

Hacker Dojo Machine Learning

Homework 7 for Chapter 8

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>

- 1) Use K-means to cluster the data. How well did K-means work on this data set?
- 2) Use Hierarchical Clustering to cluster the data. How well did this method work on this data set?
- 3) Use the Gaussian Mixture Model method to cluster the data. How well the Gaussian Mixture Model work?
- 4) Use a supervised tree method (r-part or bagged r-part) to classify these data (don't forget to regularize the data). Compare and contrast this supervised method with the unsupervised clustering methods from the previous questions.
- 5) Write a wrapper function around the r function for K-means which will implement the Bisecting K-means algorithm discussed in Chapter 8 of the text. Use bisection K-means to cluster the Iris data into 3 clusters and then into 6 clusters.