Linear Regression

Devesh

November 25, 2024

1 Introduction

Linear regression is a statistical method for modeling the relationship between a dependent variable and one or more independent variables. It is one of the simplest and most widely used techniques in machine learning.

2 Simple Linear Regression

Simple linear regression models the relationship between two variables by fitting a linear equation to observed data. The equation of a simple linear regression line is:

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

where:

- y is the dependent variable.
- x is the independent variable.
- β_0 is the y-intercept.
- β_1 is the slope of the line.
- ϵ is the error term.

3 Multiple Linear Regression

Multiple linear regression models the relationship between a dependent variable and multiple independent variables. The equation of a multiple linear regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{2}$$

where:

- y is the dependent variable.
- $x_1, x_2, \ldots, x_n$ are the independent variables.
- β_0 is the y-intercept.
- $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of the independent variables.
- ϵ is the error term.

4 Assumptions of Linear Regression

Linear regression makes several key assumptions:

- Linearity: The relationship between the dependent and independent variables is linear.
- Independence: The observations are independent of each other.
- Homoscedasticity: The variance of the error terms is constant across all levels of the independent variables.
- Normality: The error terms are normally distributed.

5 Evaluating the Model

The performance of a linear regression model can be evaluated using several metrics, including:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R-squared (R^2)

6 Linear Regression Algorithm

The linear regression algorithm can be summarized in the following steps:

1. Initialize the parameters β_0 and β_1 to random values.
2. For each iteration:
   (a) Compute the predicted values $\hat{y} = \beta_0 + \beta_1 x$.
   (b) Calculate the cost function, typically Mean Squared Error (MSE):

   $$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

   (c) Update the parameters using gradient descent:

   $$\beta_0 := \beta_0 - \alpha\frac{\partial\text{MSE}}{\partial\beta_0}$$

   $$\beta_1 := \beta_1 - \alpha\frac{\partial\text{MSE}}{\partial\beta_1}$$

3. Repeat until convergence.

7 Conclusion

Linear regression is a powerful and widely used technique for modeling the relationship between variables. By understanding its assumptions and how to evaluate its performance, we can effectively apply linear regression to a variety of problems.