Ahsanullah University of Science and Technology (AUST)
Department of Computer Science and Engineering

LABORATORY MANUAL

Course No. : CSE 4214
Course Title: Pattern Recognition Lab

For the students of 4th Year, 2nd semester of
B.Sc. in Computer Science and Engineering program

Table of Contents

COURSE OUTCOMES

1. Implement a variety of pattern recognition algorithms and their application
2. Explain traditional statistical techniques and their relationship to Machine Learning algorithms, particularly for the pattern recognition
3. Apply each algorithm for dataset analysis including
    - Appropriate data conditioning
    - Common failure modes and their detection
    - Scope of application of the algorithm results
4. Apply engineering solutions based on pattern recognition professionally in societal and environmental contexts
5. Communicate with team members to make a sustainable fruitful project

PREFFERED TOOL(S)

- Python
- Matlab

TEXT/REFERENCE BOOK(S)

- Pattern Classification (2nd Edition) by R. O. Duda, P. E. Hart and D. Stork, Wiley 2002.
- Matlab: A Practical Approach by Stormy Attaway, Elesevier, Canada, 2009.

ADMINISTRATIVE POLICY OF THE LABORATORY

Students will have to carry out their experiments individually. Each week a task will be given where every student will be asked to implement and discuss on their own assignments. Therefore, there will be two sorts of jobs to complete.

i. Lab Implementation.
ii. Report Writing.

In the implementation part, students have to implement the given algorithm with suitable data. Whereas, in report writing, they have to write a formal report on the given task along with finding and discussion sections that must highlight the working principles and effects of changing different parameters or data set. Report should be written in LaTex according to the structure of a technical report following any standard (IEEE conference paper) format. In general each lab report may contain the following sections:

i. Objective of the experiment
ii. Introduction
iii. Implementation
iv. Result analysis
v. Conclusion and future work.

Session 1

OBJECTIVE: Designing a minimum distance to class mean classifier

TASKS:

Given the following two-class set of prototypes

> **ω_1 = {(2,2), (3,1), (3,3), (-1,-3), (4,2), (-2,-2)}**
> **ω_2 = {(0,0), (-2,2), (-1,-1), (-4,2), (-4,3), (2,6)}**

1. Plot all sample points from both classes, but samples from the same class should have the same color and marker.
2. Using a minimum distance classifier with respect to 'class mean', classify the following points by plotting them with the designated class-color but different marker.

X_1 = (-1,-1)
X_2 = (3, 2)
X_3 = (-2, 1)
X_4 = (8, 2)

Linear Discriminant Function: $g_i(X) = X^T {}_i\overline{Y} - \frac{1}{2} {}_i\overline{Y}^T {}_i\overline{Y}$

3. Draw the decision boundary between the two-classes.
4. Write the report and explain the experimental results for different dataset containing varying training and test points.

Note: Useful Matlab functions: plot, mean, sqrt, hold.

Session 2

OBJECTIVE: Implementing the Perceptron algorithm for finding the weights of a Linear Discriminant function.

TASKS:
You are given the following sample points in a 2-class problem:

ω_1 = (1,1), (1,-1), (4,5)

ω_2 = (2,2.5), (0,2), (2,3)

1. Plot all sample points from both classes, but samples from the same class should have the same color and marker. Observer if these two classes can be separated with a linear boundary.
2. Consider the case of a second order polynomial discriminant function. Repose the problem of finding such a nonlinear discriminant function, as a problem of finding a linear discriminant function for a set of sample points of higher dimension. Generate the high dimensional sample points. [Hint: Φ-machine.]
3. Use Perceptron Algorithm (one at a time) for finding the weight-coefficients of the discriminant function boundary for your linear classifier in question 2.

