**Worksheet\_set\_1**

**MACHINE LEARNING**

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

**Ans: Least Square Error**

1. 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

**Ans: linear regression is not sensitive to outliers**

1. A line falls from left to right if a slope is \_\_\_\_\_\_? A) Positive B) Negative C) Zero D) Undefined

**Ans: Negetive**

1. 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

**Ans: Both of them**

1. 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

**Ans: Low bias and high variance**

1. If output involves label then that model is called as: A) Descriptive model B) Predictive modal C) Reinforcement learning D) All of the above

**Ans: Predictive model**

1. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_? A) Cross validation B) Removing outliers C) SMOTE D) Regularization

**Ans: Regularization**

1. To overcome with imbalance dataset which technique can be used? A) Cross validation B) Regularization C) Kernel D) SMOTE

**Ans: Regularization**

1. 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

**Ans: Recall and presicion**

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less. A) True B) False

**Ans: True**

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 In

**Ans: Apply PCA to project high dimensional data**

**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.

**Ans: It becomes slow when number of features is very large**

**ASSIGNMENT – 39 MACHINE LEARNING**

13. Explain the term regularization?

**Ans:** Regularization in machine learning is the concept by using which we can avoid or prevent the model from overfitting (i.e low biasedness and high variance).

When the model performs well with train data but does not perform well with our test or unseen data then the model is said to be overfitted which can handled by reducing the magnitude of the variable.

14. Which particular algorithms are used for regularization?

**Ans:** Ridge & Lasso

Ridge: It is used to analyze any data that suffers from multicollinearity where least square doesnot works well.

Lasso: it’s the analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model

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

**Ans:** An error term in linear regression equation refers to the sum of the deviations within the regression line( difference between the theoretical value of the model and actual observed value).