question of ML for exam :

D.I → Entropy (measure of impurity of attribute)
→ Gini

a). Original dataset

	Yes	NO	total
count	8	10	18
p_i	$\frac{8}{18} = 0.44$	$\frac{10}{18} = 0.56$	
$-p_i \log_2 p_i$	$-0.44 \log_2 0.44$ $= 0.52$	$-0.56 \log_2 0.56$ $= 0.47$	

Total Entropy = 0.99

(b) splitted dataset (based on the feature (CGPA))

~~High~~ ~~medium~~ ~~Low~~ ~~total~~
~~18~~

CGPA = High

	Yes	NO	total
count	4	2	6
p_i	$\frac{4}{6} = 0.67$	$\frac{2}{6} = 0.33$	
$-p_i \log_2 p_i$	0.39	0.53	

Entropy = 0.92

CGPA = medium

	Yes	NO	total
count	4	3	7
p_i	$4/7$	$3/7$	
$-p_i \log_2 p_i$	$-4/7 \log_2 \frac{4}{7}$ $\Rightarrow 0.46$	$-\frac{3}{7} \log_2 \frac{3}{7}$ $\Rightarrow 0.52$	

Entropy = 0.99

CGPA = Low

	Yes	NO	total
Count	0	5	5
p_i	0	1	
$-p_i \log_2 p_i$	0	0	

Entropy = 0

Combined Entropy = $P(X) E(X)$

$$\frac{6}{18} \times 0.92 + \frac{7}{18} \times 0.99 + \frac{5}{18} \times 0$$

$$0.30 + 0.385 + 0 \Rightarrow 0.685$$

Page No. Date: / /

Information gain = Original data set - Combined Entropy
0.99~~7~~ - 0.685
$\Rightarrow$ 0.30

Inf gain for comm^n

comm = good

	Yes	NO	total
count	7	2	9
p_i	$\frac{7}{9}$ = 0.77	$\frac{2}{9}$ = 0.22	
$-p_i \log_2 p_i$	$-0.77 \log_2 0.77$	$-0.22 \log_2 0.22$	
	$\Rightarrow$ 0.290	0.480	$\Rightarrow$ 0.77

Comm = Bad

	Yes	NO	total
count	1	8	9
p_i	0.11	0.88	
$-p_i \log_2 p_i$	$-0.11 \log_2 0.11$	$-0.88 \log_2 0.88$	
	$\Rightarrow$ 0.350		

Info Gain (CGPA) = 0.30
Info Gain (Communication) = 0.36
——— (Aptitude) = 0.47
——— (Prog Skills) = 0.04

(3marks) (5marks)

ya toh information gain aayega ya fir two feature

Gini Index ÷

$$Gini(t) = 1 - \sum_j [p(j/t)]^2$$

C1	0
C2	6

$P(C_1) = \frac{0}{6} = 0$

$P(C_2) = \frac{6}{6} = 1$

$Gini = 1 - [P(C1)^2 + P(C_2)^2]$

$= 1 - [0 + 1]$

$= 1 - 1 = 0$

C_1	1
C_2	5

$P(C_1) = 1/6 = 0.166$

$P(C2) = 5/6 = 0.833$

$Gini = 1 - [0.166 \times 0.166 + 0.833 \times 0.833]$

$\Rightarrow 1 - [0.0275 + 0.69139]$

$1 - 0.719$

$\Rightarrow 0.28$

C_1	2
C2	4

$P(C_1) = 0.333 = 0.110$ (0.333×0.333)

$P(C2) = 0.666 = 0.443$

$Gini = 1 - [\cancel{0.108} + \cancel{0.435}] \Rightarrow 1 - 0.543$

$= \cancel{0.457}$ 0.44

Highest value of Gini = 0.5

Entropy = 1

Anand
Page No.
Date / /

Important Aayega Q. Cluster the following 8 points with x,y representing location into three clusters.

