KEY

CSECE8735 Sample test problems Fall 2020

Name ______________ ______________ **Student ID** ______________ **Total score** ___________
Last First

(8 pts) Problem 1

(2 pts) **1.1**
Which one of the clustering algorithms below may use ring-shaped cluster representatives?

A. K-means
B. GMM
C. Fuzzy clustering
D. Spectral clustering

(2 pts) **1.2** Which one of the following methods can learn parts-based object representation:

A. PCA
B. K-means
C. NMF
D. SVD

(2 pts) **1.4.** Assume that in MDS, the eigen decomposition on $B_\Delta = -\frac{1}{2}JD^{(2)}J$ gives the following eigenvalue matrix:

$$\Lambda = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Up to how many dimensions can we keep for data sample coordinates?

A. 1
B. 2
C. 3
D. 4
E. None of the above

(2 pts) **1.4.** Assume that in MDS, the eigen decomposition on $B_\Delta = -\frac{1}{2}JD^{(2)}J$ gives the following eigenvalue matrix:

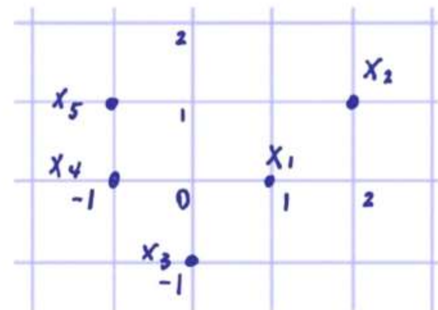
$$\Lambda = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Is the distance matrix D a Euclidean distance matrix?

A. No
B. Yes
C. None of the above

(21 pts) Problem 2

2.1 Referring to the figure on the left.
Compute $d_1(x_1, x_5)$, $d_\infty(x_1, x_5)$, $S_{\cos ine}(x_2, x_3)$.



$|x_{11} - x_{51}| + |x_{12} - x_{52}|$

$= 2 + 1 = 3$

(4 pts) $d_1(x_1, x_5) =$ 3

$\max\{|x_{11} - x_{51}|, |x_{12} - x_{52}|\}$

$= 2$

(4 pts) $d_\infty(x_1, x_5) =$ 2

$x_2^T x_3 = [2 \quad 1]\begin{bmatrix} 0 \\ -1 \end{bmatrix} = -1$ $\quad \|x_2\| = \sqrt{5}, \quad \|x_3\| = 1$

(4 pts) $S_{\cos ine}(x_2, x_3) =$ $\frac{-1}{\sqrt{5} \times 1} = -1/\sqrt{5}$

2.2 Assume that the five data samples are partitioned into two sets as $C_1 = \{x_1, x_2\}$ and $C_2 = \{x_3, x_4, x_5\}$. Compute $d^{ps}_{min}(x_2, C_2)$, and $d^{ss}_{max}(C_1, C_2)$ by using $d_2(x_i, x_j)$ for sample-pair distance.

$\sqrt{2^2 + 2^2} = \sqrt{8}$

(4 pts) $d^{ps}_{min}(x_2, C_2) =$ $\sqrt{8}$

$\sqrt{1 + 3^2} = \sqrt{10}$

(5 pts) $d^{ss}_{max}(C_1, C_2)$ $= \sqrt{10}$

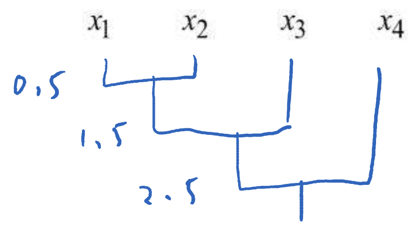
(24 pts) Problem 3 Consider the following dissimilarity matrix P.

(a) (10 pts) Use the complete-linkage clustering algorithm to generate the dendrogram.

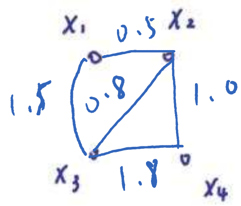
$$P = \begin{bmatrix} 0 & 0.5 & 1.5 & 2.5 \\ 0.5 & 0 & 0.8 & 1.0 \\ 1.5 & 0.8 & 0 & 1.8 \\ 2.5 & 1.0 & 1.8 & 0 \end{bmatrix}$$

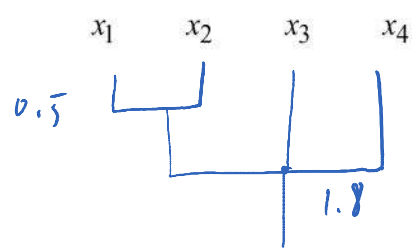
$$\begin{bmatrix} 0 & 1.5 & 2.5 \\ 1.5 & 0 & 1.8 \\ 2.5 & 1.8 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 2.5 \\ 2.5 & 0 \end{bmatrix}$$



(b) (10 pts) Use the graph theory based agglomerative clustering algorithm with h(k) the node degree of k = 2 to generate the dendrogram.





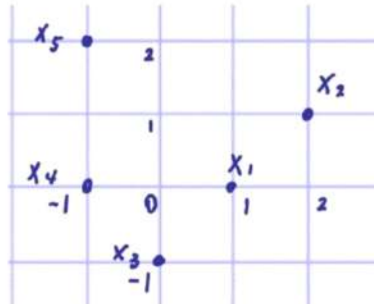
(c) (4 pts) What is the best choice of clustering in the criterion of "the longest survival time" based on the dendrogram you've derived in (b)?

