ANSWER SHEET FOR MACHINE LEARNING

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

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

A) Least Square Error

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 _____?

B) Negative

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

C) Both of them (Regression and Correlation)

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

C) Low bias and high variance

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

B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

D) SMOTE

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

C) Sensitivity and Specificity

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

B) False

11. Pick the feature extraction from below:

A) Construction bag of words from a email

B) Apply PCA to project high dimensional data

C) Removing stop words

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

12. 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.

C) We need to iterate.

D) It does not make use of dependent variable.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Every time we will have data and we will try to fit machine learning model on that, if the model is not suitable for data then our model will not learn and that is known as under fitting scenario. On the other hand, data and model combination is not good, then it will be less flexible to handle the data and scenario is known as over fitting. In order to avoid both the cases, we need to achieve sweet spot on the line, and this can be achieve by regularization through balancing bias and variance.

14. Which particular algorithms are used for regularization?

Lasso and Ridge regularization algorithms used for regression models for data regularization.

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

The distance calculated between the observed data point and predicted data point is called as error. we assume that the relation between the feature and label to be linear and we find a straight line that fits the relation well. The error term in regression is a compensate for what we miss out with this model, as predicted values are not always exact to actual values. And error represent the difference in it.