UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Bioinformatica

2526-2-F0901D043-F0901D087M

Obiettivi

Lo studente deve conoscere le basi della bioinformatica necessarie allo svolgimento della sua professione.

Contenuti sintetici

Il corso introduce la bioinformatica applicata alle biotecnologie mediche. Si approfondiscono concetti relativi all'allineamento di sequenze, l'analisi filogenetica, l'analisi dei dati di espressione genica e l'integrazione omica. Si introduce il linguaggio R per l'analisi dei dati, applicandolo a casi studio su espressione genica e bioinformatica del cancro. Infine, si esplorano il machine learning e i modelli evolutivi del cancro, inclusa l'analisi delle firme mutazionali. Ogni blocco include esercitazioni pratiche per consolidare le competenze.

Programma esteso

Il corso di Bioinformatica ha l'obiettivo di fornire una solida panoramica delle principali tecniche e degli strumenti utilizzati in bioinformatica, con particolare attenzione alle applicazioni in ambito medico e biotecnologico. Gli studenti acquisiranno competenze pratiche nell'analisi dei dati biologici, come sequenze genetiche, espressione genica, varianti genomiche, e apprenderanno l'uso di tecniche di machine learning per applicazioni in oncologia e genomica. Il corso sarà strutturato in tre parti da quattro lezioni ciascuna. Si descrivono ora nel dettaglio le tematiche trattate.

Parte 1: Introduzione alla Bioinformatica e Analisi di Sequenze
Nella prima parte, gli studenti saranno introdotti ai fondamenti della bioinformatica, esplorando la sua storia, le sue applicazioni e le risorse principali utilizzate nel campo. La bioinformatica è una disciplina fondamentale per l'analisi dei dati biologici, che include la gestione e l'interpretazione di enormi quantità di dati provenienti da esperimenti

scientifici. Si discuterà di come le basi di dati biologiche come NCBI, Ensembl, e UCSC Genome Browser siano utilizzate per raccogliere e organizzare i dati, e gli studenti apprenderanno come utilizzare software e strumenti di analisi per affrontare problemi bioinformatici.
La seconda lezione si concentrerà sull'allineamento delle sequenze, una tecnica centrale per il confronto di sequenze genetiche, utilizzata per identificare somiglianze tra geni di specie diverse. Gli studenti esploreranno vari algoritmi di allineamento, come quelli utilizzati in BLAST e FASTA, per confrontare sequenze e analizzare le loro relazioni evolutive. Si approfondirà anche l'analisi filogenetica, che permette di ricostruire la storia evolutiva delle specie attraverso l'analisi di sequenze omologhe. In questa lezione, gli studenti impareranno anche a costruire alberi filogenetici applicando i metodi di massima parsimonia e massima verosimiglianza.
La terza lezione del modulo tratterà l'analisi dei dati di espressione genica, una componente cruciale per comprendere il funzionamento dei geni in diverse condizioni biologiche. Si esploreranno tecniche di analisi come il microarray e l'RNA-Seq, mettendo in evidenza come i dati di espressione possano essere utilizzati per capire la regolazione genica e identificare i geni coinvolti in patologie. Inoltre, si discuteranno le varianti genomiche, come SNP, indels e CNV, e il loro impatto sulle funzioni biologiche. Il modulo si concluderà con un'introduzione all'analisi omica integrativa, che combina informazioni provenienti da diverse tecniche omiche, come genomica, trascrittomica e proteomica, per ottenere una visione più completa dei sistemi biologici.
L'ultima lezione di questo modulo sarà dedicata a un'esercitazione pratica in cui gli studenti avranno l'opportunità di valutare le tecniche di allineamento delle sequenze, analisi filogenetica e analisi dei dati di espressione su dataset reali, consolidando le conoscenze teoriche acquisite nelle lezioni precedenti.

