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 :- (A) Least Square Error

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 :- (A) Linear regression is sensitive to outliers

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

Answer :- (B) Negative

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 :- (C) Both of them

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) Low bias and high variance

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) Predictive modal

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

Answer :- (D) Regularization

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

Answer :- (A) Cross validation

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 :- (C) Sensitivity and Specificity

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) False

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) (D)

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 :- (B) (C) (D)

MACHINE LEARNING

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

13. Explain the term regularization?

Answer :- Regularization is one of the most important concepts of Machine Learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Machine Learning model performs will the training data but does not perform will with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output.

Two types of regularization techniques-

1.) Ridge Regression
2.) Lasso Regression

14. Which particular algorithms are used for regularization?

Answer :- Regularization Algorithms –

1.) Ridge Regression
2.) LASSO (Least Absolute Shrinkage and Selection Operator) Regression
3.) Elastic-Net Regression

1.) Ridge Regression – Ridge Regression is a method for analyzing data that suffer from multi-collinearity. Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty, so that we can get better long-term predictions. In statistics, it is known as the L-2 norm.
When we have the independent variables which are having high collinearity between them, at that time general linear or polynomial regression will fail so to solve such problems, Ridge Regression can be used.

2.) Lasso Regression – Lasso Regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator. If the number of predictors is greater than the number of data points, Lasso will pick at most n predictors are relevant. If there are two or more highly collinear variables, then LASSO regression selects one of them randomly which is not good for the interpretation of our model.

3.) Elastic-Net Regression – Elastic – Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

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

Answer :- An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

The error term in a regression equation represents the effect of the variables that were omitted from the equation. This is unsatisfactory, even in simple contexts, as the following discussion should indicate. Suppose subjects are IID, and all variables are jointly normal with expectation 0.

Suppose the explanatory variables have variance 1. The explanatory variables may be correlated amongst themselves, but any p of them have a non-singular p-dimensional distribution.