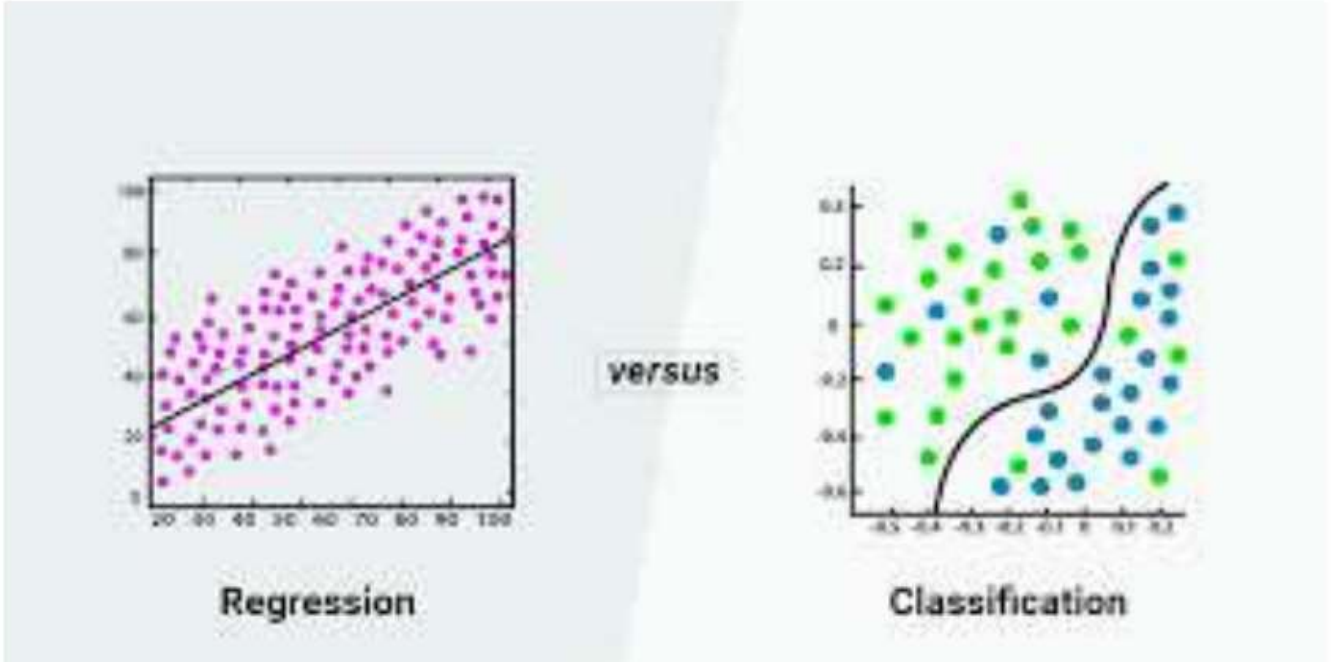
Bayesian Linear Regression

Regression Vs Classification

- Regression is a Machine Learning task to predict continuous values (real numbers), as compared to classification, that is used to predict categorical (discrete) values.



Linear Regression

- Linear Regression is a very simple machine learning method in which each data points is a pair of vectors: the input vector and the output vector.
- Linear regression is a popular regression approach in machine learning.
- Linear regression is based on the assumption that the underlying data is normally distributed and that all relevant predictor variables have a linear relationship with the outcome.
- But In the real world, this is not always possible, it will follows these assumptions, Bayesian regression could be the better choice.

Bayesian Linear Regression

- Bayesian regression employs prior belief or knowledge about the data to “learn” more about it and create more accurate predictions.
- Bayesian Regression methods is a very powerful method because they provide us with the entire distribution over regression parameters.
- In order to calculate inadequate data or unequal distributed data, Bayesian Linear Regression provides a natural mechanism.
- Statistical analysis is conducted under the conditions of Bayesian interface in Bayesian linear regression in statistics.

Bayesian Regression

- Bayesian regression uses a Bayes algorithm to estimate the parameters of a linear regression model from data, including prior knowledge about the parameters.

