Statistical Learning I (Math 540)
Final Exam
Fall 2019
Due December 5, 2019 noon

Methodology

Select 3 from Problems **M1 – M4.**

M1. An example of confusion that results when there is an attempt to interpret regression coefficients can be found in the article:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1371554/pdf/brjgenprac00071-0025.pdf

The authors state on page 70: "This interpretation immediately raised a question as to why certain coefficients were statistically significant and yet had unrealistically large or even negative values. There probably exists a strong linear relation connecting many, if not all, the predictor variables."

Read this portion of the article and comment on this remark relative to what you see in their Table 3. What seems to be the problem? Would you have performed the analysis differently? Explain.

M2. In discussing the use of variable selection as an aid in the development of theoretical models, Luo, Stefanski, and Boos in their paper

https://www4.stat.ncsu.edu/~boos/papers/lsb2581.pdf

"For such applications of regression, the ability to distinguish between important and unimportant variables is paramount. This aspect of variable selection is not as crucial when prediction is the sole or primary objective of modeling."

Do you agree with the last statement? Explain.

M3. Explain why ridge regression would never be used when there is only a single predictor variable.

M4. Consider the article on ridge regression by Meisner:

http://journals.ametsoc.org/doi/pdf/10.1175/1520-0450(1979)018%3C0904%3ARRTEAT%3E2.0.CO%3B2

Is Meisner using ridge regression for its intended purpose? If not, explain why not. If so, what do you recommend to the author regarding his method of choosing the ridge parameter, k?

Theory

Select 5 from Problems **T1 – T6.**

T1. Exercise 2.9 from Hastie/Tibshirani/Friedman.

T2. Exercise 3.12 from Hastie/Tibshirani/Friedman.

T3. Exercise 3.29 from Hastie/Tibshirani/Friedman.

T4. Prove Scheffé's Identity: $\sup_{A \in Borel} |P(A) - Q(A)| = \frac{1}{2}\int |f(x) - g(x)| dx$ where $P(A) = \int_A f(x)dx$ and $Q(A) = \int_A g(x)dx$.

T5. Exercise 6.2 from Hastie/Tibshirani/Friedman.

T6. Exercise 6.3 from Hastie/Tibshirani/Friedman.