

Machine Learning-hw6  
Jaskaran Singh(2011cs1012)

Ans1

a)

Distance Matrix

	A	B	C	D
A	0	0.3	0.4	0.7
B	0.3	0	0.5	0.8
C	0.4	0.5	0	0.45
D	0.7	0.8	0.45	0

Complete Linkage(Max distanced clusters are combined)

Step 0	Step1	Step2	Step3
A	A	ABD	ABCD
B	C	C	
C	BD		
D			

Step0-B and D have max distance of 0.8.Combined to give cluster BD.

Step1-A and BD have distance of  $\max(d(A,B),d(A,D))=0.7$ .Combined to give cluster ABD

Step2-ABD and C combine to give cluster ABCD.

b)

Single Linkage(Min distanced clusters are combined)

Step 0	Step1	Step2	Step3
A	AB	ABC	ABCD
B	C	D	
C	D		
D			

Step0-A and B have min distance of 0.3.Combined to give cluster AB.

Step1-AB and C have distance of  $\min(d(A,C),d(B,C))=0.4$ .Combined to give cluster ABC

Step2-ABC and D combine to give cluster ABCD.

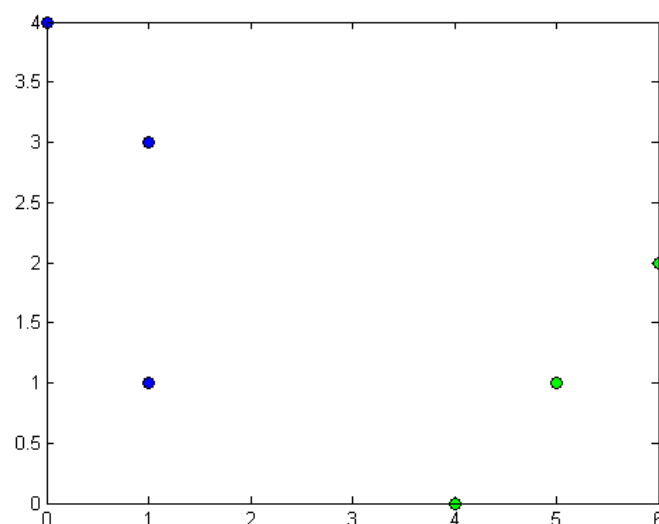
c)

2 clusters in complete linkage are ABD and C.

d)

2 clusters in single linkage are ABC and D.

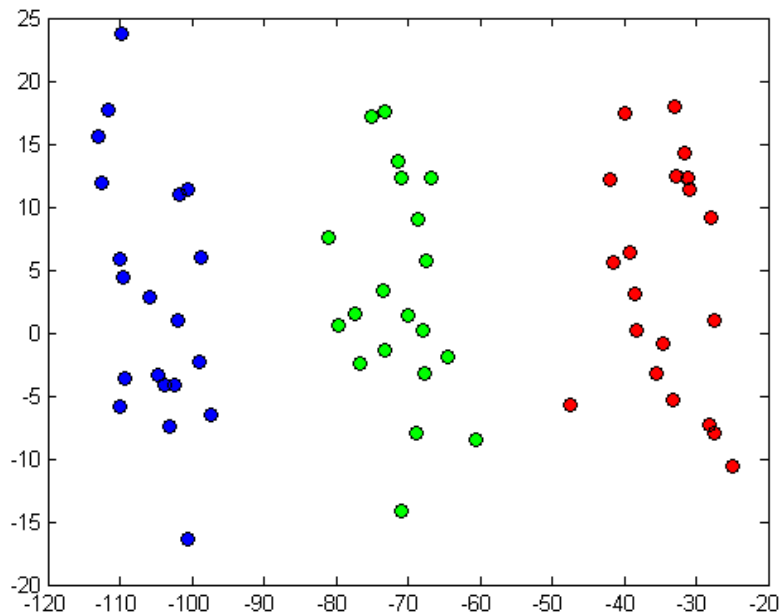
Ans2



Cluster of points after running the K Means algorithm.

For step by step run the Ans2.m file.

Ans3



**Red class-Label 1(data points 1-20)**

**Green class-Label 2(data points 21-40)**

**Blue class-Label 3(data points 41-60)**

**3 classes of data projected in 2 dimensions using PCA.** Each is normally distributed with a certain mean and variance in the original dimensions.

**c)K=3 KMeans clustering on original data(50 dim)**

The original labels and cluster labels differ. But the points clustered are the same, just the label of points has changed due to random initialization of labels in KMeans.

Points 1-20->label 1

Points 21-40->label 3

Points 41-60->label 2

**d)K=2 KMeans clustering on original data(50 dim)**

The points in the blue and green clusters combine(21-60) to form a single cluster. The red cluster(1-20) remains the same.

**e)K=4 KMeans clustering on original data(50 dim)**

Cluster 4- data points 1 to 20

Cluster 1- data points 21,22,23,24,29,32,33,35,37,39,40

Cluster 2-data points 25,26,27,28,30,31,34,36,38

Cluster 3-data points 41-60

**f) K=3 KMeans clustering on projected data(2 dim)**

No mismatches between the original and the clustered labels. This is because they're already clustered as can be seen from the plot.

