Q1. In a linear equation, what is the difference between a dependent variable and an independent variable?

* Dependent Variable: Also known as the response variable or outcome variable, it is the variable being predicted or explained in a study. It depends on the values of one or more independent variables.
* Independent Variable: Also known as the predictor variable, it is the variable that is manipulated or controlled in a study. It is used to explain changes in the dependent variable.

Q2. What is the concept of simple linear regression? Give a specific example.

**Concept of Simple Linear Regression:** Simple linear regression is a statistical method used to model the relationship between a dependent variable (Y) and a single independent variable (X). It assumes a linear relationship between the variables and seeks to find the best-fitting line (regression line) that minimizes the sum of squared residuals.

Example: Predicting a student's exam score (Y) based on the number of hours they studied (X).

Q3. In a linear regression, define the slope.

**Slope in Linear Regression:** The slope (often denoted as "m") in linear regression represents the change in the dependent variable (Y) for a unit change in the independent variable (X). It determines the steepness of the regression line.

Q4. 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).

**Calculating Slope from Two Points:**

Given points (3, 2) and (2, 2)

the slope (m) can be calculated as:

m=2−23−2

=01=0

m=3−22−2​

=10​=0

Q5. In linear regression, what are the conditions for a positive slope?

**Conditions for a Positive Slope:** In linear regression, the slope is positive when an increase in the independent variable corresponds to an increase in the dependent variable. Points on the scatterplot move upward from left to right.

Q6. In linear regression, what are the conditions for a negative slope?

**Conditions for a Negative Slope:** The slope is negative when an increase in the independent variable corresponds to a decrease in the dependent variable. Points on the scatterplot move downward from left to right.

Q7. What is multiple linear regression and how does it work?

**Multiple Linear Regression:** Multiple linear regression extends simple linear regression to include multiple independent variables. It models the relationship between a dependent variable (Y) and two or more independent variables (X₁, X₂, ... Xₖ). It aims to find the best-fitting hyperplane that minimizes the sum of squared residuals.

Q8. In multiple linear regression, define the number of squares due to error.

**Sum of Squares Due to Error (SSE) in Multiple Linear Regression:** SSE represents the sum of squared differences between the observed values of the dependent variable and the predicted values obtained from the regression equation. It measures the total unexplained variation in the dependent variable.

Q9. In multiple linear regression, define the number of squares due to regression.

**Sum of Squares Due to Regression (SSR) in Multiple Linear Regression:** SSR represents the sum of squared differences between the predicted values and the mean of the dependent variable. It measures the explained variation in the dependent variable due to the regression model.

Q10.In a regression equation, what is multicollinearity?

**Multicollinearity in Regression Equation:** Multicollinearity refers to a situation in which two or more independent variables in a regression equation are highly correlated. This can make it challenging to interpret the individual effects of these variables on the dependent variable.

Q11. What is heteroskedasticity, and what does it mean?

**Heteroskedasticity:** Heteroskedasticity occurs when the variability of the residuals (the differences between observed and predicted values) is not constant across all levels of the independent variable(s). It violates the assumption of homoscedasticity (constant variance) in regression analysis.

Q12. Describe the concept of ridge regression.

**Concept of Ridge Regression:** Ridge regression is a regularization technique used to address multicollinearity and overfitting in regression models. It adds a penalty term to the ordinary least squares cost function, constraining the coefficients to be small. This helps stabilize the model and prevent large coefficient values.

Q13. Describe the concept of lasso regression.

**Concept of Lasso Regression:** Lasso regression, short for "Least Absolute Shrinkage and Selection Operator," is another regularization technique that adds a penalty term to the cost function. It not only helps in addressing multicollinearity and overfitting but also performs feature selection by forcing some coefficients to be exactly zero.

Q14. What is polynomial regression and how does it work?

**Polynomial Regression:** Polynomial regression is a form of regression analysis in which the relationship between the dependent variable and the independent variable(s) is modeled as an nth-degree polynomial. It can capture non-linear relationships between variables by adding polynomial terms to the regression equation.

Q15. Describe the basis function.

**Basis Function:** A basis function is a mathematical function used to transform the original input features into a new space, often to capture non-linear relationships or interactions. In polynomial regression, polynomial terms are examples of basis functions.

Q16. Describe how logistic regression works.

**Logistic Regression:** Logistic regression is a statistical method used for binary classification tasks. It models the probability of an instance belonging to a particular class using the logistic function (sigmoid function). It estimates the odds of an event occurring, making it suitable for classification problems with binary outcomes.