# Hacker Dojo Machine Learning

## Homework 3

Mike Bowles, PhD & Patricia Hoffman, PhD

1) Check out the web page which describes a wine quality data set:

http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality.names

Use the Red Wine data set: winequality-red.csv  This data set contains 1599 observations of 11 attributes.

Explore both the Wine data set and the Glass data set using various visualization techniques.

2) The median score of the wine tasters is given in the last column.  (Note also that the delimiter used in this file is a semi colon and not a comma.) Remove this last column and then cluster this data set using your technique of choice.  Evaluate the results using measures discussed in class.

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

3)  Cluster the glass data set with the Expectation-Maximization algorithm. Try another clustering algorithm  on the glass data set.  Use cluster evaluations tools on these methods.  Which method worked better for that data set?

4) Challenge question: Find a method that improves upon the classification techniques used in Homework 2 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