MACHINE LEARNING

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

A) Least square error

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

A) Linear regression is sensitive to outliers

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

B) Negative

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

B) Correlation

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

C) Low bias and high variance

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

B) Predictive model

Q7. Lasso and Ridge regression techniques belong to

A) Cross validation

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

D) SMOTE

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

C) Sensitivity and Specificity

Q10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

B) False

Q11. Pick the feature extraction from below:

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?

A) We don't have to choose the learning rate.
B) It becomes slow when number of features is very large.
D) It does not make use of dependent variable.

Q13. Explain the term regularization?

Regularization is a technique which is used to solve the overfitting problem of the machine learning models. Regularization is a technique used to avoid this overfitting problem. The idea behind regularization is that models that overfit the data are complex models that have for example too many parameters. In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables.

There are two types of regularization as follows:

- **L1 Regularization or Lasso Regularization**
- **L2 Regularization or Ridge Regularization**

Q14. Which particular algorithms are used for regularization?

L2 and L1 are the most common types of regularization. Regularization works on the premise that smaller weights lead to simpler models which in results helps in avoiding overfitting. So to obtain a smaller weight matrix, these techniques add a 'regularization term' along with the loss to obtain the cost function.

logloss(N=1)=y log(p) + (1 - y) log(1-p)

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

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

error = | model estimate - true value|

The error term includes everything that separates your model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables . the terms error and residual are often interchanged, there is an important formal difference. While an error term represents the way observed data differs from the actual population, a residual represents the way observed data differs from sample population data.