$$\underline{\mathbf{w}}(\mathbf{i+1}) = \underline{\mathbf{w}}(\mathbf{i}) + \alpha\, \underline{\tilde{y}}_m^{(k)} \qquad \text{if } \underline{\mathbf{w}}^{\mathbf{T}}(\mathbf{i})\ \underline{\tilde{y}}_m^{(k)} \leq 0$$

$$\text{(i.e., if } \underline{\tilde{y}}_m^{(k)} \text{ is misclassified)}$$

$$= \underline{\mathbf{w}}(\mathbf{i}) \qquad \text{if } \underline{\mathbf{w}}^{\mathbf{T}}(\mathbf{i})\ \underline{\tilde{y}}_m^{(k)} > 0$$

4. Draw the decision boundary between the two classes.
5. Write the report and show the experimental results explaining the effect of phi-function, learning rate, and varying datasets.

Note: Useful MATLAB Functions: **sym, solve, subs**

```
% Plotting the discriminant function and the prototypes.
syms x1 x2;
s=sym(10*x1*x1-6*x2*x2+24*x1*x2-24*x1-68*x2+65);
s2=solve(s,x2);
xvals1=[-10:0.01:10];
xvals2(1,:)=subs(s2(1),x1,xvals1);
plot(xvals1,xvals2(1,:),'k');
grid, hold,
```

```
%Class S1
plot(1,1,'ro');
plot(1, -1,'ro');
plot(4,5,'ro');
%Class S2
plot(2,2,'gs');
plot(0,2,'gs');
plot(2,3,'gs');
```

Session 3

OBJECTIVE: Classify a set of unknown samples using minimum error rate classifier for a two-class problem.

TASKS:

Design a Minimum Error Rate Classifier for a two-class problem with the given data below:

$P(x|\omega_1)=N(\mu_1, \Sigma_1)$
where, $\mu_1=[0\ 0]$
and $\Sigma_1= [.25\ .3;\ .3\ 1];$
$P(\omega_1) = 0.5$

$P(x|\omega_2)= N(\mu_2, \Sigma_2)$
where, $\mu_2=[2\ 2]$
and $\Sigma_2= [.5\ 0;\ 0\ .5];$
$P(\omega_2) = 0.5$

1. Classify the following unknown samples: (1,1), (1,-1), (4,5) (-2,2.5), (0,2), (2,-3)
2. Classified samples should have different colored markers according to the assigned class label.
3. Change the parameter values (μ_i, Σ_i) to see different recognition results.
4. Label the regions R_i for each of those two classes.
5. Submit your sessional report with necessary illustrations and analysis.

Note: Useful MATLAB Code:

```
x1 = -7:.2:7; x2 = -7:.2:7;
[X1,X2] = meshgrid(x1,x2);

mu1 = [0 0];
Sigma1 = [.25 .3; .3 1];
F1 = mvnpdf([X1(:) X2(:)],mu1,Sigma1);
F1 = reshape(F1,length(x2),length(x1));
%surfc(x1,x2,F1);
meshc(X1,X2,F1);
axis([-7 7 -7 7 -1.0 .6])
xlabel('x1'); ylabel('x2'); zlabel('Probability Density');

hold on;

mu2 = [2 2];
Sigma2 = [.5 .0; 0 .5];
F2 = mvnpdf([X1(:) X2(:)],mu2,Sigma2);
F2 = reshape(F2,length(x2),length(x1));
%surfc(x1,x2,F2);
meshc(X1,X2,F2);
axis([-7 7 -7 7 -1.0 .6])
```

```
xlabel('x1'); ylabel('x2'); zlabel('Probability Density');

caxis([min(F2(:))-.5*range(F2(:)),max(F2(:))]);

%% Write Your CODE here

plot3(1,1,-1.0,'rx');
```

Session 4

OBJECTIVE: Implement the basic k-means clustering algorithm applying Euclidean distance for a given dataset. Measure the clustering performance by generating confusion matrix.

