

Aman Menta -

Roll No - 212119010

Sub :- CSOE18

CT2

1. (a) only the direction of relationship.

4 (b) Spearman Correlation.

5 (b) facilitate to plot the dataset in lower dimensional space

(c) facilitate to identify the direction of spread of data / information and include only these axes in the plot.

6 (b) overfitting

(c) having low bias and high variance.

8 (a) Yes.

2 No,

9.

Given Total Sample provided = 150
No of apple = 100. (actually)

No of orange = 50. (actually)

aim is to find apple so, Apple is positive (True)

Orange is negative

Actual class

predicted class.		Actual class	
		apple	orange.
	apple.	80.	10.
	orange	20.	40

$$TP = 80$$

$$TN = 40$$

$$FN = 20$$

$$FP = 10$$

7.

The classification model to detect the criminal record is.

		predicted class.	
		public	criminal
Actual class	criminal	10	113
Actual class.	public	60	8

↓

Convert this in the form of

		Actual class	
		public criminal	criminal public
predicted class	criminal criminal	113	8
	public public	10	60

so,

no of classes = 2.

Total no of sample. = total actual

$$= 8 + 60 + 113 + 10$$

$$= 68 + 123$$

$$= 191$$

Sample in public = 68

Sample in criminal class = 123

iv) since criminal is positive.

$$TP = 113.$$

$$TN = 60$$

$$FN = 10$$

$$FP = 8.$$

v)
$$\text{accuracy} = \frac{TP + TN}{\text{Total},..}$$

$$= \frac{113 + 60}{191}$$

$$= 90.575$$

which indicate it is a good one.

3.

vi) — Yes, age column need to be feature scaled, so that the data doesnot recult in SKEWED distribution becaus of the outlier '1000'

3.

iii)

we can identified whether speed and acceleration are related or not by using pearson coorelation after finding the covariance for finding the direction. If the graph comes nearly linear or value of pearson correlation is near to 1 or -1. then they are highly correlated.

(X value)
speed

(Y value)
acceleration

100.

60.

120.

70

140

75.

160.

85

180.

100.

u_{speed} = 140

218

$$\sum (X - \bar{X})^2 = 4000 = SS_x$$

$$\sum (Y - \bar{Y})^2 = 930 = SS_y$$

X and Y Combination.

$$N = 5.$$

$$\sum (x - \bar{x})(y - \bar{y}) = 1900.$$

R correlation

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{SS_x SS_y}}$$

$$= \frac{1900}{\sqrt{4000 \times 930}}$$

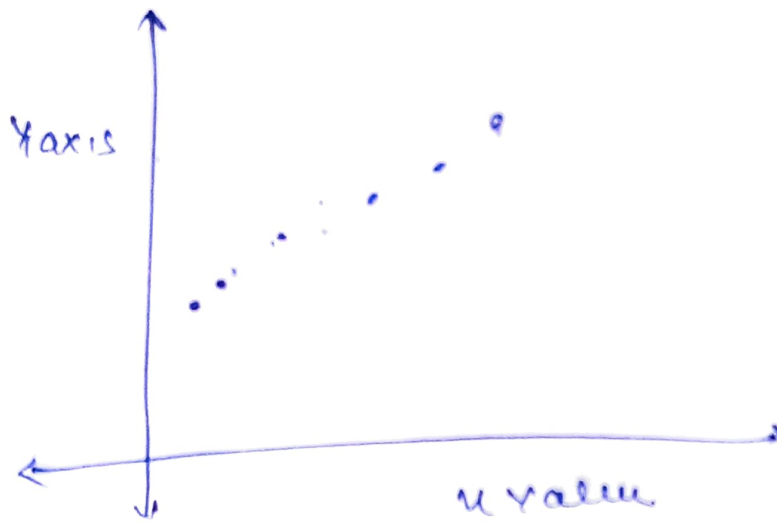
$$= 0.9851.$$

Meta numerical cross check 0.9851

It has Pearson correlation

$$= \underline{\underline{0.9851}}$$

So, hence it is correlated



10. (c) Recall vs false positive rate

11. (b) Only the number of cluster to be formed.