CMSE 381: HW8

1 (20 pts) Exercise 6.8.4

2 (20 pts) Exercise 6.8.6

3 (10 pts) Exercise 6.8.9 (e)

4 We will use ridge and lasso regressions to estimate the salary of various baseball players based on numerous predictor measurements. Download 'MLB_test.csv' and 'MLB_training.csv' file from d2l, which contains Major League Baseball (MLB) data from the 1986 season. In this dataset, each observation represents one MLB player, and we want to build a model to predict the player's logsalary using the other features in this dataset. We will first define the grid of λs using

$$\text{grid} = 10\hat{}\ \text{seq(10, -2, length=100).}$$

   a. (10 pts) For both ridge and Lasso regressions, perform 5-fold cross-validation on the training data to deter the best value of λ using the one standard error rule. Report the best λs.

   b. (10 pts) Report the coefficients estimates at the appropriate values of λ for both ridge and Lasso regression using training data.

   c. (20 pts) Evaluate the predictive capacity of ridge and lasso regression by computing R^2 for the testing data.

   d. (10 pts) Suppose that you were coaching a young baseball player who wanted to strike it rich in the major leagues. What handful of attributes would you tell this player to focus on?