1. In a linear equation, what is the difference between a dependent variable and an independent variable? Answer:--->The dependent variable is the one that depends on the value of some other number. If, say, y = x+2, then the value y can have depends on what the value of x is. Another way to put it is the dependent variable is the output value and the independent variable is the input value. The independent variable is the cause. Its value is independent of other variables in your study. The dependent variable is the effect. Its value depends on changes in the independent variable.A dependent system of equations has infinite solutions, and an independent system has a single solution.In a well-designed experimental study, the independent variable is the only important difference between the experimental (e.g. treatment) and control (e.g. placebo) groups. The dependent variable is the variable being tested and measured in an experiment, and is 'dependent' on the independent variable

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

Answer:--->The main aim of linear regression is the find out relation ship predict the dependent and independent variable with the help of best fit line. the best fit line is predicted the dependent and independent variable with y = mx+c y=independent x=dependent m=slope c=intercept

3. In a linear regression, define the slope.

Answer:--->slope = m = rise/run = dy/dx = y/ x = Parallel lines have equal slopes. In summary, if y = mx + b, then m is the slope and b is the y-intercept (i.e., the value of y when x = 0). Often linear equations are written in standard form with integer coefficients (Ax + By = C).In a regression line passing through a set of data points in data sets Argument1 and Argument2, the slope is the vertical distance divided by the horizontal distance between any two points on the line. This ratio is also known as the rate of change along the line.

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).

Answer:--->The slope of a line is calculated as the change in y-value divided by the change in x-value between two points on the line. In this case, the change in y-value between the two points is 0 (both points have a y-value of 2) and the change in x-value is -1 (the x-value of the lower point is 3 and the x-value of the higher point is 2). Therefore, the slope of the line is 0/-1 = 0.

Alternatively, you could say that the line is horizontal and has no slope. A horizontal line has a slope of 0 because it has no change in y-value as you move along it.

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

Answer:---> if the slope is positive, y increases as x increases, and the function runs "uphill" (going left to right). If the slope is negative, y decreases as x increases and the function runs downhill. If the slope is zero, y does not change, thus is constant—a horizontal line.A positive slope means that two variables are positively related—that is, when x increases, so does y, and when x decreases, y decreases also. Graphically, a positive slope means that as a line on the line graph moves from left to right, the line rises.07-Aug-2020

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

Answer:--->If the slope is negative, y decreases as x increases and the function runs downhill. If the slope is zero, y does not change, thus is constant—a horizontal line. Vertical lines are problematic in that there is no change in x.A negative slope means that two variables are negatively related; that is, when x increases, y decreases, and when x decreases, y increases. Graphically, a negative slope means that as the line on the line graph moves from left to right, the line falls.In general, straight lines have slopes that are positive, negative, or zero. If we were to examine our least-square regression lines and compare the corresponding values of r, we would notice that every time our data has a negative correlation coefficient, the slope of the regression line is negative.

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

Answer:-->Multiple linear regression refers to a statistical technique that uses two or more independent variables to predict the outcome of a dependent variable. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance.Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. Multiple regression is an extension of linear (OLS) regression that uses just one explanatory variable.Multiple Linear Regression Analysis consists of more than just fitting a linear line through a cloud of data points. It consists of 3 stages – (1) analyzing the correlation and directionality of the data, (2) estimating the model, i.e., fitting the line, and (3) evaluating the validity and usefulness of the model.

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

Answer:--->The Mean Squared Error measures how close a regression line is to a set of data points. It is a risk function corresponding to the expected value of the squared error loss. Mean square error is calculated by taking the average, specifically the mean, of errors squared from data as it relates to a function.

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

Answer:--->The multiple regression equation explained above takes the following form: y = b1x1 + b2x2 + … + bnxn + c. Here, bi's (i=1,2…n) are the regression coefficients, which represent the value at which the criterion variable changes when the predictor variable changes.In other words, the total sum of squares measures the variation in a sample. Sum of squares regression: Sum of squares due to regression is represented by SSR. It is the difference between the predicted value and the sample mean.The sum of squares is a form of regression analysis to determine the variance from data points from the mean. If there is a low sum of squares, it means there's low variation. A higher sum of squares indicates higher variance.