- Because of its probabilistic character, it can produce more accurate estimates for regression parameters than ordinary least squares (OLS) linear regression, provide a measure of uncertainty in the estimation, and make stronger conclusions than OLS.

- The output is achieved from a probability distribution, rather than usual regression techniques.
- The goal of Bayesian linear regression is to find Posterior instead of model parameters.
- Model parameters are supposed to occur from a distribution.

posterior expression

- The posterior expression is

  Posterior=(Likelihood*Prior)/Normalization where

- Posterior: It is the probability of an event to occur; say, H, given that another event; say, E has already occurred. i.e., P(H | E).

- Prior: It is the probability of an event H has occurred prior to another event. i.e., P(H)

- Likelihood: It is a likelihood function in which some parameter variable is marginalized. (Chance of something to occur)

Bayes' Theorem

- The above equation is similar to Bayes' Theorem, which is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- where A and B are events
- P(A) is the probability of occurrence of A
- P(A|B) is the probability of A to occur given that event B has already occurred.
- P(B), the probability of event B occurring cannot be 0 since it has already occurred.

Advantages of Bayesian Regression

Advantages of Bayesian Regression:

- With the help of Bayesian processing, we can retrieve the complete variety of inferential(characterized) solutions instead of a point estimate.
- It works efficiently with the small size of the dataset.
- It is very suitable for the online form of learning, whereas, in the form of batch learning, we have the whole dataset.
- It is a very powerful and tested approach.

