CSECE 8735 Fall 2015
Unsupervised Learning

Assignment 3

due Tuesday 10/20/2015

Problem 1 Textbook 13.2 (page. 693)

Problem 2 Consider the following dissimilarity matrix

$$P = \begin{bmatrix} 0 & 4 & 3 & 11 & 12 & 5 \\ 4 & 0 & 2 & 10.5 & 13 & 6.5 \\ 3 & 2 & 0 & 8 & 9 & 6 \\ 11 & 10.5 & 8 & 0 & 1 & 10 \\ 12 & 13 & 9 & 1 & 0 & 7 \\ 5 & 6.5 & 6 & 10 & 7 & 0 \end{bmatrix}$$

According to the graph theory based agglomerative clustering algorithm with h(k) being the node degree property, determine the dendrograms for

(a) k = 2
(b) k = 3.

Note: you need to label the data samples and dissimilarity levels clearly, see the dendrogram examples in pages 674-675 of textbook).

Problem 3 Textbook Computer Experiments 13.1 (page 696)

Problem 4 On the dataset GMD that was used for HW1, perform fuzzy-c-means clustering with the number of clusters $m = 4$, the fuzzifier $q = 2$, the distance measure of $d(x_i, \theta_j) = (x_i - \theta_j)^T A(x_i - \theta_j)$ with $A = I$. Include in your report 1) the values of the cluster centroids, 2) the samples plotted in 2-D with different symbols or colors for different clusters, and 3) the total distortion in each iteration.

Problem 5 Implement the Gustafson-Kessel (G-K) algorithm by using MATLAB (see textbook p. 715 and p. 728)

(1) Use your code to carry out clustering on the dataset GKlines.dat , where the following parameter setting should be used:

Number of clusters = 2;

Fuzzifier = 2;
$l = 2$;
Representative initialization:
$c_1(0) = [-1\ 0]^T$, $c_2(0) = [0\ 1]^T$,

$$\Sigma_1(0) = \Sigma_2(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};$$

Terminate the G-K clustering at the 5th iteration, i.e., use $\theta(5)$ as the representatives for the clusters.

(2) Include the following items in your report:

a) Your code (with clear comments);

b) The estimated cluster represetatives $\theta(5)$ (mean vectors, covariance matrices).

c) Use the minimum distance rule to assign each data sample x_n to one of the two clusters $j = 1,2$ by using the parameters estimated from the first iteration

$$j^* = \arg\min_j d_{GK}^2\left(x_n, \theta_j(1)\right).$$

Plot the data samples with the class information (either use different colors for different classes, or use different symbols for different classes).

d) Use the minimum distance rule to assign each data sample x_n to one of the two clusters $j = 1,2$ by using the parameters estimated from the 5th iteration

$$j^* = \arg\min_j d_{GK}^2\left(x_n, \theta_j(5)\right).$$

Plot the data samples with the class information (either use different colors for different classes, or use different symbols for different classes).

e) For each iteration of the G-K clustering, compute the total distance

$$D(i) = \sum_{n=1}^{N} \min_j d_{GK}^2\left(x_n, \theta_j(i)\right),$$

Show the total distance $D(i)$ values for the five iterations in a table.

This assignment is complete.