Machine Learning Bootcamp

Project Report

Winter of Code 3.0

Aakash Rana

IIT (ISM) Dhanbad

Description-

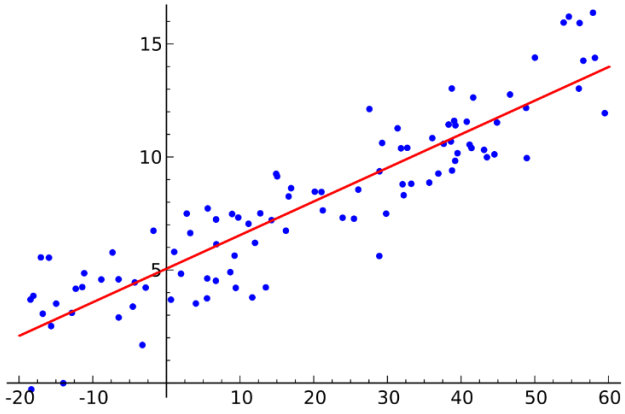
Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn without being explicitly programmed. In the span of one month, I have an implemented four algorithms from scratch using only basic python libraries like NumPy,

Pandas, matplotlib, and scikit-learn. I have also compared my algorithm's result with those of scikit's model.

Project Implementation-

1.Linear Regression

Linear regression is an algorithm through which we can relate two variables with the help of a line which represents the given data set in the most accurate way. The aim is to minimize the cost function by training on the data sets so as to get accurate prediction.

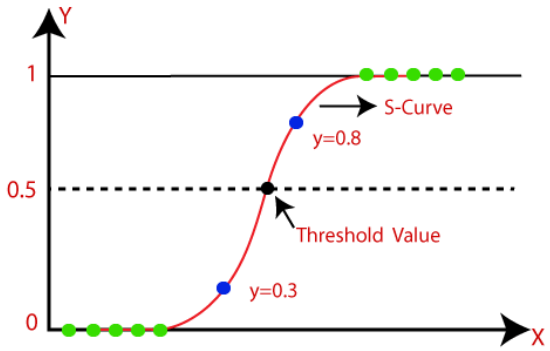


An implementation of linear regression.

In the implementation of this algorithm, my model has attained an accuracy of around 24% while the scikit's model gives an accuracy of around 21%.

2.Logistic Regression

Linear regression fails in the case of categorical variable, 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 here to make predictions accurate.



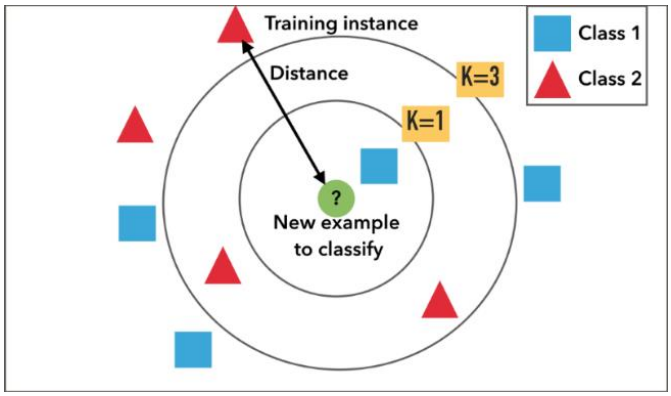
Logistic or Sigmoid Function

On implementing the algorithm through scikit learn, it ran out of iterations and was not able to predict even after scaling the features.

The model which I have implemented shows an accuracy of 89%.

3.KNN (K-Nearest Neighbors)

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

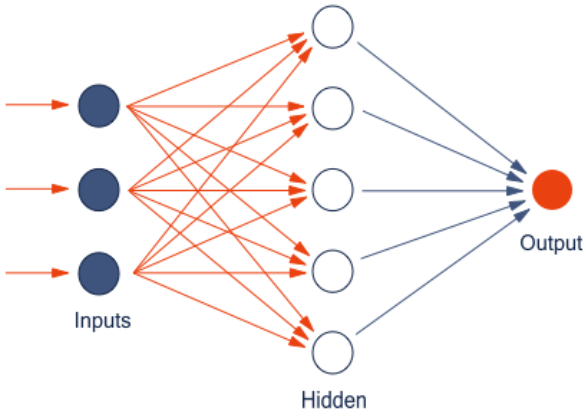


Representation of neighboring points of respective groups(class)

The scikit's model has an accuracy of 95.4% on the test set while my model showed an accuracy of 95.5% on the same test set

4.Neural Networks

It is a class of algorithms that takes inspiration from neural structure of human brain. 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 a number of layers. The weights and bias are optimized after training on different data sets. Sigmoid function is also used to produce output as probability.



A Simple Neural Network

I

Tech Stack-

- Python
- Libraries- NumPy, Pandas, Scikit-Learn and Matplotlib
- Jupyter
- Google Colab

About Me-

Name:- Aakash Rana

Place:- Palwal, Haryana

Email:- 20je0006@ece.iitism.ac.in, aakashrana000q@gmail.com

GitHub:- https://github.com/AakashRana000q

LinkedIn:- linkedin.com/in/aakash-rana-a93531200

University:- Indian Institute of Technology (Indian School of Mines), Dhanbad

Major:- Electronics and Communication Engineering

Year of study:- First Year