Lasso regression

Lasso regression, also known as L1 regularization, is a type of linear regression that adds a penalty term to the cost function to prevent overfitting. This penalty term is the sum of the absolute values of the regression coefficients multiplied by a tuning parameter alpha.

In traditional linear regression, the objective is to minimize the sum of squared errors between the predicted and actual values. However, this can lead to overfitting when the number of predictors is large compared to the number of observations. Lasso regression helps to address this issue by shrinking the coefficients of less important predictors towards zero, effectively performing feature selection and reducing the complexity of the model.

Lasso regression is useful when dealing with high-dimensional datasets that have a large number of features. It can also be used for feature selection when trying to identify the most important predictors in a dataset. Additionally, Lasso regression can be used for data preprocessing before applying other machine learning algorithms.

A practical real-life example of Lasso regression is predicting house prices based on various features such as the number of bedrooms, square footage, and location. In this scenario, Lasso regression can help to identify the most important features that contribute to the price of the house, while also reducing the influence of less important features.

The working of Lasso regression involves the addition of the L1 penalty term to the cost function of traditional linear regression. The cost function can be written as:

Cost = (1/2m) * sum((y - y_predicted)^2) + alpha * sum(|w|)

where m is the number of observations, y is the actual value, y_predicted is the predicted value, w is the regression coefficient, and alpha is the tuning parameter that controls the strength of the penalty term.

The L1 penalty term adds a constant multiple of the absolute value of the regression coefficients to the cost function, forcing some of the coefficients to be zero. This results in a sparse model where only the most important features are included in the regression equation.