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:B)Correlation

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:D)

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:C)Reinforcement learning

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:D)SMOTE

v 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)TPR and FPR

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:A)B) & C)

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)B) & C)

ASSIGNMENT – 39 MACHINE LEARNING Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Answer: When the model starts making assumptions bias variance tradeoff happens. Under this over fitting and underfitting happens to the model. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

14. Which particular algorithms are used for regularization?

The commonly used regularization techniques are :
L1 regularization

L2 regularization

Dropout regularization

L1 regularization is called as LASSO(Least absolute shrinkage and selection operator) . Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function.

L2 regularization is called as Ridge regression. Ridge regression adds "squared magnitude" of coefficient as penalty term to the loss function.

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

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

The error term provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

$$Y=aX+bc+e$$

where: a,b=Constant parameters

X,c=Independent variables

e=Error term

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.