

Linear regression is a statistical method that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. This technique is commonly used in statistics and machine learning to understand relationships between variables.

Objective

The primary goal of linear regression is to determine the best-fitting line, minimizing the sum of squared differences between the observed and predicted values. In other words, the model aims to capture the underlying linear relationship between variables while accounting for variability.

Advantages of Linear Regression

- Linear regression is a relatively simple algorithm, making it easy to understand and implement.
- It is computationally efficient and can handle large datasets effectively. It can be trained quickly on large datasets, making it suitable for real-time applications.
- It often serves as a good baseline model for comparison with more complex machine learning algorithms.
- It is a well-established algorithm with a rich history and is widely available in various machine learning libraries and software packages.

Disadvantages of Linear Regression

- Linear regression assumes a linear relationship between the dependent and independent variables. If the relationship is not linear, the model may not perform well.
- It is sensitive to multicollinearity, which occurs when there is a high correlation between independent variables. Multicollinearity can inflate the variance of the coefficients and lead to unstable model predictions.
- It assumes that the features are already in a suitable form for the model. Feature engineering may be required to transform features into a format that can be effectively used by the model.
- It is susceptible to both overfitting and underfitting.