Stat 479/579: Machine Learning, Fall 2022

Instructor: Dr. Yan Lu, `yanlu@unm.edu`

Teaching Assistant: , Md Rashidul Hasan, `mdhasan@unm.edu`

Class website: `https://canvasinfo.unm.edu`

Time and Location: 08:30am - 09:45am MW, Anderson graduate school of Mgmt 230

Office hours: MW 10:30pm-11:30am or by appointment (by zoom)

`https://unm.zoom.us/j/97558415018`

Meeting ID: 975 5841 5018

Passcode: 123

Textbook: Christensen, Ronald (2020). Statistical Learning: A Second Course in Regression. `https://math.unm.edu/~fletcher/SL.pdf` Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, An Introduction to Statistical Learning with Applications in R

`https://trevorhastie.github.io/ISLR/ISLR%20Seventh%20Printing.pdf`

Prerequisites: Stat 345 and Stat 427/527 Advanced Analysis 1. Stat 440/540 (Regression) is a helpful course but is not required.

Topics: The field of statistical learning encompasses a variety of computational tools for modeling and understanding complex data. In this introductory course, we will explore many of popular tools, such as multiple linear regression, classification trees, boosting and support vector machines.

The course starts with a review of multiple linear regression including its matrix version. It then introduces ideas of nonparametric regression and alternatives to least squares estimation, e.g. regularization. Next we examine variable selection such as the traditional method stepwise selection. Modern methods include bagging, boosting, and the general idea behind random forests. We then return to a more detailed look at nonparametric regression involving polynomials, trig functions, wavelets, splines, other local regression ideas like loess, reproducing kernel functions, and regression trees. To employ these nonparametric methods successfully in unsupervised learning requires the use of penalized estimation (regularization), which is revisited in more depth. The nonparametric structures for regression can be applied to dependent variables with 0-1 outcomes (classification) which is the next subject considered. This discussion includes

the use of support vector machines.

Computing:

R and Rstudio will be used for about the first half semester, and Python will be introduced for the second half semester (R code will still be provided).

R:

`http://cran.r-project.org`

Rstudio:

`https://www.rstudio.com/products/rstudio/download`

To see how to install R and Rstudio in windows, visit

`https://www.youtube.com/watch?v=eD07NznguA4`

for Mac

`https://www.youtube.com/watch?v=GFImMj1lMRI`

Anaconda python

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single platform

Check the following youtube to install Anaconda Python and JupyterLab (you will use Jupyternotebook to write your code and run code)

`https://www.youtube.com/watch?v=daVgEXjv6DE`

Grading: Homework will be given about every 2 weeks for 7-8 assignments. Homework 30%, Midterm Exam, 30%; Final Project, 40%, Grading for graduate students and undergraduates will be separate.

	Stat 579	Stat 479
A	90%-100%	80%-100%
B	75%-89%	65%-79%
C	65%-74%	55%-64%
D	under 65%	under 55%

Final Project: Your project in this class will involve some theoretical or applied aspect of machine learning techniques. More detailed information will be given in class later.