Machine Learning – WORKSHEET 1

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

   Ans A) Least Square Error

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

   Ans A) Linear regression is sensitive to outliers

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

   Ans B) Negative

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

   Ans C) Both of them (Regression & Correlation)

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

   Ans C) Low bias and high variance

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

   Ans B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

   Ans D) Regularization

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

   Ans D) SMOTE

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

   Ans A) TPR and FPR

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

    Ans B) False

11. Pick the feature extraction from below:

    Ans B) 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?

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

C) We need to iterate.

13. Explain the term regularization?

Ans Understanding of the terms generalization, underfitting and overfitting would be helpful before explanation of regularization.

Underfitting: Choosing too simple a model is called underfitting. Underfit model performs badly even on training data. The prediction score is very less.

Overfitting: Overfitting occurs when one fits the model too closely to the particularities of the training set and obtain a model that works well with training set but not able to generalize new data.

Regularization: The term regularization means to explicitly restricting a model to avoid overfitting. The problem with overfitting model is that the model may predict with great accuracy for training data, but on test data the model may not perform well. This discrepancy between performance on the training set and the test set is a clear sign of overfitting. That is the model is biased towards training data and may have high variance with respect to test data. On the other hand, if the prediction is low on training data as well as on the test data then it is a case of under fitting the model.

Generally, in ML the objective must be to arrive at a generalized model. Usually, we have to build models that make accurate prediction using training set. A generalized model is a good model when the model has low bias and low variance. This may be achieved by introducing a small amount of bias in the train model by readjusting the slope of the line. Lasso and Ridge regression techniques may be used in regularization.

14. Which particular algorithms are used for regularization?

Ans There are two algorithms that are extensively used for regularization, namely,

a. Lasso regression (uses L1-norm)(least absolute shrinkage and selection operator)

b. Ridge regression (uses L2-norm): Each feature should have as little effect on outcome as possible (.i.e the line should have small slope).

c. Elasticnet (combination of Lasso and Ridge).

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

Ans The linear regression equation is given by

y=mx+c+e …………..(i)

where m is slope of line, c is the y-intercept and e is the error term.

The error is defined as the difference between the actual value and the predicted value.

The objective is to have minimum error (or to minimize the error)