Harbirat Ram (1)

GURU NANAK DEV ENGINEERING COLLEGE, LUDHIANA

(AN AUTONOMOUS COLLEGE U/S 2(f) & 12(B) OF UGC ACT - 1956

AICTE Approved, Punjab Govt. Aided Status ISO : 9001:2008 Certified Affiliated to I.K. Gujral PTU Jalandhar

IEI Accredited UG Programmes, Institute Accredited by NAAC (A Grade) & TCS

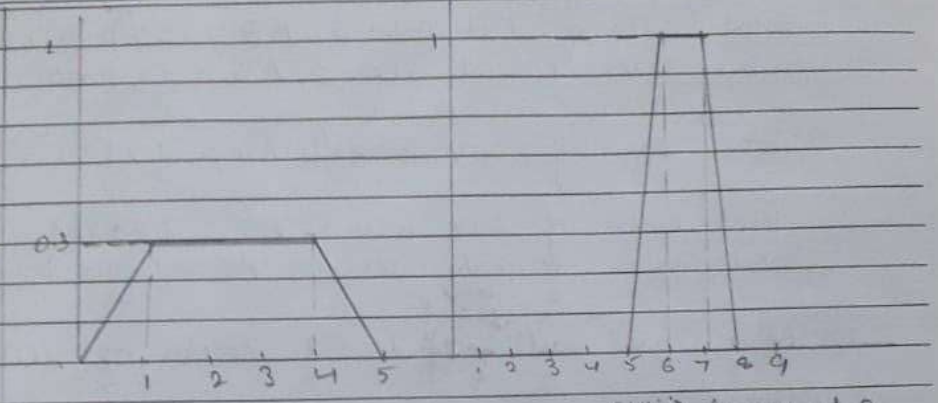
Dated 30-05-22 Class Btech Sec IT A2

Class Roll No. 1921036 Subject IML

University Roll. No.	Signature of Invigilator
1905334	

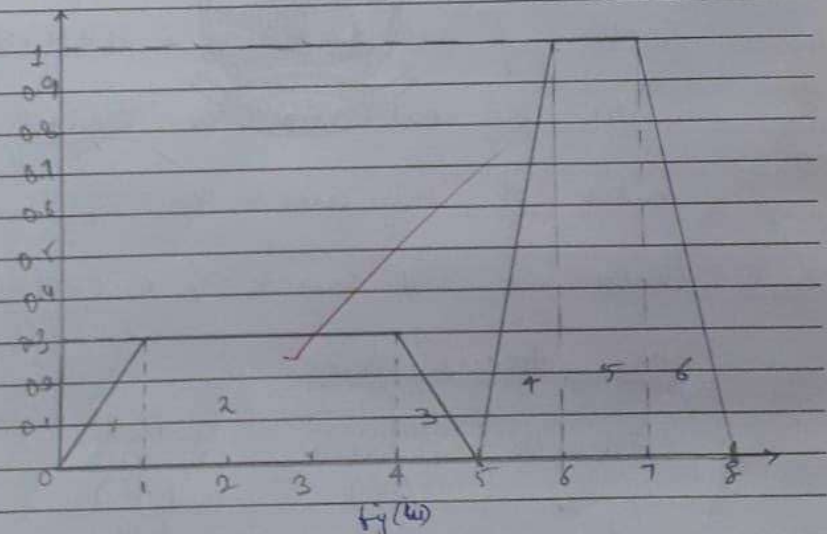
Q. No.	1	2	3	4	5	6	7	Total Marks	Sig. of Examiner
Marks	2	2	4	4	4	8		24	

Q6.



Fig(i) Fuzzy set 1 Fig(ii) fuzzy set 2

Combining fig(i) & fig(ii)



Fig(iii)

After combining both the fuzzy sets, we will divide the area into sub areas.

Here after division, we have 6 sub areas.

Now, we will calculate the total area

of all sub areas.

