Date: 3-3-25

M T W T F S S

Lecture no:02:

Machine Learning: we make the machine to learn from the data continously we make the machine to predict that thing that will happen in the future. By the given data. There are many methods that we use for predicting the data

- ✓ Supervised Learning
- ✓ Unsupervised Learning
- ✓ Reinforcement Learning

Supervised Learning: → when data set is given

Where we know what the output is, or the characterstics and feature are known to predict the dataset. The dataset is given in supervised Learning.

Dataset — ML — Output

train / test

∴ Before applying any algo on to the data, the dataset must be complete.

Unsupervised Learning:

we don't know what the Output is, we have some data to learn and predict about the dataset but we don't actially know the answer.

Data — ML — Group

Reinforcement Learning: → To learn from the environment..

There is no dataset in Reinforcement Learning There are some reward ~~which~~ & punishment which we make the machine to learn from it. Learning from the previous Learning.

ALIM NOTES

Date:

M T W T F S S

Q: why we need to preprocess the data?) → Paper Q.s

To analyze the data Following are the steps

- Data must be completed.
- No missing values
- Data must be accurate

First preprocess the data secondly check whether the output is given or not.

Supervised Learning.

Dataset given

Features Output

1

Y1 Y2 Y3

Feather	# of legs	Tail	beak	size	Creature
					Bird
					Cat
					Cow

→ Example of Supervised Learning

it is a type of supervised

Algos for Classification:

- ✓ Decision Tree
- ✓ NAVE BAYES
- ✓ KNN
- ✓ SVM
- ✓ Logistic Regression

.Data set ka complete hona zaroori ha then analyze ker no.

If the Output is categorical then it is classification.

output gr no. of rows se kam ha to classification ho gha.

If the output is discrete then it is regression.

eg) Predict house price no of rows kah hi kam ho output se to regression means hoy

values of output is countable.

If the values are two then known as Binary classification and if more than two known as multi-class classification.

Algos for Regression:

Linear Regression

output value will be changed each time

ALIM NOTES

Date: M T W T F S S

Supervised

Categorical: categorical refers to data that is divided into categories or group not numbers.

Agr no of rows output se kam ha to classification ho ghe.

Discrete: → Regression

For example; Predict house price agr no of rows there se kam ho ghe to output se to regression ho ghe.

Two types of multiclass Classification

specific order ✓ Ordinal

↓

Strongly Disagree
Disagree
Neutral Agree

Best
average
Worst

✓ Multinomial

↓ Agr specific order nhi ha.

↓

e.g Cat, Cow, Bird

e.g Male Female

Assignment: 1.

Dataset kis type ka h

No of attributes → 15 s ab

1500 ~~ea above~~ - No of rows / coloums

Dataset ki type

Download from KAGGLE or UCI.

Date: _____ M T W T F S S

Height (cm)	Weight	Tshirt size	Distance	sorted dist
160	64	L	3.16	5
163	64	L	3.6	6
165	61	L	4	7
165	62	L	4.1	8
165	65	L	5.6	10
168	62	L	7.06	11
168	63	L	7.28	12

Q: Identify shirt size of Customer having hei 161 & wei 61kg.

Step 1: Calculate Similarity based on distance

✓ Euclidean Distance:

(use any of one) $d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

✓ Manhatan Distance:

$d(x,y) = \sum_{i=1}^{n} |x_i - y_i|$

(161, 61) (158, 58)

$d(x,y) = \sqrt{(161-158)^2 + (61-58)^2}$

$= \sqrt{9+9}$

$= \sqrt{18}$

$= 4.24$

Step 2: Find k Nearest Neighbour

k = 3 M MM

k = 5 M MM M L

∴ if the value are discrete & numerical than avoid using decision Tree.

And if the values in the colaumns are invalid then also we ignore that colaumn

if no washroom are (1, 2, 2.5)
=there is no washroom that is 2.5 value is invalid then ignore this whole colaumn.

washroom can either be 2 or 3 not 2.5 invalid value

∴ if there is only one attribute in the data that is textual then convert in to the numerical. Give

if age is given (old, young, child) convert in that colaumn (0, 1, 2)

ALIM NOTES

Date: 10-03-2025

M T W T F S S

Lecture no. 02:

Supervised learning: (classification)

we divide the data into training and testing that we call it to learner

- Lazy Learners
- Eager Learners

Training

Input	Output
70	30
60	40

→ Testing

∴ Agr apne sirf 2 classes identify krne ha to wo binary ha jase cancer patient ka check kro sirf ka cancer ha ya nhi. to ye classification ha.

