

Artificial Intelligence and Expert System Lab (CSE 404)

Department of CSE

Project: 02

|  |  |
| --- | --- |
| Topic/Question: | Multivariable Liner Regression Using Open Source Dataset (data\_monthly\_rainfall). |

Date of Submission: 22march, 2021

|  |  |
| --- | --- |
| Submitted by | Submitted to |
| Name: Niamul Hasan  Id: **17201026**  Semester: 4.1  Section: A1 | Dr. Nasima Begum  Assistant Professor  Department of CSE  UAP |

|  |  |
| --- | --- |
| **Tools:** |  |
| 1. | Language: Python 3 (for coding) |
| 2. | IDE: jupyter notebook (text editor) |
| 3. | Google Colab |

My dataset title is: Rainfall data of Bangladesh(1970-2016)

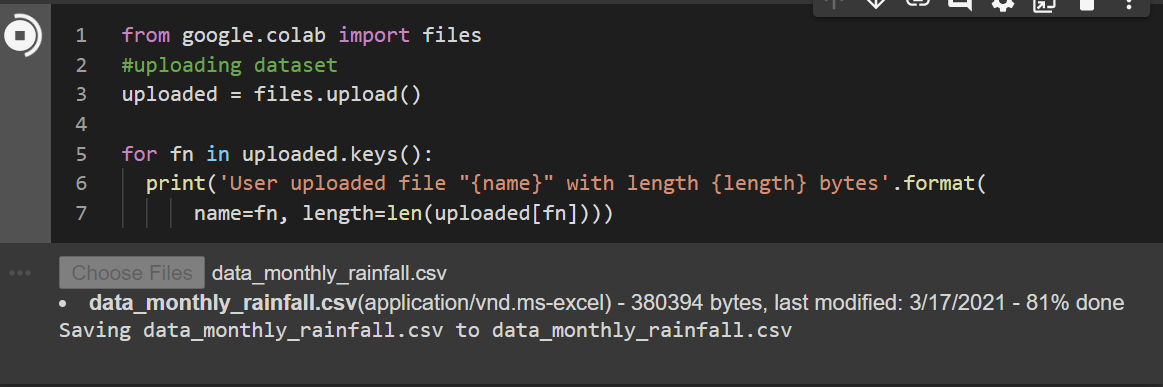
The dataset link: <https://www.kaggle.com/redikod/historical-rainfall-data-in-bangladesh>

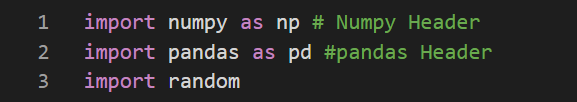
My task is to implement the Multivariable Liner Regression Using Open Source Dataset without SK-Learn.

Here the model trained by raw coding in python 3.  
I haven’t used any library function to train my model.  
But only to normalize the dataset, I used a normalized library function from Sk-learn.

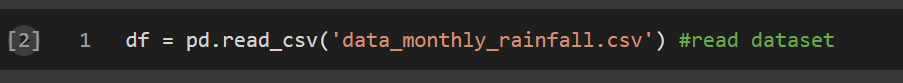
**Now, Coding: (using** Google Colab**)**

the first block of code is to upload the dataset.

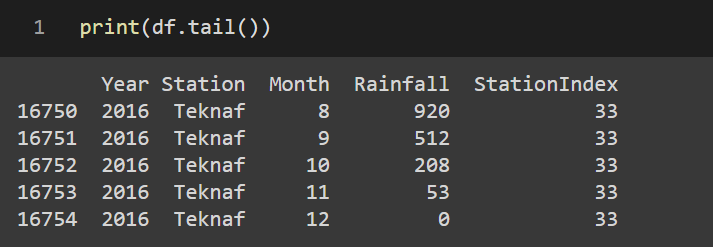
Importin some library function.



read the dataset as pandas object.