Total area of Sub area 1, $A1 = \frac{1}{2} \times 1 \times 0$

$= \frac{0.3}{2} = 0.15$

Total area of sub area 2, $A2 = 3 \times 0.3 = 0.$

Total area of sub area 3, $A3 = \frac{1}{2} \times 1 \times 0.3$

Total area of sub area 4, $A4 = \frac{1}{2} \times 1 \times 1 = \frac{1}{2}$

Total area of sub area 5, $A5 = 1 \times 1 = 1$

Total area of sub area 6, $A6 = \frac{1}{2} \times 1 \times 1 = 0.$

Now we will find the center of suba

Center of sub area 1, $x_1 = \frac{0+1+1}{3} = \frac{2}{3} = 0.66$

Centre of sub area 2, $x_2 = \frac{1+4}{2} = \frac{5}{2} = 2.5$

Center of sub area 3, $x_3 = \frac{4+4+5}{3} = \frac{13}{3} = 4.$

Center of sub area 4, $x_4 = \frac{5+6+6}{3} = \frac{17}{3} = 5.6$

Center of sub area 5, $x_5 = \frac{6+7}{2} = \frac{13}{2} = 6.5$

Center of sub area 6, $x_6 = \frac{7+7+8}{3} = \frac{22}{3} = 7.$

Putting in tabular form:-

Sub Area No.	Total area of subarea (A_i)	Center of sub area (x_i)	$A_i x_i$
1	0.15	0.66	0.099
2	0.9	2.5	2.25
3	0.15	4.33	0.64
4	0.5	5.66	2.83
5	1	6.5	6.5
6	0.5	7.33	3.66

$\sum A_i = 3.2$

$\sum A_i x_i = 15.97$

GURU NANAK DEV ENGINE

Deffuzified v

∴ The defuzzi combinat 4.99

Q5: Production b Production Defective Contribution

Contributi

P(M1) =

P(M2)

P(D) =
P(M1|D
P(M2|D
P(D|M
Probability by

∴ Probabil by

Defuzzified value, $x^* = \frac{\sum_{i=1}^{n} A_i x_i}{\sum_{i=1}^{n} A_i}$

$x^* = \frac{15.97}{3.2}$

$x^* = 4.99$

∴ The defuzzified value, x^* for the combination of 3 fuzzy sets is 4.99.

8

(ii) Production by Machine 1 = 35 screwdrivers/hr
Production by Machine 2 = 45 screwdrivers/hr
Defective screwdrivers = 1.5% in overall
Contribution of Machine 1 in defective parts = 50%
Contribution of Machine 2 in defective parts = 50%

$P(M1) = \frac{35}{80} = 0.437$

$P(M2) = \frac{45}{80} = 0.563$

$P(D) = 0.015$
$P(M1|D) = 0.5$
$P(M2|D) = 0.5$
$P(D|M1) = ?$

Probability that screwdriver is produced by machine 1 is defective,

$P(D|M1) = \frac{P(M1|D) \times P(D)}{P(M1)}$

$= \frac{0.5 \times 0.015}{0.437}$

$= 0.017$, i.e, 1.7%

4

∴ Probability that screwdriver is produced by Machine 1 is defective is 1.7%.

Q4

	Logistic Regression	K-NN	Naive Bayes theorem	S.V.M
Definition	Logistic regression is used to classify datasets with sigmoid function.	K-Nearest Neighbour looks for the similarity of data point with other clusters & assign it accordingly	Naive Bayes is based upon Bayes theorem that calculates using conditional probability. $P(A\|B) = \frac{P(B\|A) \times P(A)}{P(B)}$	Support vector classifier with the help of decision boundary or hyperplane
Advantages	Good performance - works well for linear datasets.	- can be very useful for high dimensional datasets.	- Very helpful in probabilistic analysis. - Gives more accurate results.	- Works efficiently for small or medium datasets
Disadvantages	Assumes linearity between dependent & independent variables.	- May lead to some issues if dataset is too large. - Value of k should be correct	- Sometimes it give 'Zero probability problem'	- May lead to overfitting - Not much accurate for too large dataset
Example	Spam filtration	Dividing groups, or in gender classification	In finding defective probability for manufacturing purposes.	In segmentation - Class marks of students etc.

We have 2 more classification models:-
(i) Decision Tree Classifier:-
=> It has a tree-like structure & have root nodes, branches & leaf nodes.
=> The root node contains whole dataset, branches contain subtrees decision & leaf node is

GURU NANAK DEV ENGINEERING COLLEGE, LUDHIANA
(AN AUTONOMOUS COLLEGE U/S 2(f) & 12(B) OF UGC ACT - 1956)
Accredited by NAAC (A Grade) & TCS

Date: 18/3/23 Class: IT Sec: A2
Class Roll No.: 1921036 Subject: IML
University Roll No.: 1901334