- Lazy Learners: (KNN)
  Less time in training.

- Eager Learners: (Decision Tree, Naive Bayes, ANN)
  More time in ~~training~~ ~~training~~ testing.

KNN: (Binary classification) → bcz only Finding the M & large shirts distance.

∴ if ap check krte ka cancer patient ka ~~ka~~ kn sa stage ha stag 1, 2, 3, 4 then it is regression.

example:

	Height (cm)	Weight (kg)	T shirt size	Dist
1	158	58	M	4.24
2	158	59	M	3.6
3	158	63	M	3.6
4	160	59	M	2.22
5	160	60	M	1.4
6	163	60	M	2.23
7	163	61	M	2

Short distance

ALIM NOTES

Date: 17-03-2025

M T W T F S S

Lecture no: 03.

value will predict the class (ordinal, binary)

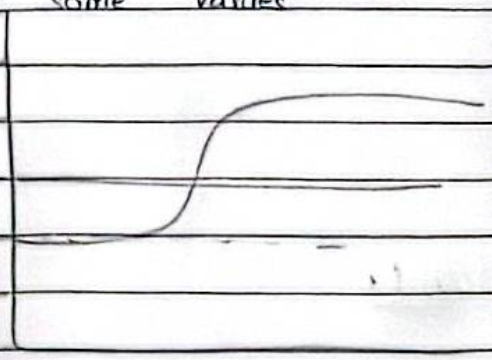
Logistic Regression - classification

- First we define a threshold like some values
- Threshold can be multiple
- Formula of logistic Regression

logistic regression with one variable or feature

$$y = \frac{1}{1+e^{f(x)}} = \frac{1}{1+e^{x}}$$

$$= \frac{1}{1+e^{-(B_0+B_1 x)}}$$



make a Sigmoid

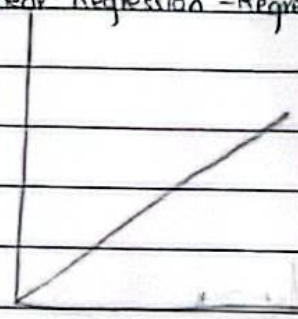
- logistic Regression with multiple variable

$$\frac{1}{1+e^{-(B_0+B_1x_1+B_2x_2)}} = \frac{1}{1+e^{-B_0-B_1x_1-B_2x_2}}$$

Input Output

x	y
0.5	0
1.2	0
2.3	0
3.1	1
4.5	1
5.0	1
5.8	1
6.3	1
7.1	1
8.0	1

Linear Regression - Regression



make a straight line

$y = B_0 + B_1 x$

- We use formula of logistic regression with one variable or feature because Input is given in one variable and this can cause classification bcz output is given.

ALIM NOTES

Date: _____

M T W T F S

We use this formula for finding the values.

$$= \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

Given

$B_0 = 3$

$B_1 = 0.8$

Put values in this formula.

$x = 0.5$ $$= \frac{1}{1+e^{-(3+(0.8)(0.5))}}$$

$$= \boxed{0.96}$$

$x = 1.2$

$$= \frac{1}{1+e^{-(3+(0.8)(1.2))}}$$

$$= \boxed{0.98}$$

$x = 2.3$

$$= \frac{1}{1+e^{-(3+(0.8)(2.3))}}$$

$$= \boxed{0.992}$$

$x = 3.1$

$$= \frac{1}{1+e^{-(3+(0.8)(3.1))}}$$

$$= \boxed{0.995}$$

$x = 4.5$

$$= \frac{1}{1+e^{-(3+(0.8)(4.5))}}$$

$$= \boxed{0.998}$$

x	Y_A	Y_P (74 3)	Y_{PL}
0.5	0	change 0.96	0.06
1.2	0	0.98	0.115
2.3	0	0.992	0.23
3.1	1	0.995	0.37
4.5	1	0.998	0.68
5.0	1	0.999	0.73
5.8	1	0.9996	0.83
6.3	1	0.9996	0.84
7.1	1	0.9998	0.93
8.0	1	0.9999	0.96

