MACHINE LEARNING

ASSIGNMENT – 10

In Q1 to Q8, only one option is correct, Choose the correct option:

1. In the linear regression equation y = θ0 + θ1x, θ0 is the:

A) Slope of the line

B) Independent variable

C) y intercept

D) Coefficient of determination

2. True or False: Linear Regression is a supervised learning algorithm.

A) True

B) False

3. In regression analysis, the variable that is being predicted is:

A) the independent variable

B) the dependent variable

C) usually denoted by x

D) usually denoted by r

4. Generally, which of the following method(s) is used for predicting continuous dependent variables?

A) Logistic Regression

B) Linear Regression

C) Both

D) None of the above

5. The coefficient of determination is:

A) the square root of the correlation coefficient

B) usually less than zero

C) the correlation coefficient squared

D) equal to zero

6. If the slope of the regression equation is positive, then:

A) y decreases as x increases

B) y increases as x increases

C) y decreases as x decreases

D) None of these

7. Linear Regression works best for:

A) linear data

B) non-linear data

C) both linear and non-linear data

D) None of the above

8. The coefficient of determination can be in the range of:

A) 0 to 1

B) -1 to 1

C) -1 to 0

D) 0 to infinity

In Q9 to Q13, more than one options are correct, Choose all the correct options:

9. Which of the following evaluation metrics can be used for linear regression?

A) Classification Report

B) RMSE

C) ROC curve

D) MAE

10. Which of the following is true for linear regression?

A) Linear regression is a supervised learning algorithm.

B) Linear regression supports multi-collinearity.

C) Shape of linear regression's cost function is convex.

D) Linear regression is used to predict discrete dependent variable.

11. Which of the following regularizations can be applied to linear regression?

A) Ridge

B) Lasso

C) Pruning

D) Elastic Net

12. Linear regression performs better for:

A) Large amount of training samples with small number of features.

B) Same number of features and training samples

C) Large number of features

D) The variables which are drawn independently, identically distributed

13. Which of the following assumptions are true for linear regression?

A) Linearity

B) Homoscedasticity

C) Non-Independent

D) Normality

Q14 and Q15 are subjective answer type questions, Answer them briefly.

14. Explain Linear Regression?

Answer:- Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the linear equation that best describes the relationship between the dependent variable and the independent variable(s).

The equation for simple linear regression is $y = \beta0 + \beta1x$, where y is the dependent variable, x is the independent variable, $\beta0$ is the intercept, and $\beta1$ is the slope of the line. The slope $\beta1$ represents the change in y for a one-unit increase in x.

Multiple linear regression extends this concept to more than one independent variable. The equation for multiple linear regression is $y = \beta0 + \beta1x1 + \beta2x2 + ... + \beta nxn$, where n is the number of independent variables.

Linear regression works by minimizing the sum of the squared errors between the predicted values and the actual values. The method used to find the best-fit line is called the least squares method.

Linear regression can be used for both simple and complex problems, including prediction, forecasting, and modeling. It is widely used in fields such as economics, finance, engineering, and social sciences.

15. What is difference between simple linear and multiple linear regression?

Answer:- The main difference between simple linear regression and multiple linear regression is the number of independent variables used to predict the dependent variable.

Simple linear regression involves modeling the relationship between a single independent variable and a dependent variable. The goal is to find a linear equation

that best describes the relationship between the two variables. The equation for simple linear regression is $y = \beta0 + \beta1x$, where y is the dependent variable, x is the independent variable, $\beta0$ is the intercept, and $\beta1$ is the slope of the line.

On the other hand, multiple linear regression involves modeling the relationship between two or more independent variables and a dependent variable. The goal is to find a linear equation that best describes the relationship between the dependent variable and all the independent variables. The equation for multiple linear regression is $y = \beta0 + \beta1x1 + \beta2x2 + ... + \betanxn$, where y is the dependent variable, x1, x2, ..., xn are the independent variables, and $\beta0$, $\beta1$, $\beta2$, ..., $\betan$ are the coefficients.

In simple linear regression, there is only one independent variable, and the relationship between the dependent variable and independent variable is modeled using a straight line. In multiple linear regression, there are two or more independent variables, and the relationship between the dependent variable and the independent variables is modeled using a plane or a hyperplane in higher dimensions.

Another difference between simple and multiple linear regression is the complexity of the model. Simple linear regression is a simpler model compared to multiple linear regression since it involves only one independent variable. As a result, the interpretation and visualization of the results are easier in simple linear regression. On the other hand, multiple linear regression can handle more complex relationships between the dependent variable and multiple independent variables.