TASKS:

Implement the basic k-means clustering algorithm applying Euclidean distance as a distance measurement formula. Do the followings:

1. An example IRIS dataset is given containing 4 dimensional feature points.
2. Load the IRIS dataset in Matlab
3. Perform the k-means clustering on IRIS dataset to generate 3 clusters.
4. Draw the clusters and corresponding points with different colors
5. Prepare the report and show experimental results using confusion matrix

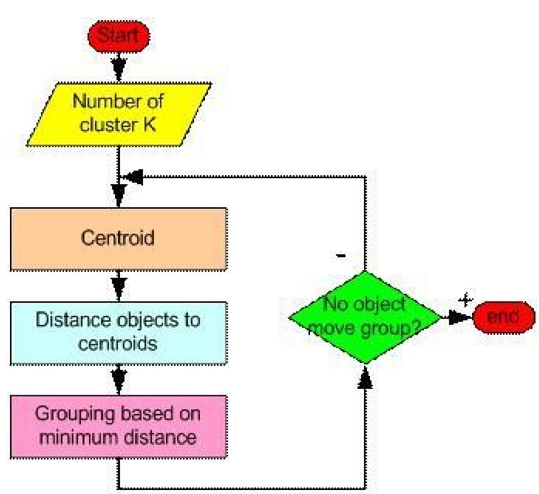
Algorithm Description:

The basic step of k-means clustering is simple. In the beginning we determine number of cluster K and we assume the centroid or center of these clusters. We can take any random objects as the initial centroids or the first K objects in sequence can also serve as the initial centroids.

Then the K means algorithm will do the three steps below until convergence

Iterate until stable (= no object move group):

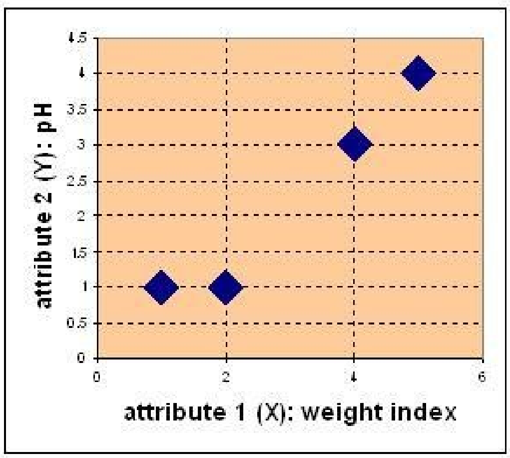
1. Determine the centroid coordinate
2. Determine the distance of each object to the centroids
3. Group the object based on minimum distance



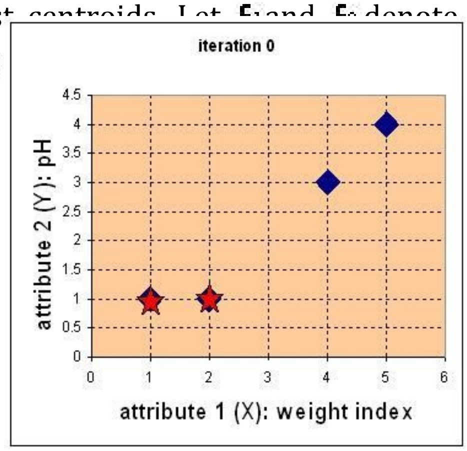
Suppose we have several objects (4 types of medicines) and each object have two attributes or features as shown in table below. Our goal is to group these objects into K=2 group of medicine based on the two features (pH and weight index).

Object	attribute 1 (X): weight index	attribute 2 (Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Each medicine represents one point with two attributes (X, Y) that we can represent it as coordinate in an attribute space as shown in the figure below.



