

Assignment 16 Solutions

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

Ans: Algebraically, a linear equation typically takes the form $y = mx + b$, where m and b are constants, x is the independent variable, y is the dependent variable. The slope tells us how the dependent variable (y) changes for every one unit increase in the independent (x) variable, on average.

The variables in a study of a cause-and-effect relationship are called the independent and dependent variables. The independent variable is the cause. Its value is independent of other variables in your study. The dependent variable is the effect

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

Ans: Simple linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative. Linear regression most often uses mean-square error (MSE) to calculate the error of the model.

For example, suppose that height was the only determinant of body weight. In this example, if an individual was 70 inches tall, we would predict his weight to be: $\text{Weight} = 80 + 2 \times (70) = 220 \text{ lbs}$. In this simple linear regression, we are examining the impact of one independent variable on the outcome.

3. In a linear regression, define the slope ?

Ans: The slope of a regression line (b) represents the rate of change in y as x changes. Because y is dependent on x , the slope describes the predicted values of y given x . The slope must be calculated before the y -intercept when using a linear regression, as the intercept is calculated using the slope.

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) ?

Ans: To determine the slope of a line given two points, we can use the formula:

$$\text{slope (m)} = (\text{change in } y) / (\text{change in } x)$$

Let's calculate the slope using the given points (3, 2) and (2, 2):

$$\text{change in } y = 2 - 2 = 0 \quad \text{change in } x = 3 - 2 = 1$$

$$\text{slope (m)} = 0 / 1 = 0$$

Therefore, the slope of the line passing through the points (3, 2) and (2, 2) is 0. This means the line is horizontal, neither rising nor falling as we move from left to right

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

Ans: If the slope is positive, y increases as x increases, and the function runs "uphill" (going left to right). If the slope is zero, y does not change, thus is constant—a horizontal line.

If these conditions are satisfied and the slope coefficient is positive, it suggests a positive relationship between the independent variable (X) and the dependent variable (Y). This means that as X increases, Y is expected to increase as well. However, it's important to note that correlation does not imply causation, and further analysis is often required to establish the direction and nature of the relationship.

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

Ans: If the slope is negative, y decreases as x increases and the function runs downhill.

If these conditions are satisfied and the slope coefficient is negative, it suggests a negative relationship between the independent variable (X) and the dependent variable (Y). This means that as X increases, Y is expected to decrease. However, as previously mentioned, correlation does not imply causation, and further analysis is often required to establish the direction and nature of the relationship.

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

Ans: 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.

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

Ans: The mean squared error (MSE) tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the "errors") and squaring them. It's called the mean squared error as you're finding the average of a set of errors.

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

Ans: Sum of squares is a statistical technique used in regression analysis to determine the dispersion of data points. In a regression analysis, the goal is to determine how well a data series can be fitted to a function that might help to explain how the data series was generated.

10. In a regression equation, what is multicollinearity ?

Ans: Multicollinearity occurs when two or more independent variables are highly correlated with one another in a regression model. This means that an independent variable can be predicted from another independent variable in a regression model.

It is important to identify and address multicollinearity in regression analysis. Some common techniques to deal with multicollinearity include removing one of the correlated variables, transforming variables, collecting more data, or using regularization techniques like ridge regression or lasso regression.

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

Ans: Heteroskedasticity (also spelled heteroscedasticity) refers to the error variance, or dependence of scattering, within a minimum of one independent variable within a particular sample. A common cause of variances outside the minimum requirement is often attributed to issues of data quality.

12. Describe the concept of ridge regression ?

Ans: Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. It is hoped that the net effect will be to give estimates that are more reliable.

13. Describe the concept of lasso regression ?

Ans: In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

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

Ans: Polynomial Regression is a form of Linear regression known as a special case of Multiple linear regression which estimates the relationship as an n th degree polynomial. Polynomial Regression is sensitive to outliers so the presence of one or two outliers can also badly affect the performance.

15. Describe the basis function ?

Ans: 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.)

The basis functions are mathematical functions used in regression analysis to transform input features into a higher-dimensional space, enabling the modeling of non-linear relationships between variables. They help capture complex patterns and non-linearities in the data. Common types of basis functions include polynomials, radial basis functions, sigmoid functions, Fourier series, and wavelet functions. The choice of basis functions depends on the specific problem and desired non-linear relationships to be represented.

16. Describe how logistic regression works ?

Ans: Logistic regression uses an equation as the representation, very much like linear regression. Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y).

The logistic regression is a statistical model used for binary classification. It estimates the probability of an event occurring based on input features. It applies a logistic/sigmoid function to a linear combination of the features and coefficients. The model is trained by optimizing the coefficients to maximize the likelihood of observed labels. A decision threshold is used to classify new instances. Logistic regression is evaluated using metrics like accuracy, precision, recall, and F1 score. It is interpretable and widely used in various

In []:

| | |
|---|--|
| 1 | |
|---|--|