No: Date: / /

Machine Learning January 2024

Q.

a) No, logistic regression is not good choice because it is meant for classification task, not for predicting continuous values like the price of laptops
A regression model, linear regression would be better choice for this problem.

b)

(i) Pruning in decision trees is a technique used to used to reduse the size of the tree. The purpose is to avoid overfitting the training set by simplifying the model. This helps to generaliza the model and get good performance by reducing the complexity of the model ~~using the~~ fro unseen data.

(ii) $y = ax + bx^2 + e$ is more appropriate to fit the training data better if the data shows a nonlinear pattern. Other one shows the linear Model. It can only used for linear data set. If data set is non linear, Model $y = ax + bx^2 + e$ provide a better fit because it can accommodate nonlinear trends.

No: Date: / /

Machine Learning January 2024

Q.

a) No, logistic regression is not good choice because it is meant for classification task, not for predicting continuous values like the price of laptops
A regression model, linear regression would be better choice for this problem.

b)

(i) Pruning in decision trees is a technique used to used to reduse the size of the tree. The purpose is to avoid overfitting the training set by simplifying the model. This helps to generaliza the model and get good performance by reducing the complexity of the model ~~using the~~ fro unseen data.

(ii) $y = ax + bx^2 + e$ is more appropriate to fit the training data better if the data shows a nonlinear pattern. Other one shows the linear Model. It can only used for linear data set. If data set is non linear, Model $y = ax + bx^2 + e$ provide a better fit because it can accommodate nonlear trends.

(c)

(i) The entropy formula is

$$H(x) = -\sum p(x) \log_2 p(x)$$

Passed	Failed
2	3

~~Pr~~ Pass Probability =

$$P(\text{passed}) = \frac{2}{5} = 0.4$$

$$P(\text{failed}) = \frac{3}{5} = 0.6$$

Entropy calculation

$$H(\text{Result}) = -0.4 \log_2(0.4) - 0.6 \log_2(0.6)$$

$$= 0.971$$

No : Date: / /

(ii) slept

No		Yes
Passed	failed	
2	1	

$H(No) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)$

$= 0.3896 + 0.5283$

$= 0.918$

$H(Yes) = \cancel{1 - 0.918}\ 0$

$= \cancel{0.082}$

$H(slept?) = \frac{2}{5} \times 0.\cancel{082} + \frac{3}{5}\ 0.918$

$= 0.5836$

Studied

No		Yes	
Passed	failed	P	F
1	1	1	2

$H(No) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) \times 2$

$= 1$

$H(yes) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)$

$= 0.918$

$H(studied) = \frac{2}{5} \times 1 + \frac{3}{5} \times 0.918$

$= 0.9508$

Attended Class

~~Passed~~

No		Yes	
Passed	Failed	Passed	Failed
0	1	2	2

$H(No) = 0 + -1 \log_2 (1)$

$= 0$

$H(yes) = -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - 1 \log_2 \left(\frac{2}{4}\right)$

$= 1$

$H(\text{Attended}) = 0 + \frac{4}{5} \times 1$

$= 0.8$

$\text{Gain}(\text{Slept}) = H(\text{Result}) - H(\text{Slept})$

$= 0.971 - 0.5536$

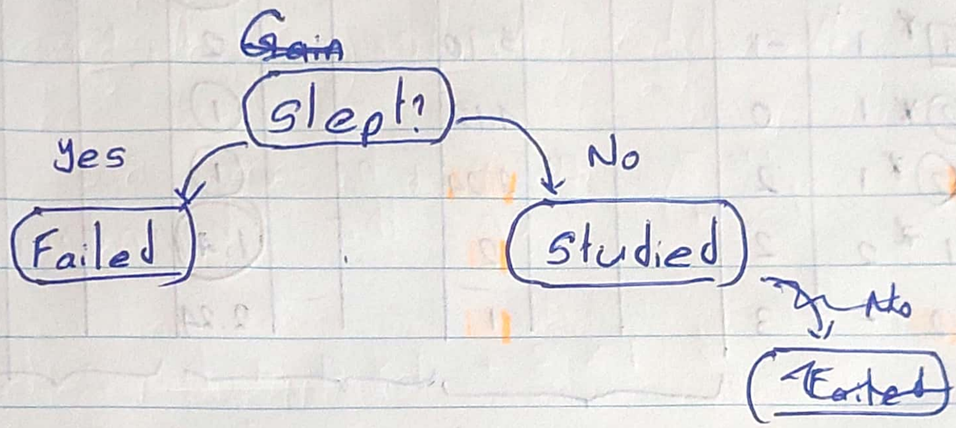
$= 0.4174$

$\text{Gain}(\text{Studied}) = H(\text{Result}) - H(\text{Studied})$

$= 0.971 - 0.9508$

$= 0.0202$

$$\text{Gain(Attended Class)} = H(\text{Result}) - H(\text{Attended})$$
$$= 0.971 - 0.8$$
$$= 0.171$$



