Linear Regression

Introduction

Linear Regression is a fundamental supervised learning algorithm used for modeling the relationship between a dependent variable and one or more independent variables. It aims to find the best-fitting straight line that describes the linear relationship between the input features and the target variable.

Model Representation

1. **Simple Linear Regression: In simple linear regression, there is only one independent variable used to predict the dependent variable. The model can be represented as: \(y = mx + b \), where \(y \) is the predicted value, \(x \) is the input feature, \(m \) is the slope, and \(b \) is the intercept.**

2. **Multiple Linear Regression: In multiple linear regression, there are multiple independent variables used to predict the dependent variable. The model can be represented as: \(y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n \), where \(y \) is the predicted value, \(x_i \) are the input features, and \(b_0, b_1, ..., b_n \) are the coefficients.**

Training a Linear Regression Model

1. **Objective Function:** **The goal of linear regression is to minimize the mean squared error (MSE) between the predicted values and the actual values in the training data.**
2. **Optimization Algorithm:** **Gradient Descent is commonly used to iteratively update the coefficients in order to minimize the MSE and find the best-fitting line.**

Evaluation Metrics

1. **R-squared (R^2):** **R-squared measures the proportion of variance in the dependent variable that is predictable from the independent variables. A higher R-squared value indicates a better fit.**
2. **Mean Squared Error (MSE):** **MSE calculates the average squared difference between the predicted values and the actual values. It is used to evaluate the overall performance of the model.**

Assumptions of Linear Regression

1. **Linearity:** **The relationship between independent and dependent variables is linear.**
2. **Independence:** **The residuals (errors) of the model are independent of each other.**
3. **Homoscedasticity:** **The variance of residuals should be constant across all levels of the independent variables.**

Regularization Techniques

To prevent overfitting in linear regression models, regularization techniques such as Lasso (L1 regularization) and Ridge (L2 regularization) can be applied to penalize large coefficients.

Applications of Linear Regression

Linear regression is widely used in various fields such as economics, finance, biology, and social sciences for tasks like predicting sales, analyzing trends, and understanding relationships between variables.

Conclusion

Linear Regression is a powerful and widely used algorithm in regression analysis due to its simplicity, interpretability, and efficiency in modeling linear relationships between variables. By fitting a straight line to the data, linear regression provides valuable insights into the underlying patterns and trends in the dataset.

Producer: Elham Jafari

Computer Engineering