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: 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 model

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: 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) Apply PCA to project high dimensional data

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

13.) Explain the term regularization?

Answer: Regularization refers to make things acceptable & regular. It is used in machine learning. The term regularization is a technique which shrinks or regularizes the coefficients towards zero. In more easy words, it discourages learning a more flexible model so that overfitting can be prevented. The idea is to reprimand the complex model.

There are basically two main techniques of regularization i.e. Lasso regression & Ridge regression.

1) Lasso regression: It is also known as L1 regularization. This technique simplifies the RSS by adding the penalty equivalent to the sum of the absolute value of coefficients. It is not similar to ridge regression as it uses absolute coefficient values for normalization.
2) Ridge Regression: It is also known as L2 regularization. This technique simplifies the RSS by adding the penalty equal to the square of magnitude of coefficients.

14.) Which particular algorithms are used for regularization?

Answer: there are basically three regularization techniques which can be used i.e. Ridge regression, Lasso and Dropout.

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

Answer. The term ‘error’ is a residual variable that is produced by a mathematical or statistic model, which is generated when the model doesn’t show the actual relationship between the dependent & independent variables. Because of this incomplete relation, error is the amount in which the equation differs during experiential analysis.

Error is also called as the disturbance, remainder or residual term & in model it is represented by letters u, e or ε. The error means the sum of the deviations within the line of regression and gives an explanation for the difference between actual results and theoretical value of the model.