STAT 6306 – Some Topics

In no particular order, here are some important concepts (membership on this list is neither necessary nor sufficient for inclusion on the exam). The exam will be closed book/notes/computer.

- Why can't I just use the training error to choose tuning parameters?

- Describe any tuning parameters in all subsets regression and ridge regression

- As I increase λ in ridge, what happens to the overall size of the coefficient estimates? Under what circumstances would any particular coefficient be set to zero?

- T/F: Forward selection, backward selection, and all-subsets regression will all pick the same model.

- T/F: all-subsets regression will provide a lower training error solution than forward selection or backward selection

- T/F: all-subsets regression will provide a lower risk solution than forward selection or backward selection

- Suppose the Bayes' rule has the form $f_*(X) = X^\top \beta$. What happens to the bias^2 of ridge regression as λ decreases? In particular, what is the bias^2 when $\lambda = 0$? What is the name of the $\lambda = 0$ procedure?

- Suppose the Bayes' rule has the form $f_*(X) = X^\top \beta$. What happens to the variance of ridge regression as λ decreases? In particular, what is the variance when $\lambda = 0$?

- Suppose the Bayes' rule has the form $f_*(X) = X^\top \beta$. What happens to the bias^2 of ridge regression as λ increases? In particular, what is the bias^2 when $\lambda \to \infty$?

- Suppose the Bayes' rule has the form $f_*(X) = X^\top \beta$. What happens to the variance of ridge regression as λ increases? In particular, what is the variance when $\lambda \to \infty$?

- Answer the above four bullet points, exchanging t for λ.

- What would have to happen for ridge to be preferred over linear regression for prediction?

- Accurately produce a bias/variance decomposition plot including things like: risk, Bayes' rule, risk of the Bayes' rule, bias^2, variance, and training error. Where would a no bias procedure be? What is an example of a no variance procedure?