

Hacker Dojo Machine Learning

Homework 2

Mike Bowles, PhD & Patricia Hoffman, PhD

For this homework set use the Synthetic Control Chart Time Series Data Set from the University of California at

Irvine: <http://archive.ics.uci.edu/ml/datasets/Synthetic+Control+Chart+Time+Series> As these are unsupervised methods, you must remove the target column before you invoke the clustering method. Do each of these problems on first the un-scaled data set and then for the scaled data set.

- 1) Use K-means to cluster the data set. How well did K-means work on this data set? What effect did scaling the data have on the results?
- 2) Use Divisive Hierarchical Clustering to cluster the data set. How well did this method work on this data set? What effect did scaling the data have on the results?
- 3) Now use Agglomerative Hierarchical Clustering to cluster this data set. What effect did scaling the data have on the results?
- 4) Compare the three clustering methods. Which one did the best? Which one did the worst?