UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Informatics

2122-1-E0201Q046

Obiettivi

L'insegnamento si propone di spiegare i concetti base dell'Informatica e della Statistica, con una particolare attenzione alle applicazioni di carattere biotecnologico. Durante l'insegnamento verranno illustrate le relazioni fra le discipline informatiche e le discipline biologiche: la trattazione dei concetti base dell'Informatica sarà affiancata alla spiegazione dell'importanza che tali concetti rivestono per l'analisi di dati biologici (ad es. ricerca in banche dati biologiche, problemi di Bioinformatica, Biologia Computazionale e Biologia dei Sistemi).

1. Conoscenza e capacità di comprensione.

Al termine dell'insegnamento lo studente dovrà sviluppare la capacità di:
- "computational thinking", al fine di utilizzare in modo appropriato gli strumenti dell'Informatica (algoritmi, metodi computazionali, software) per la soluzione di un dato problema;
- analisi critica relativamente alla scelta dei metodi statistici più adeguati per l'analisi di dati in ambito biologico/clinico, e per l'interpretazione dei rispettivi risultati.

2. Capacità di applicare conoscenza e comprensione.

Al termine dell'insegnamento lo studente dovrà essere in grado di applicare le conoscenze acquisite al punto 1 per la soluzione di problemi in ambito computazionale e statistico:
- organizzare e trattare automaticamente i dati o risolvere semplici problemi tramite l'utilizzo di fogli di calcolo (esercitazioni in laboratorio);
- sviluppare semplici algoritmi, programmare in linguaggio Python e utilizzare l'ambiente di sviluppo R per analisi statistiche (esercitazioni in laboratorio).

3. Autonomia di giudizio.

Al termine dell'insegnamento lo studente dovrà essere in grado di elaborare quanto appreso, e saper riconoscere le situazioni e i problemi in cui le metodologie informatiche e statistiche apprese possano essere utilizzate.

4. Abilità comunicative.

Al termine dell'insegnamento lo studente dovrà essere in grado di esprimersi in modo appropriato nella descrizione delle tematiche affrontate, con proprietà di linguaggio e sicurezza di esposizione.

5. Capacità di apprendimento

Al termine dell'insegnamento lo studente dovrà essere in grado di consultare la letteratura sugli argomenti trattati, nonché analizzare, applicare, integrare e collegare le conoscenze acquisite con quanto verrà appreso in insegnamenti correlati all'applicazione delle scienze computazionali e statistiche per l'analisi dei dati.

Contenuti sintetici

Informatica:
Nozioni di base sui calcolatori e sulla codifica dei dati.
Algoritmi, computational thinking, ed elementi di programmazione strutturata.
Nozioni di complessità computazionale.
Fondamenti di Bioinformatica, Biologia Computazionale, Biologia dei Sistemi.
Metodi computazionali euristici di ispirazione biologica.

Statistica:
Statistica descrittiva.
Statistica inferenziale.

Esercitazioni in laboratorio:
Fogli di calcolo e ricerche in banche dati biologiche.
Elementi di programmazione in Python e R.

Programma esteso

Informatica

1) Nozioni di base sui calcolatori e sulla codifica dei dati. Com'è fatto e come funziona un computer (architettura di von Neumann, ciclo macchina). Codifica dell'informazione.
2) Pensare in modo algoritmico ed elementi di programmazione. Definizione di algoritmo. Dal problema all'algoritmo, dall'algoritmo al programma. Programmi e linguaggi di programmazione. Programmazione strutturata e pseudo-codice. Problemi "facili" e problemi "difficili", efficienza degli algoritmi. Fare calcoli con il DNA: l'esperimento di Adleman.
3) Fondamenti di Bioinformatica, Biologia Computazionale e Biologia dei Sistemi. Le banche dati biologiche (banche dati primarie, secondarie, specializzate). Algoritmi di allineamento fra sequenze e tecniche euristiche per ricerche in banche dati. Protein folding, molecular docking. Approcci computazionali per sistemi biologici complessi.
4) Dalla Biologia all'Informatica: metodi computazionali di ispirazione biologica per la soluzione di problemi "difficili".

Statistica

1) Statistica descrittiva. Campioni e popolazioni, tipologie di dati e variabili Disegno degli esperimenti (cieco, replicazione, strategie di campionamento). Il concetto di frequenza (assoluta, relativa, cumulativa).

Rappresentazione grafica dei dati (istogrammi, diagrammi a torta, diagrammi a dispersione). Misure di centralità (media, mediana, moda). Misure di dispersione (range, deviazione standard, varianza). Misure di posizione (quantili, percentili). Analisi esplorativa dei dati (outlier, boxplot).
2) Statistica inferenziale. Nozioni di teoria della probabilità. Distribuzioni di probabilità (uniforme, binomiale, normale, distribuzione t di Student). Metodi di stima con un campione (intervalli di confidenza, stima con un campione). Verifica di ipotesi con un campione. Correlazione e regressione.

Esercitazioni in laboratorio

I fogli di calcolo per l'elaborazione dei dati. Ricerca in banche dati biologiche. Elementi di programmazione in Python. Elementi di statistica descrittiva ed inferenziale con R.

Prerequisiti

Prerequisiti: nessuno.
Propedeuticità: nessuna.

Modalità didattica

Lezioni frontali in aula (40 ore) ed esercitazioni (30 ore), supportate da presentazioni PowerPoint.
Tutta l'attività didattica verrà videoregistrata e resa disponibile tramite la piattaforma Moodle.

Materiale didattico

Slide delle lezioni e delle esercitazioni reperibili sulla pagina Moodle dell'insegnamento.

Testi consigliati:

- M.G. Schneider, J.L. Gersting. Informatica. Algoritmi, architetture, linguaggi, applicazioni. Maggioli Editore - Apogeo Education, 2020
- P.J. Deitel, H.M. Deitel. Introduzione a Python. Per l'informatica e la data science. Pearson, 2020
- M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole. Fondamenti di Bioinformatica. Zanichelli, 2018
- M.M. Triola, M.F. Triola. Fondamenti di statistica per le discipline biomediche, Pearson, 2017
- M.C. Whitlock, D. Schluter. Analisi statistica dei dati biologici, Zanichelli, 2010

Periodo di erogazione dell'insegnamento

Primo semestre.

Modalità di verifica del profitto e valutazione

Esame scritto della durata di 2 ore, con 9 domande a risposta multipla e 1 domanda a risposta aperta, riguardanti gli argomenti trattati durante le lezioni frontali e le esercitazioni in laboratorio.

Non verranno svolte prove in itinere.

Orario di ricevimento

Su appuntamento previa richiesta via e-mail a daniela.besozzi@unimib.it