**g)K=3 KMeans clustering on original data after scaling each feature to standard deviation 1.**

No mismatches between the original and the clustered labels.Only the label names have changed.

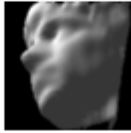
Points 1-20->label 2

Points 21-40->label 3

Points 41-60->label 1

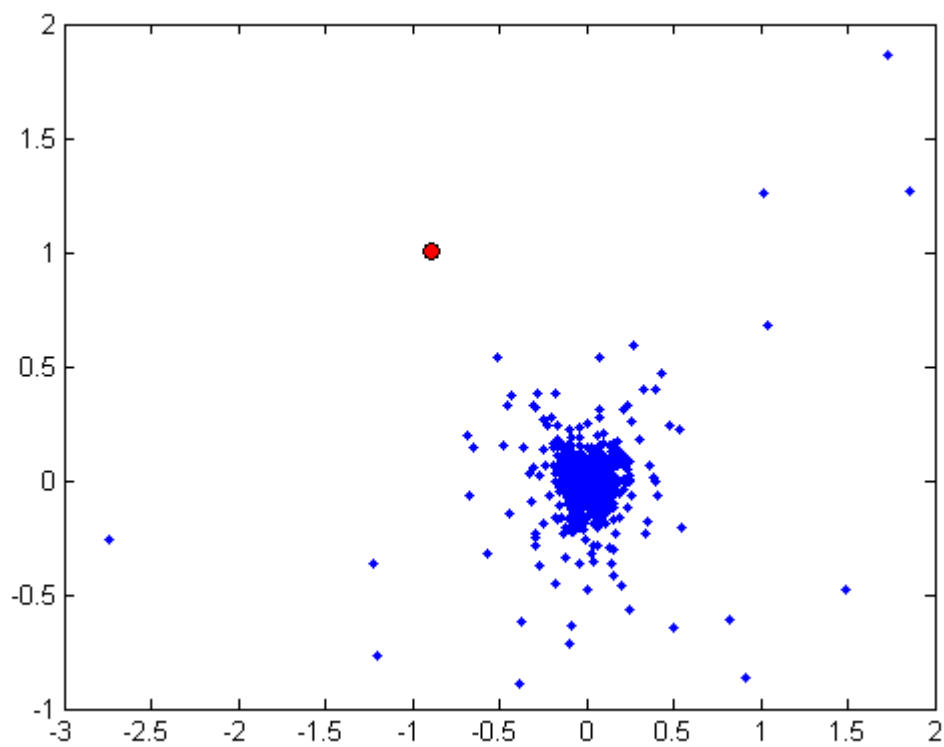
Ans4

For the image:

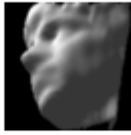


Isomap( $K=6$ ) with Euclidean distance

(Red point is the corresponding image)

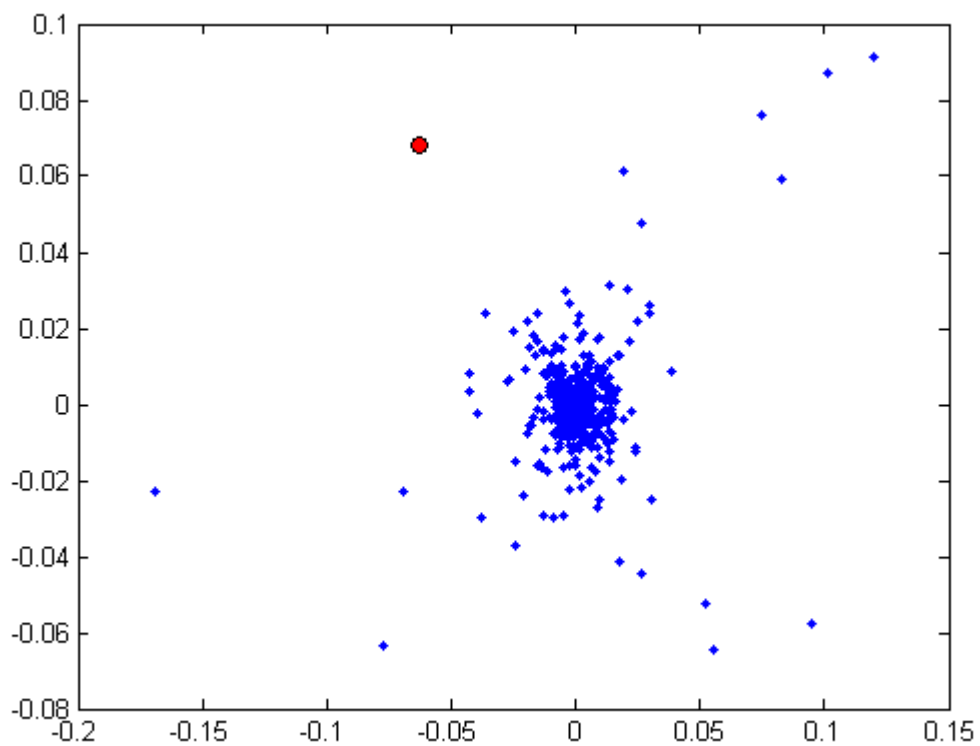


For the image:



Isomap( $K=6$ ) with cosine distance

(Red point is the corresponding image)



The cosine and the Euclidean embedding are quite similar in their appearance and show very good results.