Syllabus for STAT 452: Introduction to Statistical Learning
Spring 2021

Instructor: Dr. Eric Fox
Email: `eric.fox@csueastbay.edu`
Office: SC-N 303A

Lecture: Mon/Wed 12:30-1:45 over Zoom

Office Hours: Tu/Th 1-2 and Wed 2-4, or by appointment
Zoom link: `https://csueb.zoom.us/j/502694714`

Website: Course materials will be posted on Blackboard.

Textbooks: James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.
Free PDF version: `https://statlearning.com/`

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 or 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:

- Simple and Multiple Linear Regression

- Logistic Regression and Classification Metrics

- K-Nearest Neighbors Algorithm

- Cross-Validation: Training and Test Set

- Variable Selection and Regularization: Stepwise Selection, Ridge Regression, The Lasso

- Decision Trees and Random Forests

- Regression Splines, Local Regression, and Generalized Additive Models*

- Unsupervised Learning: K-Means Clustering, Hierarchical Clustering*

* Indicates optional topics

Grading: Homework will be worth 100% of your grade, and will be assigned once a week, or once every two weeks. All assignments should be completed using R Markdown, rendered to PDF or HTML format, and then submitted to Blackboard.

Policy on Late Assignments: Late homework will generally not be accepted. I may agree to extensions on due dates if you are experiencing an emergency or illness.

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

- Apply fundamental methods for statistical learning, including (a) simple and multiple 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.

Technology Requirements: This course will use the web conferencing software Zoom. To participate you will need a stable internet connection, and a laptop or desktop computer equipped with a webcam, microphone, and speakers. Please refer to the Zoom system requirements here.

Course Policies and Zoom Etiquette:

- All lectures will be delivered live during the scheduled class time, and attendance is highly recommended. Recordings of the sessions will be posted on Blackboard for students that cannot attend or have connectivity issues.

- Make sure that your audio is muted upon entry into the class.

- You may ask questions by using the chat function or by unmuting yourself. Please try to not disrupt the instructor or other students.

Common Syllabus Items: Items such as policies on academic dishonesty, disability, and handling emergency situations can be found under "University Policies" on Blackboard.