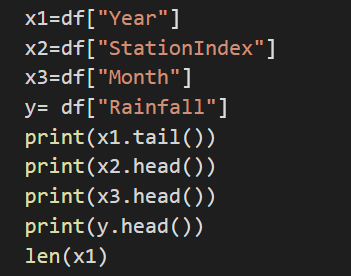


checking the data.



Here ‘m’ is the length of dataset.

Initializing the features Xs and label data Y.



Preprocessing of dataset/ Normalization of dataset.

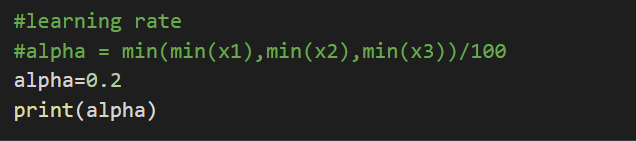
Normalizing the dataset between -1 and 1

Code Block 9.

Setting the value of learning rate (alpha)

Here, as we know that the dataset is huge. To make it time efficient we are using the value of alpha 0.2.

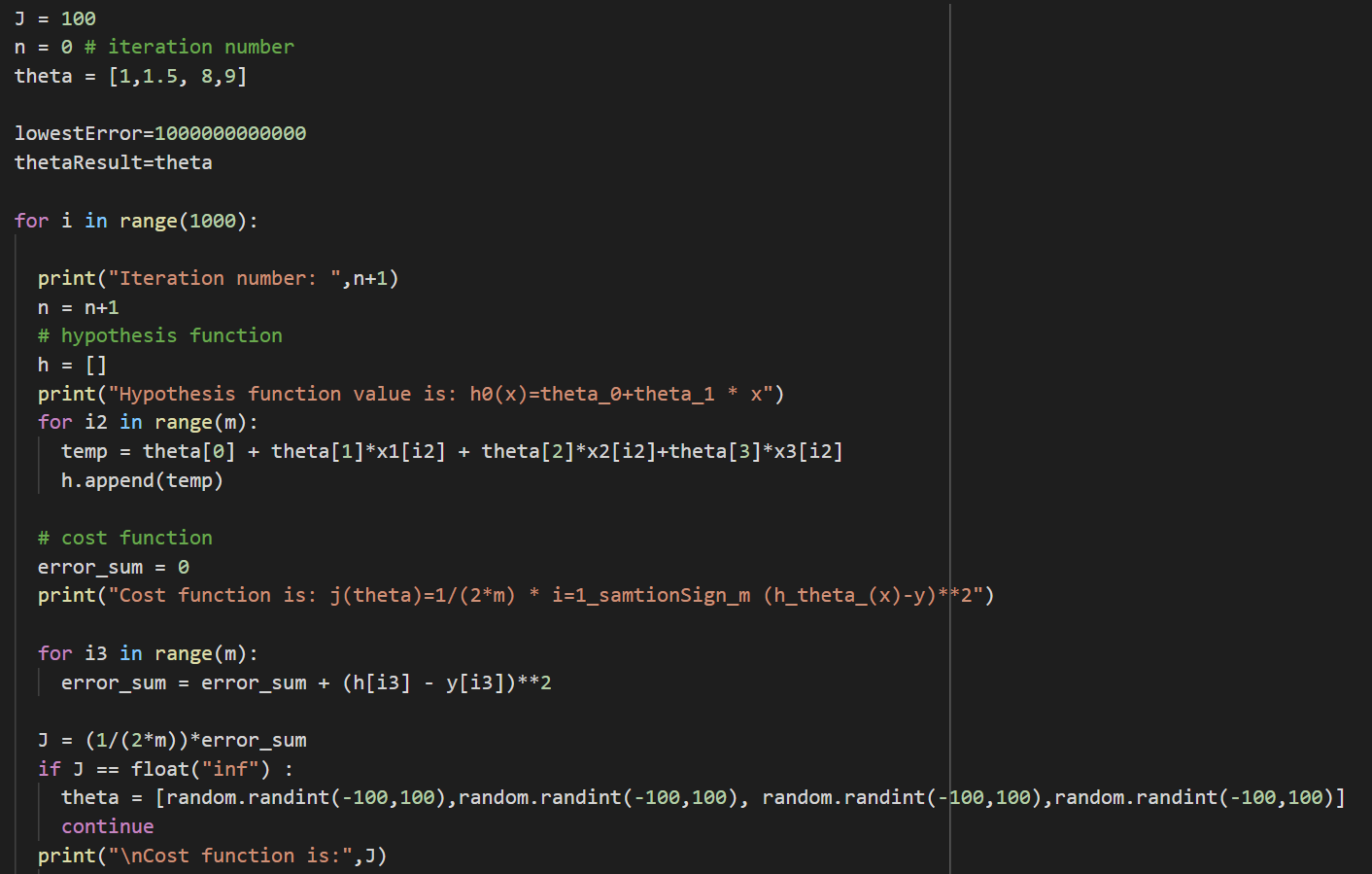
But the perfect way to do this: take the lowest value of all features and divide it by 100 then take the value. But by this the optimization process will take more than day in my computer.  
So, I had to make it 0.2.

