INTRODUCTION TO ARTIFICIAL INTELLIGENCE
Lab 1 – Clustering

1) **K-MEANS - ALGORITHM:** Given is the following data set. Assume that K = 2. Initially, use objects D and F as the centroids. Then, assign objects to different groups. To compute the similarity, use Euclidean distance. However, in the 1st iteration, no calculations need to be performed (use the plot to compare distances between objects).

ID	X	Y
A	0.1	0.6
B	0.2	0.9
C	0.2	0.4
D	0.7	0.5
E	0.8	0.1
F	0.9	0.3

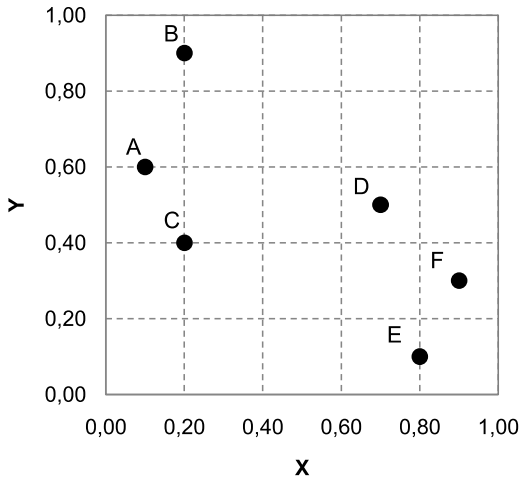
1st ITERATION
Centroids:
C1 = [ ]
C2 = [ ]

Group assignment:
G1 = [ ]
G2 = [ ]

2nd ITERATION
Centroids:
C1 = [X=
Y=
C2 = [X=
Y=

Mark C1 and C2 on the plot. Where D should be assigned to?
dist(D, C1) =
dist(D, C2) =

Group assignment:
G1 = [ ]
G1 = [ ]
Is it necessary to perform 3rd iteration?



2) **K-MEANS - NORMALIZATION:** Given is the following data set. Normalize the data using min-max normalization. Then, depict the points on the plot. How many groups (K) do you see?

Input:

ID	X	Y
A	30	20
B	10	80
C	40	60
D	30	120
E	80	60
F	110	120
G	100	140
H	90	100
I	60	180
J	70	220

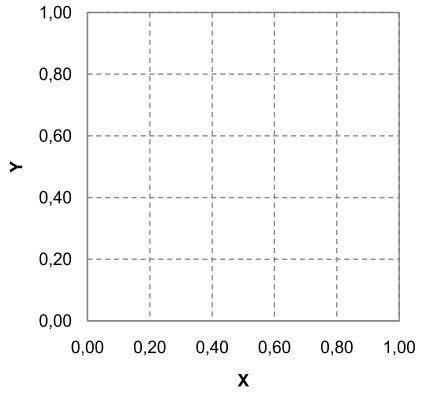
Min and Max:

	Min	Max
X		
Y		

Normalized:

ID	X	Y
A		
B		
C		
D		
E		
F		
G		
H		
I		
J		

The best K =



3) **K-MEANS – THE BEST K:** Consider the data set given in (2). Assess the quality of final clusters for different K. For this reason, compute a total distance **(TD)** between objects within groups and the corresponding centroids. Mark the results on the plot. Decide which K gives the best results (find the "elbow"). Distances between objects and corresponding centroids are given in tables.

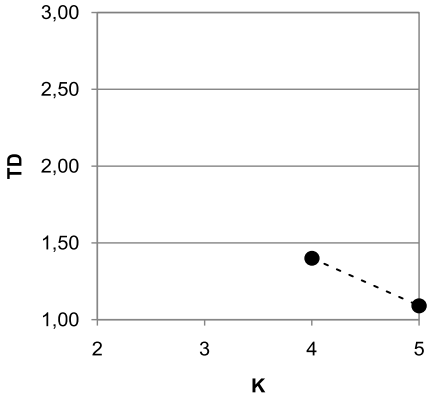
K=2	C1	C2
A	0.25	-
B	0.18	-
C	0.14	-
D	0.25	-
E	-	0.39
F	-	0.26
G	-	0.15
H	-	0.19
I	-	0.33
J	-	0.45

K=3	C1	C2	C3
A	0.25	-	-
B	0.18	-	-
C	0.14	-	-
D	0.25	-	-
E	-	0.27	-
F	-	0.16	-
G	-	0.18	-
H	-	0.05	-
I	-	-	0.11
J	-	-	0.11

TD(K=2) = .

TD(K=3) =

The best K =



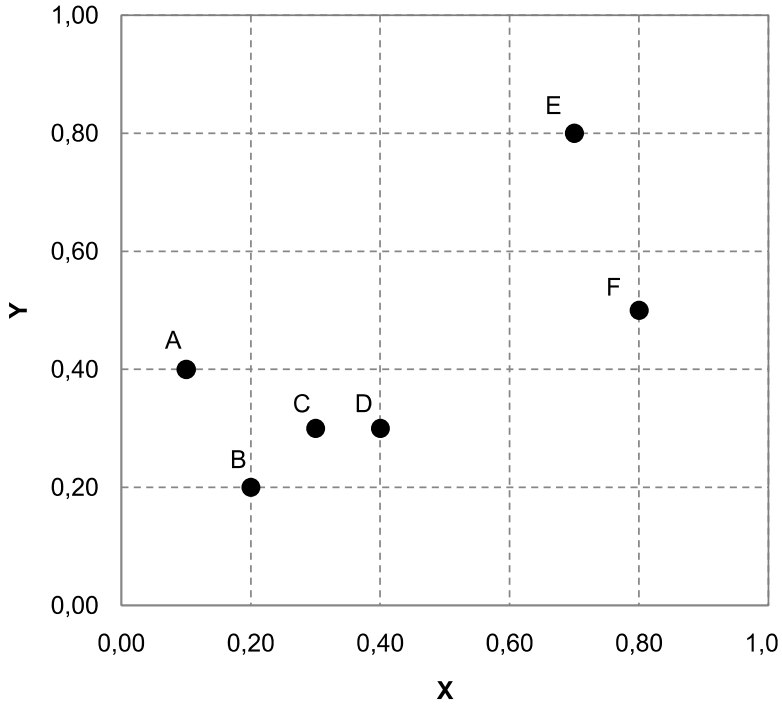
4) **AHC ALGORITHM:** Given is the following data set and the corresponding distance matrix. Use AHC algorithm to group the data. Consider a **complete-linkage** approach (a distance between clusters = a distance between two objects – one in each cluster – being the farthest away from each other). In case of ambiguity, merge two candidate clusters randomly. During each step (a) update the dendrogram and (b) illustrate groups in the plot. Where do you think the dendrogram should be cut?

Objects:

ID	A	B	C	D	E	F
X	0.1	0.2	0.3	0.4	0.8	0.7
Y	0.4	0.2	0.3	0.3	0.5	0.8

Distance matrix:

	A	B	C	D	E	F
A	-	0.22	0.22	0.31	0.71	0.72
B	-	-	0.14	0.22	0.67	0.78
C	-	-	-	0.10	0.54	0.64
D	-	-	-	-	0.45	0.58
E	-	-	-	-	-	0.31
F	-	-	-	-	-	-



Some calculations:

DENDROGRAM:

A

B

C

D

E

F