ALIM NOTES

Date: _____

M T W T F S S

$x = 5.0$

$$= \frac{1}{1+e^{-(3+(0.8)(5.0)}}$$

$$= \boxed{0.999}$$

$x = 5.8$

$$= \frac{1}{1+e^{-(3+(0.8)(5.8)}}$$

$$= \boxed{0.9995}$$

$x = 6.3$

$$= \frac{1}{1+e^{-(3+(0.8)(6.3)}}$$

$$= \boxed{0.9996}$$

$x = 7.1$

$$= \frac{1}{1+e^{-(3+(0.8)(7.1)}}$$

$$= \boxed{0.9998}$$

$x = 8.0$

$$= \frac{1}{1+e^{-(3+(0.8)(8.0)}}$$

$$= \boxed{0.9999}$$

Date: _____ M T W T F

Formulas.

$$\beta_0 = \frac{\Sigma x^2 \Sigma y - \Sigma x \Sigma xy}{n\Sigma x^2 - (\Sigma x)^2}$$

$$\beta_1 = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

Find Accuracy

$\beta_0 = -3$, $\beta_1 = 0.8$ } 90%

$\beta_0 = 3$, $\beta_1 = 0.8$ } 70%

Where the value of β is not known

X	Y	X^2	XY
0.5	0	0.25	0
1.2	0	1.44	0
2.3	0	5.29	0
3.1	1	9.61	3.1
4.5	1	20.25	4.5
5.0	1	25	5.0
5.8	1	33.64	5.8
6.3	1	39.69	6.3
7.1	1	50.41	7.1
8.0	1	64	8.0

$\Sigma x = 43.80$ $\Sigma y = 7$ $\Sigma x^2 = 250$ $\Sigma xy = 39.8$

e.g. $x = 5.2$

$y = ?$

$$y = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1)}}$$

$$y = \frac{1}{1+e^{-(-3+0.8(5.2))}}$$

$y =$

a jis ki accuracy ziada percent ho ghe hm wo us ki values use kren ghe. means hm 90% vale ki value use kren gho. Jb hm ne 'y' find kna ho.

We find β when β is not given?!

$$\beta_0 = \frac{\Sigma x^2 \Sigma y - \Sigma x \Sigma xy}{n\Sigma x^2 - (\Sigma x)^2}$$

Put values in this formula

$$\beta_0 = \frac{(249.58)(7) - (43.80)(39.8)}{10(249.58) - (43.80)^2} = \boxed{0.006}$$

$$\beta_1 = \frac{10(39.8) - (43.80)(7)}{10(249.58) - (43.80)} = \boxed{0.157}$$

Suggest a visible a polynomial to fit and read and display the dataset? check the binary for logistic regression write binary used to input logistic regression.

ALIM NOTES

Date:

M T W T F S S

Quiz examples:

student	study hours (X)	Pass(1)/Fail(0) (Y)	X^2	XY
1	1.5	0	2.25	0
2	2.0	0	4	0
3	2.5	0	6.25	0
4	3.0	0	9	0
5	3.5	1	12.25	3.5
6	4.0	1	16	4.0
7	4.5	1	20.25	4.5
8	5.0	1	25	5.0
9	5.5	1	30.25	5.5
10	6.0	1	36	6.0
	37.5	6	151.25	28.5

Finding β_0 & β_1

$$\beta_0 = \frac{\Sigma x^2 \Sigma y - \Sigma x \Sigma xy}{n \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{(151.25)(37.5)(28.5)}{(10)(151.25) - (37.5)^2}$$

$$\beta_0 = -1.5176$$

$$\beta_1 = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{10(28.5) - (37.5)(6)}{10(151.25) - (1406.25)}$$

$$\beta_1 = 0.5647$$

ALI
NOT

Date: _____

$X = 3.2$

$$y = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \Rightarrow \frac{1}{1+e^{-(-1.5176+(0.5647)(3.2))}}$$

$$y = 0.5719$$

X	Y		
1.5	0	0.3384 ≈ 0	
2.0	0	0.4042 ≈ 0	
2.5	0	0.4736 ≈ 0	
3.0	0	0.5440 ≈ 1	→ 1 change & remain are same so its accuracy is 90%. 1 change = 10%.
3.5	1	0.6127 ≈ 1	
4.0	1	0.6773 ≈ 1	
4.5	1	0.7357 ≈ 1	
5.0	1	0.7868 ≈ 1	
5.5	1	0.8304 ≈ 1	
6.0	1	0.8665 ≈ 1	

$$y = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

$$y = \frac{1}{1+e^{-(-1.5176+(0.5647)(1.5))}}$$

$$y = 0.3383$$

Now ~~so~~ put remaining values according to this formula

Accuracy: $9/10$

$= 0.9 \rightarrow 90\%$

$= 90\%$ Accuracy.

