

## CMSE 381: HW3

- 1 Exercise 5.4.1
- 2 Exercise 5.4.6
- 3 Exercise 5.4.7
- 4 Exercise 5.4.8
- 5 We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain  $p + 1$  models, containing  $0, 1, 2, \dots, p$  predictors. Which of the three models with  $k$  predictors has the smallest *training RSS*?
- 6 Exercise 6.6.3
- 7 Exercise 6.6.4
- 8 Exercise 6.6.8
- 9 Exercise 6.6.9 (a-d)
- 10 We consider the following model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

where  $\mathbb{E}(\epsilon) = 0$ ,  $\text{Var}(\epsilon) = 1$ ,  $\text{Cov}(X_1, X_2) = \begin{pmatrix} 1 & 0.9999 \\ 0.9999 & 1 \end{pmatrix}$ ,  $\beta_1 = -1$ , and  $\beta_2 = 2$ . We repeat the following 4000 times

- a. Generate 50 data points  $\{y_1, \dots, y_{50}\}$  and  $\{(x_{1,1}, x_{1,2}), \dots, (x_{50,1}, x_{50,2})\}$ .
- b. Calculate the  $\hat{\beta}_1^{ols}$  and  $\hat{\beta}_1^{ridge}$  (with  $\lambda = 0.01$ ) using the simulated 50 data points.
- c. Among the 4000 simulations, how many times  $\hat{\beta}_1^{ridge}$  is closer to the true  $\beta_1$ , which is  $-1$  than  $\hat{\beta}_1^{ols}$ ?