Distance $= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

d) given

(a)	x	y	Point 1 x	Point 1 y	D	Point 2 x	Point 2 y	D	Point 3 x	Point 3 y	D	Point 4 x	Point 4 y	D
class 1 *	-1	1	2	4	4.24	1	1	2	1	1	2*	2	-1	3.61
class 2 *	0	1			3.61			1			1*			2.83
class 1 *	0	2			2.83			1.41			1.41*			3.61
class $\boxed{1}$ *	1	-1			5.10			2			2*			$\boxed{1}$
class 2 *	1	0			4.12			1			1*			1.41
class 2 *	1	2			2.24			1			1*			3.16
class 1 *	2	2			2			1.41			1.41*			3
class 2	2	3			1			2.24			2.24			4

3 nearest neighbour — 5 nearest neighbour — 7 — 1

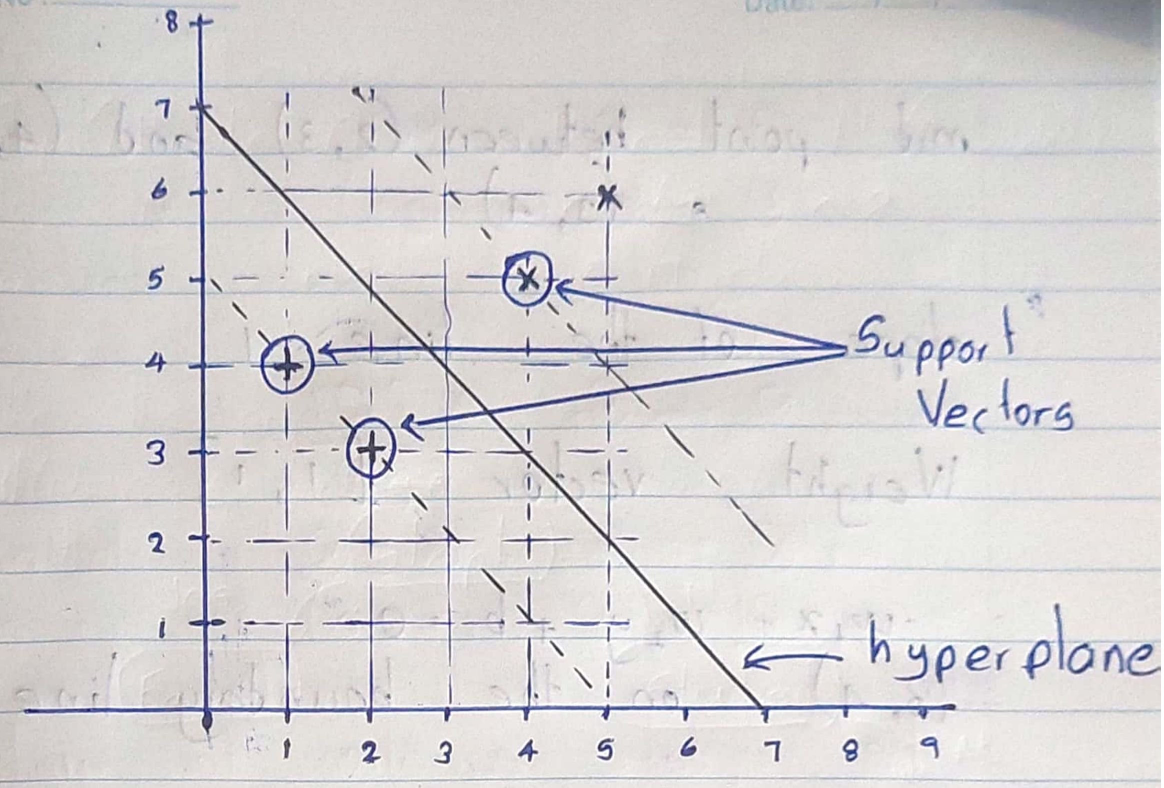
(i) Class 2

(ii) Class 2

(iii) Class 1

(iv) Class 1

e)