ALIM NOTES

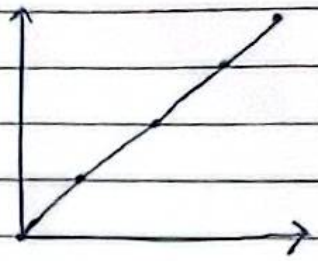
Date: 30.4.25

M T W T F S S

Final Syllabus:

Lecture no: 4 (Regression)

Linear Regression:



Linear Regression with one variable:

$$Y = \beta_0 + \beta_1 x_1$$

constant — Constant — attribute

Linear Regression with multiple variable

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n$$

Example:

X (Hours studied)	Y (marks obtained)
1	50
2	55
3	60
4	65
5	70
6	75
7	80
8	85
9	90
10	95
11	100
12	105

$\Sigma X = 78$ $\Sigma Y = 935$

ALIM NOTES

Date: _____

x^2	xy	yp	Ep_1 $\lvert YA-YP \rvert$	yp^2_2	Ep^2_2	yp^3_3	Ep^3_3	$(Ep_1)^2$
1	50	51.27	1.27	50.43	0.43	50.5	0.5	
4	110	56.11	1.11	54.43	0.57	55	0	
9	195	60.95	4.05	58.43	6.57	59.5	5.5	
16	280	65.7	4.3	62.43	7.57	64	6	
25	325	70.63	5.63	66.43	1.43	68.5	3.5	
36	450	75.47	0.47	70.43	4.50	73	2	
49	560	80.31	0.30	74.43	5.57	77.5	2.5	
64	680	85.15	0.15	78.43	6.57	82	3	
81	810	89.99	0.01	82.43	7.57	86.5	3.5	
100	950	94.83	0.17	86.43	8.57	91	4	
121	1100	99.67	0.33	90.43	9.57	95.5	4.5	
144	1260	104.5	0.49	94.43	10.57	100	5	
$\Sigma x^2 = 650$	$\Sigma xy = 6770$		$\Sigma \lvert YA-YP \rvert = 18.29$	$\Sigma 69.56$ $\Sigma YA+YP^2$		$\Sigma YA-YP) = 40$		Σ

$YA-YP$ → $YA-YP$

$(YA-YP)$ ki jgha (Ep_1) ka total sum put krna.

Put values in β_0 and β_1

$$\beta_0 = \frac{(650)(935) - (78)(6770)}{12(650) - (78)^2}$$

$$\beta_0 = \boxed{46.43}$$

$$\beta_1 = \frac{12(6770) - 78(935)}{12(650) - (78)^2}$$

$$\beta_1 = \boxed{4.84}$$

Put the value in y.

$y = \beta_0 + \beta_1 x_1$

$46 + 4.5(3)$

∴ When you are finding yp^2 you will take a random value which is nearest to β_1. After finding this then you will check the distance b/w yp' & yp'' wh after checking the distance when yp' and yp^2 distance is nearest then that one is your ~~nearest~~ nearest point

ALIM NOTES

Date:

M T W T F S S

Performance Calculation of Regression: 1st Iteration:

- Mean Square Error = $\Sigma (YA - YP)^2 / n = (18.29)^2 / 12 = 27.87$

- Root Mean Square = $\sqrt{\Sigma (YA - YP)^2 / n} = \sqrt{27.87} = 5.2$

∴ iteration:
We do the iterations for algo

2nd Iteration:

mean square error:

$$\text{Error} = \Sigma (YA - YP^2)^2 / n$$
$$= (69.49)^2 / 12$$
$$= 403.2$$

Root mean square:

$$\sqrt{\Sigma (YA - YP^2)^2 / n}$$
$$= 20.07$$

3rd Iteration:

Mean Square error:

$$\text{Error} = \Sigma (YA - YP^3)^2 / n$$
$$= (40)^2 / 12 = 133$$

root mean square:

$$\sqrt{\Sigma (YA - YP^3)^2 / n} = \sqrt{133}$$
$$= 11.5$$

ALIM NOTES

Date: _____ M T W T F S S

Q. If the student studied for 6.5 h, then how many marks will be obtained.

So, lower mean square value is 5.2

$= 46.43 + 4.84 (6.5)$ — given

$= \boxed{77.89}$

~~We can~~ Firstly, we can find the iterations then we will check which is nearest to the 0 if the value is fully 0 then your ans/iteration is wrong. If the ans is nearest to 0 then we will take that values of β_0 and β_1 and put in the given condition. and you also change the value of β_0 and β_1.