1. **Initial value of centroids** : Suppose we use medicine A and medicine B as the first centroids. Let c_1 and c_2 denote the coordinate of the centroids, then c_1



2. **Objects-Centroids distance:** we calculate the distance between cluster centroid to each object. Let us use Euclidean distance, then we have distance matrix at iteration 0 is

$$\mathbf{D}^0 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 1 & 0 & 2.83 & 4.24 \end{bmatrix} \quad \begin{matrix} \mathbf{c}_1 = (1,1) & group-1 \\ \mathbf{c}_2 = (2,1) & group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \quad \begin{matrix} X \\ Y \end{matrix}$$

Each column in the distance matrix symbolizes the object. The first row of the distance matrix corresponds to the distance of each object to the first centroid and the second row is the distance of each object to the second centroid. For example, distance from medicine C = (4, 3) to the first centroid $\mathbf{c}_1 = (1,1)$ is $\sqrt{(4-1)^2 + (3-1)^2} = 3.61$, and its distance to the second centroid $\mathbf{c}_2 = (2,1)$ is $\sqrt{(4-2)^2 + (3-1)^2} = 2.83$, etc.

3. **Objects clustering** : We assign each object based on the minimum distance. Thus, medicine A is assigned to group 1, medicine B to group 2, medicine C to group 2 and medicine D to group 2. The element of Group matrix below is 1 if and only if the object is assigned to that group.

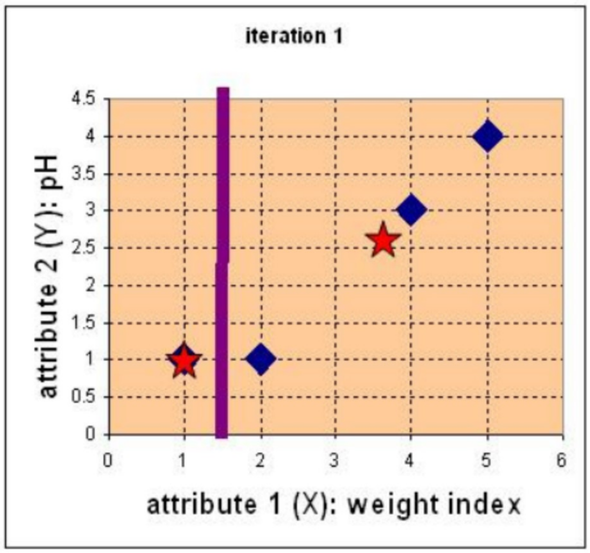
$$\mathbf{G}^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad \begin{matrix} group-1 \\ group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

4. Iteration-1, determine centroids : Knowing the members of each group, now we compute the new centroid of each group based on these new memberships.

Group 1 only has one member thus the centroid remains in $\mathbf{c}_1 = (1,1)$. Group 2 now has three members, thus the centroid is the average coordinate among the three members.

$$\mathbf{c}_2 = \left(\frac{2+4+5}{3}, \frac{1+3+4}{3}\right) = \left(\tfrac{11}{3}, \tfrac{8}{3}\right)$$



5. **Iteration-1, Objects-Centroids distances** : The next step is to compute the distance of all objects to the new centroids. Similar to step 2, we have distance matrix at iteration 1 is

$$\mathbf{D}^1 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 3.14 & 2.36 & 0.47 & 1.89 \end{bmatrix} \begin{matrix} \mathbf{c}_1 = (1,1) & group-1 \\ \mathbf{c}_2 = (\tfrac{11}{3}, \tfrac{8}{3}) & group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \begin{matrix} X \\ Y \end{matrix}$$

6. **Iteration-1, Objects clustering:** Similar to step 3, we assign each object based on the minimum distance. Based on the new distance matrix, we move the medicine B to Group 1 while all the other objects remain. The Group matrix is shown below