Q. No.	1	2	3	4	5	6	7	Total Marks	Sig.
Marks									

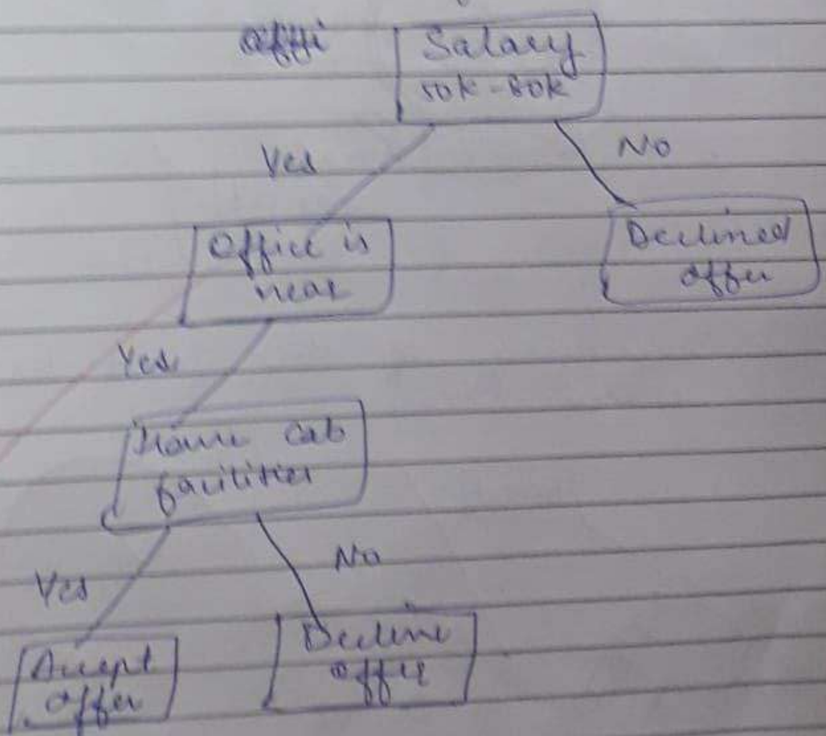
final decision.

Advantages :- Very helpful in classifying
true/false etc datasets
· less complex than random f
· efficient & easy to implement

Disadvantages :- May lead to overfitting,
· Not much suitable for large datasets.

Example :- To classify if the candidate
Random will win elections or if
student will get admission in
the reputed college.

Visualization: lets take an example of some
job offer :-



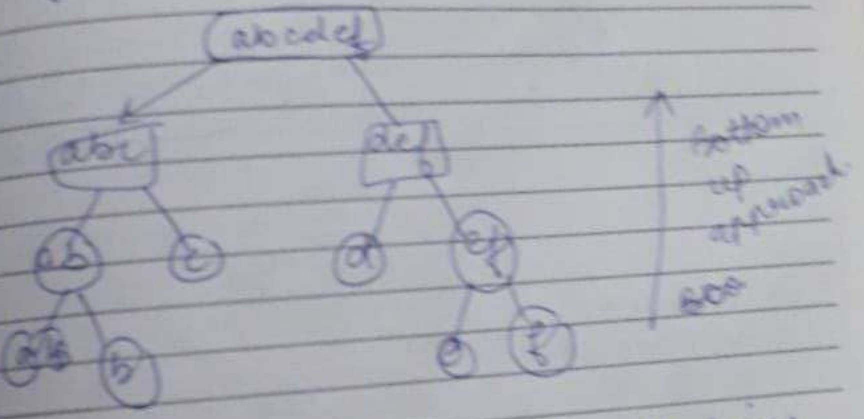
#2: Random Forest Classifier:
: It is madeup of a lot of decision
trees and is very huge.
: It has a tree like structure.

Hierarchical Agglomerative clustering

It comes under unsupervised machine learning technique i.e. the dataset provided is unlabelled. In this we form a tree-like or dendrogram in form of hierarchy structure.

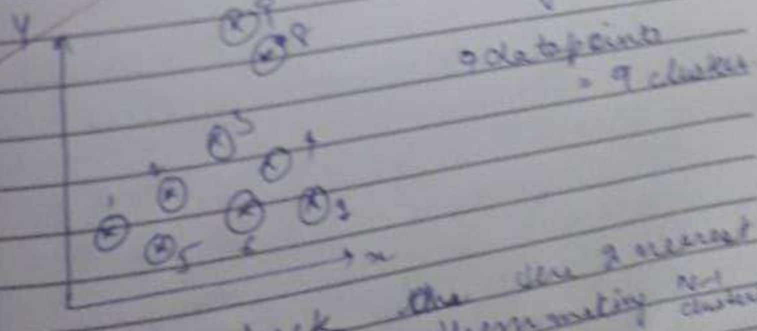
This is a bottom up approach i.e. it makes a so only one cluster at the end.

