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

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

B) Correlation

5. Which of the following is the reason for over fitting condition?

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

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

8. To overcome with imbalance dataset which technique can be used?

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

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

B) False (AUC should be high)

11. Pick the feature extraction from below:

B) Apply PCA to project high dimensional data

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.

13. Explain the term Regularization.

Regularization is considered as one of the key concepts of machine learning. It helps in decreasing the complexity of the model and also prevents the problem of overfitting. To improve the performance of the machine learning model, it is very necessary to prevent overfitting. There often comes a situation, where your machine learning model performs very well on training data, but fails to perform well on testing data. Such situations can be dealt with the regularization.

14. Which particular algorithms are used for regularization?

There are two techniques used for regularization, namely LASSO Regularization and RIDGE Regularization.

1. LASSO – It modifies the residual sum of square by adding the penalty equivalent to the sum of absolute values of coefficients. It uses absolute coefficient values for normalization. This regularization technique performs L1 Regularization.
2. RIDGE – It modifies the residual sum of square by adding the penalty equivalent to the square of the magnitude of coefficients. This regularization technique performs L2 Regularization.

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

The term error is defined as a residual variable, which is created when the model does not fully represent the actual relationship between independent and dependent variables. The error term is the amount at which the equation may differ. The error term is also known as residual and is represented in the model by letter 'e'. An error term indicates the uncertainty in the model.

The equation as follows:

$$Y=\alpha X+\beta\rho+\epsilon$$

Where,

α, β=Constant parameters

X, ρ=Independent variables

ϵ=Error term