1. Using a graph to illustrate slope and intercept, define basic linear regression.

2. In a graph, explain the terms rise, run, and slope.

3. Use a graph to demonstrate slope, linear positive slope, and linear negative slope, as well as the different conditions that contribute to the slope.

4. Use a graph to demonstrate curve linear negative slope and curve linear positive slope.

5. Use a graph to show the maximum and low points of curves.

6. Use the formulas for a and b to explain ordinary least squares.

7. Provide a step-by-step explanation of the OLS algorithm.

8. What is the regression's standard error? To represent the same, make a graph.

9. Provide an example of multiple linear regression.

10. Describe the regression analysis assumptions and the BLUE principle.

11. Describe two major issues with regression analysis.

12. How can the linear regression model's accuracy be improved?

13. Using an example, describe the polynomial regression model in detail.

14. Provide a detailed explanation of logistic regression.

15. What are the logistic regression assumptions?

16. Go through the details of maximum likelihood estimation.

Answer:

1. Basic linear regression is a statistical method used to model the linear relationship between a dependent variable (Y) and one or more independent variables (X). The regression model can be represented by a straight line equation of the form Y = a + bX, where "a" is the intercept and "b" is the slope. The intercept "a" represents the value of Y when X is zero, while the slope "b" represents the change in Y for every one-unit change in X.
2. In a graph, the rise is the vertical change between two points on the y-axis, while the run is the horizontal change between the corresponding points on the x-axis. The slope is the ratio of the rise to the run and represents the steepness of the line.
3. A positive linear slope is one in which the line rises from left to right, while a negative linear slope is one in which the line falls from left to right. The slope of a line is determined by the change in Y divided by the change in X. A slope of zero indicates a horizontal line, while an undefined slope indicates a vertical line.
4. A curve linear positive slope is one in which the line rises at an increasing rate from left to right, while a curve linear negative slope is one in which the line falls at an increasing rate from left to right.
5. The maximum point of a curve is the highest point on the curve, while the low point is the lowest point on the curve.
6. The formulas for a and b in ordinary least squares (OLS) represent the values that minimize the sum of the squared differences between the predicted values and the actual values. The formula for the intercept "a" is a = Y - bX, while the formula for the slope "b" is b = Σ((Xi - X)(Yi - Y))/Σ(Xi - X)^2.
7. The OLS algorithm involves the following steps: a) Collect data on the independent and dependent variables. b) Calculate the means of the independent and dependent variables. c) Calculate the variance and covariance of the independent variable. d) Calculate the slope and intercept of the regression line using the OLS formula. e) Use the regression equation to predict the dependent variable values for given independent variable values. f) Evaluate the goodness of fit of the regression line using metrics such as R-squared and residuals.
8. The regression's standard error is a measure of the accuracy of the regression line's predictions. It represents the standard deviation of the residuals, or the differences between the predicted and actual values. A graph of the residuals against the predicted values will show the spread of the residuals around the regression line.
9. Multiple linear regression is a statistical method used to model the relationship between a dependent variable and two or more independent variables. For example, a multiple regression model could be used to predict a person's weight based on their height, age, and gender.
10. The regression analysis assumptions include linearity, independence, homoscedasticity, normality, and absence of multicollinearity. The BLUE principle states that the best linear unbiased estimator of the coefficients in a linear regression model is one that is both unbiased and has the smallest variance.
11. Two major issues with regression analysis are multicollinearity and overfitting. Multicollinearity occurs when the independent variables are highly correlated, which can lead to unstable and unreliable estimates of the coefficients. Overfitting occurs when the model is too complex and fits the noise in the data rather than the underlying relationship.
12. The linear regression model's accuracy can be improved by using more relevant and informative independent variables, transforming the variables to fit a more appropriate functional form, removing outliers, and checking for and addressing violations of the regression assumptions.
13. Polynomial regression is a form of linear regression that models the relationship between the dependent variable and one or more independent variables by fitting a polynomial equation to the data. The polynomial equation can take the form Y = a + b1X + b2X^2 + ... + bnx^n, where "n" is the degree of the polynomial. For example, a quadratic polynomial regression model for the relationship between temperature (X) and ice cream sales (Y) might take the form Y = a + b1X + b2X^2, where "a" is the intercept, "b1" and "b2" are the coefficients of the linear and quadratic terms, respectively.
14. Logistic regression is a statistical method used to model the relationship between a binary dependent variable and one or more independent variables. The logistic regression model estimates the probability of the dependent variable taking the value 1 (or yes) given the values of the independent variables. The model takes the form log(p/(1-p)) = a + b1X1 + b2X2 + ... + bnxn, where "p" is the probability of the dependent variable taking the value 1, "a" is the intercept, and "b1" to "bn" are the coefficients of the independent variables.
15. The logistic regression assumptions include linearity of the logit, independence of observations, absence of influential outliers, absence of multicollinearity, and a large enough sample size.
16. Maximum likelihood estimation is a method used to estimate the parameters of a statistical model that maximize the likelihood of the observed data. The likelihood function is a function of the parameters that measures how likely it is to observe the data given a set of parameter values. The maximum likelihood estimate of the parameters is the set of values that maximizes the likelihood function. In logistic regression, the maximum likelihood estimation is used to estimate the coefficients of the independent variables that best predict the probability of the dependent variable taking the value 1. The likelihood function takes the form L = ∏(p^yi)(1-p)^(1-yi), where "p" is the probability of the dependent variable taking the value 1, "yi" is the observed value of the dependent variable, and the product is taken over all observations. The maximum likelihood estimate of the coefficients is obtained by maximizing the log-likelihood function using iterative numerical methods such as Newton-Raphson or gradient descent.