MACHINE LEARNING WORK SHEET SET 03 Assignment 10

Q1 to Q8 have only one correct answer. Choose the correct option to answer your question.

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

Answer: C) y intercept

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

    A) True   B) False

Answer: A) True

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

Answer: B) the dependent variable

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

Answer: B) Linear Regression

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

Answer: C) the correlation coefficient squared

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

Answer: B) y increases as x increases

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

Answer: A) linear data

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

Answer: A) 0 to 1

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

Answer: B) RMSE, 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.

Answer: A) Linear regression is a supervised learning algorithm, C) Shape of linear regression's cost function is convex.

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

A) Ridge   B) Lasso   C) Pruning   D) Elastic Net

Answer: A) Ridge, B) Lasso, 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

Answer: A) Large amount of training samples with small 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

Answer: A) Linearity, B) Homoscedasticity, D) Normality

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

14. Explain Linear Regression?

Answer:

Linear regression is a statistical method used to establish a relationship between a dependent variable and one or more independent variables. The objective of linear regression is to find the best-fitting straight line that describes the relationship between the independent variable(s) and the dependent variable. The resulting line is called the regression line.

The dependent variable is also called the response variable or target variable, and the independent variable(s) are also called predictor variable(s) or feature(s). In simple linear regression, there is only one independent variable, while in multiple linear regression, there are more than one independent variables.

The equation for a simple linear regression model is as follows:

$Y = a + bX$

where Y is the dependent variable, X is the independent variable, a is the intercept or constant term, and b is the slope coefficient, which represents the change in Y for a one-unit increase in X.

The goal of linear regression is to find the values of a and b that minimize the sum of the squared errors (SSE) between the predicted values and the actual values of the dependent variable. This is typically done using the method of least squares.

Linear regression is commonly used in various fields, such as economics, finance, social sciences, and engineering, to model and predict relationships between variables. It can also be used for causal inference and hypothesis testing.

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.

Simple linear regression involves only one independent variable and one dependent variable, whereas multiple linear regression involves two or more independent variables and one dependent variable.

In simple linear regression, the relationship between the dependent variable and the independent variable is modeled as a straight line, and the equation of the line is given by $Y = a + bX$, where Y is the dependent variable, X is the independent variable, a is the intercept, and b is the slope coefficient.

In multiple linear regression, the relationship between the dependent variable and the independent variables is modeled as a linear equation, and the equation is given by $Y = a + b_1X_1 + b_2X_2 + ... + b_kX_k$, where Y is the dependent variable, $X_1, X_2, ..., X_k$ are the independent variables, a is the intercept, and $b_1, b_2, ..., b_k$ are the slope coefficients.

The goal of both simple and multiple linear regression is to find the values of the intercept and slope coefficients that best fit the data and minimize the sum of squared errors. However, in multiple linear regression, there can be interactions between the independent variables, which can complicate the analysis and interpretation of the results.

In summary, the main difference between simple linear regression and multiple linear regression is the number of independent variables, and multiple linear regression is used when there are two or more independent variables that may affect the dependent variable.