**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 B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

**ANSWER: A) Least Square Method**

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

A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can’t say D) none of these

**ANSWER: A) Linear regression is sensitive to outliers**

3. A line falls from left to right if a slope is \_\_\_\_\_\_?

A) Positive B) Negative

C) Zero D) Undefined

**ANSWER: B) Negative Slope**

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

A) Regression B) Correlation

C) Both of them D) None of these

**ANSWER: B) Correlation**

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

A) High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none of these

**ANSWER:** **A) High bias and high variance**

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

A) Descriptive model B) Predictive model

C) Reinforcement learning D) All of the above

**ANSWER: B) Predictive modal**

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

**ANSWER: D) Regularization**

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

A) Cross validation B) Regularization

C) Kernel D) SMOTE

**ANSWER: D) SMOTE**

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 B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

**ANSWER:**

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

curve should be less.

A) True B) False

11. Pick the feature extraction from below:

A) Construction bag of words from an email

B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection

**ANSWER:**

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.

**ANSWER: Option A, B, C are correct.**

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

13. Explain the term regularization?

**Answer:** Regularization

When it comes to training models, there are two major problems one can encounter: overfitting and underfitting. Overfitting happens when the model performs well on the training set but not so well on unseen (test) data. Underfitting happens when it neither performs well on the train set nor on the test set. Particularly, regularization is implemented to avoid overfitting of the data, especially when there is a large variance between train and test set performance.

14. Which particular algorithms are used for regularization?

**Answer:** Following are the algorithms used for regularisation:

1. **Ridge Regression:** Ridge regression is the method used for the analysis of multicollinearity in multiple regression data. It is most suitable when a data set contains a higher number of predictor variables than the number of observations.
2. **LASSO (Least Absolute Shrinkage and Selection Operator) Regression:** Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

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

**Answer**: Error term in linear equationexpresses the difference between the actual outcome variables and the outcome variables predicted by the statistical model.