Econ 108 FAll 2022
Problem Set 6

This problem set is due at 7pm, Friday, November 11, 2022.

1. Continue with question 1 from the last problem set using the spam.csv data.

   (a) Report the AIC, AICc, and BIC for both the full model and the cut model. Which model is prefered by each of the model selection information criterion?

   (b) Fit a LASSO model to the spam dataset, and plot the LASSO path. I suggest changing the lambda minimum ratio, for example to 0.0001 or even smaller.

   (c) How many features are selected (to be nonzero) by AICc? How many features are selected by the Baysian information criterion? What are the corresponding values of log lambda chosen by AICc and BIC?

   (d) Cross-validate the LASSO, and plot the cross-validation average deviance and its standard deviation as a function of the log of lambda.

   (e) What is the value of log lambda that minimizes the average cross validation deviance? How many features are selected at this deviance minimizing log lambda?

   (f) What is the value of log lambda that corresponds to the one-standard deviation rule using the average cross validation deviance? How many features are selected at this log lambda?