

Machine Learning assignment

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?

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 modal

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: B) Sensitivity and precision

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., D) It does not make use of dependent variable.

13. Explain the term regularization?

Ans: Regularisation is a set of machine learning techniques used to prevent models from fitting the training data too closely, which can lead to overfitting. When a model learns not only the underlying patterns in the data but also the noise and unpredictability, it performs poorly on unseen data. Regularisation approaches constrain or penalise the model during training, preventing it from becoming unduly complicated and encouraging a better balance between fitting the training data and retaining generalisation.

14. Which particular algorithms are used for regularization?

Ans: Ridge regression: It is a sort of linear regression in which a little amount of bias is introduced to improve long-term predictions. Ridge regression is a regularisation approach that is used to reduce model complexity. L2 regularisation is another name for it. The penalty term is added to the cost function in this technique. Ridge Regression penalty is the degree of bias introduced into the model. We can compute it by multiplying the squared weight of each individual feature by the lambda.

Lasso regression: It is another regularisation technique for reducing model complexity. It is an abbreviation for Least Absolute and Selection Operator. It is identical to Ridge Regression except that the penalty term only contains absolute weights rather than a square of weights. Because it uses absolute data, it can decrease the slope to zero, whereas Ridge Regression can only get close. It is also referred to as L1 regularisation.

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

Ans: The difference between the dependent variable's actual value and the value predicted by the regression model is represented statistically by the term "error" in linear regression. It is sometimes referred to as the residue. The dependent variable's variability that cannot be explained by the independent variables in the model is captured by the error term. The error term is a critical component of regression analysis because it indicates the presence of unobserved factors or random variation that affect the dependent variable. It is commonly considered to have certain statistical qualities, such as being regularly distributed and having a mean of zero.