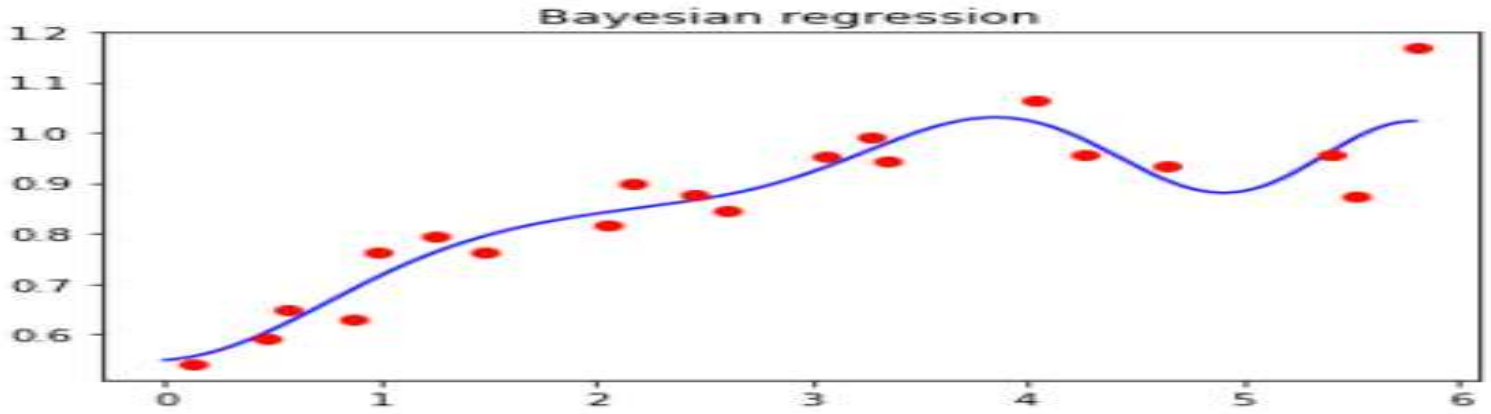
Disadvantages of Bayesian Regression

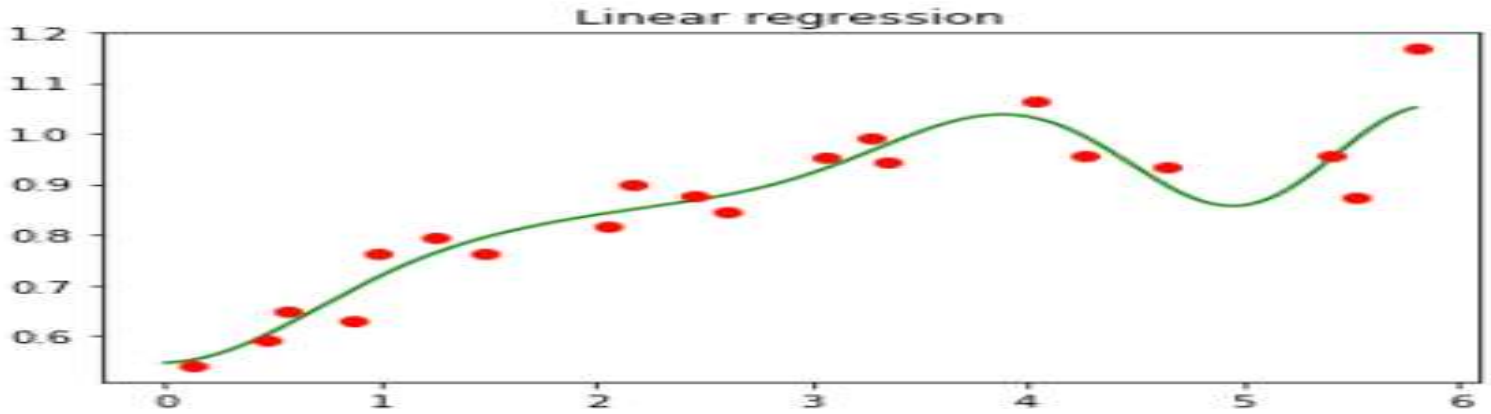
Disadvantages of Bayesian Regression:

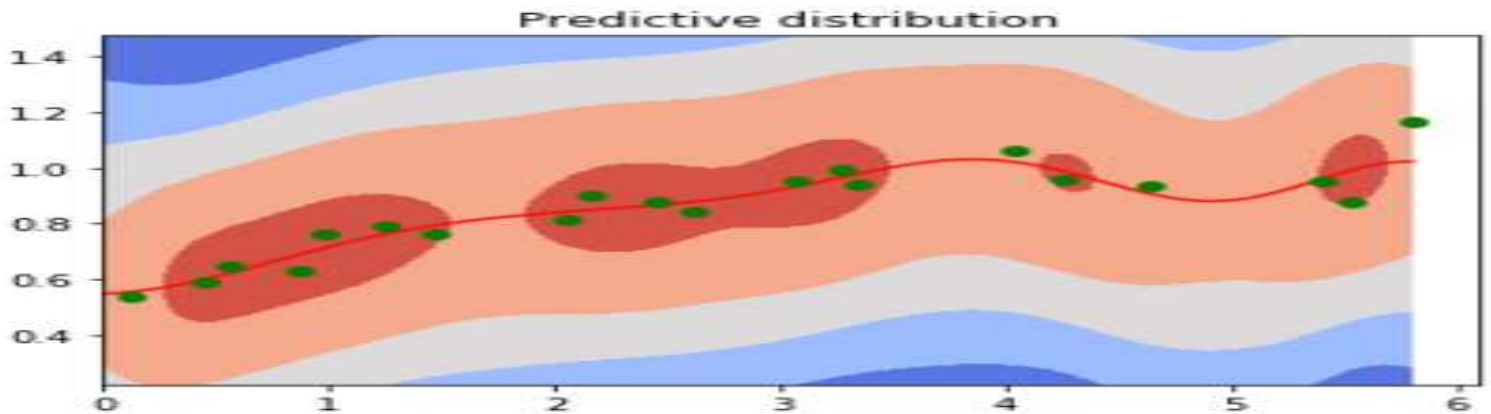
- It does not work efficiently if the dataset contains a huge amount of data. The conjecture of the model can be time-consuming.

Real-life Application Of Bayesian Linear Regression

- Using Priors: Consider a scenario in which your supermarkets carry a new product, and we want to predict its initial Christmas sales.


Bayesian regression
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6


Linear regression
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6


Predictive distribution
1.4
1.2
1.0
0.8
0.6
0.4
0
1
2
3
4
5
6

- According to 3, the predictive distribution can give the confidence on the prediction if it is within the dense-color area because of the data is dense, but not in sparse area,
- eg. prediction at x=5 may not be trustworthy.