Machine Learning Bootcamp

Project Report
Winter of Code 4.0

Vaibhav Jain
IIT ISM Dhanbad

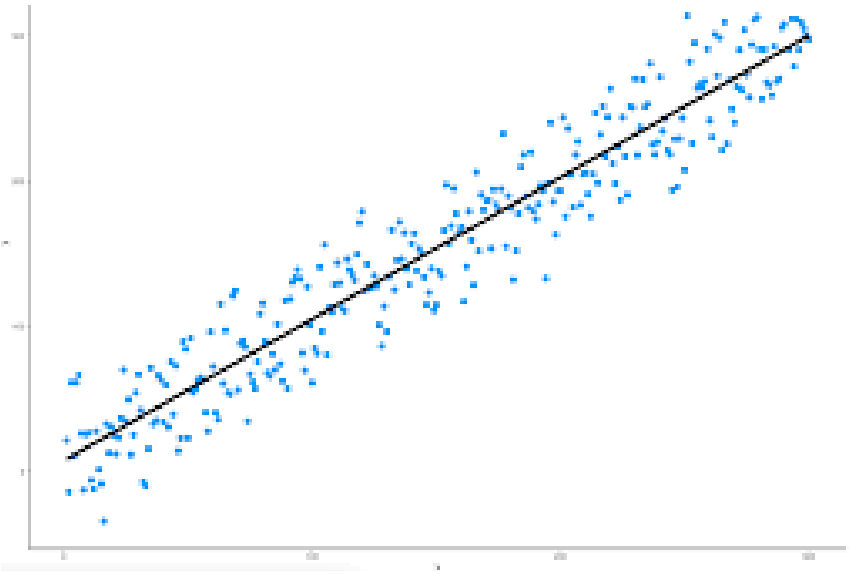
Description:

Machine learning is an application of artificial intelligence that allows systems to learn without being explicitly programmed automatically. In the given time, I have implemented five algorithms from scratch using only basic python libraries like NumPy, Pandas, and matplotlib.

Implementation:

1.Linear Regression

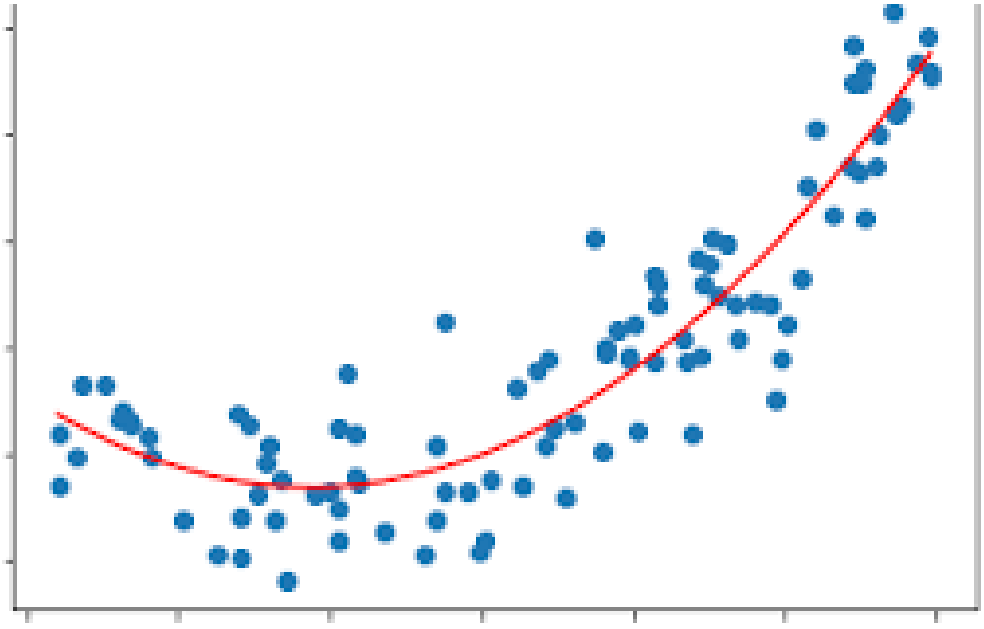
Linear regression is an algorithm through which we can relate two variables with the help of a line that represents the given data set most accurately. The aim is to minimize the cost function by training the data sets to predict accurately.



