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

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

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

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

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

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

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

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

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

11. Pick the feature extraction from below:
    A) Construction bag of words from an email
    B) Apply PCA to project high dimensional data
    C) Removing stop words
    D) Forward selection

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.

MACHINE LEARNING

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

13. Explain the term regularization?

Answer: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or under fitting. It is a technique to prevent the model from overfitting by adding extra information to it. In regularization technique, magnitude of the features is reduced by keeping the same number of features.

14. Which particular algorithms are used for regularization?

Answer: There are three main regularization techniques

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

Ridge Regression: It is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.

LASSO Regression: It is another regularization technique to reduce the complexity of the model.

Elastic-Net Regression: It 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: Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. An error term is generally unobservable and a residual is observable and calculable, making it much easier to quantify and visualize.