ALIM NOTES

Date:

M T W T F S S

Logistic Regression with multiple variable:

X_1	X_2	Y_A	Y_P'	EP' $(YA-YP)$	YP^2 $\rightarrow YA \cdot YP^2$	EP^2
1	0.10	1.5	1.6	0.1	1.4	0.1
2	0.12	2	2.68	0.68	2.4	0.4
3	0.14	3.5	3.71	0.21	3.4	0.1
4	0.16	4.0	4.74	0.74	4.4	0.4
5	0.18	5.5	5.77	0.27	5.5	0
6	0.20	6	6.8	0.8	6.5	0.5

Suppose the value of β_0 & β_1 & β_2.

$\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 1.5$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

$= 0.5 + (1)(1) + (1.5)(0.10)$

$= 1.6$.

continue till $x_1 = 6$ & $x_2 = 0.20$

$EP' = (YA - YP')$

$= 1.5 - 1.6$

$= 0.1$

Continue till $Y_1 = 6$ & $YP = 6.8$

Firstly, finding the values of β_0 then find mean sq error & root mean sq. then change the value of β_0

Change the value of $\beta_0 = 0.25$

Mean square root $= \Sigma (YA - YP')^2 / n = (1.5)^2 / 6$

$= 0.375$

Root Mean Square $= \sqrt{\Sigma (YA - YP)^2 / n} = \sqrt{0.375}$

$= 0.6$.

ALIM NOTES

Date: _____ M T W T F S S

Formula for β_0 for multivariable.

$$\beta_0 = \frac{\Sigma(x_1x_2)^2\Sigma y - \Sigma(x_1x_2)\Sigma(x_1x_2)y}{n\Sigma(x_1x_2)^2 - (\Sigma x_1x_2)^2}$$

Lecture no:05:

Performance Metrics of Classification:

Accuracy: Prediction must be correct.

$$8/12 = \frac{TP + TN}{TP + TN + FN + FP}$$

Precision:

$$\frac{TP}{TP + FP}$$

Recall:

$$\frac{TP}{TP + FN}$$

F_1 Square:

$$2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

ALIM NOTES

Date:

M T W T F S S

Data set:

S-No		Actual A	Actual YP	
1	A_1	Positive	P	TP
2	A_2	Positive	N	FN
3	A_3	Negative	N	
4	A_4	Positive	P	
5	A_5	Negative	P	
6	A_6	Negative	N	
7	A_7	Positive	P	
8	A_8	Positive	P	
9	A_9	Negative	N	
10	A_{10}	Negative	P	
11	A_{11}	Positive	N	
12	A_{12}	Negative	N	

True Positive
Actual P, Predicted P

True Negative
Actual N, Predicted N

False Positive
Actual N, Predicted P

False Negative
Actual P, Predicted N

Confusion Matrices

	Actual	
	Positive	Negative
Positive	TP = 4	FP = 2
Negative	FN = 2	TN = 4

Precision:

Proportion of Correctly predicted +ve cases out of all predicted +ve cases.

Recall:

Proportion of Correctly predicted +ve cases out of all actual +ve cases.

Date:

F_1 Square:

Harmonic mean of precision and recall:

It balances both metrics especially useful when classes are imbalanced.

3x3 Matrices:

	cat	Dog	Rabbit
cat	1 TP	2	5
Dog	3	4 TP	2
Rabbit	1	6	TP 10

Confusion Matrix w.r.t cat:

Actual	cat	others → Dog, Rab
cat	1	7
others	4	22

∴ To check the performance of Regression there is a suitable method named as root mean square and mean square error.

Confusion Matrix w.r.t Dog

Actual	Dog	cat or Rabbit
Dog	4	5
cat or Rabbit	8	17

∴ To check the performance of classification there is a suitable method named as Confusion matrix.