In my implementation of linear regression, my model has obtained the root mean square error equal to 52, and the cost function converges to 1907. To get a more generalized model, I have also standardized the training and testing datasets to get close to the Gaussian distribution curve.

2.Polynomial Regression

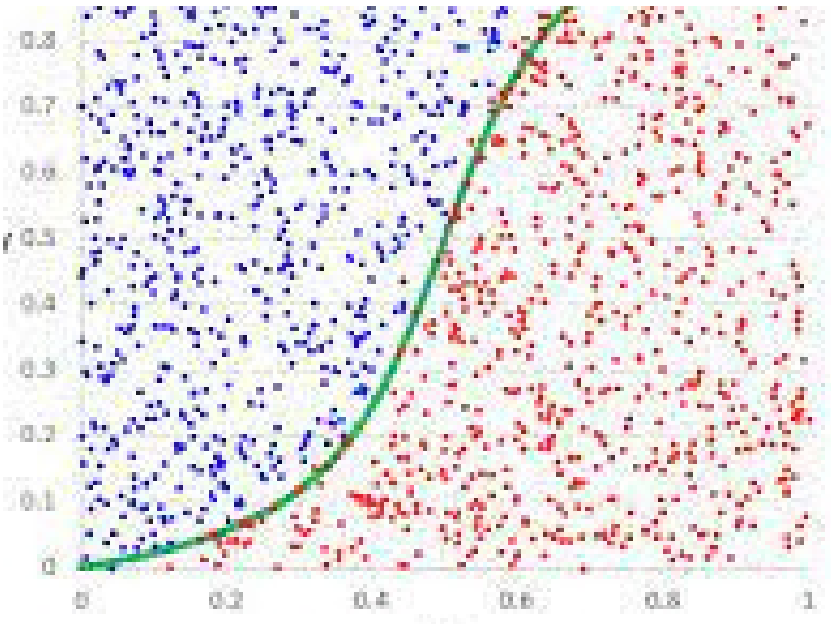
Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial.



In my implementation of polynomial regression, my model has obtained the root mean square error (RMS error) equal to 7.855, and the cost function converges to 36.71. I have taken the hypothesis function as a cubic polynomial and have standardized the datasets of both testing and training sets.

3.Logistic Regression

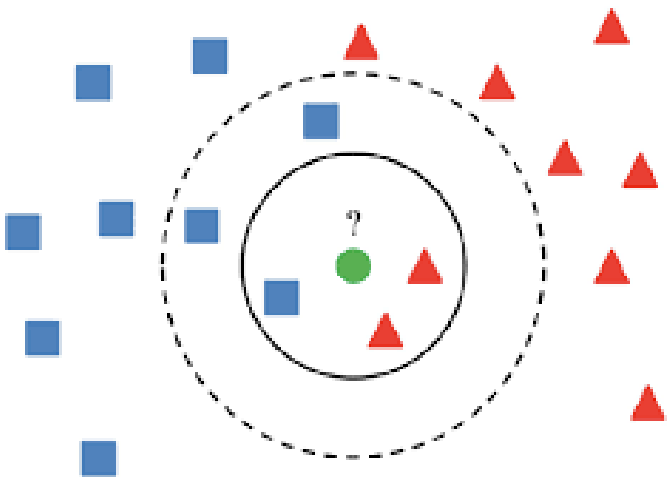
Linear regression fails in the case of classification, so we use logistic regression instead. This algorithm predicts the value of the categorical variable based on input. The outputs are brought in the range (0,1) using the sigmoid function to predict the probability. Based on the output it divides the data set into two classes. As in linear regression, the gradient descent method can also be applied to make accurate predictions.



