Machine learning

In Q1 to Q11, only one option is correct, choose the correct option:

Q1. Which of the following method do we use to find the best fit line for data in linear regression?

Ans – A) Least square method

Q2. Which of the following statement is true about outliers in the linear regression?

Ans – A) Linear regression is sensitive to outliers.

Q3. A line falls from left to right if a slope is _____?

Ans – B) Negative

Q4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans – B) Correlation

Q5. Which of the following is the reason for overfitting condition?

Ans – C) Low bias and high variance

Q6. If output involves label, then that model is called as:

Ans – B) predictive model

Q7. Lasso and Ridge regression techniques belong to ____?

Ans – D) Regularization

Q8. To overcome with imbalance data set which technique can be used?

Ans – D) SMOTE

Q9. The AUC receiver operator characteristic curve is an equation metric for Binary Classification problems. It uses _____ to make a graph.

Ans – A) TPR and FPR

Q10. In AUCROC curve for the better model area under the curve should be less.

Ans – B) False

Q11. Pick the feature extraction from below:

Ans – B) Apply PCA to project higher dimensional data.

In Q12, more than one options are correct choose all the correct options:

Q12. Which of the following is true about normal equation used to compute the coefficient of the linear regression?

Ans – A) We don't have to choose the learning rate.

B) It becomes slow when a number of features is very large.

D) It does not make use of dependent variable.

Machine learning

Q13. Explain the term Regularization.

When we use regression model to train some data, there is a good chance that the model will overfit the given training data set. Regularisation helps sort this over fitting problem by restricting the degrees of freedom of a given equation that is simply reducing the number of degrees of a polynomial function by reducing their corresponding weights. In a linear equation we do not want huge weights or coefficients as a small change in weight can make a large difference for dependent variable(Y). So, regularisation constraints the weights of such features to avoid overfitting.

Two regularise the model, a shrinkage penalty is added to the cost function. Let's see different types of regularisations in regression:

1.Lasso

2.Ridge

3.Elasticnet

Q14. Which particular algorithms are used for regularization?

Regularization is a technique which reduces the overfitting in machine learning models by adding penalty term to the loss function. There are various types of regularization methods which are mentioned below:

1.Lasso regularization- It uses the L1 norm of the coefficients as the penalty term. This method tends to produce sparse solutions, meaning that some of the coefficients are shrunk to 0.

2.Ridge regularization- It uses the L2 norm of the coefficients as the penalty term. This method tends to shrink all the coefficients by the equal amount but does not eliminate them completely to 0.

3.Elastic net regularization – It combines both L1 and L2 norms as the penalty term.

Q15. Explain the term error present in linear regression equation.

The term error in linear regression equation is the difference between the actual value of the dependent variable and the predicted value of the dependent variable in the equation. The error represents the variation in the data that is not explained by the equation, or the unwanted noise in the equation. The error is usually denoted by e,or ε and it has a mean of zero and a constant variance. The main moto of linear regression is to minimize the sum of squared errors, or the residual sum of squares.