Name: Shubham Dutta

Year : 2nd

Stream : CST

Enrollment: 12019009022112

Registration: 304201900900752

Paper Name : Artificial Intelligence & Machine Learning Advanced.

Paper Code : PCCCS405

Signature : Shubham Dutta

Date : 16th March 2021

Answer

1. A)

ID	Student	Credit Rating	Actual	Predicted	Outcome
1	0	Fair	0	1	FN
2	0	Excellent	0	0	TN
3	0	Fair	1	1	TP
4	1	Excellent	0	0	TN
5	1	Excellent	1	0	FP
6	0	Fair	0	0	TN
7	0	Excellent	1	1	TP
8	1	Fair	1	0	FP
9	0	Excellent	0	1	FN

So, Total

$FN = 2 \quad TN = 3$

$FP = 2 \quad TP = 2$

So, Accuracy $= \dfrac{TP + TN}{TP + FP + TN + FN}$

$= \dfrac{5}{9} = 0.55 = 55\%$

$\therefore$ Precision $= \dfrac{TP}{TP + FP} = \dfrac{2}{4} = 0.5 = 50\%$

$\therefore$ Recall $= \dfrac{TP}{TP + FN} = \dfrac{2}{4} = 0.5 = 50\%$

Here,
TP → True Positive
TN → True Negative
FP → False Positive
FN → False Negative

2. B)

x	y	x^2	xy
95	85	4025	8075
85	95	7225	8075
80	70	6900	5600
70	65	4100	4550
60	70	3600	4200

$\Sigma x = 390$ $\Sigma y = 385$ $\Sigma x^2 = 31.15$ $\Sigma xy = 30,50$

To find the value of coefficient we need to minimize the object the fn using the derivative

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

$$= \frac{5 \times 30500 - 390 \times 385}{5 \times (31,1500) - (390)^2}$$

$$= \frac{(152,500 - 150,150)}{(155,750 - 152,100)}$$

$$= \frac{2350}{3650} = 0.644$$

$$\beta_0 = \frac{\Sigma y - \beta_1 \Sigma x}{n}$$

$$= \frac{385 - 0.644 \times 390}{10}$$

$$= 13.384$$

so, linear regression : y is

$y = \theta_0 + \theta_1 x$

$\Rightarrow y = \beta_1 \cdot x + \beta_0$

$= 0.6447 x + 13.384$

$\therefore$ Pearson's coefficient is that

$\beta_1 = 0.644$ & $\beta_0 = 13.384$.

3b)

Underfitting	Overfitting
i) A statistical model on machine is said to have underfitting when it cannot capture the underlying trend of data	i) A statistical model is said to be overfitted when we train it with lot of data.
ii) underfitting destroys the accuracy of our machine learning model.	ii) In overfitting, model starts learning from the noise & inaccurate data enters in dataset
iii) It has high bias & low variance	iii) It has high variance & low bias
iv) It usually happens when we have less data to built an accurate model & also when we try to build a linear model with non-linear data	iv) It usually happens because of non-parametric & non-linear methods as these type of ML algorithm have more freedom in building model based on dataset

3b) 2nd part

Popular Methods of evaluating classifier

1. **Holdout method**

- Split data into train & test set (usually 2/3 for train & 1/3 for test). Learn model using train set & measure performance over test set.
- Usually used when there is sufficiently large data, since both train & test data will be a part.

2. **k-fold cross-validation**

- 1st step :- data is split into k subsets of equal size
- 2nd step :- each subset in turn is used for testing & the remainder for training.
- Performance measure averaged over all folds.

4. A) By the method of least squares,

by given data $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$

$y = ax + b$ (a, b are parameters)

$$E(a,b) = \sum_{n=1}^{N} \{y_n - (ax_n + b)\}^2$$

now,

$$\frac{\partial E}{\partial a} = 0, \quad \frac{\partial E}{\partial b} = 0$$

$$\therefore \frac{\partial E}{\partial a} = \sum_{n=1}^{N} 2(y_n - (ax_n + b)).(-x_n) = 0$$

$$\& \frac{\partial E}{\partial b} = \sum_{n=1}^{N} 2(y_n - ax_n + b).1 = 0$$

Now the eq.

$$\left(\sum_{n=1}^{N} x_n^2\right).a + \left(\sum_{n=1}^{N} x_n\right)b = \sum_{n=1}^{N} x_n y_n$$

By minimising error

$$\begin{bmatrix} \sum_{n=1}^{N} x_n^2 & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & \sum_{n=1}^{N} 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{pmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{pmatrix}$$

If matrix is invertible then,

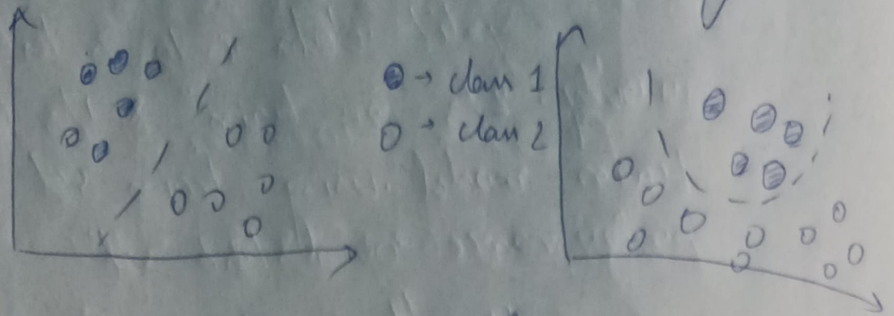
$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} x_n^2 & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & \sum_{n=1}^{N} 1 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{bmatrix}$$

where a, b are parameters of regression eq^n

5. b) Sigmoid Function

- It acts as an activation function in machine learning which is used to add non-linearity in ML model.
- It decides which value to pass as output & what not to pass.
- There are mainly 7 types of Activation f^n which are used in ML & Deep Learn

6 A) A decision boundary is the region of a problem space in which the output label ~~of a~~ is classified as ambiguous.



Here, dashed line is decision boundary

Decision boundary ~~sp~~ ~~sep~~ separates different classes.

i) An optimal decision boundary is one that minimizes the distance from points on both the classes of data. This classification observed by using 2 individu binary classifiers & considers different input features

- For binary classification, one decision boundary is the ~~s~~ surface in the feature space that separates the test inp. into 2 classes.

ii) In a statistical classification problem, with more than ~~...~~ 1 classes or multiple classes, a decision boundary is hyper surface that partitions underlying vector space into 2 or more sets, one for each class.

7. B) Decission trees are composed of 3 ma[illegible] parts :-

i) Decision Nodes
ii) Chance Nodes
iii) End Nodes

⇒ **Decision Nodes** :- It is a node in one activity at which the flow branches into several optional flows

⇒ **Chance Nodes** :- It is represented by a circle & shows the probability of certain results.

⇒ **End Nodes** - It shows the final outcome of a decision path.

→ There are 2 types of decision Trees. They are :-

i) Categorical variable decision Tree
ii) Continuous variable decision Tree.

→ There are 3 popular attribute selection measures

i) **Information Grain** :- Attribute with highest gain is chosen as the splitting attribute

ii) **Gain Ratio** - Ratio of information gain to the intrinsic information

iii) Gini Index :- Calculated by subtracting the sum of squared probabilities of each class from one.