Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram

Class Assignment, 2019

Subject: **Machine Learning** Subject Code: **COM510**

Assignment-I: Clustering

Date: 16-01-2019

1. **K-means clustering:** Use the K-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters:

$$\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \\ A_8 \end{array} \begin{array}{c} \begin{array}{cc} X_1 & X_2 \end{array} \\ \begin{pmatrix} 2 & 10 \\ 2 & 5 \\ 8 & 4 \\ 5 & 8 \\ 7 & 5 \\ 6 & 4 \\ 1 & 2 \\ 4 & 9 \end{pmatrix} \end{array}$$

   Suppose the initial seeds (centers of each cluster) are A_1, A_4, and A_7. Run the K-means algorithm for one epoch only. At the end of this approach show:

   (a) The new clusters (i.e., the examples belonging to each cluster)

   (b) The centers of the new clusters

   (c) Draw a 10 by 10 space with all the 8 points and show the clusters after the first epoch and the new centroids.

   (d) How many more iterations are required to converge? Draw the result for each epoch.

   (e) Find the clusters and their centers using Manhattan distance measure after one epoch.

2. **Hierarchical clustering** Given a data set of five objects characterised by a single continuous feature:

	a	b	c	d	e
Feature	1	2	4	5	6

   Apply the agglomerative algorithm with complete-link and averaging cluster distance measures to produce two separate dendrogram trees.

$$\begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{pmatrix} 0 & 1 & 3 & 4 & 5 \\ 1 & 0 & 2 & 3 & 4 \\ 3 & 2 & 0 & 1 & 2 \\ 4 & 3 & 1 & 0 & 1 \\ 5 & 4 & 2 & 1 & 0 \end{pmatrix} \end{array}$$

———— x ————