per approfondire autonomamente metodi di ottimizzazione e analisi dei dati, seguendo l'evoluzione degli algoritmi e delle tecniche nell'ambito dell'ottimizzazione numerica.

Programma	Il programma del corso è composto dai seguenti contenuti didattici: Fondamenti teorici dell'ottimizzazione per la Data Science (2CFU/16h) • Introduzione ai metodi di ottimizzazione e problemi di minimizzazione • Proprietà delle funzioni: convessità, forte convessità, smoothness e proprietà Polyak–Łojasiewicz • Discesa del gradiente (Gradient Descent): definizione, analisi della convergenza e complessità • Struttura a somma di funzioni e problemi su dati di grandi dimensioni Metodi di ottimizzazione stocastica (1CFU/8h) • Stochastic Gradient Descent (SGD): algoritmo, scelte dei passi e proprietà di convergenza • Analisi del ruolo della varianza e delle condizioni di interpolazione • Estensioni e varianti dello SGD per problemi convessi e fortemente convessi Metodi per funzioni non lisce (1CFU/8h) • Funzioni non differenziabili e rappresentazione tramite funzioni a valori estesi • Subdifferenziale e subgradiente per funzioni convesse • Algoritmi di subgradient descent e loro proprietà di convergenza Metodi prossimali e problemi compositi (1CFU/8h) • Operatore prossimale e sue proprietà • Proximal Gradient Descent per problemi della forma somma di funzioni • Proximal Stochastic Gradient Descent e analisi della complessità • Applicazioni a modelli con vincoli o termini di regolarizzazione Sviluppo e analisi di casi applicativi (1CFU/8h) • Formalizzazione di problemi di analisi dati come problemi di ottimizzazione • Implementazione e analisi del comportamento degli algoritmi studiati • Valutazione delle prestazioni in termini di convergenza e complessità computazionale
Tipologie di attività didattiche previste e relative modalità di svolgimento	L'insegnamento è strutturato in lezioni di didattica frontale e laboratoriali, incoraggiando l'interazione e la partecipazione attiva degli studenti. È previsto l'utilizzo di tecnologie digitali per l'erogazione delle lezioni e delle esercitazioni.
Metodi e criteri di valutazione dell'apprendimento	La valutazione dell'apprendimento (sia per studenti frequentanti che non frequentanti) consiste nello svolgimento di una prova scritta contenente quesiti teorici e pratici relativi ai contenuti del corso.

Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	La valutazione dell'apprendimento prevede l'attribuzione di un voto finale espresso in trentesimi. Il voto finale sarà determinato attraverso l'esame scritto sopra dettagliato. La lode sarà valutata per i soli studenti che abbiano raggiunto la valutazione complessiva di 30/30 analizzando la capacità di proporre soluzioni corrette ed efficienti nella risoluzione di problemi.
Propedeuticità	È propedeutica la frequenza dei seguenti insegnamenti: • Analisi Matematica (96 ore – 12 CFU – I anno – Attività caratterizzante) • Fisica (96 ore – 12 CFU – I anno – Attività caratterizzante) • Fondamenti di Informatica (48 ore – 6 CFU – I anno – Attività caratterizzante) • Fondamenti di Algebra Lineare e Geometria (96 ore – 12 CFU – I anno – Attività caratterizzante) • Machine learning e AI (48 ore – 6 CFU – III anno – Attività caratterizzante)
Materiale didattico utilizzato e materiale didattico consigliato	Testi di riferimento • G. Garrigos, R. M. Gower, <i>Handbook of Convergence Theorems for (Stochastic) Gradient Methods</i>, 2024 • D. P. Bertsekas, <i>Convex Optimization Theory</i>, Athena Scientific