(x,y) A1 (2,10) A_2 (2,5) A3 (8,4), A4 (5,8)
A5 (7,5) A6 (6,4) A7 (1,2), A8 (4,9)

three clusters means $K=3$
if Paper not tells you about cluster toh Aapne hisab se lo.
but now given Intial cluster are :
A1 (2,10)
A4 (5,8)
A7 (1,2)
two choice distance fn given out you need to choose.
in this equestion $f(a,b) = |x_1 - x_2| + |y_1 - y_2|$
→ use Kmean clustering Algo. to find out three clusters.

Point	C1 A1 (2,10)	C2 A4 (5,8)	C3 A7 (1,2)	Cluster
A1 (2,10)	0	5	9	C1
A2 (2,5)	5	6	4	C3
A3 (8,4)	12	7	9	C2
A4 (5,8)	5	0	10	C2
A5 (7,5)	10	5	9	C2
A6 (6,4)	10	5	7	C2
A7 (1,2)	9	10	0	C3
A8 (4,9)	3	2	10	C2

C1 → A1 ⟶ A1 (2,10)
C2 → A3, A4, A5, A6, A8
C3 → A2, A7

(8,4)
(5,8) $\Rightarrow$ $8+5+7+6+4 \left(\frac{30}{5}\right)$ $\quad$ $4+8+5+4+9 \left(\frac{30}{5}\right)$
(7,5)
(6,4) $\Rightarrow (6,6)$
(4,9)

C3 $\Rightarrow$ A2 = (2,5) $\quad \left(\frac{3}{2}, \frac{7}{2}\right) \Rightarrow (1.5, 3.5)$
A7 = (1,2)

Point	4 (2,10)	C2 (6,6)	C3 (1.5,3.5)	Cluster
A1(2,10)	0	8	7	C1
A2(2,5)	5	5	2	C3
A3(8,4)	12	4	7	C2
A4(5,8)	5	3	8	C2
A5(7,5)	10	2	7	C2
A6(6,4)	10	2	5	C2
A7(1,2)	9	9	2	C3
A8(4,9)	3	5	2.5+5.5 (8)	C1

C1 $\rightarrow$ A1, A8 $\quad \left(\frac{6}{2}, \frac{19}{2}\right) \Rightarrow (3, 9.5)$

C2 $\rightarrow$ A3, A4, A5, A6 $\quad \frac{8+5+7+6}{4}, \frac{4+8+5+4}{4}$

$\Rightarrow \left(\frac{26}{4}, \frac{21}{4}\right) \rightarrow 6.5, 5.25$

C3 $\rightarrow$ A2, A7 $\quad \frac{2+1}{2}, \frac{5+2}{2} \Rightarrow (1.5, 3.5)$

Page No. Anand
Date / /

Point	(3, 9.5)	(6.5, 5.25)	(1.5, 3.5)	Cluster
A1 (2,10)	1.5	9	7	
A2 (2,5)	5.5	4.75	2	
A3 (8,4)	10.5			
A4 (5,8)	3.5			
A5 (7,5)	8.5			
A6 (6,4)	8.5			
A7 (1,2)	9.5			
A8 (4,9)				

21/11/22 Region What is a Region?
→ group of connected Pixel with similar property.

Basic formulation

a). $\bigcup_{i=1}^{n} R_i = R$

b). R_i is a connected region $i = 1, 2, \ldots, n$

c). $R_i \cap R_j = \emptyset$ for all i and j, $i \neq j$

d). $P(R_i) = TRUE$ for $i = 1, 2, \ldots, n$

e).

Region Growing :-

Theory question on this
↓
* Region growing
* Region splitting.

$$P(R_i(K) \cup \{x y\}) = TRUE$$

⇒ if we add different colour to some field we find
Arithmetic deviation,
Standard mean.

Problems :- Large execution time.
The election of the property to be satisfy

KNN is subervised and K-Mean is unsubervised.
In both algorithm we can findout value of K using Elbow method.