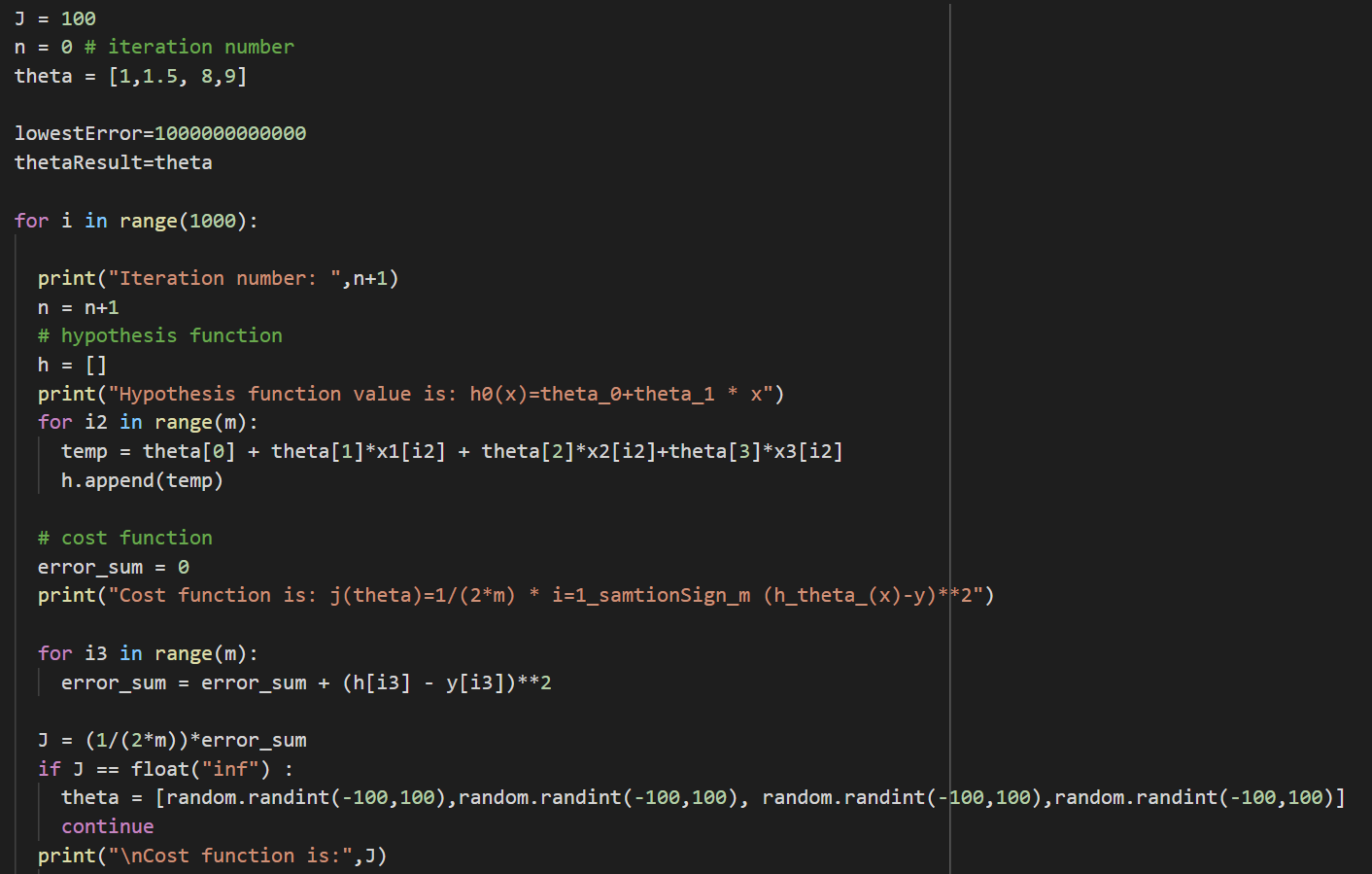


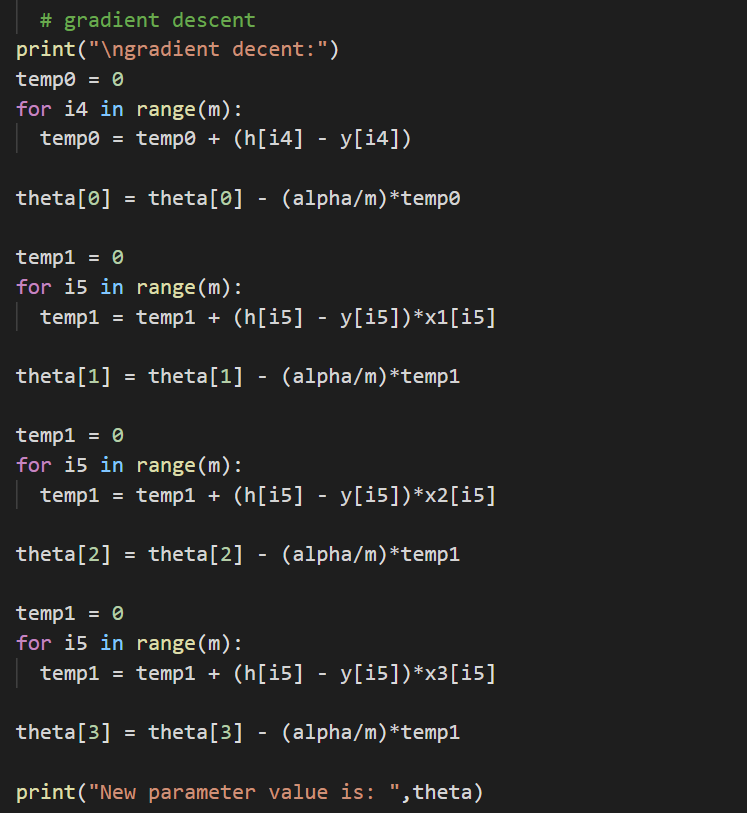
Now the Code block 10 is the optimization process.  
where we are initializing the values of coefficients (theta)

Then following the step:

1. Hypothesis function (to calculate hypothetical values)
2. Calculating cost
3. Gradient decent (changing the value of coefficients)

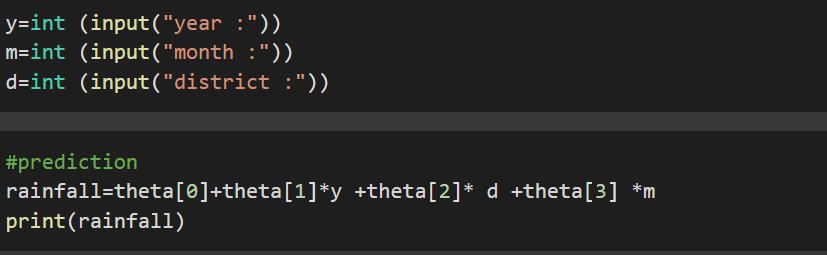








Now after optimization we can predict the result of given inputs:



The model works fine.