Machine Learning Bootcamp

Project Report

Winter of Code 4.0

Shobhit Gupta

IIT (ISM) DHANBAD

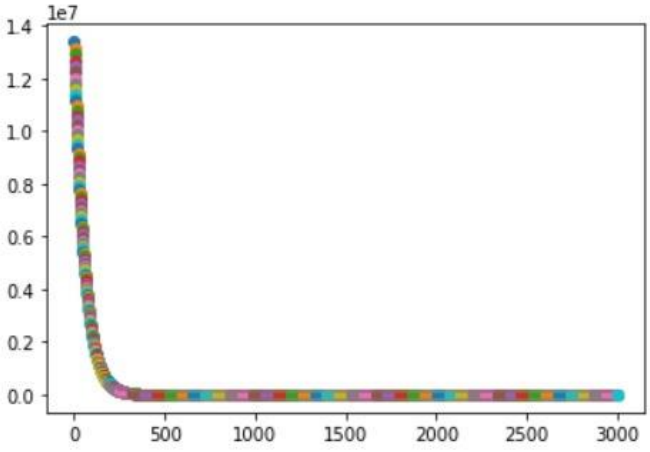
Description:

This is my project report for the WOC 4.0 for Machine Learning bootcamp. I implemented various machine learning algorithms such as linear regression, polynomial regression, logistic regression, KNN, K-Means and Neural Networks. I used the datasets which were provided to me by my mentors who helped me throughout this time. I have used various python libraries such as numpy, pandas and matplotlib. I spent nearly 30 days on this project and here is the result of all the implementations I did.

1.Linear Regression

I spent nearly 3 days on this algorithm and used the dataset given to me by my mentors. There were some errors initially but thanks to my mentors who helped me to figure out those errors and then my algorithm worked well. The number of iterations and the learning rate I used were 3000 and 0.009 respectively.

The graph of cost vs iterations is shown below:



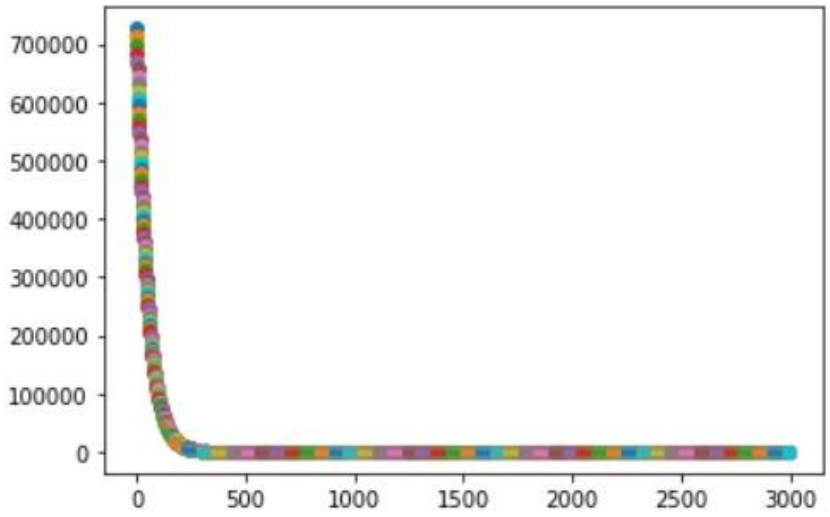
The final value of the cost function converged to 1907.0986.

The RMS error obtained was1.29307.

2.Polynomial Regression

It took me about 1 day to implement this algorithm. I used the polynomial test and train dataset given to me. The number of iterations and the learning rate I used were 3000 and 0.01 respectively.

The graph of cost vs iterations is shown below:



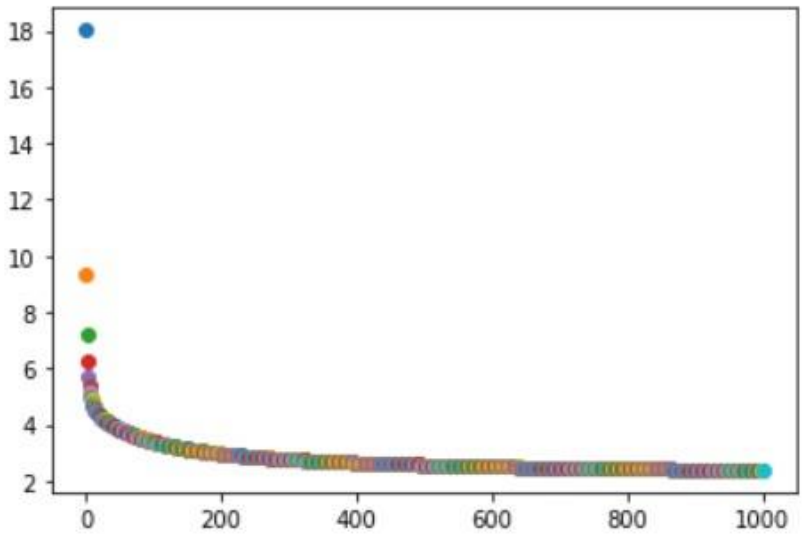
The final value of the cost function converged to 124.619.

The RMS error obtained was 9.904.

3.Logistic Regression

It took me about a week to implement this algorithm. I used the emnist dataset given to me. It was fun doing this one because I learnt many new things while implementing this algorithm. The number of iterations and the learning rate I used were 200 and 0.05 respectively.

The graph of cost vs iterations is shown below:



The final value of the cost function converged to 2.418.

The accuracy is 63.83%.

4.KNN

K-Nearest neighbours or KNN is a supervised machine learning algorithm which can solve both classification and regression problem statements. It took me about 3 days to implement this algorithm. I used the same emnist dataset used while implementing logistic regression algorithm. I repeated the process of making centroids and grouping 100 times.

The accuracy is about 84.28%.

5.K-Means Clustering

This method is used for grouping various points into ‘K’ different clusters. This took me about 3 days to finish it and finally I grouped the various data points of the emnist dataset into 26 clusters which are the 26 alphabets of the English vocabulary.

6.Neural Networks

This class of algorithm takes inspiration from the neural structure of the human brain. This was the toughest algorithm and it took me about a week to implement this algorithm on the same emnist dataset. The number of iterations and the learning rate I used were 200 and 0.05 respectively.

The value of cost obtained is: 2.403

The accuracy obtained when tested on Y_test is: 61.24%

Tech stack used:

- Jupyter Notebook
- Pandas
- Numpy
- Matplotlib

About Me:

Name: Shobhit Gupta

Place: Jhansi, Uttar Pradesh

Contact Number : 8922910350

Email: shobhitnote10@gmail.com , 21je0890@iitism.ac.in

Branch: Integrated Master of technology in Mathematics and Computing

College: Indian Institute of Technology(Indian School of Mines) Dhanbad

Year: First Year

LinkedIn: https://www.linkedin.com/in/shobhit-gupta-4281a8231

Github: https://github.com/shobhit9sept