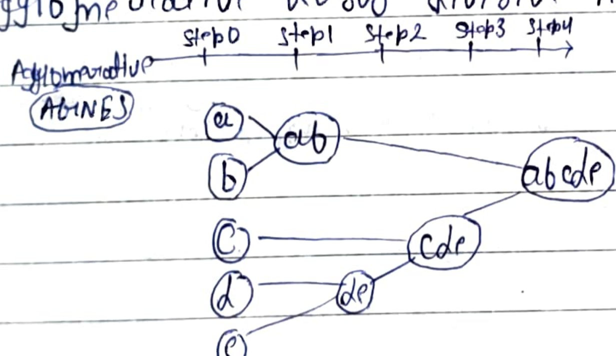
Basic clustering methods :-

1. Partitioning methods
   K-means.

2. Hierarchical methods.
   Agglomerative (bottom-up) or divisive (topdown),

3. Density-based methods } not in syllabus.
4. Grid-based

Agglomerative versus divisive Hierarchical clustering.



dendrogram

steps to perform Hierarchical clustering.

Distance-based Algorithm :-

~~Proximity~~ Proximity matrix

ID	1	2	3	4	5
1	0	3	18	10	25
2	3	0	21	13	28
3	18	21	0	8	7
4	10	13	8	0	15
5	25	28	7	15	0

Student_ID	Marks
1	10
2	7
3	28
4	20
5	35

(max)
(Min)
(Avg.)

(1,2)
(3,5) } merging

Student-ID	Marks
(1,2)	10
3	28
4	20
5	35

ID	(1,2)	3	4	5
(1,2)	0	18	10	25
3	18	0	8	7
4	10	8	0	~~15~~
5	25	7	~~15~~	0

Q- proximetrix draw kejege.
difference b/w Algomerittive and dendogram.

22/11/22

Setting the DOM :-

-> value

-> innerHTML

-> InnerTEXT

-> name

-> setAttribute   document.getElementbyId. Set Attribute ('class', 'a'); .removeattribute( );

-> Attributes

190 ReverseBits :-

Automata :-

Imp. Church Turing Thesis :-

-> 1913

Some common ...

Machine Algorithm

Imp

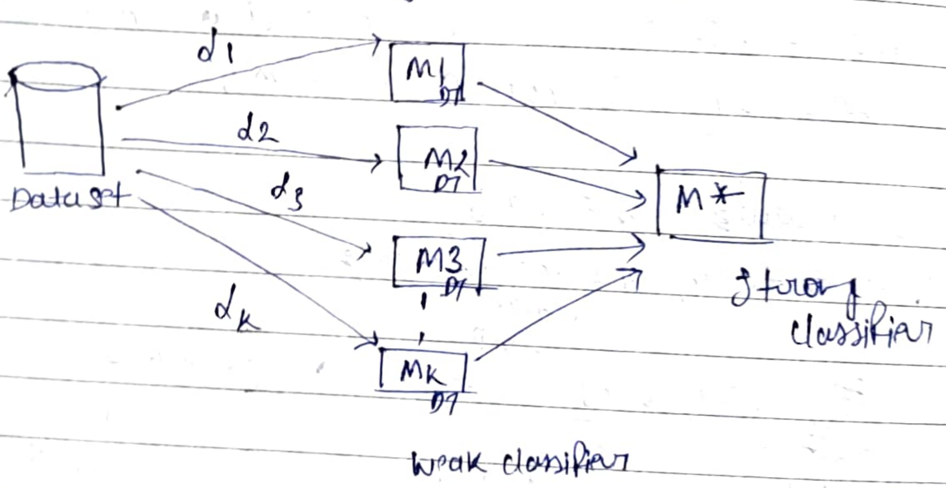
Q. Diff. Algomaritic and devise clustering
What is devisue clustering

⇒ Ensemble model :-

Ensemble methods uses collation of models to imbrove the Accuracy.

