

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: B) 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 model

7. Lasso and Ridge regression techniques belong to _____?

Ans: D) Regularization

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

Ans: A) Cross validation

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?

Ans: A) We don't have to choose the learning rate.

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

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

13.Explain the term regularization?

As the term regularize suggests it means to make regular by conformance to law, rules or customs. Particularly in machine learning term regularization is associate with training a machine learning model which moderates & normalised the features so that algorithms do not rely on just features to predict the results. This can avoid problem of over-fitting or underfitting. To avoid or to reduce the generalization (prevent over-fitting or under fitting) in machine earning model & to minimise the loss, we need regularization techniques. Using regularization, we can run and train our model appropriately and reduce the errors.

An optimal model is one such model where Bias and Variance are both optimal. This is called Bias-Variance Trade-off, and this can be achieved in any over or under fitting model with the help of regression.

14.Which particular algorithms are used for regularization?

There are three main types of regularization techniques:

1. **Ridge Regression:** also known as L2 norm or Regularization. This adds a penalty to the error functions. The penalty is equivalent to the sum of the squares of absolute values. This means that mathematical function representing our model is minimised and coefficient is calculated. Ridge regression performs regularization by shrinking the coefficient present. Here penalty term is represented by term Lambda.
2. **Lasso Regularization:** also known as L1 Regression. This technique is different from ridge regression. But it also uses Lambda function for tuning parameters. As compare to ridge regression lasso eliminates less important features and sets respective weight values to zero unlike ridge regression where optimised algorithm brings parameters

near to zero but not actually zero. Thus, lasso also performs feature selection along with regularization.

3. **Dropout:** it is used in many kinds of neural networks like ANN, DNN, CNN, to moderate the learning

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

In regression analysis there are two important variables. We have dependent variable which we want to predict and independent variable which we think/assume can have an impact on our dependent variable. In regression line we will always have error term because in real-life scenario, independent variables are never perfect predictors of the dependent variables. But the line is actually the estimation based on the available data. It means the error term tells you how sure or certain you can be about the formula. The larger it is, the less certain the regression line is.