ML_ASSISMENT_16

1. In a linear equation, what is the difference between a dependent variable and an independent variable?
Sol:

- In a linear equation, **the dependent variable** is the one we trying to predict or explain, while

- the **independent variable** is the one we using to make that prediction.

- The dependent variable's value depends on the values of one or more independent variables.

- For example, in the equation **y = mx + b,** where y is the dependent variable, x is the independent variable, m is the slope, and b is the intercept.

2. What is the concept of simple linear regression? Give a specific example.
Sol:

- Simple linear regression is a statistical method used to model the relationship between two continuous variables by fitting a linear equation to observed data.

- It involves finding the best-fitting line (linear equation) that minimizes the difference between the observed data points and the predicted values from the line.

- For instance, predicting a person's weight (dependent variable) based on their height (independent variable) is an example of simple linear regression.

3. In a linear regression, define the slope.
Sol:

- In a linear regression equation of the form **y = mx + b**.
- The **slope (m)** represents the change in the **dependent variable (y)** for a unit change in the **independent variable (x)**.
- It indicates the rate at which the dependent variable changes with respect to the independent variable.

4. Determine the graph's slope, where the lower point on the line is represented as (3, 2) and the higher point is represented as (2, 2).
Sol:

- The slope of a line passing through two points (x1, y1) and (x2, y2) is given by the formula:
  slope (m) = (y2 - y1) / (x2 - x1).

- Using the points , the calculation would be: slope = (2 - 2) / (2 - 3)
  $$= 0 / -1$$
  $$= 0$$

5. In linear regression, what are the conditions for a positive slope?

Sol:

- A positive slope in linear regression means that as the independent variable increases, the dependent variable also tends to increase.
- The conditions for a positive slope involve having a positive correlation between the independent and dependent variables, indicating that higher values of the independent variable are associated with higher values of the dependent variable.

6. In linear regression, what are the conditions for a negative slope?

Sol:

- A negative slope in linear regression means that as the independent variable increases, the dependent variable tends to decrease.
- The conditions for a negative slope involve having a negative correlation between the independent and dependent variables, indicating that higher values of the independent variable are associated with lower values of the dependent variable.

7. What is multiple linear regression and how does it work?

Sol:

- Multiple linear regression extends simple linear regression to model the relationship between a dependent variable and multiple independent variables.
- It aims to find the best-fitting linear equation that accounts for the combined effects of these independent variables on the dependent variable

8. In multiple linear regression, define the number of squares due to error.

Sol:

- The sum of squares due to error (SSE) in multiple linear regression quantifies the total unexplained variation in the dependent variable by the regression model.
- It represents the sum of the squared differences between the actual dependent variable values and the values predicted by the regression equation.

9. In multiple linear regression, define the number of squares due to regression.

Sol:

- The sum of squares due to regression (SSR) in multiple linear regression measures the explained variation in the dependent variable by the regression model.
- It represents the sum of the squared differences between the predicted values and the mean of the dependent variable.

10.In a regression equation, what is multicollinearity?

Sol:

- Multicollinearity refers to a situation in multiple linear regression where two or more independent variables are highly correlated with each other.
- This can cause problems in regression analysis because it becomes difficult to isolate the individual effects of these variables on the dependent variable.
- It can lead to unstable coefficient estimates and difficulty in interpreting the importance of each independent variable.

11. What is heteroskedasticity, and what does it mean?

Sol:

- Heteroskedasticity refers to a situation where the variability of the residuals (the differences between the observed and predicted values) in a regression model is not constant across all levels of the independent variable(s).
- In other words, the spread of residuals changes as the values of the independent variable(s) change.
- Heteroskedasticity can affect the accuracy of statistical inference and lead to biased standard errors of coefficient estimates

12. Describe the concept of ridge regression.

Sol:

- Ridge regression is a technique used to address multicollinearity and overfitting in multiple linear regression.
- It adds a penalty term to the regression equation, which discourages the model from assigning large coefficients to any particular independent variable. This helps to stabilize the coefficient estimates and can lead to a more robust model.

13. Describe the concept of lasso regression.

sol:

- Lasso (Least Absolute Shrinkage and Selection Operator) regression is another method to handle multicollinearity and perform variable selection in regression analysis.
- Like ridge regression, it adds a penalty term, but in lasso, the penalty is based on the absolute values of the coefficients. This has the effect of shrinking some coefficients to exactly zero, effectively performing variable selection and creating a sparse model.

14. What is polynomial regression and how does it work?

Sol:

- Polynomial regression is a type of regression analysis that allows for nonlinear relationships between the independent and dependent variables.
- It involves fitting a polynomial equation (quadratic, cubic, etc.) to the data. This can capture more complex patterns that a simple linear regression cannot.

15. Describe the basis function.

Sol:

- Despite its name, logistic regression is used for binary classification problems rather than regression tasks. It models the probability that a given input point belongs to a certain class. It uses the logistic (sigmoid) function to transform a linear combination of the input features into a value between 0 and 1, which represents the probability. If the probability is above a certain threshold (usually 0.5), the data point is classified into one class; otherwise, it's classified into the other class.

16. Describe how logistic regression works.

Sol:

- In the context of regression, a basis function is a mathematical function used to transform the original independent variables into a different space. This transformation can help capture nonlinear relationships between variables. For instance, in polynomial regression, the basis functions are powers of the original independent variable.