WU #20 - Lasso variable selection

Math 158 - Jo Hardin

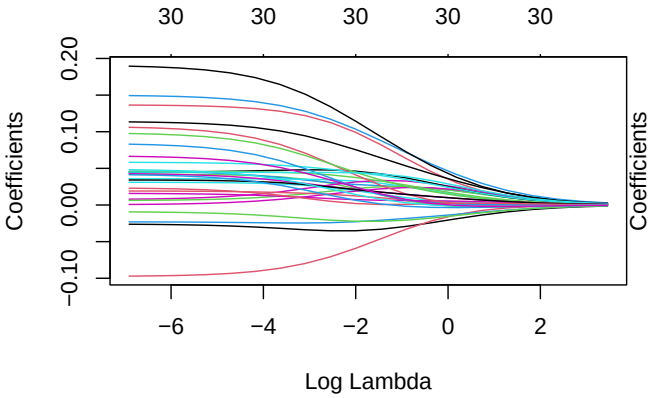
Tuesday 4/19/2022

Name: _____

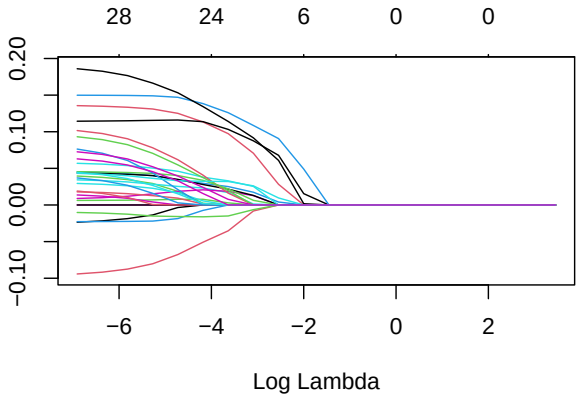
Names of people you worked with: _____

The dataset we will be working with has `imdb_rating` as the response variable. The predictor (explanatory) variables are: `season`, `episode` and 28 columns representing the number of lines of a particular character.

Ridge regression



Lasso



1. For what value(s) of λ do ridge regression and lasso produce the same coefficient values?
2. What is the number at the top of the graph? Why is it always the same for ridge regression? Why is it zero for some values of lambda on the lasso plot?

Solution:

1. When $\lambda = 0$ the problem boils down to OLS, so ridge regression and lasso will have the same coefficients (which will also be the same as OLS).
2. The number at the top of the plot is the number of non-zero coefficients in the model. For ridge regression, the estimates for beta coefficients never goes to zero, so there will always be 30 non-zero coefficients. For lasso, the optimization solution will hit a corner of the boundary conditions, causing coefficients to be shrunk all the way to zero. In this way, lasso is a variable selection technique.