MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A) Least Square Error

B) Maximum Likelihood
C) Logarithmic Loss

D) Both A and B

Answer-D

2. Which of the following statement is true about outliers in linear regression?
A) Linear regression is sensitive to outliers

B) linear regression is not sensitive to outliers
C) Can't say D) none of these

Answer-B
3. A line falls from left to right if a slope is _____?
A) Positive B) Negative
C) Zero D) Undefined

Answer-A

4. Which of the following will have symmetric relation between dependent variable and independent
variable?
A) Regression B) Correlation
C) Both of them D) None of these

Answer-B

5. Which of the following is the reason for over fitting condition?

A) High bias and high variance B) Low bias and low variance
C) Low bias and high variance D) none of these

Answer-C

6. If output involves label then that model is called as:
A) Descriptive model B) Predictive modal
C) Reinforcement learning D) All of the above

Answer-B

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization

Answer-D

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE

Answer-D

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary
classification problems. It uses _____ to make graph?
A) TPR and FPR B) Sensitivity and precision
C) Sensitivity and Specificity D) Recall and precision

Answer-A

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the
curve should be less.
A) True B) False

Answer-B

11. Pick the feature extraction from below:
A) Construction bag of words from a email
B) Apply PCA to project high dimensional data

C) Removing stop words
D) Forward selection

Answer-B

In Q12, more than one options are correct, choose all the correct options:
12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear
Regression?
A) We don't have to choose the learning rate.
B) It becomes slow when number of features is very large.
C) We need to iterate.
D) It does not make use of dependent variable.

Answer-A

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

ANSWER

Regularization is a technique in machine learning used to prevent overfitting. Overfitting occurs when a model captures too much detail from the training data, including noise and irrelevant information. This makes it perform well on the training data but poorly on unseen data.

14. Which particular algorithms are used for regularization?

Answer-

Lasso and Ridge regression algorithms are used for regularization.

15. Explain the term error present in linear regression equation?

In linear regression, the error term, often denoted as epsilon, represents the difference between the observed value of the dependent variable (y) and the value predicted by the model's equation. It captures the inherent variability and uncertainty that cannot be explained by the linear relationship between the independent variables (x) and the dependent variable.