Hacker Dojo Machine Learning

Homework 5

Mike Bowles, PhD & Patricia Hoffman, PhD.

1) In homework 3 problem 4 you performed ridge regression on the wine quality data set : winequality-red.csv. Now use a support vector machine to classify these data.

1a) First classify the data treating the last column as an ordered factor (the wine tasters score). Next treat the last column as a numeric. Which SVM implementation is better? Why do you think it is better?

1b) Using the best version choose two attributes and a slice through the data to plot. Choose another slice. Choose a different set of attributes and another set of slices to plot.

1c) Compare and contrast the best version of the SVM with the ridge regression from homework 3.

2) Classify the sonar data set.

2a) Use a support vector machine to classify the sonar data set. First tune an SVM employing radial basis function (default). Next tune an SVM employing a linear kernel. Compare the results.

2b) In past homework, trees were used to classify the sonar data. Compare the best result using trees with the best result using SVM.

3) The in class example (svm1.r) used the glass data set. Use the Random Forest technique on the glass data. Compare the Random Forest results with the results obtained in class with SVM.

4) Classify the wine quality data using Ada Boost.

4a) Consider the classification target as a numeric and then as a factor. Compare the results and explain why one of the methods used works better than the others.

4b) Use the Random Forest technique on the wine quality data.

4c) Compare the results of Ada Boost and Random Forest.