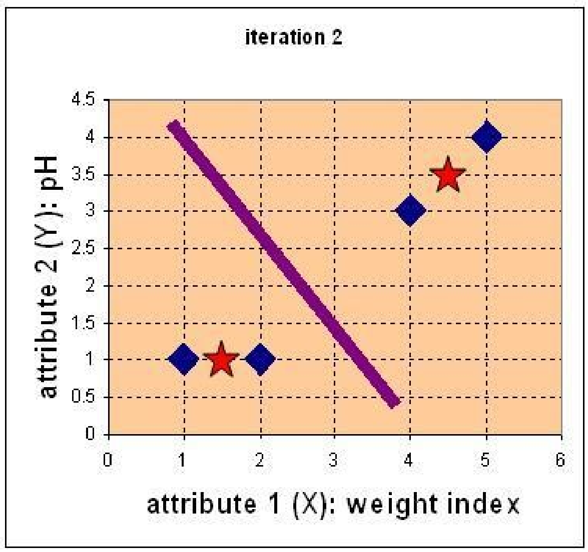
$$\mathbf{G}^1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} group-1 \\ group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

7. **Iteration 2, determine centroids:** Now we repeat step 4 to calculate the new centroids coordinate based on the clustering of previous iteration. Group1 and group 2 both has two members, thus the new centroids are

$$\mathbf{c}_1 = \left(\frac{1+2}{2}, \frac{1+1}{2}\right) = (1\tfrac{1}{2}, 1) \qquad \mathbf{c}_2 = \left(\frac{4+5}{2}, \frac{3+4}{2}\right) = (4\tfrac{1}{2}, 3\tfrac{1}{2})$$

and



8. **Iteration-2, Objects-Centroids distances:** Repeat step 2 again, we have new distance matrix at iteration 2 as

$$\mathbf{D}^2 = \begin{bmatrix} 0.5 & 0.5 & 3.20 & 4.61 \\ 4.30 & 3.54 & 0.71 & 0.71 \end{bmatrix} \begin{matrix} \mathbf{c}_1 = (1\tfrac{1}{2}, 1) & group-1 \\ \mathbf{c}_2 = (4\tfrac{1}{2}, 3\tfrac{1}{2}) & group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \begin{matrix} X \\ Y \end{matrix}$$

9. **Iteration-2, Objects clustering:** Again, we assign each object based on the minimum distance.

$$\mathbf{G}^2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} group-1 \\ group-2 \end{matrix}$$

$$\begin{matrix} A & B & C & D \end{matrix}$$

We obtain result that $\mathbf{G}^2 = \mathbf{G}^1$. Comparing the grouping of last iteration and this iteration reveals that the objects does not move group anymore. Thus, the computation of the k-mean clustering has reached its stability and no more iteration is needed. We get the final grouping as the results

Object	Feature 1 (X): weight index	Feature 2 (Y): pH	Group (result)
Medicine A	1	1	1
Medicine B	2	1	1
Medicine C	4	3	2
Medicine D	5	4	2

Session 5

OBJECTIVE: Implement single and complete link agglomerative clustering algorithm for given distance matrix.

TASKS:

- Use single and complete link agglomerative clustering to group the data described by the following distance matrix.

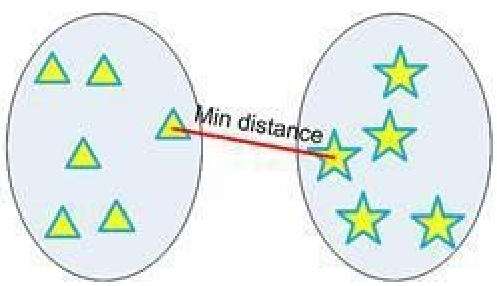
	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

Figure: Distance Matrix

- Show the Dendrogram

Algorithm Description:

Minimum distance clustering is also called as single linkage hierarchical clustering or nearest neighbor clustering. Distance between two clusters is defined by the minimum distance between objects of the two clusters, as shown below.



For example, we … 6. This distance matrix was calculated ba… …ious section.

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

We have 6 objects and we put each object into one cluster (analogue to put a ball into a basket). Instead of calling them as objects , now we call them clusters . Thus, in the beginning we have 6 clusters. Our goal is to group those 6 clusters such that at the end of the iterations, we will have only single cluster consists of the whole six original objects.