Clustering of (b) $\{X_1\ X_2\}\ \{X_3\}\ \{X_4\}$

(27 pts) Problem 4

4.1. Five samples are shown in the figure:

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & -1 & -1 \\ 0 & 1 & -1 & 0 & 2 \end{bmatrix},$$



from which the EM algorithm is used to estimate parameters of a Gaussian mixture density (GMD) with two Gaussian components.

Assume that at one step of Expectation, the posterior probabilities of the Gaussian sources are obtained as:

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
$P(z_t=1 \mid x_t)$	0.8	1.0	0.4	0.2	0.0
$P(z_t=2 \mid x_t)$	0.2	0.0	0.6	0.8	1.0

(a) (10 pts) For the subsequent Maximization step, compute the mixture weights of the two Gaussian sources, π_1^{new} and π_2^{new}, and the mean vector μ_2^{new} (show your calculations clearly).

$$\pi_1^{now} = \frac{1}{5}(0.8 + 1.0 + 0.4 + 0.2 + 0.0) = \frac{2.4}{5} = 0.48$$

$$\pi_2^{now} = 0.52$$

$$N_2 = 2.6, \quad \mu_2 = \frac{1}{N_2}\left(0.2\begin{bmatrix}1\\0\end{bmatrix} + 0.6\begin{bmatrix}0\\-1\end{bmatrix} + 0.8\begin{bmatrix}-1\\0\end{bmatrix} + \begin{bmatrix}-1\\2\end{bmatrix}\right)$$

$$= \frac{1}{2.6}\begin{bmatrix}-1.6\\1.4\end{bmatrix} = \frac{1}{13}\begin{bmatrix}-8\\7\end{bmatrix}$$

$\pi_1^{new} =$ 0.48 $\qquad \pi_2^{new} =$ 0.52

$\mu_2^{new} = \frac{1}{13}\begin{bmatrix}-8\\7\end{bmatrix}$

(b) (5 pts) Cluster the data samples to the Gaussian sources (C1 or C2) according to the maximum posterior probability criterion.

Data samples	x_1	x_2	x_3	x_4	x_5
Gaussian sources	C_1	C_1	C_2	C_2	C_2

4.2 Refer to the same figure in 4.1 but perform k-means clustering on the five data samples instead. Assume that at the current iteration the cluster assignment is

$$C_1 = \{x_1, x_2, x_3\} \quad C_2 = \{x_4, x_5\} \ .$$

(c) (6 pts) Compute the cluster centroids μ_1 and μ_2:

$$\mu_1 = \frac{1}{3}\left(\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}2\\1\end{bmatrix} + \begin{bmatrix}0\\-1\end{bmatrix}\right) = \begin{bmatrix}1\\0\end{bmatrix}$$

$$\mu_2 = \frac{1}{2}\left(\begin{bmatrix}-1\\0\end{bmatrix} + \begin{bmatrix}-1\\2\end{bmatrix}\right) = \begin{bmatrix}-1\\1\end{bmatrix}$$

(d) (6 pts) Compute the total distortion J based on the current cluster assignment and the cluster centroids.

$$J = \left\|\begin{bmatrix}1-1\\0-0\end{bmatrix}\right\|^2 + \left\|\begin{matrix}2-1\\1-0\end{matrix}\right\|^2 + \left\|\begin{matrix}0-1\\-1-0\end{matrix}\right\|^2 + \left\|\begin{matrix}-1+1\\0-1\end{matrix}\right\|^2 + \left\|\begin{matrix}-1+1\\2-1\end{matrix}\right\|^2$$

$$= 0 + 2 + 2 + 1 + 1$$

$$= 6$$

2

(20 pts) Problem 5

Assume that for latent semantic analysis a word-document matrix W has been decomposed by SVD as $W = USV^T$. Let

$$S = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}, \quad V = \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} \\ v_{21} & v_{22} & v_{23} & v_{24} \\ v_{31} & v_{32} & v_{33} & v_{34} \\ v_{41} & v_{42} & v_{43} & v_{44} \end{bmatrix}$$

and U be represented by its column vectors as $U = [u_1\ u_2\ u_3\ u_4]$ with its size being 10x4.

(8 pts) (a) Consider a rank-1 approximation for W. Determine the approximated document-pair distance $\|d_1 - d_3\|_2^2$ in terms of the relevant elements of V:

$$W \approx u_1 \times 10 \times [V_{11}\ \ V_{21}\ \ V_{31}\ \ V_{41}]$$

$$= u_1 \times [10V_{11}\ \ 10V_{21}\ \ 10V_{31}\ \ 10V_{41}]$$

$$\|d_1 - d_3\|^2 \approx (10V_{11} - 10V_{31})^2 = 100(V_{11} - V_{31})^2$$

The approximation error $\|W - \hat{W}\|_F^2 =$ $5^2 + 1^2 + 0.1^2 = 26.01$

(12 pts) (b) Consider a rank-2 approximation for W. Determine the approximated document-pair distance $\|d_1 - d_3\|_2^2$ in terms of the relevant elements of V:

$$W \approx [u_1\ \ u_2] \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} V_{11} & V_{21} & V_{31} & V_{41} \\ V_{12} & V_{22} & V_{32} & V_{42} \end{bmatrix}$$

$$= [u_1\ \ u_2] \begin{bmatrix} 10V_{11} & 10V_{21} & 10V_{31} & 10V_{41} \\ 5V_{12} & 5V_{22} & 5V_{32} & 5V_{42} \end{bmatrix}$$

$$\|d_1 - d_3\|^2 \approx 100(V_{11} - V_{31})^2 + 25(V_{12} - V_{32})^2$$

The approximation error $\|W - \hat{W}\|_F^2 =$ $1^2 + 0.1^2 = 1.01$