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

Dependent Variable: The outcome or response variable we are trying to predict or explain.

Independent Variable: The predictor or explanatory variable used to predict the dependent variable.

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

Simple Linear Regression: Models the relationship between two variables by fitting a linear equation to observed data.

Example: Predicting a student's final exam score (dependent variable) based on hours studied (independent variable).

3. In a linear regression, define the slope.

Slope: The rate of change of the dependent variable for a one-unit change in the independent variable, indicating the strength and direction of their relationship.

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).

Slope: \(\frac{2 - 2}{2 - 3} = \frac{0}{-1} = 0\). The slope is 0, indicating a horizontal line.

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

Positive Slope: Occurs when the dependent variable increases as the independent variable increases, indicating a direct relationship.

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

Negative Slope: Occurs when the dependent variable decreases as the independent variable increases, indicating an inverse relationship.

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

Multiple Linear Regression: Models the relationship between one dependent variable and two or more independent variables by fitting a linear equation to the observed data.

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

Sum of Squares Due to Error (SSE): Measures the total deviation of the observed values from the fitted values, indicating the model's error.

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

Sum of Squares Due to Regression (SSR): Measures the total deviation of the fitted values from the mean of the dependent variable, indicating the explained variation.

10. In a regression equation, what is multicollinearity?

Multicollinearity: Occurs when independent variables in a regression model are highly correlated, making it difficult to isolate their individual effects on the dependent variable.

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

Heteroskedasticity: Refers to the condition where the variance of the errors in a regression model is not constant across all levels of the independent variable, violating one of the assumptions of linear regression.

12. Describe the concept of ridge regression.

Ridge Regression: A technique used to address multicollinearity by adding a penalty term to the linear regression cost function, shrinking the regression coefficients.

13. Describe the concept of lasso regression.

Lasso Regression: A regression technique that adds a penalty term to the cost function to enforce sparsity, effectively shrinking some coefficients to zero, thereby performing variable selection.

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

Polynomial Regression: Models the relationship between the dependent and independent variables as an nth-degree polynomial, allowing for non-linear relationships.

15. Describe the basis function.

Basis Function: Transforms the original input variables into a new set of features to capture the non-linear relationships between the independent and dependent variables in models like polynomial regression.

16. Describe how logistic regression works.

Logistic Regression: Models the probability that a given input belongs to a particular class by using the logistic function to map predicted values to probabilities, used for binary classification.