MACHINE LEARNING

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: **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: **D) SMOTE**

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: **A) True**

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) Apply PCA to project high dimensional data**

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) We don't have to choose the learning rate.**
B) It becomes slow when number of features is very large.

ASSIGNMENT – 39

MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?
Answer:

a. It is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding extra information to it.
b. It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting.
c. Now, let's understand the "How flexibility of a model is represented?"
d. For regression problems, the increase in flexibility of a model is represented by an increase in its coefficients, which are calculated from the regression line.
e. In simple words, "In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables". It maintains accuracy as well as a generalization of the model.

14. Which particular algorithms are used for regularization?
Answer:
The commonly used regularization techniques are :

a. L1 regularization
b. L2 regularization
c. Dropout regularization

A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage and Selection Operator) regression.
A regression model that uses L2 regularization technique is called Ridge regression.
Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function(L).

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

Answer:
A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.