## Q1

Basic Linear Regression Graph:

Linear regression is a method to model the relationship between a dependent variable (Y) and an independent variable (X) by fitting a linear equation (Y = aX + b) to the observed data points.

In a graph, this is represented by a straight line that best fits the data points. The line is characterized by its slope (a) and intercept (b).

## Q2

Graphical Explanation of Terms:

Rise: The vertical distance between two points on a line, typically represented on the y-axis.

Run: The horizontal distance between two points on a line, typically represented on the x-axis.

Slope: The ratio of the rise to the run, indicating how steep or shallow the line is. Slope (a) = Rise / Run.

## Q3

Graphs for Linear Slopes:

Linear Positive Slope: In this graph, as X increases (positive run), Y also increases (positive rise). The line slopes upward from left to right, indicating a positive relationship between X and Y.

Linear Negative Slope: In this graph, as X increases (positive run), Y decreases (negative rise). The line slopes downward from left to right, indicating a negative relationship between X and Y.

## Q4

Graphs for Curve Linear Slopes:

Curve Linear Positive Slope: In this graph, the line starts with a gentle slope, becomes steeper, and then levels off. It indicates a positive relationship where the rate of increase in Y with respect to X changes.

Curve Linear Negative Slope: In this graph, the line starts steep, becomes less steep, and then levels off. It indicates a negative relationship where the rate of decrease in Y with respect to X changes.

## Q5

Graph for Maximum and Minimum Points:

In a curve, the maximum point is the highest point on the curve, and the minimum point is the lowest point. These points represent extremes in the data.

## Q6

OLS is a method to find the best-fitting line by minimizing the sum of the squared differences between the observed and predicted values.

Intercept (a):

To find the intercept (a), calculate the average of all the Y-values (dependent variable) and subtract the product of the slope (b) and the average of all the X-values (independent variable) from it.

Formula for Intercept (a):

Intercept (a) = Mean(Y) - (Slope (b) \* Mean(X))

Slope (b):

To calculate the slope (b), find the difference between the average of the product of X and Y (X\*Y) and the product of the averages of X and Y. Then, divide this difference by the difference between the average of X-squared (X^2) and the squared average of X.

Formula for Slope (b):

Slope (b) = ((Mean(X\*Y) - (Mean(X) \* Mean(Y))) / (Mean(X^2) - (Mean(X))^2))​

## Q7

OLS Algorithm:

Step 1: Calculate the means of X and Y.

Step 2: Calculate the sums of XY, X^2, and Y^2.

Step 3: Use the formulas for a and b to find the slope and intercept.

Step 4: Construct the linear equation (Y = aX + b).

Step 5: Fit the line to the data points.

## Q8

Regression's Standard Error Graph:

The standard error is a measure of the spread of data points around the regression line. It is represented by the vertical distance between data points and the regression line in a scatterplot.

## Q9

Multiple Linear Regression Example:

In multiple linear regression, there are multiple independent variables. For example, predicting house prices (dependent variable) based on square footage, number of bedrooms, and neighborhood (independent variables).

## Q10

Regression Analysis Assumptions and BLUE Principle:

Assumptions include linearity, independence of errors, constant variance (homoscedasticity), and normality of residuals.

BLUE stands for Best Linear Unbiased Estimators, emphasizing the goal of finding the best-fitting linear model.

## Q11

Issues with Regression Analysis:

Overfitting: Creating a model too complex for the data.

Underfitting: Creating a model too simple to capture relationships.

## Q12

Improving Linear Regression Model Accuracy:

Feature engineering: Selecting relevant independent variables.

Regularization: Using techniques like ridge or lasso regression.

Checking assumptions: Ensuring regression assumptions hold.

## Q13

Polynomial Regression:

Polynomial regression fits a polynomial equation to data instead of a linear equation. For example, predicting a person's weight (Y) based on their age (X) using a quadratic equation.

## Q14

Logistic Regression:

Logistic regression is used for binary classification problems. It models the probability of an event occurring (e.g., spam or not spam) based on independent variables.

## Q15

Logistic Regression Assumptions:

Linearity in the log-odds.

Independence of observations.

No multicollinearity.

Large sample size.

## Q16

Maximum Likelihood Estimation (MLE):

MLE is a method for estimating the parameters of a statistical model. In logistic regression, it finds the parameters that maximize the likelihood of the observed data given the model.