Student Answer Script View

MIT MPL-BTech-M Sc - MCA - 1st-3rd-5th and 7th Semester - Mid Term Examination - Sep 2024

Answer Sheet

Student Name:	HARSH GONDAL . .
Roll Number:	220962334
Course:	Computer Science and Engineering - Artificial Intelligence and Machine Learning
Year/Sem:	Semester 5
Subject Name:	MACHINE LEARNING
Exam Date:	23-Sep-2024

Q.No : 1)

Which type of linear regression method can be used if we have a training set with millions of features

Normal equation method [✓ Gradient descent method] Polynomial regression method Matrix method

Q.No : 2)

Suppose the features in a training set have very different scales. Identify the correct option regarding regression

[✓ Gradient Descent algorithm will converge fast]	Normal equation method requires scaling	Feature scaling is applicable for multiple polynomial regression	None of the above

Q.No : 3)

Considering that the cost function used in logistic regression is the logistic loss for binary classification problem and assuming the learning rate is not too high then Gradient Descent algorithm

can get stuck in a local minimum when training a Logistic regression model	approach towards the local optimum and quit processing without finding the minimum	never converges	[✓ is guaranteed to find the minimum.]

Q.No : 4)

Suppose we are using Polynomial Regression and notice that there is a large gap between the training error and the test error. How to solve this?

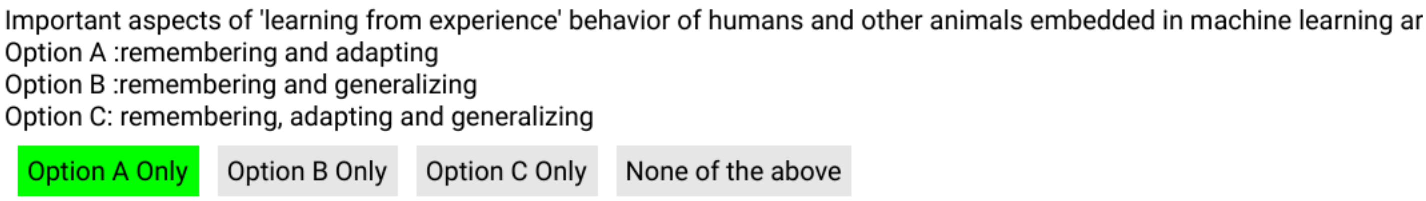
- Increase the polynomial degree
- Decrease the size of the training set
- [✓ Improve the scaling and polynomial degree]
- None of the above

Q.No : 5)

To classify pictures as outdoor/indoor and daytime/night-time, assuming that all four combinations are possible we should train

[✓ two logistic regression models]	four logistic regression models	eight logistic regression models	None of the above

Q.No : 6)



Important aspects of 'learning from experience' behavior of humans and other animals embedded in machine learning ar

Option A :remembering and adapting

Option B :remembering and generalizing

Option C: remembering, adapting and generalizing

[✓ Option A Only] Option B Only Option C Only None of the above

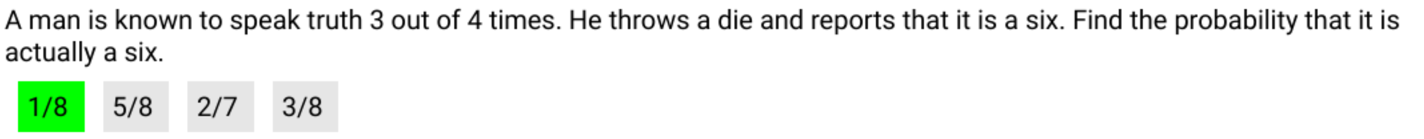
Q.No : 7)

Given the confusion matrix below, what is the accuracy of the model?

	PredictedPositive	PredictedNegative
ActualPositive	30	10
Actual Negative	05	55

75% [✓ 85%] 90% 95%

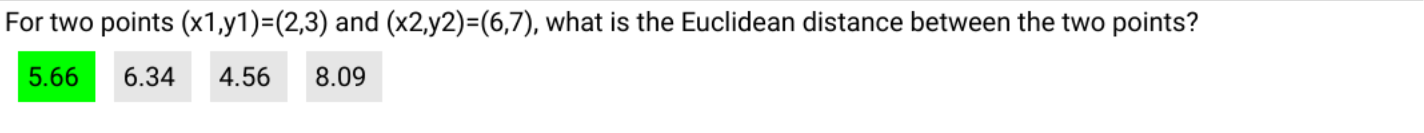
Q.No : 8)



A man is known to speak truth 3 out of 4 times. He throws a die and reports that it is a six. Find the probability that it is actually a six.

