National University of Singapore
School of Computing

SWS3018 Predictive Analytics
Lab 7

Learning Objectives

- Perform k-means clustering on R
- Perform hierarchical clustering on R

1. In this exercise, we will work with the Breast Cancer Wisconsin dataset. It can be download from:
   Data:
   https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data
   Description of the data:
   https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.names
   a) Generate a K-means model for doing clustering

   b) Calculate the accuracy of the model generated in (a)

   c) Do you get different accuracy for different k-means model (different seed values)?

   d) Generate the hierarchical clustering models the different dissimilarity measures (complete, single, average, centroid). Show the dendrogram for each case

   e) Generate the clusters for 4 different hierarchical clustering models in (d) and measure the accuracy of the models for each case. Which linkage approach is better?

   f) Determine the classification accuracy for the decision tree, bagging, and forest approach and compare the accuracy with the above clustering approaches. How about Logistic Regression and SVM?