Instructor. Daniel J. McDonald

Time and place. Tue/Thu 0800h – 0930h, Earth Sciences Building, 1012

Website. https://ubc-stat.github.io/stat-406/

Prerequisites. STAT 306 or CPSC 340. These prerequisites are strict.

Course description. This is a course in statistical learning methods. Based on the theory of prediction models covered in STAT 306, this course will focus on applying many techniques of data analysis to interesting datasets. There is also a strong emphasis on data science technologies and practices, includin version control, reproducible research, and data visualization.

The course combines analysis with methodology and computational aspects. It treats both the "art" of understanding unfamiliar data and the "science" of analyzing that data in terms of statistical properties. The focus will be on practical aspects of methodology and intuition to help students develop tools for selecting appropriate methods and approaches to problems in their own lives.

This is not a "how to program" course, nor a "tour" of machine learning methods. Rather, this course is about how to understand some ML methods. STAT 306 tends to give background in many of the tools of understanding as well as working with already-written `R` packages. On the other hand, CPSC 340 introduces many methods and focuses on from-scratch implementation (in Julia or Python). This course will bridge this gap. Depending on which course you took, you may be more or less skilled in some aspects than in others. That's okay and expected.

Topics. assess the prediction properties of the supervised learning methods covered in class; correctly use regularization to improve predictions from linear models, and also to identify important explanatory variables; explain the practical difference between predictions obtained with parametric and non-parametric methods, and decide in specific applications which approach should be used; select and construct appropriate ensembles to obtain improved predictions in different contexts; select sensible clustering methods and correctly interpret their output; correctly utilize and interpret principal components and other dimension reduction techniques; employ reasonable coding practices and understand basic `R` syntax and function; write reports and use proper version control; engage with standard software

Textbook. *An Introduction to Statistical Learning*, James, Witten, Hastie, Tibshirani, 2013, Springer, New York. Available free at https://www.statlearning.com.

The Elements of Statistical Learning, Hastie, Tibshirani, Friedman, 2009, Second Edition, Springer, New York. Also available free at https://web.stanford.edu/~hastie/ElemStatLearn/

Assessment. Assignments; Labs; Student response devices; Final Exam

Student engagement. Lecture recordings will be available; some policies in place for late/dropped assessments; some office hours will be held remotely; note that regular class attendance is required.