STAT460 – Midterm Topics

In no particular order, here are some concepts that I feel are good test questions.

1. Why can't I use the training data to choose tuning parameters.

2. As I increase λ in ridge, what happens to the overall size of the coefficient estimates? Are any coefficients set to zero?

3. As I increase λ in lasso, what happens to the overall size of the coefficient estimates? Are any coefficients set to zero?

4. Sketch the optimization problems behind ridge and lasso. What does λ correspond to? How does the geometry of the problem inform you answers to the previous questions?

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

6. Give me an example of using AIC and why would you use it?

7. Describe to me the idea behind bootstrap.

8. Which procedure is more biased: Multiple linear regression or Ridge? How do you know?

9. Which procedure has higher variance: Multiple linear regression or Ridge? How do you know?

10. What would have to happen for Ridge to be preferred over Multiple linear regression for prediction?

11. What does it mean qualitatively for an estimator to have high variance?

12. If I gave you some R output from one of the methods we have discussed so far, could you interpret the results such as which variables are selected?

13. Given a confusion matrix, explain the properties of a given classifier. What are its specificity? Sensitivity?

14. What is the difference between Random Forest and Bagging Trees? Is one a special case of the other?

15. If I wrote down some sample data, could you produce the tree that would be grown with a particular criterion and stopping criterion?

16. Suppose we produce 10 bootstrapped samples from a data set with two classes $Y \in \{0, 1\}$. We grow a classification tree to each bootstrap sample and for a specific value X, produce 10 estimates of $P(Y = 0|X)$:

$$0.1, 0.15, 0.2, 0, 2, 0.55, 0.6, 0.6, 0.65, 0.7, 0.75$$

There are two common methods for generating classifications given this data:

   (a) Take the average of these probabilities and round that to 0 or 1
   (b) Take a majority vote for the rounded probabilities of each tree.

What would the classification based on each of these methods?