

CAP 6546 Data Mining for Bioinformatics

Credits: 3 credits

Textbook, title, author, and year: No required textbook for this course

Reference materials: N/A

Specific course information

Catalog description: Course focuses on the principles of data mining as it relates to bioinformatics. Topics covered include gene selection, class imbalance, classification, biomarker discovery and prediction models. No prior knowledge of biology is required.

Prerequisites: *Graduate standing or permission of instructor*

Specific goals for the course:

- Learn fundamental principles of bioinformatics and data mining.
- Use the R programming language to wrangle, visualize, and analyze a diversity of large and complex biological datasets.
- Write a report detailing a data analysis project in R.
- Describe key terminology and concepts in bioinformatics and data mining.
- Apply data mining techniques to bioinformatics problems in the R programming language.
- Analyze findings from applications of data mining techniques to bioinformatics problems.

Brief list of topics to be covered:

- Fundamentals of bioinformatics
- Introduction to the R programming language
- Statistical inference in bioinformatics
- Exploratory data analysis in bioinformatics
- Predictive modeling in bioinformatics
- Pre-processing and aligning high-throughput sequencing reads
- Assaying genomic variation with DNA-seq data
- Quantifying gene expression with RNA-seq data
- Identifying protein-DNA interactions with ChIP-seq data
- Analyzing DNA methylation with BS-seq data