Worksheet_3

MACHINE LEARNING

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

1. In the linear regression equation $y = \theta_0 + \theta_1 X$, θ_0 is the:
A) Slope of the line B) Independent variable
C) y intercept D) Coefficient of determination

Ans: C

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

Ans: A

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

Ans: B

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

Ans: B

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

Ans: C

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

Ans: B

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

Ans: A

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

Ans: A

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

Ans: B, D

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.

Ans: A, C

11. Which of the following regularizations can be applied to linear regression?
A) Ridge B) Lasso
C) Pruning D) Elastic Net

Ans: A,B,D

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

Ans: A,D

13. Which of the following assumptions are true for linear regression?
A) Linearity B) Homoscedasticity
C) Non-Independent D) Normality

Ans:A,B

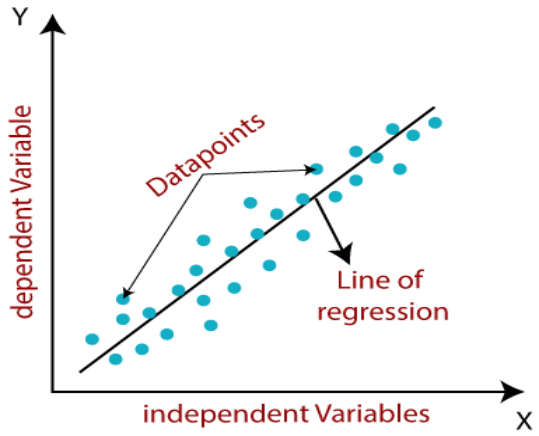
Q14 and Q15 are subjective answer type questions, Answer them briefly.

14. Explain Linear Regression?

Ans: Linear Regression is a machine learning algorithm based on supervised learning and also one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.
Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.
The relation is usually a straight line that best fits the different data points as close as possible. The output is of a continuous form, i.e., numerical value

The linear regression model provides a sloped straight line representing the relationship between the variables as shown below:



Mathematically, we can represent a linear regression as:

$$y= a_0+a_1x+e$$

where:

y=DependentVariable(TargetVariable)
x=IndependentVariable(predictorVariable)
a0=interceptoftheline
a1=coefficient
e= random error
The values for x and y variables are training datasets for Linear Regression model representation.

Linear regression can be further divided into two types of the algorithm:

- **Simple**

  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- **Multiple**

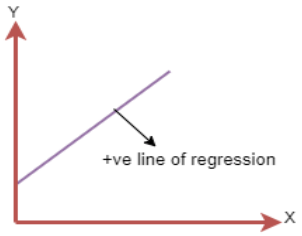
  If more than one independent variable is used to predict the value of a numerical

dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

A linear line showing the relationship between the dependent and independent variables is called a **regression line**. A regression line can show two types of relationship:

- **Positive**

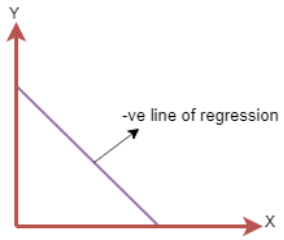
  If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.



The line equation will be: $\mathbf{Y= a_0+a_1x}$

- **Negative**

  If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.



The line of equation will be: $\mathbf{Y= -a_0+a_1x}$

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

Ans:

Simple linear regression:

- ✓ Simple linear regression has only one x and one y variable.
- ✓ For straight-forward relationships, simple linear regression may easily capture the relationship between the two variables.
- ✓ For instance, when we predict rent based on square feet alone that is simple linear regression.

Multiple linear regression:

- ✓ Multiple linear regression has one y and two or more x variables.
- ✓ Multiple linear regression is a more specific calculation than simple linear. For more complex relationships requiring more consideration, multiple linear regression is often better.
- ✓ When we predict rent based on square feet and age of the building that is an example of multiple linear regression.
