CSECE 8735 Fall 2015
Unsupervised Learning

Assignment 2

due Thursday 10/1/2015

Problem 1 Textbook 11.2

Problem 2 Textbook 11.6 (note the inequality: $\left(\sum_{i=1}^{l}|x_i - y_i|\right)^2 \geq \sum_{i=1}^{l}|x_i - y_i|^2$)

Problem 3 Textbook 11.12 part a.

Problem 4 Given $x^T = [2 \quad 2]$, compute the point-set dissimilarity for the following two cases:

(a) The set is described by H: $a^T x + a_0 = 0$, with $a^T = [-1 \quad 2]$, and $a_0 = 2$.

(b) The set is described by Q: $\|z\|^2 - 1 = 0$,

Problem 5 Consider the sets $D_1 = \{x_1, x_2, x_3, x_4\}$ and $D_2 = \{y_1, y_2, y_3\}$, with $x_1 = [1, \quad 2]^T$, $x_2 = [1, \quad 3]^T$, $x_3 = [3, \quad 2]^T$, $x_4 = [2, \quad 2]^T$, $y_1 = [-1, \quad 1]^T$, $y_2 = [-2, \quad 1]^T$, $y_3 = [-1 \quad 0]^T$. The Euclidean distance is employed as the distance between two vectors. Compute $d_{\min}^{ss}(D_1, D_2)$, $d_{\max}^{ss}(D_1, D_2)$, $d_{avg}^{ss}(D_1, D_2)$, $d_{mean}^{ss}(D_1, D_2)$, $d_e^{ss}(D_1, D_2)$.
(**Note**: please use mean point in $d_{mean}^{ss}(D_1, D_2)$).

Problem 6 Textbook 12.3
(**Note:** please set the maximum number of allowed clusters q to 14).

Problem 7 On the dataset GMD that was used for HW1, perform k-means clustering with k = 4. 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.

Note: you are required to use MATLAB to complete the Problems 6 and 7.

This assignment is complete.