In a regression equation, what is multicollinearity?

Answer:-->Multicollinearity exists whenever an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation. Multicollinearity is a problem because it will make the statistical inferences less reliable.Multicollinearity refers to a situation in which more than two explanatory variables in a multiple regression model are highly linearly related. There is perfect multicollinearity if, for example as in the equation above, the correlation between two independent variables equals 1 or −1.Multicollinearity reduces the precision of the estimated coefficients, which weakens the statistical power of your regression model. You might not be able to trust the p-values to identify independent variables that are statistically significant.

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

Answer:--->In statistics, heteroskedasticity (or heteroscedasticity) happens when the standard deviations of a predicted variable, monitored over different values of an independent variable or as related to prior time periods, are non-constant.Heteroskedasticity refers to situations where the variance of the residuals is unequal over a range of measured values. When running a regression analysis, heteroskedasticity results in an unequal scatter of the residuals (also known as the error term).Homoscedasticity, or homogeneity of variances, is an assumption of equal or similar variances in different groups being compared. This is an important assumption of parametric statistical tests because they are sensitive to any dissimilarities. Uneven variances in samples result in biased and skewed test results.

12. Describe the concept of ridge regression. Answer:-->Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.Ridge regression is the method used for the analysis of multicollinearity in multiple regression data. It is most suitable when a data set contains a higher number of predictor variables than the number of observations. The second-best scenario is when multicollinearity is experienced in a set.Ridge regression adds a ridge parameter (k), of the identity matrix to the cross product matrix, forming a new matrix (X`X + kI). It's called ridge regression because the diagonal of ones in the correlation matrix can be described as a ridge.

13. Describe the concept of lasso regression.

Answer:Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. Lasso is a modification of linear regression, where the model is penalized for the sum of absolute values of the weights. Thus, the absolute values of weight will be (in general) reduced, and many will tend to be zeros.The goal of lasso regression is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. The lasso does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero.

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

Answer:-->In polynomial regression, the relationship between the independent variable x and the dependent variable y is described as an nth degree polynomial in x. Polynomial regression, abbreviated E(y |x), describes the fitting of a nonlinear relationship between the value of x and the conditional mean of polynomial is an expression consisting of indeterminates (also called variables) and coefficients, that involves only the operations of addition, subtraction, multiplication, and positive-integer powers of variables. An example of a polynomial of a single indeterminate x is x2 − 4x + 7.Polynomial Regression is a form of Linear regression known as a special case of Multiple linear regression which estimates the relationship as an nth degree polynomial. Polynomial Regression is sensitive to outliers so the presence of one or two outliers can also badly affect the performance.In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x.

15. Describe the basis function.

Answer:--->Basis functions (called derived features in machine learning) are building blocks for creating more complex functions. In other words, they are a set of k standard functions, combined to estimate another function—one which is difficult or impossible to model exactly.This is a generalization of linear regression that essentially replaces each input with a function of the input. (A linear basis function model that uses the identity function is just linear regression. In mathematics, a basis function is an element of a particular basis for a function space. Every function in the function space can be represented as a linear combination of basis functions, just as every vector in a vector space can be represented as a linear combination of basis vectors.

16. Describe how logistic regression works.

Answer:--->Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of certain classes based on some dependent variables. In short, the logistic regression model computes a sum of the input features (in most cases, there is a bias term), and calculates the logistic of the result.Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.Logistic Regression is a classification technique used in machine learning. It uses a logistic function to model the dependent variable. The dependent variable is dichotomous in nature, i.e. there could only be two possible classes (eg.:

either the cancer is malignant or not).logistic regression is also used to estimate the relationship between a dependent variable and one or more independent variables, but it is used to make a prediction about a categorical variable versus a continuous one