Model Selection and Averaging
Summer School in Economics and Econometrics
University of Crete
July 22-26, 2019

Exercise 5

Use the CPS (cps09mar) dataset posted at
https://www.ssc.wisc.edu/~bhansen/econometrics/

1) Use the same data file as in the first three exercises, but now use all observations. Make similar transformations to obtain indicators for education and power of experience
2) Create a large list of regressors. Notice that previously your dataset was Hispanic women. Now you have both genders and all race groups. Create dummy variables as needed for the flexibility you desire. Include interactions if desired.
3) Estimate the regression using least squares
   a) OLS
   b) Ridge Regression
   c) Lasso
   d) Elastic Net with alpha=1/2
4) Report your coefficient estimates in a large table
5) Comment on your findings

Computational notes:

For OLS, use the *lm(y~X)* command where X contains your regressors but no intercept. This will eliminate collinear regressors if needed

For the other three, use glmnet

a) Activate the library *glmnet*
b) The command *name <- cv.glmnet(X,y,family="gaussian",nfolds=K)* computes the model with lambda selected by K-fold CV
c) If you want, use the command *plot(name)* to plot the CV function
d) The command *coef(name,name$lambda.min)* extracts the coefficient estimates obtained with the CV-minimizing value of lambda