# DATA 221
# Homework 5 (rev 0)

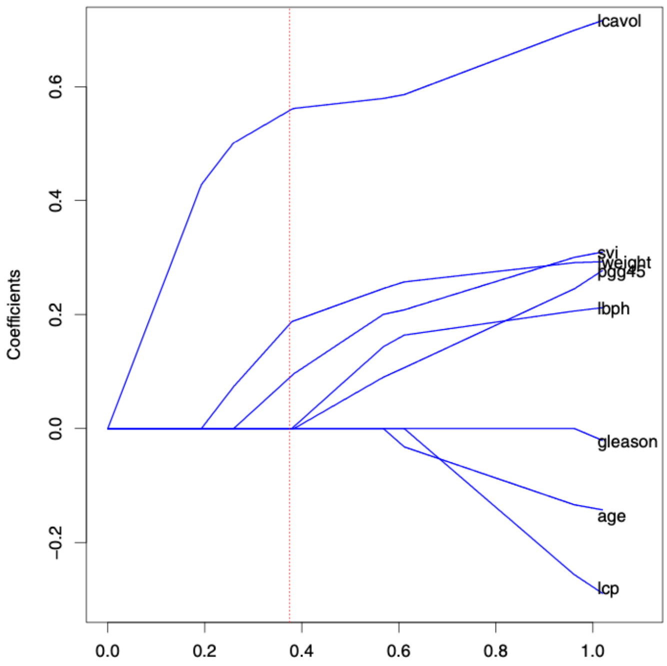
**W. Trimble**

Due: Friday 2022-05-06 - 11:59pm

1. Use a logisitc regression model to fit the UCI Credit Default dataset. You should standardize the inputs and prepare a testing set separate from the training set. Report the column names and coefficients ordered by the absolute value of the regression coefficient.

2. Regularization

   Find the logistic regression coefficients for this dataset with L1 regularization. Plot the regression coefficients as a function of the (logarithm of the) regularization parameter. Find the optimum regularization parameter by optimizing for minimum error on the test set.

   (Here is an example of L1, also called LASSO, regularization from Hastie Elements of Statistical Learning Figure 3.10. High regularization is to the left. Low regularization reproduces problem 1, high regularization eventually sets all coefficients to zero.)

3. SVD Perform Singular Value Decomposition on all the features of the UCI Credit Default dataset. Display scatter plots of PC1 vs PC2 for default and non-default classes. Label the axes with the fraction of the variance in PC1 and PC2. Compute a table with the fraction of the variance in each of the first few principal components.

4. Perform L1-regularized logistic regression on the SVD-transformed features and plot the feature coefficients as a function of the logarithm of the regularization parameter. Find the optimum regularization parameter by minimizing error on the test set. Does the optimum include more or fewer features than the standardized (but not coordinate-transformed) model?