MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A. Both A and B

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 _____?
A. Positive

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

5. Which of the following is the reason for over fitting condition?
A. Low bias and high variance

6. If output involves label then that model is called as:
A. Predictive model

7. Lasso and Ridge regression techniques belong to _____?
A. Regularization

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

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.
A. False

11. Pick the feature extraction from below:
A. Apply PCA to project high dimensional data

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
A1. We don't have to choose the learning rate.
A2 .It becomes slow when number of features is very large.
A3. We need to iterate.

13. Explain the term regularization?
A. Reglarization is the technique used to remove bias and variance or we can say that for removing Underfitting and Overfitting. Let us understand about Bias and Variance. Bias means the error in training data and Variance means the error in testing data. This the basic understanding of Bias and Variance.When this Bias and Variance are low and high respectively, such situation or scenario is called Overfitting.ie the accuracy of training data is high and accuracy of testing data is low. In other words,the model which have low bias and high variance it is overfitted. And in other cases, the accuracy of both training and testing data is high such scenario is known as Underfitting.ie the model have High bias and high variance. In this both situation Regularization technique is used. Regularization will help he model to remove Overfitting and Underfitting . ie the model will become with low variance and low bias.This is all about Regularization.

14. Which particular algorithms are used for regularization?

A. Algorithms used in regularization are:

a). Lasso (L1) Regression

b). Ridge (L2) Regression

c). Elastic Net Regression

In L1 the unwanted features where eliminated which is not useful for the prediction.The parameters where added in to traditional equation. 'Lambda' and 'magnitude of the slope'. The parameter adds the penalty to the traditional equation. In L1 we are using the absolute value, so it try to move towards zero.

In L2 regression it will takes all features which is not useful or not. The parameters where added to the traditional equation. Two parameters are added 'square of the slope' and 'lambda'. In L2 regression it gets near to zero. It will select all the features and shrink it for the prediction. If we give less value for the lambda the model try to remove overfit and try to get the best fit.

In Elastic Net Regression , in simple words it is the combination of lasso and ridge. When the model have high data, we do not know which features are use or usefull for prediction, in such cases we use elastic Net regression. The parameters of L1 and L2 are combined. And also elastic net is good at dealing with situations when there are correlations between parameters.

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

A. While predicting the model using linear regression there may be occurs the error. Error simply means the difference or distance between the actual value and predicted value. It means when we draw the line through the points there is some distance between the actual point and the predicted point. To find the error we want to find difference between the actual value and predicted value and want o square it. When we got the small variant between the testing and training data, regression is fitted. So we can conclude that less difference in error we got the more line is fitted.