Name – Sachin Subhash Awalkar
Batch No – DS2312
Subject – MACHINE LEARNING
Date – 12-01-2024

MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
   A) Least Square Error B) Maximum Likelihood
   C) Logarithmic Loss D) Both A and B

Answer – A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?
   A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers
   C) Can't say D) none of these

Answer – A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?
   A) Positive B) Negative
   C) Zero D) Undefined

Answer – C) Zero

4. Which of the following will have symmetric relation between dependent variable and independent variable?
   A) Regression B) Correlation
   C) Both of them D) None of these

Answer – B) Correlation

5. Which of the following is the reason for over fitting condition?
   A) High bias and high variance B) Low bias and low variance
   C) Low bias and high variance D) none of these

Answer – C) Low bias and high variance

6. If output involves label then that model is called as:
   A) Descriptive model B) Predictive model
   C) Reinforcement learning D) All of the above

Answer – B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?
   A) Cross validation B) Removing outliers
   C) SMOTE D) Regularization

Answer – D) Regularization

8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation B) Regularization
   C) Kernel D) SMOTE

Answer – D) SMOTE

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

A) TPR and FPR B) Sensitivity and precision
C) Sensitivity and Specificity D) Recall and precision

Answer – A) TPR and FPR

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

A) True B) False

Answer – B) False

11. Pick the feature extraction from below:

A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Answer – B) Apply PCA to project high dimensional data

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

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.
C) We need to iterate.
D) It does not make use of dependent variable.

Answer – 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?

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.

14. Which particular algorithms are used for regularization?

Answer –

LASSO (Least Absolute Shrinkage and Selection Operator) Regression L1 Form
LASSO regression penalizes the model based on the sum of magnitude of the coefficients.

RIDGE Regression L2 Form
Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients.

ELASTICENT (Less Popular)

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

Answer –

The regression model is linear in terms of coefficients and error term. The mean of the residuals is zero. The error terms are not correlated with each other, i.e. given an error value; we cannot predict the next error value. No Multicollinearity, i.e. no independent variables should be correlated with each other or affect one another. If there is multicollinearity, the precision of prediction by the OLS model decreases. The error terms are normally distributed.

End of Assignment