

Machine Learning Worksheet – 1

Submission by: Sahil Kumar

Project: Internship 24

Date: 15 January 2022

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

Answer: **(A) Least Square Error**

2. Which of the following statement is True about outliers in Linear Regression?

Answer: **(A) Linear regression is sensitive to outliers**

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

Answer: **(B) Negative**

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

Answer: **(B) Correlation**

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

Answer: **(C) Low bias and high variance**

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

Answer: **(B) Predictive model**

7. Lasso and Ridge Regression techniques belong to _____.

Answer: **(D) Regularization**

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

Answer: **(D) SMOTE**

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

Answer: **(C) Sensitivity and Specificity**

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

Answer: **(B) False**

11. Pick the feature extraction from below:

Answer: **(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?

Answer: **(A) We don't have to choose the learning rate**

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

13. Explain the term regularization.

Answer: When we use regression models to train some data, there is a good chance that the model will overfit the given training data set. Regularization helps sort this overfitting problem by restricting the degrees of freedom of a given equation, i.e., simply reducing the number of degrees of a polynomial function by reducing their corresponding weights.

In a linear equation, we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependent variable (Y). So, regularization constraints the weights of such features to avoid overfitting.

To regularize the model, a Shrinkage penalty is added to the cost function. The different types of regularizations are as follows:

- LASSO
- RIDGE
- ELASTICNET (less popular)

14. Which particular algorithms are used for regularization?

Answer: The different algorithms used for regularization are primarily Lasso, Ridge and Elasticnet. However, Lasso and Ridge are the widely used techniques.

- LASSO (Least Absolute Shrinkage and Selection Operator) Regression (L1 Form): LASSO regression penalizes the model based on the sum of magnitude of the coefficients. The regularization term is given by: **Regularization = $\lambda * \sum |\beta_j|$** --- where, λ is the shrinkage factor.
- Ridge Regression (L2 Form): Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients. The regularization term is given by: **Regularization = $\lambda * \sum |\beta_j|^2$** --- where, λ is the shrinkage factor.

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

Answer: The error term is a random variable with a mean of zero and a constant variance. The meaning of this is that the variances of the independent variables are independent of the value of the variable.

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. It represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.