MACHINE LEARNING

Q1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans.- A) Least Square Error

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

Ans-A) Regression

Q5. Which of the following is the reason for over fitting condition?

Ans- C) Low bias and high variance

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

Ans- B) Predictive modal

Q7. Lasso and Ridge regression techniques belong to _____?

Ans-D) Regularization

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

Ans- D) SMOTE

Q9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans- A) TPR and FPR

Q10. In AUC Receiver Operator Characteristic (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 high dimensional data

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 number of features is very large, C) We need to iterate.

Q13. Explain the term regularization?

Ans.- In machine learning, regularisation is a technique for preventing overfitting. Random noise from training data is more likely to be included into complex models, which could hide any patterns that might otherwise be there. Regularization helps to reduce the effect of noise on the forecasting capabilities of the model. The main concept underlying regularisation in machine learning is to offer greater loss values to complex models and award losses to difficult models with the addition of a complexity term. The Lasso and Ridge Regression algorithms for regularisation in machine learning are the two methods used.

Q14. Which particular algorithms are used for regularization?

Ans. Since its loss function is based on the weights or absolute coefficients, lasso regression differs from ridge regression in that it uses lasso regularisation to normalise the absolute values of coefficients.

Limitation of Lasso regression

- Issues with certain dataset types: If the number of predictors is more than the number of data points, Lasso will only choose at most n predictors as non-zero. This is true even if all predictors are meaningful.
- Multicollinearity Issue: LASSO regression randomly selects one of the highly correlated variables when there are two or more of them, which is detrimental to the interpretation of our model.

Ridge regression- Ridge regression is a regularisation method for machine learning that alters the amount of shrinkage by adding a penalty that is determined by the square of the coefficient's value.

Limitation of Ridge regression

- It reduces the complexity of a model but does not reduce the number of independent variables because it never minimises a coefficient to zero, which is not useful for feature selection. As a result, this approach is inadequate for feature selection.
- Model Interpretability: Its disadvantage is that it will only ever significantly reduce the coefficients for the least important predictors, never fully eliminating them. In other words, the final model will take into account all independent variables, often known as predictors.

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

Ans.- The term error present in linear regression equation represents the difference between the actual value and Predicted value and the goal is to reduce this difference. Mean-square error (MSE), which is most frequently used in linear regression, is used to determine the model's error. MSE is determined by:

1. At each value of x, the difference between the observed and anticipated y-values is measured
2. squared, and the mean of these squared distances is calculated.
3. By identifying the regression coefficient that produces the smallest MSE, linear regression finds a line that fits the data.