Hacker Dojo Machine Learning

Homework 1

Mike Bowles, PhD & Patricia Hoffman, PhD.

1) Use Random Forest to classify the sonar data. Use rpart to generate trees with a depth of one on randomly selected attributes. Use ridge regression to combine these trees to make predictions.

2) Use cross validated ridge regression to classify the sonar data. Create and plot a ROC curve for this classification method.

3) See if you can improve on regression-based classification of the iris data that we did in class. Classify the iris data set with second degree terms added using a ridge regression. (ie supplement the original 4 attributes x1, x2, x3, and x4 by including the 10 second degree terms (x1*x1, x1*x2, x1*x3, …) for a total of 14 attributes.) Use multiclass to classify the data and then compare the results with the results obtained in class.

4) This is a multi-class problem. Consider the Glass Identification Data Set from the UC Irvine Data Repository. The Data is located at the web site: http://archive.ics.uci.edu/ml/datasets/Glass%2BIdentification This problem will only work with building and vehicle window glass (classes 1,2 and 3), so it only uses the first 163 rows of data. (Ignore rows 164 through 214) With this set up this is a three class problem. Use ridge regression to classify this data into the three classes: building windows float processed, building windows non float processed, and vehicle windows float processed.