In each step of the iteration, we find the closest pair clusters. In this case, the closest cluster is between cluster F and D with shortest distance of 0.5. Thus, we group cluster D and F into cluster (D, F). Then we update the distance matrix (see distance matrix below). Distance between ungrouped clusters will not change from the original distance matrix. Now the problem is how to calculate distance between newly grouped clusters (D, F) and other clusters?

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	?	4.24
B	0.71	0.00	4.95	?	3.54
C	5.66	4.95	0.00	?	1.41
D, F	?	?	?	0.00	?
E	4.24	3.54	1.41	?	0.00

That is exactly where the linkage rule comes into effect. Using single linkage, we specify minimum distance between original objects of the two clusters.

Using the input distance matrix, distance between cluster (D, F) and cluster A is computed as

$$d_{(D,F)\mapsto A} = \min\left(d_{DA}, d_{FA}\right) = \min\left(3.61, 3.20\right) = 3.20$$

Distance between cluster (D, F) and cluster B is

$$d_{(D,F)\mapsto B} = \min\left(d_{DB}, d_{FB}\right) = \min\left(2.92, 2.50\right) = 2.50$$

Similarly, distance between cluster (D, F) and cluster C is

$$d_{(D,F)\mapsto C} = \min\left(d_{DC}, d_{FC}\right) = \min\left(2.24, 2.50\right) = 2.24$$

Finally, distance between cluster E and

$$d_{E\to(D,F)} = \min\left(d_{ED}, d_{EF}\right) = \min\left(1.00, 1.12\right) = 1.00$$

Then, the updated distance matrix becomes

Looking at the lower triangular updated distance matrix, we found out that the closest distance between cluster B and cluster A is now 0.71. Thus, we group cluster A and cluster B into a single cluster name (A, B).

Now we update the distance matrix. Aside from the first row and first column, all the other elements of the new distance matrix are not changed.

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	3.20	4.24
B	0.71	0.00	4.95	2.50	3.54
C	5.66	4.95	0.00	2.24	1.41
D, F	**3.20**	**2.50**	**2.24**	0.00	1.00
E	4.24	3.54	1.41	**1.00**	0.00

Dist	A,B	C	(D, F)	E
A,B	0	?	?	?
C	**?**	0	2.24	1.41
(D, F)	**?**	2.24	0	1.00
E	**?**	1.41	1.00	0

Using the input distance matrix (size 6 by 6), distance between cluster C and cluster (D, F) is computed as $d_{C\rightarrow(A,B)} = \min(d_{CA}, d_{CB}) = \min(5.66, 4.95) = 4.95$

Distance between cluster (D, F) and cluster (A, B) is the minimum distance between all objects involves in the two clusters

$$d_{(D,F)\rightarrow(A,B)} = \min(d_{DA}, d_{DB}, d_{FA}, d_{FB}) = \min(3.61, 2.92, 3.20, 2.50) = 2.50$$

Similarly, distance between cluster E and (A, B) is

$$d_{E\rightarrow(A,B)} = \min(d_{EA}, d_{EB}) = \min(4.24, 3.54) = 3.54$$

Then the updated distance matrix is

Min Distance (Single Linkage)

Dist	A,B	C	(D, F)	E
A,B	0	4.95	2.50	3.54
C	**4.95**	0	2.24	1.41
(D, F)	**2.50**	2.24	0	1.00
E	**3.54**	1.41	1.00	0

Observing the lower triangular of the updated distance matrix, we can see that the closest distance between clusters happens between cluster E and (D, F) at distance 1.00. Thus, we cluster them together into cluster ((D, F), E).

The updated distance matrix is given below.

