Roll: 1820416

CSE-6110

Ans! to the Ques 02.

2 (a)

We can tell the right numbers of cluster using Elbow method.

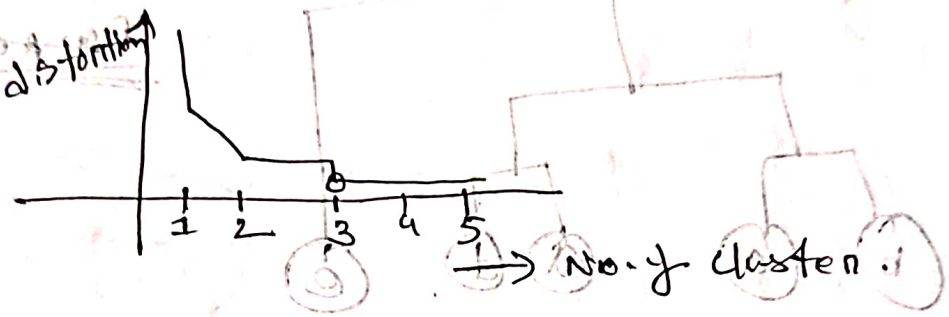


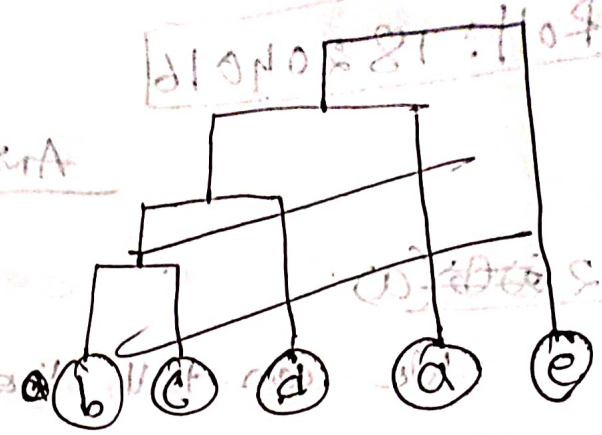
Fig: Elbow finder method.

In Elbow finder method we take no. of cluster as when the Elbow is found & from where the remaining no. of cluster gives no Elbows & it gradually reaches into 0.

In the fig we can see 3 is the no. of cluster we can use.

Agglomerative cluster.

	Age	Income
a	27	70
b	29	90
c	29	61
d	28	60
e	42	15



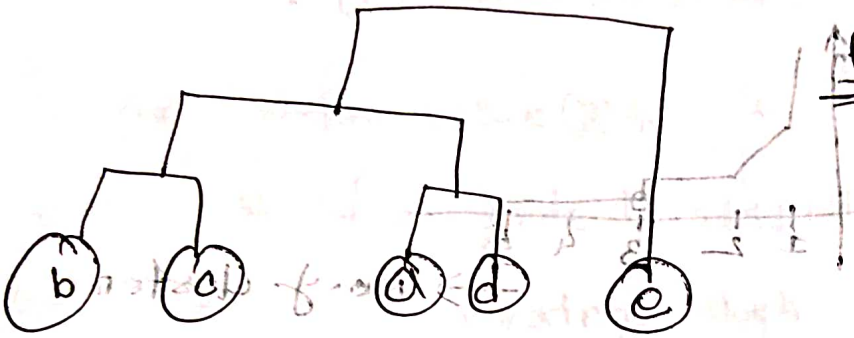


Fig: Agglomerative cluster.

2(b)

1) density reachability: density reachability is called if we can reach to through one core point 'S' to 'R' by core point in it.

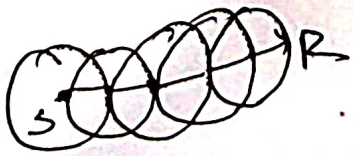


Fig: density reachability.

ii) density connectivity:

If a point 'P' connect to a point 'Q' through the border point 'm' from both 'P' & 'Q' then this is called density connectivity.

