Linear Regression
Epoch IITHyderabad

Chakka Surya Saketh
AI22BTECH11005

Introduction

Linear regression is a supervised machine learning algorithm used to predict the relationship between two variables.it can also be extended to multiple linear regression.The aim is to find a best fitting line that explains the relationship.It is used for predicting a continuous response.
When it comes to simple linear regression

$$y_(pred) = b + WX_i$$

describes the equation of the line which is used to predict the output for the samples. The aim of the algorithm is to find the best values of $b(bias)$ and $W(weights)$ so that we have a minimum residual sum of squares.
Similarly we extend the equation to multiple dependent variables in multiple linear regression.

Cost function

$$J = (1/2n)\sum_{i=1}^{n}(y_{pred} - y_i)^2$$

where n is the number of data points.

Gradient Descent

It is the algorithm generally used to optimise the cost function.
The main parameter of it is the learning rate(η)
If η is high the algorithm just oscillates and if it is low then it takes a low of computational power to reach the ideal parameters.
A good η can be chosen with the help of grid search cv

Implementation

While implementing linear regression we first initialise the weights and bias and begin the prediction.
Then we calculate the cost of the prediction.
If the cost is more than a suitable value we run the gradient descent step and update the weights and biases

$$W = W - \eta\frac{\partial J}{\partial W}$$

and

$$b = b - \eta\frac{\partial J}{\partial b}$$

On calculation

$$\frac{\partial J}{\partial W} = 1/n \sum_{i=1}^{n} y_{pred} - y_i$$

and

$$\frac{\partial J}{\partial W} = 1/n \sum_{i=1}^{n} (y_{pred} - y_i) X_i$$

We can keep running these steps for a specified number of times or till the cost function reaches a specified value.

Assumptions

1) Linearity: The relationship between the dependent and independent variables should be linear.

2) Independence: Statistical analyses rely on the assumption that observations are independent and not influenced by one another to avoid biased results i.e there should not be any visible patterns in the error terms.The absence of this phenomenon is known as Autocorrelation.

3) Homoscedasticity: The error terms must have constant variance. This phenomenon is known as Homoscedasticity. Generally, non-constant variance arises in the presence of outliers or extreme leverage values.

4) Normality: The mean of residuals should follow a normal distribution with a mean equal to zero or close to zero. If the error terms are non-normally distributed, suggests that there are a few unusual data points that must be studied closely to make a better model.

5) No Multicollinearity: To avoid redundancy and multicollinearity issues, independent variables should not be strongly correlated with each other, preserving the stability of multiple regression models.

Evaluation Metrics

1) Root Mean Squared Error (RMSE): It is the square root of the MSE and provides a more interpretable measure of the model's predictive accuracy. For this reason it is the most commonly used metric. It specifies the absolute fit of the model to the data.

2) Mean Squared Error (MSE): It calculates the average squared difference between the actual and predicted values. Lower MSE values indicate better model performance.

3) R-squared (R2): It explains the amount of variation that is captured by the developed model. It always ranges between 0 and 1. A higher R-squared value indicates a better fit.