[✓ 1/8] 5/8 2/7 3/8

Q.No : 9)



For two points (x1,y1)=(2,3) and (x2,y2)=(6,7), what is the Euclidean distance between the two points?

[✓ 5.66] 6.34 4.56 8.09

Q.No : 10)

In Naive Bayes numerical variable must be binned and converted to ______.

Categorical Values Numerical Values Both 1 and 2 [✓ None of these]

Q.No : 11)

Consider the following dataset representing the relationship between advertising and sales.

Advertising (X)	Sales (Y)
1	4
2	5
3	7
4	8
5	10

a) Calculate the slope and intercept of the simple linear regression line for the given dataset. Use matrix method.

b) Add an outlier point (20, 25) to the dataset and recalculate the slope and intercept with the outlier included to the dataset. Use matrix method.

c) Compute the RMSE for the regression model of the dataset without and with outlier.

d) How do you analyse the impact of outlier on model performance? Explain.

Page:1

Page:2

$$\begin{bmatrix} \Sigma' y \\ \Sigma' xy \end{bmatrix} = \begin{bmatrix} n & \Sigma' x_1 \\ \Sigma' x_1 & \Sigma x_1^2 \end{bmatrix} \begin{bmatrix} b \\ b_1 \end{bmatrix}$$

$$\begin{bmatrix} 34 \\ 117 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 55 \end{bmatrix} \begin{bmatrix} b \\ b_1 \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 2.3 \\ 1.5 \end{bmatrix}$$

$$\boxed{\begin{array}{l} \text{Intercept} = 2.3 \\ \text{slope} = 1.5 \end{array}}$$

(no,

b) $$\begin{bmatrix} 59 \\ 417 \end{bmatrix} = \begin{bmatrix} 6 & 35 \\ 35 & 455 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$b_0 = 3.4883$$

$$b_1 = \underline{\underline{1.0677}}$$

d) The outlier changes the best fit lin and rmse is increased due to outl

Q.No : 12)

A binary classification model has to be developed to predict whether a customer will churn (leave) or not. The model outputs probabilities between 0 and 1, where valu higher likelihood of churn. Calculate the log loss for each of the four scenarios. Explain which scenario is penalized more heavily because the model is very confident i Use your calculations to support your answer.

For a customer who actually churned (true label = 1):

a) Model predicts a probability of 0.9

b) Model predicts a probability of 0.1

For a customer who did not churn (true label = 0):

c) Model predicts a probability of 0.2

d) Model predicts a probability of 0.8

Page:1

where truth label is 1 —

a) log loss = $-y \log(\hat{y})$

$= -1 \times \log(0.9)$

$= 0.10536$

b) logloss = $-1 \times \log(0.1)$

$= 2.30258$

So, in this case model which predicts probability of 0.1 will be penalized mor its log loss is significantly higher than model which predicts probability 0.9

for truth label = 0 :

c) log loss = $1 \log(0.8)$

$= -0.22314$

d) log loss = $1 \log(0.2)$

a) log loss =

$= -1.60943$

In this scenario model which predicts probab
0.8 will be penalized more as we can

through the log loss.

Q.No : 13)

A logistic regression model is available to predict whether a customer will purchase a product (1) or not (0) based on their Age (x1) and Income (x2). Given the follow model: intercept (b0) = -3, coefficient for Age (b1) = 0.04, coefficient for Income (b2) = 0.0005.

Calculate the log-odds for a customer aged 30 with an income of Rs 50,000. Calculate the probability that a customer aged 30 with an income of Rs 50,000 will purcha happens if the relationship between features and the target variable of the model is non-linear? Provide an example to support your answer.

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$z = \frac{1}{1+e^{-y}}$$

$$y = -3 + 0.04 \times 30 + 0.0005 \times 50$$

$$= 23.2$$

$$z = 0.9999$$

$\boxed{\therefore \text{ The log odds is } 0.9999}$

If the relationship between features and the variable of the model is not linear, then best fit line will not be a straight it will be a curve who power will be which may quadratic, cubic etc.

