

1. (4 pts) True or false

- (a) The results of a cluster analysis will likely change if the variables are first standardized. **T** or **F**
- (b) A factor analysis model can be fit using principal components. **T** or **F**
- (c) The specific variance in a factor analysis model is changed when the factors are rotated. **T** or **F**
- (d) The variance-covariance for a multivariate normal distribution has an elliptical or spherical shape. **T** or **F**

2. (4 pts) For the following data, we will conduct a  $k$ -means cluster analysis using Euclidean distance with

$k = 2$ . The two seed means are  $\bar{\mathbf{X}}_1^0 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ ,  $\bar{\mathbf{X}}_2^0 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ .

Obs Num	Var 1	Var 2	Dist from $\bar{\mathbf{X}}_1^0$	Dist from $\bar{\mathbf{X}}_2^0$	Cluster id
1	1	2	2.23		
2	2	1	3.00		
3	5	3	3.16	1.41	2
4	6	7		5.39	
5	6	5		3.61	

- (a) (1) Now using Euclidean distance compute the distances from each point to each mean, and fill in the blanks of the relevant columns of the table above.
- (b) (1) Assign the 5 points into one of the two clusters, and write this into the table above.
- (c) (2) Compute the two means that will seed the next iteration.

$$\bar{\mathbf{X}}_1^1 = \quad \quad \quad \bar{\mathbf{X}}_2^1 =$$

3. (2 pts) The following matrix,  $\mathbf{D}$ , represents shows the interpoint distances between 5 points, A-E.

$$\mathbf{D} = \begin{bmatrix} & A & B & C & D & E \\ A & 0 & 1 & 2 & 3 & 3 \\ B & 1 & 0 & 3 & 4 & 5 \\ C & 2 & 3 & 0 & 2 & 3 \\ D & 3 & 4 & 2 & 0 & 2 \\ E & 3 & 5 & 3 & 2 & 0 \end{bmatrix}$$

- (a) Points A and B would be joined at the first step of hierarchical clustering. Compute the intercluster distance between cluster AB and point C, using complete linkage. \_\_\_\_
- (b) What would be the height of the dendrogram where A and B are joined? \_\_\_\_