Bayesian Linear Regression vs Linear Regression: A comprehensive Guide

March 19, 2023

Linear regression is a statistical technique that has been used for more than a century to model the relationship between a dependent variable and one or more independent variables. This technique is widely used in various fields like economics, engineering and social sciences, computer sciences, data science, among others. However, with the advent of Bayesian statistics, a new approach to linear regression has emerged, which is Bayesian linear regression. In this blog, we will discuss the differences between Bayesian linear regression and linear regression and show some plots and posterior plots to help you understand the topic better.

What is Linear Regression?

Linear regression is a statistical technique that models the relationship between a dependent variable (Y) and one ore more independent variables (X) by fitting a linear equation to the observed data. The equation of a linear regression model is given as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

Where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficient of the independent variables, $X_1, X_2, \ldots, X_n$ and ϵ is the error term.

The goal of linear regression is to find the best values of the coefficients that minimizes the sum of squared errors between the predicted values and the actual values of the dependent variable. The technique assumes the error term is normally distributed and has a constant variance.

What is Bayesian Linear Regression ?

Bayesian Linear regression, on the other hand, is a probabilistic approach to linear regression that uses Baye's theorem to update the prior beliefs of the model parameters based on the observed data. In this approach, the model parameters are treated as random variables, and prior distributions are specified for each parameter. The observed data are then used to update the prior distributions to obtain posterior distributions, which represent our updated beliefs about the model parameters given the the data. The equation pf Bayesian linear regression is same as linear regression above.

In Bayesian linear regression, the prior distributions for the model parameters are specified as probability distributions. The posterior distributions are then obtained by applying Baye's theorem to update the prior distributions based on the observed data.

Difference between Bayesian Linear regression and Linear regression

The main differences between Bayesian linear regression and linear regression are:

1. Prior Distributions:In linear regression, the model parameters are treated as fixed values, while in Bayesian linear regression, the model parameters are treated as random variables with prior distributions.

2. Posterior Distributions: In linear regression, the model parameters are estimated using maximum likelihood estimation, while in Bayesian linear regression, the posterior distributions of the model parameters are obtained by applying Baye's theorem.

3. Interpretation: In linear regression, the coefficients of the independent variables represent the change in the dependent variable for a unit change in the independent variable, while in Bayesian linear regression, the posterior distribution represent the uncertainty in the model parameters given the observed data.

4. Complexity: Bayesian linear regression is generally more computationally intensive than linear regression since it involve updating the prior distributions based on the observed data.

Plots and Posterior plots

To illustrate the differences between Bayesian linear regression and linear regression, we will use a simple example of predicting the price of a house based on its size. We will use the Boston Housing dataset and fit both a linear regression model and a Bayesian linear regression model.

Linear Regression

First, let's fit linear regression model to the data and plot the results
The plot shows the linear regression line, which represents the relationship between the price of a house and its size. The slope of the line represents the change in the price of a house for a unit change in its size.

Bayesian linear regression

Next, let's fit a Bayesian linear regression model to the data and plot posterior distributions of the model parameters.
The plot shows the posterior distributions of the intercept and the slop of the linear regression line. The distributions represents the uncertainty in the values of the intercept and the slope given the observed data. We can see that the distributions are centered around the values obtained from linear regression model, but they have a wider spread, indicating more uncertainty in the values.

Conclusion

In conclusion, Bayesian linear regression is a powerful technique that allows us to incorporate prior beliefs or knowledge into the model and obtain posterior distribution that represent our updated beliefs about the model parameters given observed data. Although Bayesian linear regression is more computationally intensive than linear regression, it provides a more comprehensive and nuanced understanding of the relationship between the dependent variable and independent variables