It is also known as Agglomerative nesting i.e. AGNES.

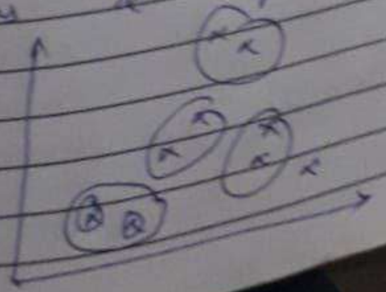


Steps/Working of Agglomerative Clustering

Step 1. All the datapoints are considered as clusters i.e. no. of datapoints = no. of clusters.



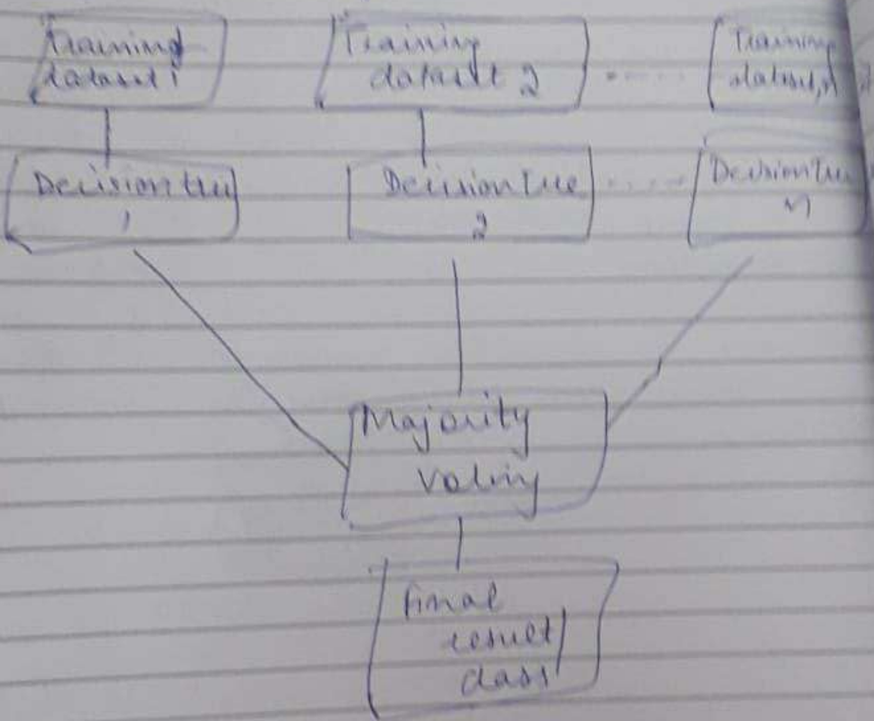
Step 2. Now we check the the 2 nearest clusters & group them making N-1 clusters.



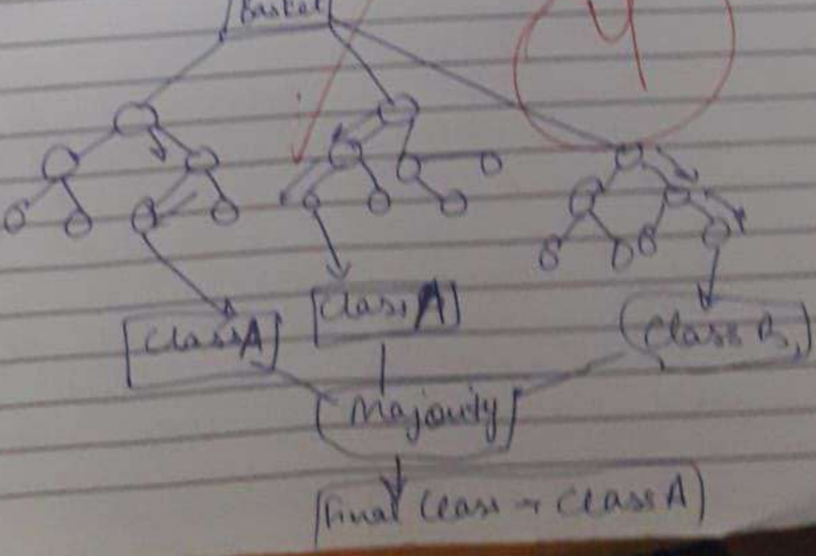
Advantages: can help overcome problems faced by decision tree algorithm
- More effective than decision tree

Disadvantages - It is very complex.
- May have chances of error during the process.
- May lead to overfitting.

