Machine Learning Worksheet

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

ANSWER : Least Square Error

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

ANSWER : Linear regression is sensitive to outliers

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

ANSWER : Negative

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

ANSWER : Correlation

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

ANSWER : Low bias and high variance

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

ANSWER : Predictive model

7.Lasso and Ridge regression techniques belong to?

ANSWER : Regularization

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

ANSWER : SMOTE

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

ANSWER : TPR AND FPR

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

ANSWER : FALSE

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

ANSWER: We don’t have to choose the learning rate ,

It becomes slow when number of features is very large

11. Pick the feature extraction from below:

ANSWER : Apply PCA to project high dimensional data

13. Explain the term regularization?

ANS : Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and “avoid overfitting”.

It introduces a cost term for bringing in more features with the objective function. Hence it tries to push the coefficients for many variables to zero and hence reduce cost term.

14. Which particular algorithms are used for regularization?

ANS:1. Ridge Regression

2. LASSO (Least Absolute Shrinkage and Selection Operator) Regression

3. Elastic-Net Regression

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

ANS: Linear regression uses one independent variable to explain or predict the outcome of the dependent variable Y

Linear regression $Y = a+bX+e$

Where

e = The Regression Residual Error.