HW 10 due Thursday, April 11

Find a data set appropriate for classification. In R if you type "data()" you will see a list of many data sets already loaded into R. University of California, Irvine has collected many data sets at: https://archive.ics.uci.edu. Or you can find a data set somewhere else.

On this data set use cross-validation to compare the classification methods we have gone over in class: logistic regression (if there are only two classes), linear discriminant analysis, quadratic discriminant analysis, naïve Bayes, and k-nearest neighbors (possibly combined with principal components). Also use cross-validation to find the optimal value for any parameters (e.g., the neighborhood size k for k-nearest neighbors).

Turn in an explanation of the data set you chose, a figure or table comparing the performance of the different classification methods, and a few sentences explaining what you found.