STAT 479/579   Statistical Learning   Spring 2020

Time: 11:00–12:15

Site: DSH–129

Professor: Ronald Christensen

Office: SMLC 304

E-mail: fletcher@stat.unm.edu

Web: www.stat.unm.edu/~fletcher   (Click on course title under teaching.)

Phone: They removed it in a pseudo budget cut.

Office Hours: T,Th 12:45–1:45 and by appointment.

Text: Christensen, Ronald (2020). *Statistical Learning: A Second Course in Regression*. Will be available online or I will email it to registered students.

Prerequisites: Stat 345 and Stat 440 or 540.

Material: 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 including model goodness criteria such as C_p, AIC, and BIC. Traditional and modern methods of stepwise selection are considered, i.e., *greedy selection algorithms*. Modern methods include *bagging*, *boosting*, and the general idea behind *random forests*. A brief digression will be made into a pair of multiple comparison methods that are appropriate for *big data*. 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*. The course concludes with discussion of topics traditionally viewed as multivariate analysis, i.e., discrimination (classification with a twist), *clustering*, *dimension reduction* (*principal components*), and multidimensional scaling.

Grading: Grading will be based on homework, quizzes, and/or exams. Grading for graduate students and undergraduates will be separate. **You are required to keep legible copies of all work,** except in-class examinations.

Virus Update

Hello:

I have intentionally left you alone this week so as not to interfere with your ability to enjoy spring break. With the rest of the world doing its best to interfere, I thought that I would resist the temptation.

Me being a contemporary of Homer and Telemachus, you will probably not be surprised if I am not on the cusp of online teaching technology. My vision for the near future is pretty simple. Lots of reading assignments and lots of exercises. I've been working on producing a series of "Quizzes" for you to perform. But I am really planning them as learning exercises more than quizzes. Because of the current situation, it will be acceptable for you to discuss these quizzes with your classmates (as long as you do not get within 6 ft of them). You should, however, write up your answers yourself, i.e., **no copying**.

I will probably just post the quizzes on the course website AS I GET THEM DONE. (Go to www.stat.unm.edu/~fletcher and figure it out.) Also, if/when I can figure out how to do it, I will try to hold open our usual class times for something like a chat room where you can come online and ask any questions that you have. Of course you can always email me BUT if I start getting inundated with email, I will probably start getting unresponsive to email. Hence my desire to do something like a chat room. If I become aware of better technology, I will let you know.

My current idea is that there will be one quiz for each chapter of the book that I want to cover. Quiz 1 covers Chapter 1 (Those in Stat Learning have already done that one) and all quizzes will be numbered by the chapter they cover. If I decide I need more than one quiz for a chapter, the number will still agree with the chapter but I will add extra identifiers, e.g. Quiz 4a would relate to Chapter 4. Remember that the online computing book also has its chapters and sections set up to agree with the text. Once I get all of the quizzes up, you can work at your own speed to finish the class. When you finish the quizzes, you will be done with the class. (In my youth this was called "programmed learning.")

For now, I don't know of a better way for you to hand in your work than to email it to me. I want to have the answers written in the spaces provided for the answers on the quiz form. Normally I would require that but I'm not sure if you all have appropriate technology. Ideally, you could type/add your answers to the pdf file and return it. I know that would be inconvenient for me to do. I would probably handwrite the answers on a printed copy of the pdf and either scan it or take a clear photo of it. Let me know if you have problems. (If worse comes to worse you can snail mail them.)

Frankly, I think you are going to learn a lot more from reading and doing exercises than you would from just listening to me spout off.

Good Luck, Ron