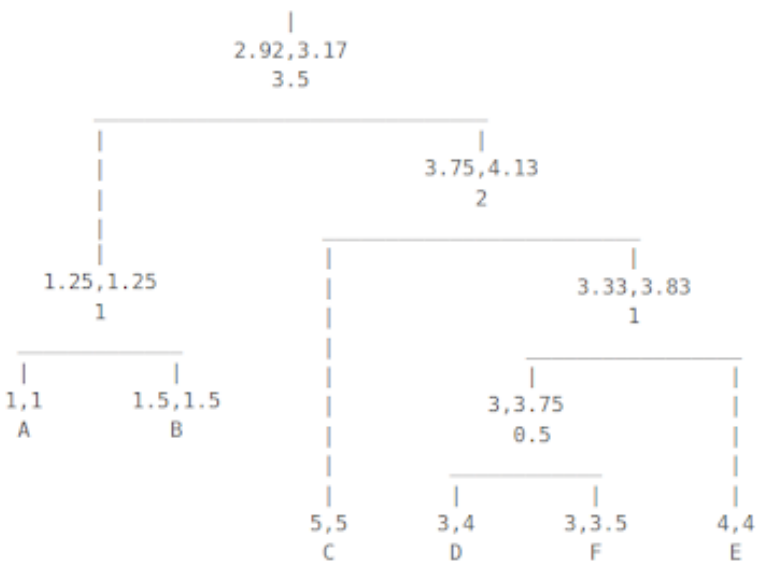
SML (CSE342/ ECE356) Quiz 1

Each question carries 2 marks.
Duration: 1 hr

Q1. Perform hierarchical clustering on the given data:

A	(1, 1)
B	(1.5, 1.5)
C	(5, 5)
D	(3, 4)
E	(4,4)
F	(3, 3.5)

Use Manhattan distance as the distance metric. Draw the dendrogram.



Q2. Find out which of the two clustering results is better using Silhouette analysis

	X	Cluster Ids	Silhouette Score
A	(4, 4)	1	0.2701
B	(0, 6)	0	0.3591
C	(-2, 4)	0	0.1464
D	(2, 4)	1	0.1796
E	(0, 2)	1	0.3095
F	(2, 6)	0	-0.04199

	X	Cluster Ids	Silhouette Score
A	(4, 4)	0	0.5154
B	(0, 6)	1	-0.0919
C	(-2, 4)	1	0.4137
D	(2, 4)	0	0.3787
E	(0, 2)	1	0.13
F	(2, 6)	0	0.3382

Avg: 0.20378 0.28068
Seconding clustering is better.

Q3. Perform Fuzzy c-means clustering on the following data:

A	(4, 4)
B	(0, 6)
C	(-2, 4)
D	(2, 4)
E	(0, 2)
F	(2, 6)

In the first iteration, assume fuzzy centroids are (-2,6) and (2,2). Compute fuzzy centroids and membership matrix in the third iteration. Assume m=2.

Iteration: 1
Cluster Centers: [[-2, 6], [2, 2]]
Membership Matrix: [[0.167, 0.833], [0.833, 0.167], [0.833, 0.167], [0.167, 0.833], [0.167, 0.833], [0.5, 0.5]]
Iteration: 2
Cluster Centers: [[-0.419, 5.064], [1.930, 3.651]]
Membership Matrix: [[0.176, 0.824], [0.898, 0.102], [0.811, 0.189], [0.018, 0.982], [0.403, 0.597], [0.451, 0.549]]
Iteration: 3
Cluster Centers: [[-0.421, 4.910], [2.205, 3.962]]
Membership Matrix: [[0.137, 0.863], [0.869, 0.131], [0.842, 0.158], [0.006, 0.994], [0.502, 0.498], [0.373, 0.627]]

Q4. Perform PCA on the following data.

X_1	1	3	0
X_2	2	4	6

where X1 and X2 are features.

Standardize:

X1	X2
-0.2182	-1
1.0911	0
-0.8729	1

The covariance matrix:

1	-0.3273
-0.3273	1

Eigenvectors:

0.7071	0.7071
-0.7071	0.7071

Parameter	PC_1	PC_2
Eigenvalue	1.3273	0.6727

Final projections:

PC_1	PC_2
0.5528	-0.8614
0.7715	0.7715
-1.3243	0.08989

Q5. Explain the following with a suitable diagram:

1. Supervised Learning
2. Unsupervised Learning

Refer to lecture slides.

Quiz-1 Solution

A1

	A	B	C	D	E	F
A	0	1	8	5	6	5.5
B	1	0	7	4	5	3.5
C	8	7	0	3	2	3.5
D	5	4	3	0	1	0.5
E	6	5	2	1	0	1.5
F	5.5	3.5	3.5	0.5	1.5	0

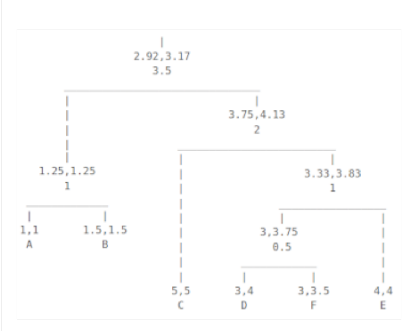
Merge D, F → 3, 3.75

$d_{DF \to B} = \min(d_{D-}, d_{F-})$

	A	B	C	D,F	E
A	0	1	8	4.5	6
B	1	0	7	3.5	5
C	8	7	0	3	2
D,F	4.5	3.5	3	0	1
E	6	5	2	1	0

Merge D,F & E, Merge AB.

Similarly, solving further we will get the following dendogram.



A2

	A	B	C	D	E	F
A	0	$2\sqrt{5}$	6	2	$2\sqrt{5}$	$2\sqrt{10}$
B	$2\sqrt{5}$	0	$2\sqrt{2}$	$2\sqrt{2}$	4	2
C	6	$2\sqrt{2}$	0	4	$2\sqrt{2}$	$2\sqrt{5}$
D	2	$2\sqrt{2}$	4	0	$2\sqrt{2}$	2
E	$2\sqrt{5}$	4	$2\sqrt{2}$	$2\sqrt{2}$	0	$2\sqrt{5}$
F	$2\sqrt{2}$	2	$2\sqrt{5}$	2	$2\sqrt{5}$	0

$$a(A) = \frac{1}{2}\left(d(A,D) + d(A,E)\right)$$

$$= \frac{1}{2}\left(2 + 2\sqrt{5}\right)$$

$$= 1 + \sqrt{5} \approx 3.23606$$

$$b(A) = \frac{1}{3}\left(d(A,B) + d(A,C) + d(A,F)\right)$$

$$= \frac{1}{3}\left(2\sqrt{5} + 6 + 2\sqrt{2}\right)$$

$$= \frac{2}{3}\left(\sqrt{5} + 3 + \sqrt{2}\right)$$

$$\approx 4.43352$$

$$s(A) = \frac{b(A) - a(A)}{\max(a(A), b(A))}$$

$$\approx 0.2701$$

Similarly for other points