Syllabus for STAT 452: Introduction to Statistical Learning
Spring 2024 (3 units)

Instructor: Eric Fox
Email: eric.fox@csueastbay.edu
Office: North Science 303A

Lecture: Tues/Thurs 2:45-4 at South Science 146

Office Hours: Tues/Thurs 1-2 at North Science 303A, or by appointment

Website: Course materials will be posted on Canvas.

Textbook: James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning with Applications in R.* Second Edition, Springer, 2021.
Free PDF version: https://www.statlearning.com/

Additional Reference: Bradley Boehmke and Brandon Greenwell. *Hands-On Machine Learning with R.* CRC Press, 2019.
Free online version: https://bradleyboehmke.github.io/HOML/

Software:
R, can be downloaded here https://www.r-project.org/
RStudio, can be downloaded here https://www.rstudio.com/
RStudio Cloud, https://rstudio.cloud/

Suggested Prerequisites: STAT 432 and STAT 450.
This is the most advanced data science course offered at the undergraduate level. It will be assumed that students have some prior computer programming experience, preferably with R.

Course Topics:

- Review linear regression

- Logistic regression and classification metrics

- Cross-validation: training and test set

- Nonlinear methods: polynomial and local regression

- k-nearest neighbors algorithm

- Decision trees and random forests

- Neural networks

- Unsupervised learning: k-means clustering, hierarchical clustering

Grading: There will be weekly homework assignments and two projects. The projects will focus on analyzing complex data sets and comparing different statistical learning methods. All assignments should be completed using Quarto, rendered to PDF or HTML format, and then submitted to Canvas.

- 5% Attendance
- 55% Homework
- 40% Two Projects

Policy on Late Assignments: Late assignments will either receive a point deduction or not be accepted. However, your lowest scoring homework will be dropped. I may agree to extensions on due dates if you are experiencing an emergency or illness.

Attendance Policy: Students are required to attend class on campus during the scheduled times and participate in class activities.

Important Dates:

- First day of classes: Tuesday, January 16
- Last day to drop: Monday, January 29
- Spring break: April 1-5
- Last day to withdraw: Friday, April 14
- Last day of classes: Friday, May 3

A complete list of important dates:
https://www.csueastbay.edu/registrar/important-dates/spring-2024.html

Student Learning Outcomes: Upon successful completion of this course, students should be able to:

- Apply fundamental methods for statistical learning, including (a) linear regression, (b) logistic regression, and (c) decision trees and random forests.
- Understand the basic theory and concepts underlying these methods.
- Select statistical models and assess model performance using cross-validation (i.e., splitting data into training and test sets)
- Use R and RStudio to implement statistical learning procedures and analyze complex data sets.
- Communicate statistical learning concepts clearly and appropriately to others.

Common Syllabus Items: Items such as policies on academic integrity, disability, handling emergency situations, and protection against discrimination, harassment, and retaliation can be found under "University Policies" on Canvas.

Student Services: To access student services offered at Cal State East Bay, click on the My-Compass icon to get you to your one-stop online student support hub for information on academic advising, tutoring, financial aid, the library, the health center, technology support, career counseling, campus life, equity programs, and more.

Grade Appeal and Academic Grievanes: If you wish to appeal your course grade at the end of the semester or have other academic concerns related to a course, please visit the Grade Appeals and Academic Grievances (GAAG) section of the catalog, which explains the process.