Visualization & Example:



Example: Classifying if the fruit belongs to which class.

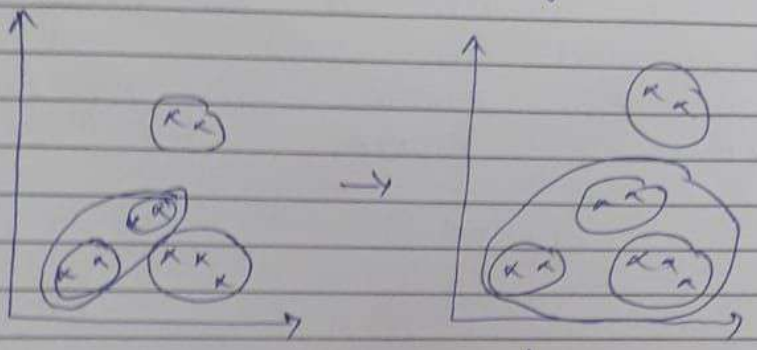


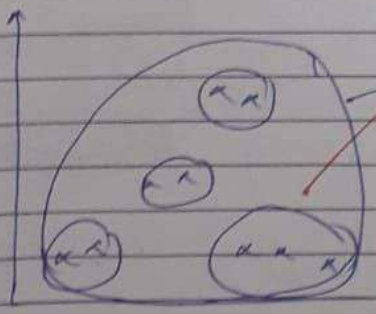
4

Step 3:- Again we will group 2 nearest clusters making no. of clusters N-1

Step 4:- Repeat step 3 until we get only one cluster.

Step 5:- All clusters are combined & made a single cluster.





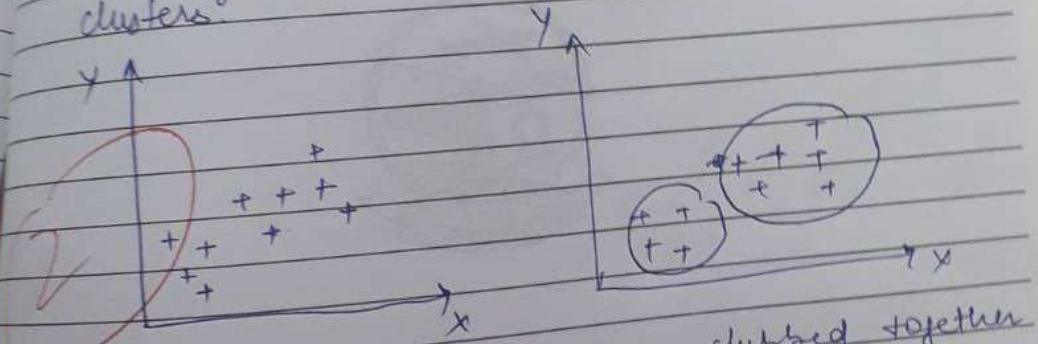
Eg. students of same standard in different class groups.

We can find the nearest clusters using diff. distance metrics like Euclidean distance.

⇒ AGNES is the inverse of Divisive Hierarchical Clustering.

⇒ AGNES is less complex & is less efficient as it doesn't take decision consider all the global datapoints & just starts combining or grouping the local datapoints. It doesn't consider or think before acting.

Clustering is called ~~unsupervised~~ unsupervised learning because of the ~~datasets~~ unlabelled datasets.
The datasets provided in unsupervised learning are not organised & labelled.
Also the process of models is done under no supervision.
As the dataset is unlabelled, it have scattered datapoints, which are grouped ~~together~~ by forming clusters.



The similar datapoints are clubbed together in clusters.

Ques 1. ⇒ K in k-mode clustering refers to the number of clusters, taken for the clustering process.
The very first step of k-mode clustering is defining the k value of 'k'.
⇒ In most cases the value of K is taken randomly at first.
⇒ At the end, ~~depending upon~~ after solving all iterations, we get the final number of clusters, which is also ~~in most cases~~ same as in

⇒ In clustering, we can also solve the value of k using the elbow method (i.e. solved using WCSS). This method is mostly used in k-means clustering.

⇒ K-Mode clustering is used for categorical data i.e. when we don't have numerical datapoints.