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 a 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 is of the important concepts of machine learning. This technique used to prevent the model from overfitting. model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data noise in the output, so the model is called overfitted.

Regularization works by adding a penalty or complexity term to the complex model. There are mainly two types of regularization techniques,

1) **Lasso regression**
2) **Ridge regression**

Ridge regression

- Ridge regression 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.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.
- It helps to solve the problems if we have more parameters than samples.

#Lasoo regression

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- It is also called as L1 regularization

14. Which particular algorithms are used for regularization?

There is three technics for regularization :

1) **Ridge** 2) **Lasso** 3)**Dropout**

Ridge regression

- Ridge regression 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.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.
- It helps to solve the problems if we have more parameters than samples.

#Lasoo regression

- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- It is also called as L1 regularization

#Dropout

- Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.

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

- An error term is a 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.

- 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