Assignment-02

Name: Nikita Patra
Class: AIML-3B
Roll No.: 2100301530067

Ques-1) Given a confusion matrix for a classification task with two classes calculate
(i) Precision
(ii) Recall
(iii) F1-Score

n=165	Predicted No	Predicted YES
Actual NO	50 (TN)	10 (FP)
Actual YES	5 (FN)	100 (TP)

(i) Precision $= \frac{TP}{TP+FP} = \frac{100}{10+100} = \frac{100}{110} = 0.90 = 90\%$

(ii) Recall $= \frac{TP}{TP+FN} = \frac{100}{100+5} = \frac{100}{105} = 0.95 = 95\%$

(iii) F1 Score $= 2 * \frac{Precision * Recall}{Precision + Recall} = 2 * \frac{0.855}{1.85} = 2 * 0.462 = 0.924$
$= 92\%$

Ques-2) Find the line of regression for the following data set. Also check whether it is a best fit line.

X	2	4	6	8
Y	3	7	5	10

X	Y	X^2	XY	Yp
2	3	4	6	3.4
4	7	16	28	5.3
6	5	36	30	7.2
8	10	64	80	9.1
$\Sigma X = 20$	$\Sigma Y = 25$	$\Sigma X^2 = 120$	$\Sigma XY = 144$	

$$\beta_0 = \frac{(\Sigma Y)(\Sigma x^2) - (\Sigma xy)(\Sigma x)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{(25)(120) - (20)(144)}{4(120) - 400} = 1.5$$

$\boxed{\beta_0 = 1.5}$

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

$$= \frac{4(144) - (20)(25)}{80} = 0.95$$

$\boxed{\beta_1 = 0.95}$

$y_{p1} = \beta_0 + \beta_1(x_1) = 1.5\ \ 0.95(2) = 3.4$
$y_{p2} = \beta_0 + \beta_1(x_2) = 1.5 -, 0.95(4) = 5.3$
$y_{p3} = \beta_0 + \beta_1(x_3) = 1.5 + 0.95(6) = 7.2$
$y_{p4} = \beta_0 + \beta_1(x_4) = 1.5 + 0.95(8) = 9.1$

∵ The error margin is not neglegible
∴ It is not the best-fit line.

Ques-3) In a neighbourhood, 90% children were falling sick due flu and 10% due to measles and no other disease. The probability of observing rashes from measles is 0.95 and for flu is 0.08. If a child develops rashes, find the child's probabily of having flu.

P(flu) = 0.90 P(measles) = 0.10

P(rashes/flu) = 0.08 P(rashes/measles) = 0.95

$$P(flu/rashes) = \frac{P(rashes/flu) \cdot P(flu)}{P(rashes)} = \frac{0.08 * 0.90}{P(rashes)}$$

$$P(\text{rashes}) = P(\text{rashes}/\text{flu}) \cdot P(\text{flu}) + P(\text{rashes}/\text{measles}) \cdot P(\text{measles})$$
$$= 0.08 \times 0.90 + 0.95 \times 0.10 = 0.167$$

$$P(\text{flu}/\text{rashes}) = \frac{0.08 \times 0.90}{0.167} = \frac{0.072}{0.167} = 0.4311377 \approx 0.43$$

Ques-4) Differentiate between linear and logistic regression along with suitable examples.

Ans-4)

Linear Regression	Logistic Regression.
1. Linear regression is used for predicting the value of a continuous dependent variable based on one or more independent variables.	Logistic Regression is used for predicting the probability of a binary outcome based on one or more independent variables.
2. It assumes a linear relationship between the independent variables and the dependent variable.	It models the probability of the dependent variable belonging to a particular category using the logistic function.
3. The output of linear regression is a continuos value.	The output of logistic regression is a probability score betwfeen 0 & 1.
4. Example: Predicting house prices based on features like square footage, no. of bedrooms and location. Here the dependent variable (house price) is continous, and the independendent variables (features) can be linearly related to the house price.	Example : Predicting whether a customer will buy a product (1) or not (0) based on factors like age, income and previous purchase history. Here the dependent variable (buying decision) is binary, and the independent variables can influence the probability of making a purchase.

Ques-5) Explain Find S algorithm with a suitable example.

Ans-5) The Find S algorithm is a concept used in machine learning for learning a hypothesis from a given set of training data. It's a simple algorithm that finds the most specific hypothesis that fits all positive instances in the training data. Finding a maximally specific hypothesis.

1. Initialise h to the most specific hypothesis in H.

$$h_0 = \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$$

2. For each positive training instance x
   - For each attribute
   - If the constraint is satisfied by x.
   - Then do nothing
   - Else, replace attribute in h by the next most general constraint that is satisfied by x.

ample:	Sky	Air Temp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Iteration-1 x_1 = <Sunny, Warm, Normal, Strong, Warm, Same>
h_1 = <Sunny, Warm, Normal, Strong, Warm, Same>

Iteration-2 x_2 = <Sunny, Warm, High, Strong, Warm, Same>
h_2 = <Sunny, Warm, ?, Strong, Warm, Same>

Iteration-3 x_3 = <Rainy, Cold, High, Strong, Warm, Change>
∵ It is negative the previous hypothesis is kept as it is $h_3 = h_2$

Iteration-4 x_4 = <Sunny, Warm, High, Strong, Cool, Change>
h_4 = <Sunny, Warm, ?, Strong, ?, ?>

h_4 is the maximally specific hypothesis.

Ques-6) Explain EM algorithm. For which type of algorithms EM algorithms does work?

Ans-6) The Expectation-Maximization (EM) algorithm is a method used to estimate model parameters, especially when dealing with hidden or missing data. It's useful for algorithms involving latent variables, like mixture models (e.g. Gaussian Mixture Models) and problems with missing data.