Modulo 2: Fondamenti di Programmazione e Analisi dei Dati di Espressione
Nella seconda parte, gli studenti inizieranno a familiarizzare con il linguaggio di programmazione R, uno degli strumenti più potenti per l'analisi statistica dei dati biologici. Le prime due lezioni saranno dedicate all'insegnamento dei fondamenti di R, partendo dalle basi della sintassi e della gestione dei dati fino ad arrivare alla manipolazione avanzata degli stessi. Impareranno a scrivere e eseguire script in R, a gestire variabili, a utilizzare funzioni e librerie, e a visualizzare i dati mediante grafici e tabelle. Inoltre, saranno introdotti all'analisi statistica dei dati, con particolare attenzione alle tecniche più comuni, come le analisi di regressione e i test statistici.
La terza lezione del modulo sarà focalizzata su un caso di studio pratico riguardante l'analisi dei dati di espressione genica. Gli studenti lavoreranno su dati reali provenienti da esperimenti di RNA-Seq o microarray e apprenderanno come eseguire il preprocessing dei dati, compresi i passaggi di normalizzazione e filtraggio, per preparare i dati per l'analisi. In seguito, si concentreranno sull'identificazione dei geni differenziali e sull'analisi multivariata dei dati, utilizzando tecniche come la PCA (analisi delle componenti principali) e il clustering per scoprire pattern nascosti nei dati.
Il modulo si concluderà con un'esercitazione pratica in cui saranno discusse le tecniche apprese durante le lezioni.

Modulo 3: Machine Learning, Cancro e Mutazioni
Nella terza ed ultima parte, gli studenti verranno introdotti al machine learning, una disciplina che sta assumendo un ruolo centrale nella bioinformatica moderna. Impareranno le basi del machine learning, distinguendo tra tecniche supervisionate e non supervisionate, e come applicare algoritmi come alberi decisionali, support vector machines (SVM) e k-NN per analizzare e fare previsioni sui dati biologici. La lezione si concentrerà anche sull'uso di machine learning per affrontare problemi complessi in bioinformatica, come la classificazione dei dati genomici e l'identificazione di pattern nei dati omici.
La lezione successiva esplorerà i modelli di evoluzione del cancro, analizzando come le cellule tumorali accumulano mutazioni nel tempo e come queste mutazioni influenzano il comportamento e la progressione del cancro. Gli studenti apprenderanno i modelli evolutivi del cancro e come le analisi genomiche possono essere utilizzate per ricostruire le traiettorie evolutive delle cellule tumorali, utilizzando strumenti computazionali per studiare la distribuzione delle mutazioni somatiche.
La penultima lezione si concentrerà invece sulle mutational signatures, un concetto fondamentale per la comprensione delle cause e della progressione del cancro. Le firme mutazionali sono modelli caratteristici delle mutazioni che si verificano in un tipo di cancro specifico, e l'analisi di queste firme è fondamentale per personalizzare i trattamenti oncologici. Gli studenti impareranno a identificare e analizzare queste firme utilizzando tecniche avanzate di bioinformatica e machine learning.
L'ultimo incontro del corso sarà un'esercitazione finale che permetterà agli studenti di mettere in pratica quanto appreso riguardo all'analisi delle mutazioni genomiche e delle mutational signatures.

Prerequisiti

Conoscenze di matematica e biologia di base.

Modalità didattica

Lezioni frontali.

Materiale didattico

Diapositive e materiale didattico fornito dal docente.

Periodo di erogazione dell'insegnamento

Primo semestre.

Modalità di verifica del profitto e valutazione

Modalità d’esame: Prova scritta.
Tipologia di quesiti: Quiz a risposta multipla con 5 opzioni di cui 1 corretta per un totale di 16 domande.
Risultati delle valutazioni: Il punteggio è espresso in trentesimi con punteggio minimo di 18/30 e punteggio massimo di 30/30 con lode.
Argomenti inclusi nei quesiti: Tutti gli argomenti delle lezioni possono essere richiesti.
Durata dell’esame: Lo studente ha 45 minuti per svolgere l’esame.

Orario di ricevimento

Su appuntamento richiesto via e-mail.

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