↳ Max voting
↳ min voting
↳ Average voting
↳ weighted Average voting.

Advance Ensemble Methods :-
i). Bagging
ii). Boosting.

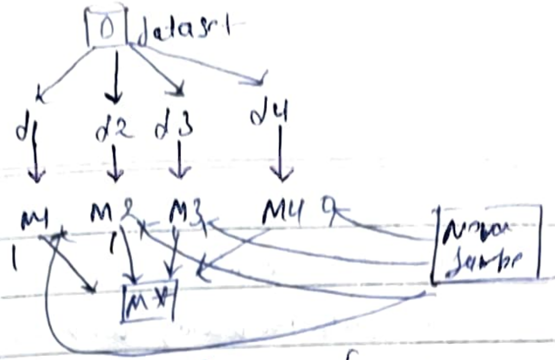


If all the models are of same type then the Ensemble classifier is homogenous. Otherwise hetrogeneous.

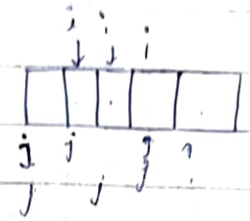
Bagging → Training a bunch of individual model in a parallal way Each model is trained by the random subset of data.

The result uses averaging the prediction over a collection of classifier.

Bosting → Training a bunch of individual model in a sequential way Each individual model learn from their mistakes made by Previous base model the result uses weighted vote with the coll^n of classifier.



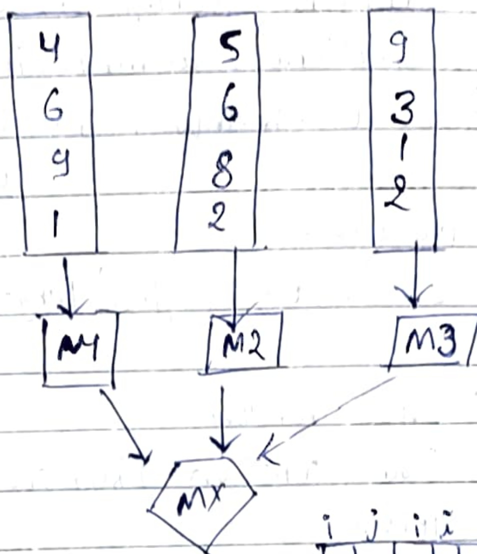
Bagging ⇒

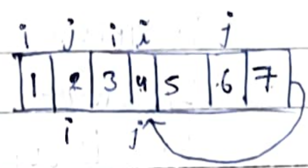


Ex:- of Bagging ⇒ Random forest.

Boosting :-

1	w_1
2	w_2
3	w_3
4	w_4
5	w_5
6	w_6
7	w_7
8	w_8
9	w_9
10	w_{10}





Ex:- Of Boosting is Adaboost Algorithm.

left SVM
regularisation

HTML ka parent Parentelement → Null
Parent Null → document

Full Stack :-

Difference b/w .parentelement & .parentNode
↓ null   ↓ document

.setAttribute
.innerHTML
.parentElement
. child. (no. of childs)
↓ Element count (gives you the count).
.children
.childNodes

Event Listener.   I Event bubbling :- Important benefit of event listener
Event handling

$A \rightarrow$ ...

$A \rightarrow az \;/\; bz \;/\; dz$

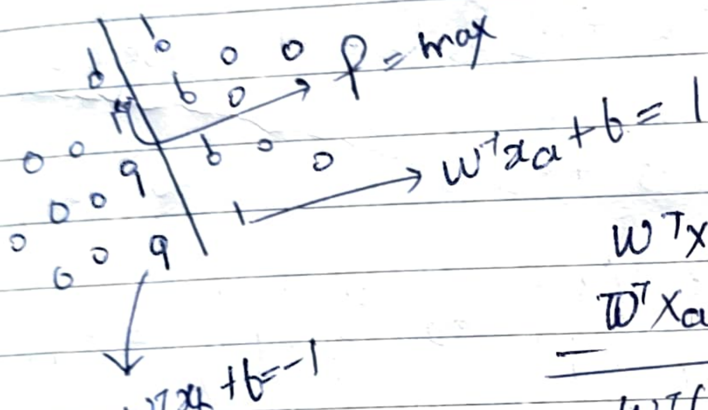
$Z \rightarrow \alpha \;/\; \alpha Z$

<u>Machine Learning</u>

SVM → classifier (to findout the hyperplane b/w the data-points it will be a line when their are two points only).