Confusion Matrix w.r.t Rabbit

Actual	Rabbit	cat or Dog
Rabbit	10	7
cat or dog	7	10

Confusion w.r.t cat

	cat	others
cat	TP	FP
others	FN	TN

Date:

M T W T F S S

Lecture no: 06:

Unsupervised Learning

Partitioning → K means

Hierarchical → AGGLOMERATIVE

Density Based → DBSCAN

Partitioning (k mean):

ID	x_1	x_2	Iteration 1 centroid	Iteration 2 centroid	K = 2
1	1.0	1.5	C_1	C_1	
2	1.2	1.3	C_1	C_1	
3	0.8	1.6	C_1	C_1	
4	1.1	1.0	C_1	C_1	
5	3.5	4.0	C_2	C_2	
6	3.8	3.7	C_2	C_2	
7	3.6	4.1	C_2	C_2	
8	3.9	3.8	C_2	C_2	
9	1.3	1.2	C_1	C_1	
10	3.7	4.2	C_2	C_2	
11	1.0	1.4	C_1	C_1	
12	3.6	3.9	C_2	C_2	

∴ if value of K = 3 then you will take 3 values if K = 2 then you will take 2 values

C_1 C_2

$d(C_1, 2) = \sqrt{(1.0-1.2)^2 + (1.5-1.3)^2} = 0.28$

$d(C_2, 2) = \sqrt{(3.6-1.2)^2 + (3.9-1.3)^2} = 3.5$

$d(C_1, 3) = \sqrt{(1.0-0.8)^2 + (1.5-1.6)^2} =$

$d(C_2, 3) = \sqrt{(3.6-0.8)^2 + (3.9-1.6)^2} =$

ALIM NOTES

Date: _____

$d(C_1, 4) = \sqrt{(1.0-1.1)^2 + (1.5-1.0)^2} =$

$d(C_2, 4) = \sqrt{(3.6-1.1)^2 + (3.9-1.0)^2} =$

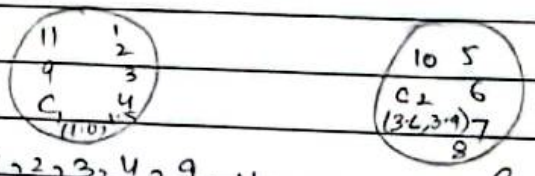
$d(C_1, 5) = \sqrt{(1.0-5.5)^2 + (1.5-4.0)^2} =$

$d(C_2, 5) = \sqrt{(3.6-3.5)^2 + (3.9-4.0)^2} =$

⋮

Continue up to $(C_1, 12)$ $(C_2, 12)$

$\text{Min}(d(C_1, 2), d(C_2, 2))$



$C_1 = 1, 2, 3, 4, 9, 11$

$C_1^2 x_1 = \frac{1.0 + 1.2 + 0.8 + 1.1 + 1.3 + 1.0}{6}$

$= 1.066$

$C_2 = 5, 6, 7, 8, 10, 12$

$C_2^2 x_2 = \frac{3.5 + 3.8 + 3.6 + 3.9 + 3.7 + 3.}{6}$

$= 3.68$

$C_1^2 y_2 = \frac{1.5 + 1.3 + 1.6 + 1.0 + 1.2 \cdot 1.4}{6}$

$= 1.3$

$C_2^2 x_2 = \frac{4.0 + 3.7 + 4.1 + 3.8 \cdot 4.2 + 3.9}{6}$

$= 3.95$

ALIM

Date:

M T W T F S S

Hierarchical (Aggecomerative).

ID	X_1	X_2
P_1	0.40	0.53
P_2	0.22	0.38
P_3	0.35	0.32
P_4	0.26	0.19
P_5	0.08	0.41
P_6	0.45	0.30

Iteration 1:

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.23	0				
P_3	0.22	0.14	0			
P_4	0.37	0.19	0.13	0		
P_5	0.34	0.14	0.28	0.23	0	
P_6	0.24	0.24	$\boxed{0.10}$	0.22	0.59	0