The best fit curve is a curve, who difference between the y-pred and y-ac is minimum. When the relation is not the minimum error difference curve will n a straight line rather it will a straight [illegible]

curve. For example, to study the de rate during a pandemic, it not alway a straight line, it may be an nth d polynomial graph.

Q.No : 14)

The following dataset contains customer's details about food and can be used to predict whether a customer's taste will default (the last column is the classification). Us
to determine whether a customer X=(Cook = Sita, Mood = Bad, Cuisine = Continental) should be classified as Tasty or not. So, determine which is larger, P(Yes|X) or

Cook	**Mood**	**Cuisine**	**Tasty**
Sita	Bad	Indian	Yes
Sita	Good	Continental	Yes
Asha	Bad	Indian	No
Asha	Good	Indian	Yes
Usha	Bad	Indian	Yes
Usha	Bad	Continental	No
Asha	Bad	Continental	No
Asha	Good	Continental	Yes
Usha	Good	Indian	Yes
Usha	Good	Continental	No

$P(Yes|X) = P(Cook|Yes) \times P(Mood|Yes) \times P(Cuisine|Y$

$P($

$$= \frac{2+1}{6+3} \times \frac{2+1}{6+2} \times \frac{2+1}{6+2} \times \frac{3}{5}$$

$$= 0.028125$$

$P(No|X) = P(Cook|No) \times P(Mood|No) \times P(Cuisine|No$

$\times P(No)$

$$= \left(\frac{1}{4+3} \times \frac{3+1}{4+2} \times \frac{3+1}{4+2} \right) \times \frac{2}{5}$$

$$= \frac{1}{7} \times \frac{4}{6} \times \frac{4}{6} \times \frac{2}{5}$$

$$= \frac{1}{7} \times \frac{4}{9} \times \frac{2}{5} = 0.02539$$

$\therefore P(Yes|X) > P(No|X)$, it will b

classified as Tasty

Q.No : 15)

We have data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a specia
not. Here is four training samples as follows. Apply the K-nearest neighbors (KNN) algorithm when K=3 to classify an instance (5, 6) as good or bad.

X1 = Acid Durability (seconds)	X2 = Strength (kg/square meter)	Y = Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Page:1

(x1, x2)	Euclidean distance	
(7, 7)	2.23	Bad
(7, 4)	2.83	Bad
(3, 4)	2.83	Good
(1, 4)	4.47	Good

formula, and values not shown with formula...

(.

Arranging in Ascending order

(x1, x2)	Distance	Classification
(7, 7)	2.23	Bad
(7, 4)	2.83	Bad
(3, 4)	2.83	Good
(1, 4)	4.47	Good

when we take $k = 3$, we get 2 Bad C

and 1 good classification

∴ (5, 6) is classified as Bad.

Q.No : 16)

It is observed that 50% of mails are spam. There is a software that filters spam mail before reaching the inbox. It accuracy for detecting a spam mail is 99% and chance
mail as spam mail is 5%. If a certain mail is tagged as spam find the probability that it is not a spam mail.

$P(S) = 0.5$

$P(T|\sim S) = 0.05$

$P(\sim S|T) = ?$

$P(\sim T|S) = 0.95$

Q.No : 17)

What is meant by regression analysis? For the following regression models, describe the regression line equations explaining the terms involved.

a. Multiple linear regression

b) Polynomial regression

Page:1

regression Analysis is finding best fit curve the model , so as minimize the differen between y prediction and y acutral

a) Multiple linear regression for 2 features

$= \quad y_{pred} = b_0 + b_1 X_1 + b_2 X_2$

for nth feature = $y_{pred} = b_0 + b_1 X_1 \cdot - - - b$

b) Polynomial for degree 2

$y_{pred} = b_0 + b_1 x + b_2 x^2$

$y_{pred_n} = b_0 + b_1 x + b_2 x^2 + - - - b_n x$

Q.No : 18)
What are the two main tasks that supervised/directed learning aims to solve? Briefly explain each of them.

Page:1

- To map input

✗

Q.No : 19)

What happens if the test set is not independent of the training set? How can this affect the error rate estimation of the model?

Page:1

if the test set is not independent, then is a chance of underfitting the data

^ Top