

CS4801: Principle of Machine Learning

Assignment 4

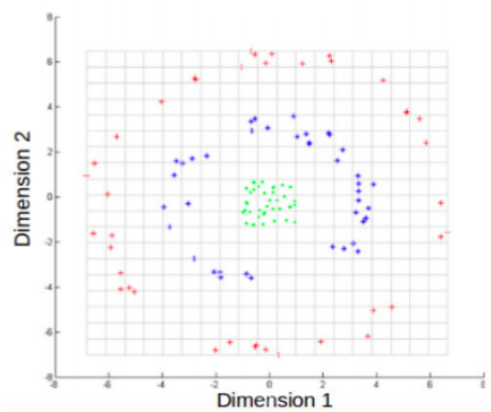
October 4, 2018

This homework consists of problems covering Clustering. A few instructions to make life easier for all of us:

- Assignment need not to be submitted.

Exercise 1: Classification

- (a) Discuss advantage and disadvantage of K-means clustering. Prove that kmeans clustering algorithm converge.
- (b) When following two clustering behave similarly and when not? [you may draw two examples in 2D]
 - K-means
 - Hierarchical Clustering when distance between two clusters measured by distance between their centred.
- (c) BIRCH also result into clusters of spherical shape. Can you change Birch algorithm so that it will result into clusters with hyper-cube shape.
- (d) (2 points each) Please go through at-least 10 data set for clustering task from UC Machine Learning Repository and Study what is feature and how the clusters should be.
- (e) Consider a 3D data-set of 3 concentric sphere, where 2D slice on any two dimensions will generate the following picture. Compare results of following clustering methods [It will be better if you actually run clustering algorithm on such data set and see what is happening]



- K-means
- Birch
- Hierarchical clattering with
 - Single Linkage
 - Complete Linkage