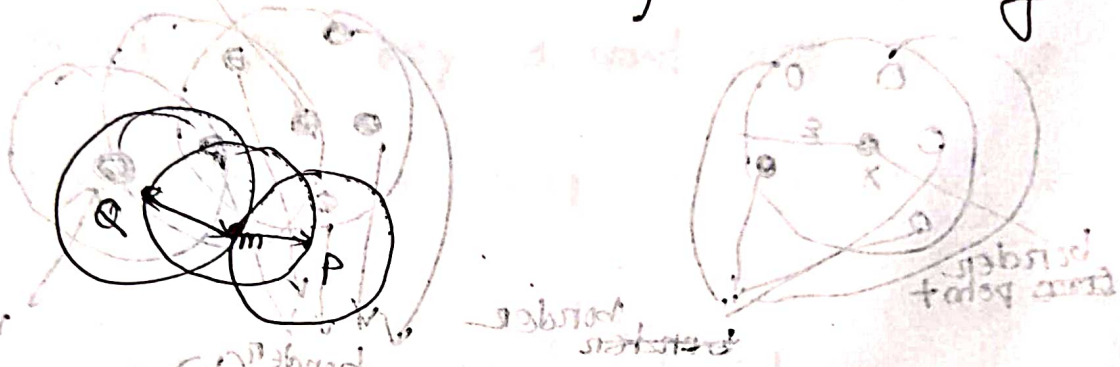


Fig: density connectivity

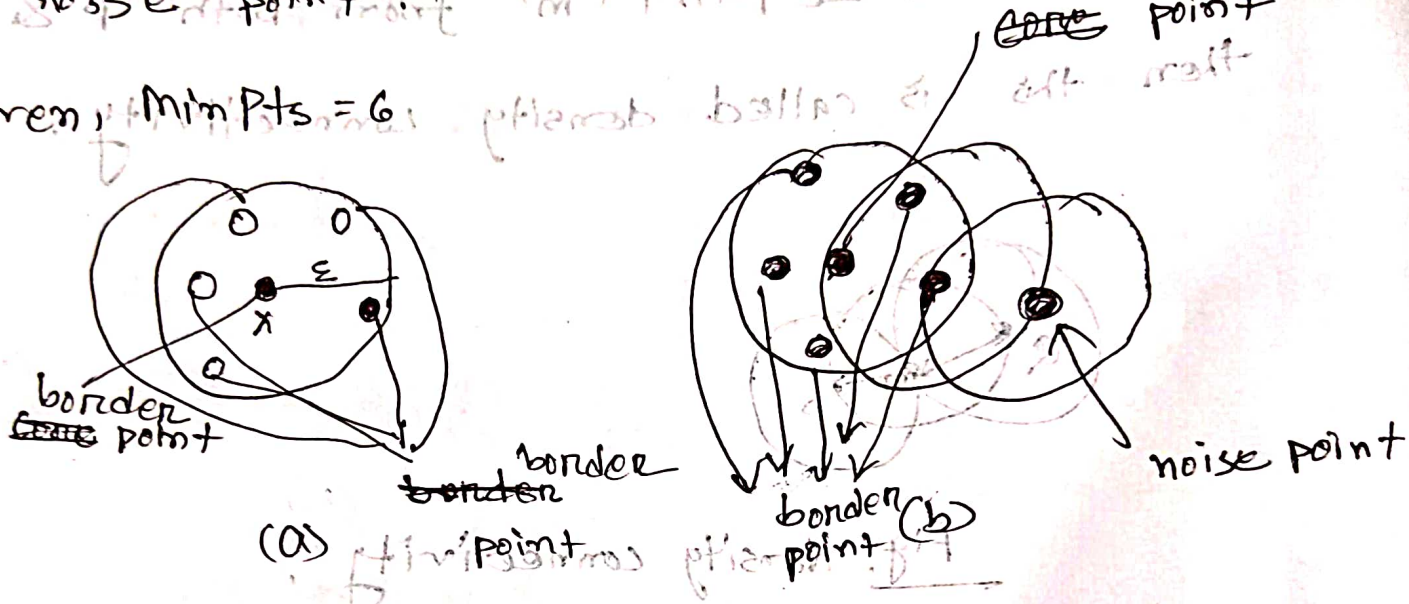
identify core point, border point, noise point —

1. core point: core point is a point from where the no of eps-neighbor found such that the no. of eps-neighbor is equal or greater than minPts.

2. border point: Border point is a point from where the no. of eps-neighbor found i.e. the no. of eps-neighbor is less than the minPts.

3. noise point: A point that is neither a core point or border point is known as noise point.

Given, MinPts = 6



Ans! to Ques! NO: 01

1(b)

Cause of over fitting in dt:

1. When decision tree generates decision node with only 1 leaf node.

2.

