

Machine Learning Assignment

Q1: Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans: A) Least Square Error

Q2: Which of the following statement is true about outliers in linear regression?

Ans: A) Linear regression is sensitive to outliers

Q3: A line falls from left to right if a slope is _____?

Ans: B) Negative

Q4: Which of the following will have symmetric relation between dependent variable and independent variable?

Ans: B) Correlation

Q5: Which of the following is the reason for over fitting condition?

Ans: C) Low bias and high variance

Q6: If output involves label then that model is called as:

Ans: B) Predictive model

Q7: Lasso and Ridge regression techniques belong to _____?

Ans: D) Regularization

Q8: To overcome with imbalance dataset which technique can be used?

Ans: A) Cross Validation

Q9: 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

Q10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Ans: B) False.

Q11. Pick the feature extraction from below:

Ans: B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:

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

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.

Q13. Explain the term regularization?

Ans : Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting by adding extra information to it.

The machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique. This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy of the model.

Q14. Which particular algorithms are used for regularization?

Ans: There are mainly three types of regularization techniques, which are given below:

- A) Ridge Regression (L2): Ridge regression is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients.
- B) Lasso Regression(L1): Least Absolute Shrinkage and Selection Operator
Lasso regression helps to reduce the overfitting in the model as well as feature selection.
- C) Elastic-Net Regression: It is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

Q15. Explain the term error present in linear regression equation?

Ans: An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model. The error term is a residual variable that accounts for a lack of perfect goodness of fit.

For example, assume there is a multiple linear regression function that takes the following form: $Y = \alpha X + \beta p + \epsilon$
where: α, β = Constant parameters