(i) $y = \underbrace{m}_{\text{weight}}x + \underbrace{c}_{\text{bias}}$

$y = -1x + 7$ hyperplane

~~weight = [-1]~~

~~bias = 7~~

$w \rightarrow$ parameters of the line

$w = \begin{bmatrix} -1 \\ 7 \end{bmatrix} \quad [-1 \quad 7]$

$x_1 = (1, 4)$

$x_2 = (2, 3)$

$x_3 = (4, 5)$

$w^T x_1 = \begin{bmatrix} -1 \\ 7 \end{bmatrix} [1 \quad 4]$

No : Date: / /

mid point between (2, 3) and (4, 5)
= (3, 4)

slope of the line = 1

Weight vector [1, 1]

$w_1 x + w_2 y + b = 0$

(3, 4) on the boundary line

$1 \times 3 + 1 \times 4 + b = 0$

$b = -7$

weight vector [1, 1]

bias $b = -7$

No :   Date: / /

Q_3
(a)

(i) Batch gradient descent
Stochastic gradient descent
Mini batch gradient descent

(ii) cost function

$$J(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2$$

we need partial derivatives of J respect to θ_1 and θ_2

$$\frac{\partial J}{\partial \theta_1} = 2\theta_1 \quad \sim \text{①}$$

$$\frac{\partial J}{\partial \theta_2} = 2\theta_2 \quad \sim \text{②}$$

The gradient of J

using ~~vector~~ matrix form

$$\nabla J(\theta_1, \theta_2) = \begin{bmatrix} \frac{\partial J}{\partial \theta_1} \\ \frac{\partial J}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} 2\theta_1 \\ 2\theta_2 \end{bmatrix}$$

using vector form

$$\nabla J(\theta_1, \theta_2) = \frac{\partial J}{\partial \theta_1} \underline{i} + \frac{\partial J}{\partial \theta_2} \underline{j}$$

$$= 2\theta_1 \underline{i} + 2\theta_2 \underline{j}$$

using gradient descent

$$\theta_i := \theta_i - \alpha \frac{\partial J}{\partial \theta_i}$$

using ①

$$\theta_1 := \theta_1 - \alpha \frac{\partial J}{\partial \theta_1}$$

$$\theta_1 := \theta_1 - \alpha\, 2\theta_1$$
$$\theta_1 := \theta_1 (1 - 2\alpha)$$

using (2)

$$\theta_2 := \theta_2 - \alpha \frac{\partial J}{\partial \theta_2}$$
$$\theta_2 := \theta_2 - \alpha\, 2\theta_2$$
$$\theta_2 := \theta_2 (1 - 2\alpha)$$

The update equation for gradient descent are

$$\theta_1 := \theta_1 (1 - 2\alpha)$$
$$\theta_2 := \theta_2 (1 - 2\alpha)$$

No:
Date: / /

b)
(i) D B
(ii) A B
(iii) A C
(iv) C D

c) Stratified sampling is technique where a dataset is divided into distinct groups, and samples are taken from each group proportionally

In machine learning, it helps ensure that both the training and testing datasets have a similar distribution, which is useful for handling imbalanced ~~bad~~ ~~to~~ data and improving model accuracy

d) Train - Test Split
It helps to assess how well a machine learning model will generalize to new unseen data while avoiding overfitting
Missing / Null Values Handling
Fills in or removes missing data to avoid biases and improve model accuracy

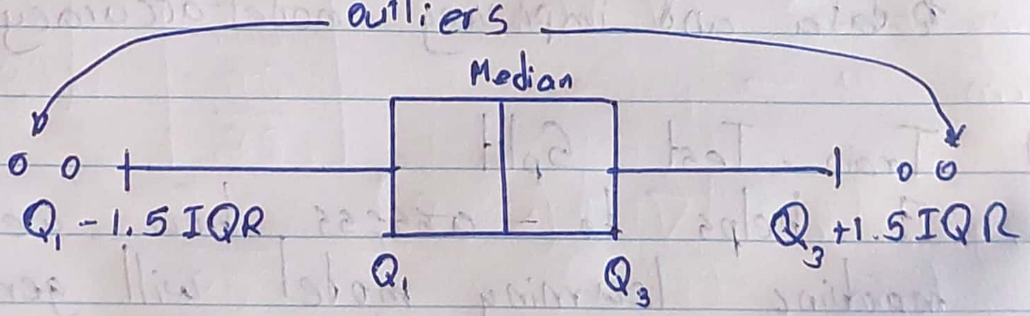
Scalling
Ensures that have similar ranges,
It importance for model performance.

~~Normalization~~
Encoding Categorical Variables
Converts categorical data into numerical
form, allowing machine learning models
to process it effectively

Treating Outliers and Duplicate
Duplicate ~~are~~ can be see every
data set removing Duplicate we
can avoid over fitting

e)

(i)



We remove the outliers to get
improve the model performance.

No : Date: / /

(ii) ~~For small data sets~~
* Anomaly Delection
* Fraud Detection