In my implementation of this algorithm, I have got an accuracy of 69.56%

4. K-Nearest Neighbours(KNN)

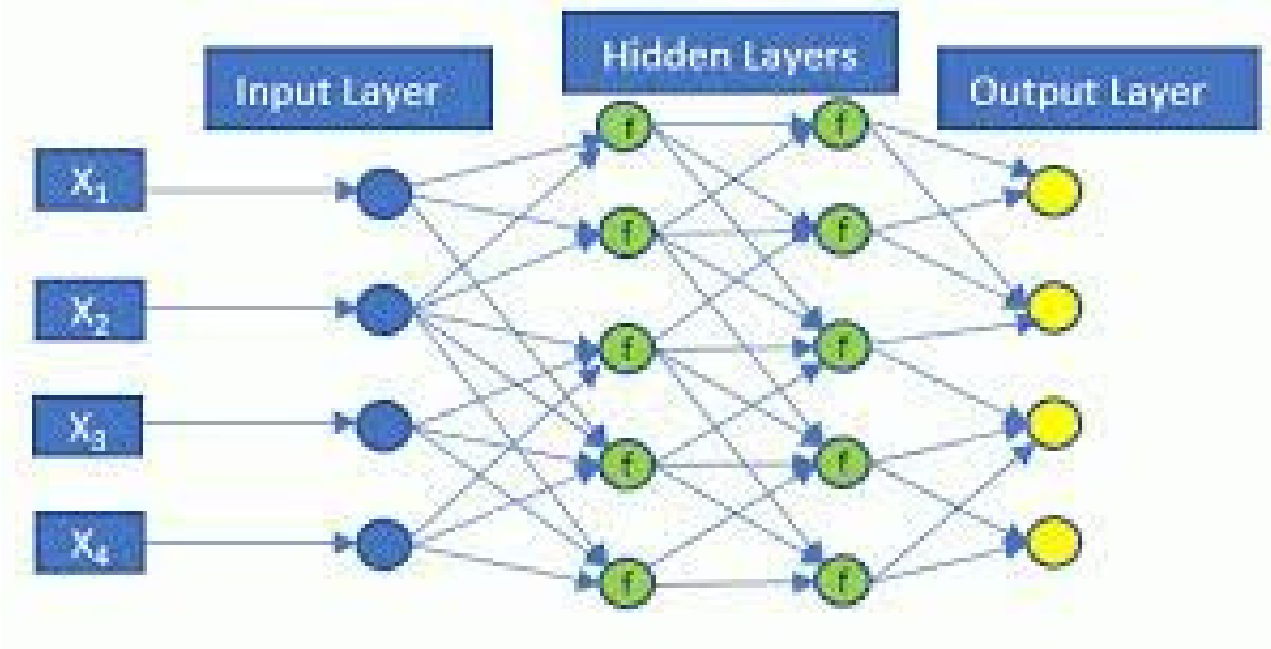
It is a supervised machine learning algorithm that divides the input points into groups based on the groups of k-nearest neighbors. Based on the given data set, we choose an appropriate value of k. When an input point is passed, k nearest points are found based on Euclidean distance, and the input point is placed in the group in which most k neighbors are present.



I have implemented it on a small data set of the original set to check if the model is working correctly. Then later, I applied it to a more extensive set of values, but it took too much time to run, so it crashed.

5. Neural Network

It is a class of algorithms that takes inspiration from the human brain's neural structure. The main aim is to develop an algorithm that takes input in various nodes of a layer and then produces an output after processing through several layers. The weights and biases are optimized after training on different data sets. The sigmoid function is also used to produce the result as a probability.



This algorithm took most of the time. I have taken the initialization of the parameter same as by Andrew Ng. I have taken two hidden layers to get a more accurate model.

Tech stack used:

- Google Colaboratory
- Pandas
- Numpy
- Matplotlib

About Me:

Name:- Vaibhav Jain
Place:- Sriganganagar , Rajasthan
Email:- 21je1014@iitism.ac.in
Major:- Integrated Master of Technology in Mathematics and Computing.
University:- Indian Institute of Technology (Indian School of Mines), Dhanbad
Year:- First Year
LinkedIn:- https://www.linkedin.com/in/vaibhav-jain-5b6231228
Github:- https://github.com/VaibhavJain05