Machine Learning

Question -1. Which of the following methods do we use to find the best fit line for data in linear regression.

Answer- A). least square Error

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

Answer- A) Linear regression is sensitive to outliers

Question-3. A line falls from left to right if a slope is

Answer- B) Negative

Question- 4. Which of the following will have symmetric relation between dependent variable and independent vriable.

Answer- C) Both of them

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

Answer- C) Low bias and high variance

Question-6. If output involves labsl then that model is called as:

Answer- D) All of the above

Question-7. Lasso and Ridge regression techniques belong to_____.

Answer- D) Regularization

Question-8. To overcome with imbalance dataset whichtechnique can be used.

Answer- D) SMOTE

Question-9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problem. It uses _____to make graph.

Answer- - C) Sensitivity and Specificity

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

Answer- B) False

Question-11. Pick the feature extraction from below.

Answer- B) Apply PCA to project high dimensional data

Question- 12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression.

Answer- A) We don't have to choose the learning rate.

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

C) We need to iterate.

Question-13. Explain the term regulazation.

Answer-13. The term regulazation refers to a set of techniques that regularizes learning from particular features for traditional algorithms or neurons in the case of neural network algorithms.

It normalize and moderates weights attached to a feature or a neuron so that helps to avoid the problem of overfiting.

Question- 14. Which particular algorithms are used for regularization.

Answer- 14. There are three main regularization techniques –Ridge Regression (L2 NORM) Lasso (L1 NORM) Dropout.

Ridge Regression (L2 Regularization)---- Ridge regression is also called L2 norm or regularization.

When using this technique, we add the sum of weight square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda \sum_{i=1}^{n} Wi^2$$

Lasso Regression(L1 Regularization)----- Loss function only considers absolute weight optimization algorithms penalize highr weight values.

In right regression loss function along with the optimization algorithm brings parameters near to zero but not actully zero , while lasso eliminates less important festures and sets repective weight values to zero. Thus ,lasso also perfoms festure selection along with regularization.

$$\text{Loss} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2 + \lambda \sum_{i=1}^{n} |Wi|$$

Dropout---- **Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.**

In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with *1-p* probability for each of the specified layers. Where *p* is called keep probability parameter and which needs to be tuned.

Question-15. Explain the term error present in linear regression equation.

Answer-15. Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. ... The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.

Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time. The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.

The two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error.