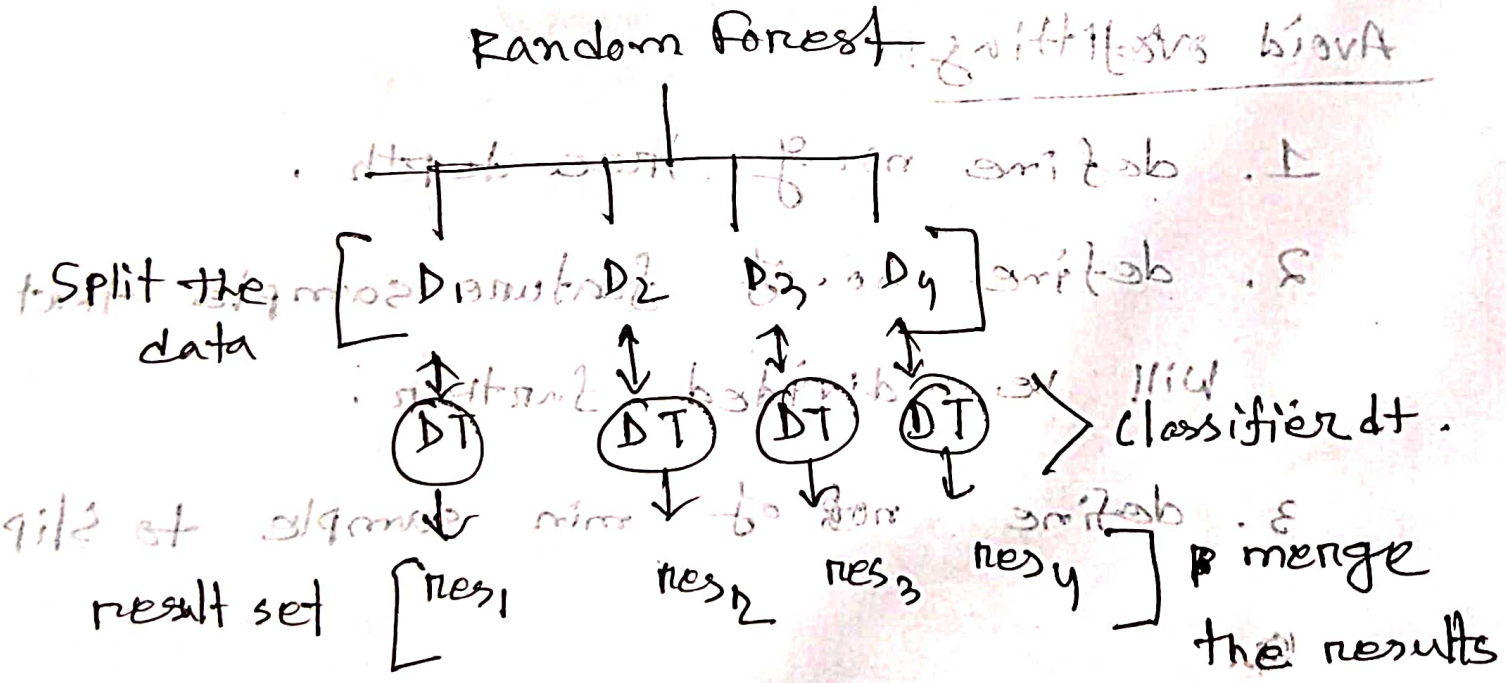
Avoid overfitting:

1. define no. of tree depth.
2. define no. of feature sample that will be divided further.
3. define no of of min sample to slipt.

Tree based dt model works on both categorical and regression data when relation between data has higher information gain, lower entropy.

1 (a)

Random forest divides data into subset of data then apply the decision tree based algorithm on each partition of the data and finally merge the output to produce model output.



A	B	C	Y
0	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

Here, 1 = T

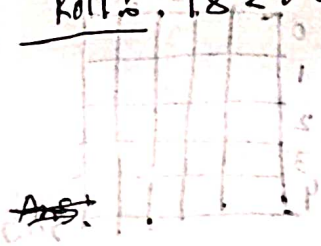
0 = F.

Gini score calulate as $(p^2 + q^2)$.

Here, p = probability of successfull

q = probability of failtoure.

Roll: 18204016 Code-CSE-6110 Date: 9.8.2021



~~Ans:~~

Ans: to the Ques.No: 04

4 (a)

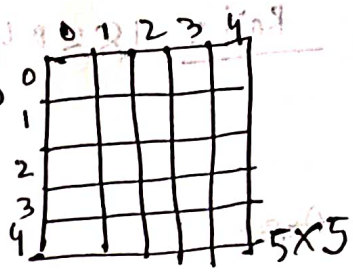
pooling on convolution Neural Network:

pooling in CNN contains few steps —

1. define window size. (usually (2x2)/(3x3))
2. define window stride (usually 2)
3. the window will pass throughout the image by the defined stride and take the maximum value every time from the window.
4. Taking the maximum value from the window it build a new smaller image compare to input image.