minimum cluster

$d(P_1, P_2) = \sqrt{(0.40-0.22)^2 + (0.53-0.38)^2} = 0.2$

$d(P_1, P_3) = \sqrt{(0.40-(0.35)^2 + (0.53-0.32)^2} = 0.21$

$d(P_1, P_4) = \sqrt{(0.40-0.26)^2 + (0.53-0.19)^2} = 0.37$

$d(P_1, P_5) = \sqrt{(0.40-0.08)^2 + (0.53-0.41)^2} = 0.34$

$d(P_1, P_6) \quad \sqrt{(0.40-0.45)^2 + (0.53-0.30)^2} = 0.29$

ALL NOTE

Date: _____

$d(P_1, P_2) = \sqrt{(0.40-0.22)^2 + (0.53-0.38)^2} = 0.2$

$d(P_2, P_3) = \sqrt{(0.22-0.35)^2 + (0.38-0.32)^2} = 0.14$

$d(P_2, P_4) = \sqrt{(0.22-0.26)^2 + (0.38-0.19)^2} = 0.19$

$d(P_2, P_5) = \sqrt{(0.22-0.08)^2 + (0.38-0.41)^2} = 0.14$

$d(P_2, P_6) = \sqrt{(0.22-0.45)^2 + (0.38-0.30)^2} = 0.2$

Continue)

$d(P_6, P$

minimum cluster

$P_1 \quad P_2 \quad P_3 \quad P_4 \quad P_5 \quad P_6$

Iteration 2:

	P_1	P_2	P_3P_6	P_4	P_5
P_1	0				
P_2	0.23	0			
P_3P_6	0.22	0.14	0		
P_4	0.37	0.19	min value $\boxed{0.13}$	0	
P_5	0.34	0.14	0.28	0.23	0

$\text{Min}((P_3, P_6), P_1) = \min(d(P_3, P_1), d(P_6, P_1))$

$= \min(0.22, 0.24)$

$= 0.22$

ALIM NOTES

Date: _____ M T W T F S S

✓ $Min((P_3, P_6), P_2) = min((P_3, P_2), (P_6, P_2))$
$= min(0.14, 0.24)$
$= 0.14$

✓ $min((P_3, P_6), P_4) = min(P_3, P_4), (P_6, P_4)$
$= min(0.13, 0.22)$
$= 0.13$

✓ $min((P_3, P_6), P_5) = min(P_3, P_5), (P_6, P_5)$
$= min(0.28, 0.39)$
$= 0.28$

P_3 P_6 P_4

Iteration 3:

	P_1	P_2	$P_3 P_6 P_4$	P_5
P_1	0			
P_2	0.23 (previous min value)	0		
$P_3 P_6 P_4$	0.22	0.14 (previous value)	0	
P_5	0.34	0.14	0.23	0

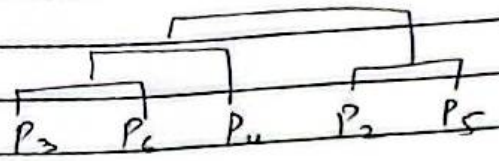
$d((P_3, P_6, P_4), P_1) = \cancel{min(P_3, P_6)} \cancel{(} min((P_3 P_6, P_1), (P_4, P_1))$
$= min(0.22, 0.37)$
$= 0.22$

$d(P_3 P_6, P_4), P_2) = min(P_3 P_6, P_2), (P_4, P_2)$
$= min(0.14, 0.19)$
$= \cancel{0.149}\ 0.14$

ALIM NOTES

Date: _____

$d((P_3 P_6, P_4), P_5) = \min((P_3 P_6, P_5), (P_4, P_5))$
$= \min(0.28, 0.23)$
$= 0.23$

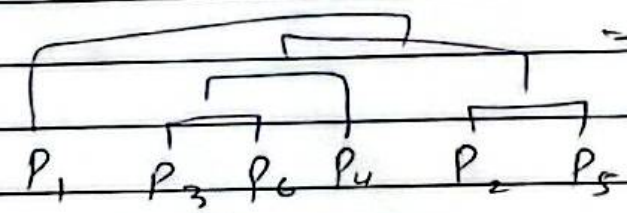


Iteration 4:

	P_1	$P_2 P_5$	$P_3 P_6 P_4$
P_1	0		
$P_2 P_5$	0.23	0	
$P_3 P_6 P_4$	0.22	$\boxed{0.14}$ min	0

$d((P_2, P_5), P_1) = \min((P_2, P_1), (P_5, P_1))$
$= \min(0.23, 0.34)$
$= 0.23$

$d((P_3, P_6, P_4), P_1) = \min((P_3, P_6, P_4, P_2), (P_3, P_6, P_4$
$= \min(0.14, 0.23)$
$= 0.14.$



choose 4 clusters

P_1 $P_3 P_6$ P_4 $P_2 P_5$

Choose 2 clusters

P_1 $P_3 P_6 P_4 P_2 P_5$

ALIM NOTES