STA 314, Review problems 4 (Lecture 5)

Multiple choice questions can have any number of correct answers (including zero).

1. The table below gives estimated values for b_1, b_2, b_3 (based on data $(x_i, y_i)_{i=1,...,n}$) in a linear model $y_i = b_1 + b_2 x_{i,1} + b_3 x_{i,2} + \varepsilon_i$ with lasso penalty.

λ	0	1	2	20	50	100
$\hat{b}_1$	3	2	1	0.3	-1	-1
$\hat{b}_2$	4	3	2	1.2	0	0
$\hat{b}_2$	5	2	0	0	0	0

   (a) Based on this table, provide the values for $\hat{b}_1, \hat{b}_2, \hat{b}_3$ that one would obtain from a simple linear regression without any penalty.

   (b) Based on this table, what is the sample mean of $y_1, ..., y_n$?

   (c) Assume that $\lambda = 20$ was chosen by cross-validation. Which model does this correspond to?

2. For each of the following, will it make the model more flexible or less flexible?

   (a) Including additional predictors in a linear model.
   (b) Increasing λ in lasso.
   (c) Decreasing λ in lasso.
   (d) Increasing k in k-nn regression.

3. Which of the following is true for the one standard error rule compared to usual cross-validation?

   (a) It will lead to models with a smaller number of predictors if applied to best subset selection.
   (b) It will lead to models with a smaller number of predictors if applied to forward stepwise selection.
   (c) It will lead to smaller values of λ in lasso.

4. Which of the following statements are true?

   (a) Lasso and linear regression will never give the same solution.

   (b) Lasso can set some regression coefficients to 0 when λ is large.

   (c) Lasso and linear regression regression will give the same solution when $\lambda = 0$.

5. Prove the statement made in lecture 5: if best subset selection is applied in the first step, then

$$\hat{b}^{AIC} = \text{argmin}_{b \in R^{p+1}} \Big\{ RSS(b) + 2\hat{\sigma}^2 \#\{k = 2, ..., p+1 : b_k \neq 0\} \Big\}$$

6. * Consider the simplified example for LASSO considered in lecture 5: one-dimensional predictors with $\bar{x} = 0$. Prove the statement made in lectures: if $n\overline{xy} < -\lambda/2$, then

$$\frac{\overline{xy} + \lambda/(2n)}{\overline{x^2}} = \text{argmin}_{b_2 \in R} \Big(\sum_{i=1}^{n} (y_i - \bar{y} - b_2 x_i)^2 + \lambda |b_2| \Big).$$