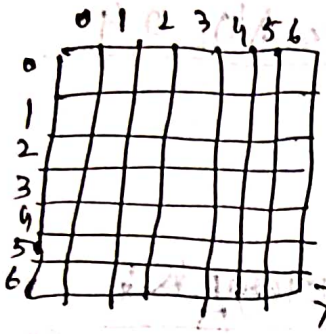
pooling uses CNN to pool out the maximum value from the defined window size in the input image.

In CNN, if input size is 5×5 that is



input

& if the filter size 7×7 that is,



filter

then the output size will be the size of input image that is 5×5 as the input image size is less than the filter.

4(b)

The difference between traditional Neural Network & Recurrent Neural Network (RNN) are written bellow:

Traditional Neural Network

1. Inputs of every layer are independent in here.
2. It does not have any memory cell.

3. As In (NN) it doesn't have any memory cell It doesn't ~~contains~~ stores output. ~~In every~~ layer.

4. Example: CNN, perceptron, backpropagation

Recurrent Neural Network (RNN)

1. Inputs of every layer are not independent as Traditional Neural Network.

2. It has memory cell to store output for next time use.

3. As it can stores the output for next time input it works well in text generation.

Ex: RNN, LSTM

RNN works better in text data.

In text, data analysing it is important to know the previous text or data to generate next data or text.

RNN ~~works~~ model has memory cell to store output from every layer, such that the next time it can be used as the input.

So In text data ~~RNN~~ unlike CNN or traditional neural Network RNN works better.

Ex: 1. "I have a bought a ____ ".

Here in the blank space RNN can generate pen or book or other ~~the things~~ related to buy. If it predicts or generate play. we can not buy play. so it is irrelevent to the context.

Ans: to the Ques NO: 03

3(a) If the learning rate is too low then the model could take longer time to train on the given trained dataset.

It will be worst it it comes in a large dataset.

If the learning rate is too ~~low~~ high then the model could overlook the optimum Error point hence it will give bad performance on the test data.

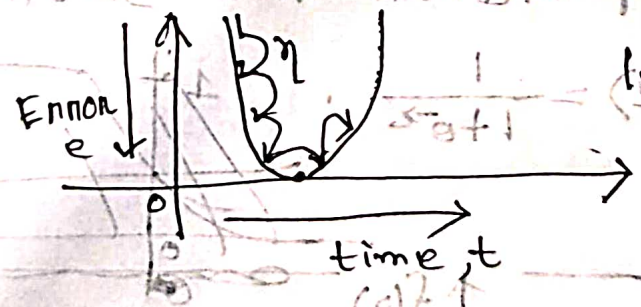
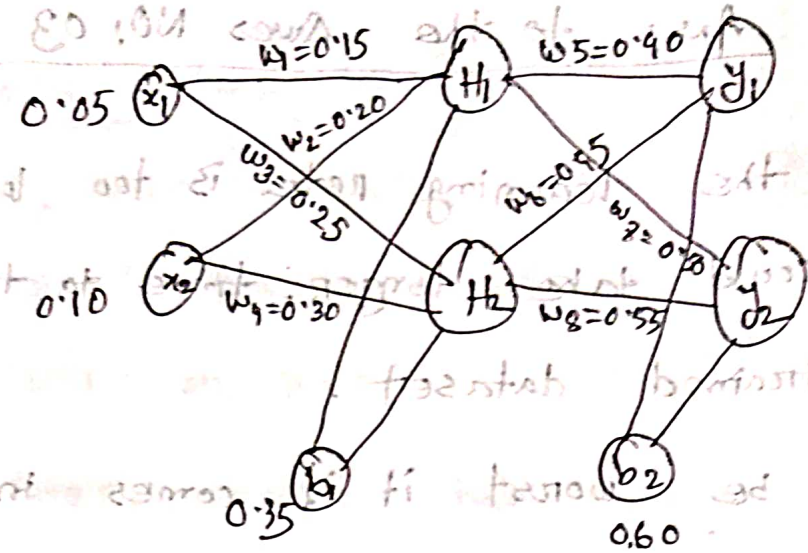


Fig: when learning rate is too high.

Given,



Target value.