Min Distance (Single Linkage)

Dist	(A,B)	C	(D, F), E
(A,B)	0.00	4.95	2.50
C	4.95	0.00	1.41
(D, F), E	**2.50**	**1.41**	0.00

Distance between cluster ((D, F), E) and cluster (A, B) is calculated as

$d_{((D,F),E)\rightarrow(A,B)} = \min\left(d_{DA}, d_{DB}, d_{FA}, d_{FB}, d_{EA}, d_{EB}\right) = \min\left(3.61, 2.92, 3.20, 2.50, 4.24, 3.54\right) = 2.50$

Distance between cluster ((D, F), E) and cluster C yields the minimum distance of 1.41. This distance is computed as $d_{((D,F),E)\rightarrow C} = \min\left(d_{DC}, d_{FC}, d_{EC}\right) = \min\left(2.24, 2.50, 1.41\right) = 1.41$. After that, we merge cluster ((D, F), E) and cluster C into a new cluster name (((D, F), E), C).

The updated distance matrix is shown in the figure below

Min Distance (Single Linkage)

Dist	(A,B)	(D, F), E),C
(A,B)	0.00	2.50
((D, F), E),C	**2.50**	0.00

The minimum distance of 2.5 is the result of the following computation

$$d_{(((D,F),E),C)\rightarrow(A,B)} = \min\left(d_{DA}, d_{DB}, d_{FA}, d_{FB}, d_{EA}, d_{EB}, d_{CA}, d_{CB}\right)$$

$$d_{(((D,F),E),C)\rightarrow(A,B)} = \min\left(3.61, 2.92, 3.20, 2.50, 4.24, 3.54, 5.66, 4.95\right) = 2.50$$

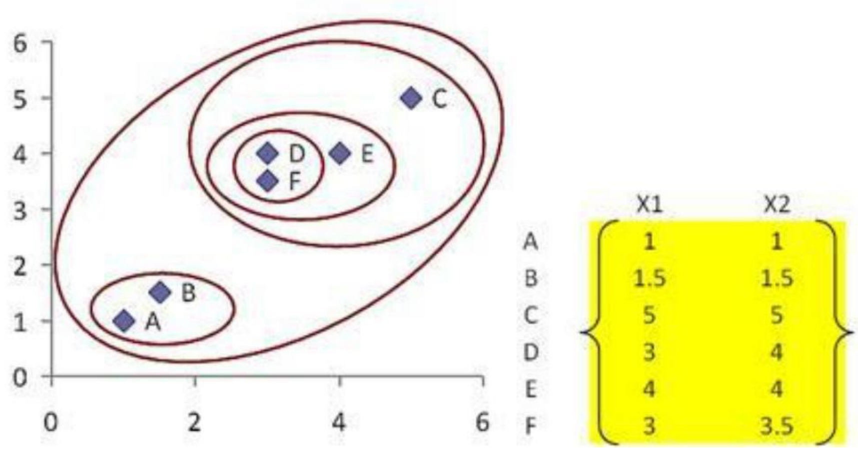
Now if we merge the remaining two clusters, we will get only single cluster contain the whole 6 objects. Thus, our computation is finished. We summarized the results of computation as follow:

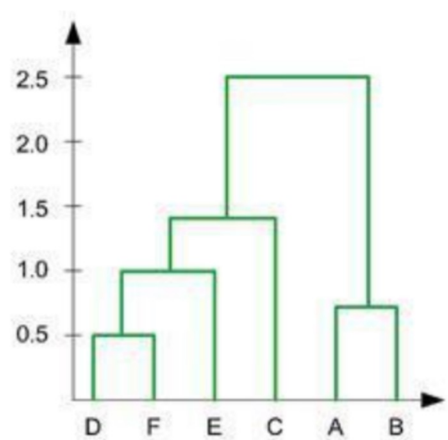
1. In the beginning we have 6 clusters: A, B, C, D, E and F
2. We merge cluster D and F into cluster (D, F) at distance **0.50**
3. We merge cluster A and cluster B into (A, B) at distance **0.71**
4. We merge cluster E and (D, F) into ((D, F), E) at distance **1.00**
5. We merge cluster ((D, F), E) and C into (((D, F), E), C) at distance **1.41**

6. We merge cluster (((D, F), E), C) and (A, B) into ((((D, F), E), C), (A, B)) at distance **2.50**
7. The last cluster contain all the objects, thus conclude the computation

Using this information, we can now draw the final results of a dendogram. The dendogram is drawn based on the distances to merge the clusters above.