$$w^T x_i + b \ge 1 \quad \text{if } y_i = 1.$$

$$w^T x_i + b \le -1 \quad \text{if } y_i = -1.$$



$$w^T x_b + b = -1$$

$$w^T x_a + b = 1$$

$$w^T (x_b - x_a) = -2$$

$$w^T (x_a - x_b) = 2$$

$$x_a - x_b = \frac{2}{w^T}$$

$$\rho = |x_a - x_b| = \frac{2}{||w||}$$

Non Linear

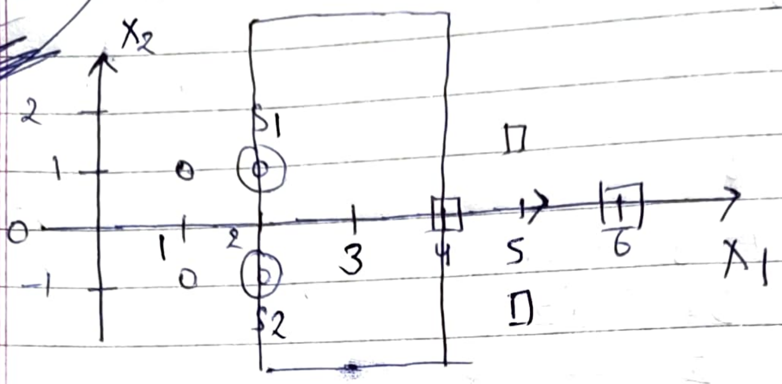
→ Kernal Trick

→ Gaussian Trick,

→

Question :-
point hai. draw karo
Equation nekalo
w and v nekalo.

Important ayega



$s_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $s_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ $s_3 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$

add bias $\Rightarrow$ $\tilde{s}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ $\tilde{s}_2 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$ $\tilde{s}_3 = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$

$\alpha_1 \tilde{s}_1 \cdot \tilde{s}_1 + \alpha_2 \tilde{s}_2 \cdot \hat{s}_1 + \alpha_3 \tilde{s}_3 \cdot \hat{s}_1 = -1$ (-ve class)

$\alpha_1 \hat{s}_1 \cdot \tilde{s}_2 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_2 + \alpha_3 \hat{s}_3 \tilde{s}_2 = -1$ (-ve class)

$\alpha_1 \tilde{s}_1 \cdot \hat{s}_3 + \alpha_2 \tilde{s}_2 \cdot \hat{s}_3 + \alpha_3 \hat{s}_3 \cdot \hat{s}_3 = +1$ (+ve class)

$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ $\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = -1$

$[2\times2+1\times1+1\times1]$ $\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = -1$

$\Rightarrow 6$

$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = +1$

After simplification we get :-

$$6\alpha_1 + 4\alpha_2 + 9\alpha_3 = -1$$
$$4\alpha_1 + 6\alpha_2 + 9\alpha_3 = -1$$
$$9\alpha_1 + 9\alpha_2 + 17\alpha_3 = 1$$

By solving the above 3,

$\alpha_1 = \alpha_2 = -3.25 \quad \alpha_3 = 3.5$

$\left(\hat{w} = \sum_i \alpha_i \hat{S}_i \right)$ w find karne ka formula.

$$\tilde{w} = \alpha_1 \tilde{S}_1 + \alpha_2 \hat{S}_2 + \alpha_3 \tilde{S}_3$$

$$\hat{w} = \alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

$$= (-3.25) \cdot \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + (-3.25) \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + (3.5) \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}$$

Last topic :- Regularization.