$T_1 = 0.01$

$T_2 = 0.99$

learning rate $\eta = 0.25$.

i) Activation function.

As the output is from 0 to 1, so we should use sigmoid activation function.

$$f(z) = \frac{1}{1+e^{-z}}$$

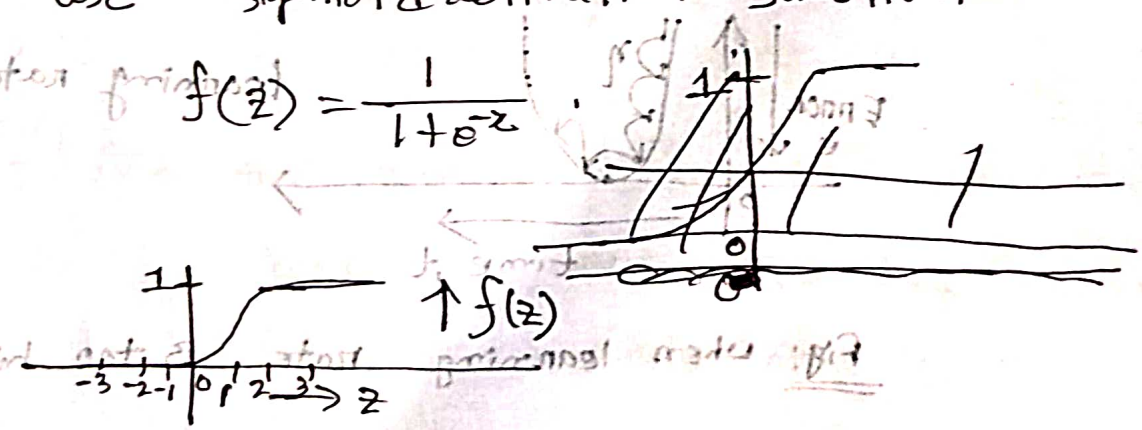


Fig: sigmoid Activation function.

for hidden layer we can use ReLu activation function. $A(x) = \max(0, x)$

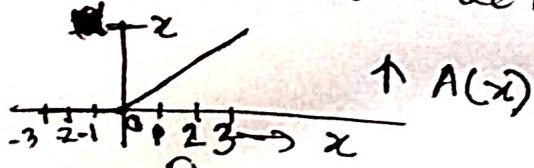


Fig:

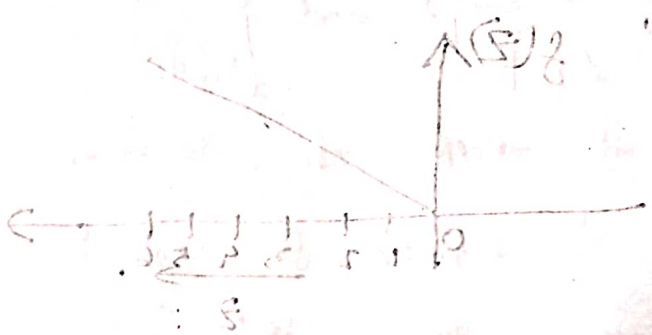
As the ReLu function gives the max value from (0, x). It could help to get better result.

ii) Calculation of Error:

$$E = \sum \frac{(\text{Actual output} - \text{Expected output})^2}{\text{Expected output}}$$

$$E = \sqrt{\frac{\sum (\text{Actual output} - \text{Expected output})^2}{\text{Expected output}}}$$

iii)



3(b) **Relu Activation!**

Relu Activation gives output range from (0 to ∞). It defines as ——

$$F(z) = \text{Max}(0, z).$$

if the given value(z) to it is less of equal to 0 then it will give output = 0; if the given value (z) is greater than 0, it will give output z.

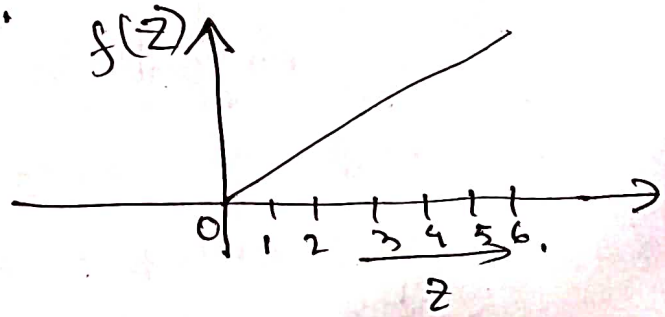


Fig: Relu Activation function.