Hacker Dojo Machine Learning

Homework 8 for Chapter 9

Mike Bowles, PhD & Patricia Hoffman, PhD

1) First use Naive Bayes to cluster the red wine data set.  Next follow the in class example (NaiveBayesMore).  Use different versions of a  covariance matrix on a normalized version of these data to classify the data set.  Was there an improvement?

2) Next use the Expectation-Maximization algorithm on the wine data set. Which method worked better for that data set?

3) First use Naive Bayes to cluster the glass data set.  Next follow the in class example (NaiveBayesMore).  Use different versions of a  covariance matrix on a normalized version of these data to classify the data set.  Was there an improvement?

4) Next use the Expectation-Maximization algorithm on the glass data set. Which method worked better for that data set?

5) Use Gaussian Discriminate Analysis to classify the sonar data set.  How does it compare with other methods we've used?

6) Challenge question: Find a method that improves upon the classification techniques used in Homework 7 for the Synthetic Control Chart Time Series Data Set.  Recall that the methods used so far do not take advantage of the fact that this is time series data.  You are free to create a method or find one in any research paper. One paper that discusses time series data written by R. J. Alcock and Y. Manolopoulous can be found here http://machinelearning2010fall.pbworks.com/w/file/32772288/TimeSeriesData10.1.1.79.1572.pdf
 The data set is here
http://archive.ics.uci.edu/ml/datasets/Synthetic+Control+Chart+Time+Series