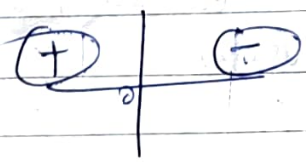
$y = wx + b$

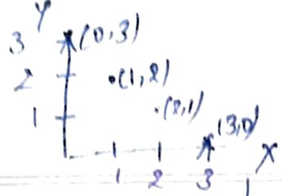
$w \Rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad b = -3$

$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow$ vertical line.

$\rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow$ horizontal line.

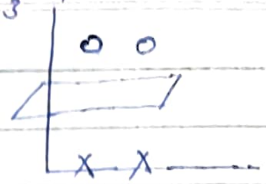
$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow$ /

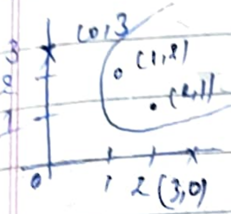




Machine Learning :-

Kernel Trick





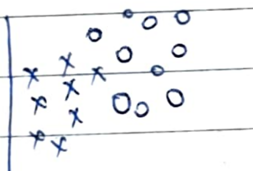
	X	O	O	X
	(0,3)	(1,2)	(2,1)	(3,0)
$x+y$	3	3	3	3
$x*y$	0	2	2	0
x^2	0	1	4	9

Q. what is the uses of kernel trick
↳ to transformation of the value.
to linear Separate.

- polynomial kernel fn.
- Gaussian Kernel fn.
- Radial Bases fn.

Q. What is SVM
Q. types Of kernel tricks
Q. Use of SVM
Q. type of SVM
Q. support vector
Hyper plain
Q. why kernel tricks needed.
Q. numerical.

Regularization :-



Technique is to prevent the model from over fitting by adding Extra information.

like :- Panalatity term (Extra Info^n) / our complexity term

$$y = B_0 + B_1 x_1 + B_2 x_2 + \dots + B_n x_n + b$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{n} B_j * X_{ij} \right)^2$$

Type of Regularization :-

i). Ridge (L2 - Regularization)
ii). Lasso (L1 - regularization)

DEVESH SRIVASTAV

Page No.
Date: / /

Ridge → In this technique the cost fn is altered by adding the penality term to it the amount of baise added to the model is called Ridge penality, we can calculate it by multiply λ to the squared weight of each individual feature.
The equation of cost function for Ridge is :-

$$\sum (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{n} \beta_j X_{ij}\right)^2 + \lambda \sum_{j=0}^{n} \beta_j^2$$

weights

Lasso → Least Absolute and selection operator.

It is similar to Ridge Except that the penality term contains only absolute weights instead of square of weight

$$\sum (y_i - \hat{y}_i)^2 = \sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{n} \beta_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{n} \beta_j$$

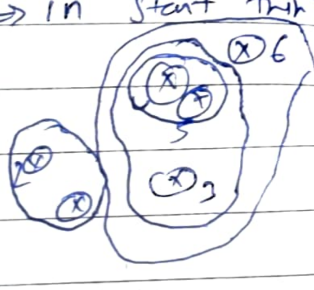
Q. Regularization, types of regularization
trade off baise v Variance.

Agglomerative Hierarchical clustering :-

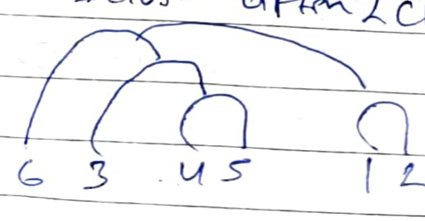
Drawback K

K mean depends upon the distance to clustroid.

Agglomerative clustering :-
⇒ in start think every point a cluster.



1 cluster after 2 cluster.



← Dendogram
↓
store Info how to merge.