The hierarchy is given as (((D, F), E),C), (A,B). We can also plot the clustering hierarchy into XY space





Session 6

OBJECTIVE: Implement K-nearest neighbor (K-NN) classifier for a given dataset and measure performance of the algorithm by generating F-measure score.

TASKS:

- Use the Nearest Neighbor clustering algorithm and Euclidean distance to design a classifier to classify the feature points of IRIS dataset.
- Determine the accuracy of IRIS dataset using F-Measure score for individual classes. Consider the threshold value 4 as the nearest neighbor. Use the lower 10% data of IRIS dataset as the test data.

Algorithm Description:

Here is step by step on how to compute K-nearest neighbors KNN algorithm:

1. Determine parameter K = number of nearest neighbors
2. Calculate the distance between the query-instance and all the training samples
3. Sort the distance and determine nearest neighbors based on the K-th minimum distance.
4. Gather the category Y of the nearest neighbors.
5. Use simple majority of the category of nearest neighbors as the prediction value of the query instance

Example

We have data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here is four training samples

X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	Y = Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Now the factory produces a new paper tissue that pass laboratory test with X1 = 3 and X2 = 7. Without another expensive survey, can we guess what the classification of this new tissue is?

1. Determine parameter K = number of nearest neighbors

Suppose use K = 3

2. Calculate the distance between the query-instance and all the training samples

Coordinate of query instance is (3, 7), instead of calculating the distance we compute square distance which is faster to calculate (without square root)

X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	**Square Distance to query instance (3, 7)**
7	7	
7	4	
3	4	
1	4	

$$(7-3)^2+(7-7)^2=16$$
$$(7-3)^2+(4-7)^2=25$$
$$(3-3)^2+(4-7)^2=9$$
$$(1-3)^2+(4-7)^2=13$$

3. Sort the distance and determine nearest neighbors based on the K-th minimum distance

X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	Square Distance to query instance (3, 7)	**Rank minimum distance**	**Is it included in 3-Nearest neighbors?**
7	7		**3**	**Yes**
7	4		**4**	**No**
3	4		**1**	**Yes**
1	4		**2**	**Yes**

$$(7-3)^2+(7-7)^2=16$$
$$(7-3)^2+(4-7)^2=25$$
$$(3-3)^2+(4-7)^2=9$$
$$(1-3)^2+(4-7)^2=13$$

4. Gather the category Y of the nearest neighbors. Notice in the second row last column that the category of nearest neighbor (Y) is not included because the rank of this data is more than 3 (=K).

X1 = Acid Durability (seconds)	X2 = Strength (kg/square	Square Distance to query instance (3, 7)	Rank minimum distance	Is it included in 3-Nearest	**Y = Category of nearest**

	meter)			neighbors?	**Neighbor**
7	7		3	Yes	**Bad**
7	4		4	No	**-**
3	4		1	Yes	**Good**
1	4		2	Yes	**Good**

$$(7-3)^2+(7-7)^2=16$$
$$(7-3)^2+(4-7)^2=25$$
$$(3-3)^2+(4-7)^2=9$$
$$(1-3)^2+(4-7)^2=13$$

5. Use simple majority of the category of nearest neighbors as the prediction value of the query instance

We have 2 good and 1 bad, since 2>1 then we conclude that a new paper tissue that pass laboratory test with X1 = 3 and X2 = 7 is included in **Good** category.

MID TERM EXAMINATION

There will be a 40-minutes written mid-term examination. Different types of questions will be included such as MCQ, mathematics, writing code fragments etc.

FINAL TERM EXAMINATION

There will be a one-hour written examination. Different types of questions will be included such as MCQ, mathematics, write a program etc.