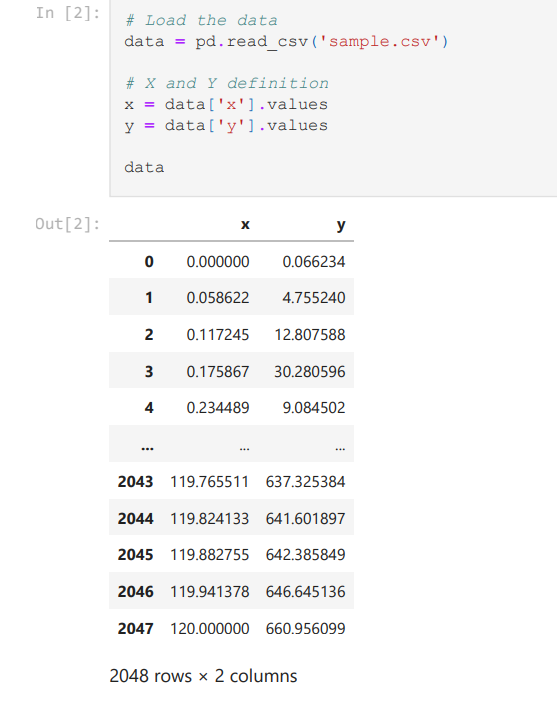
**Machine Learning Assignment 1:**

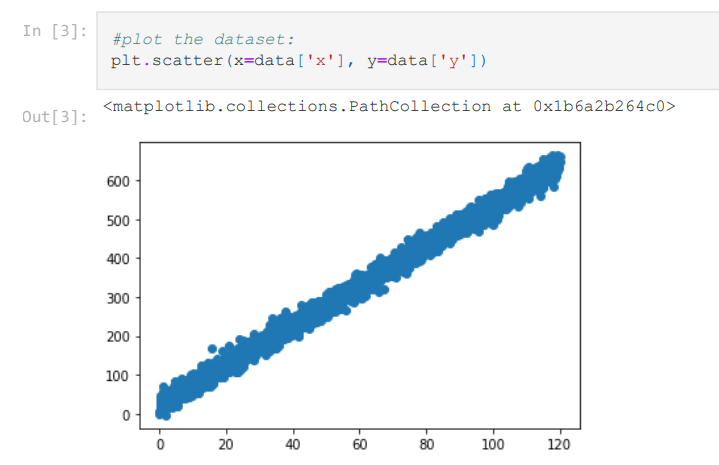
**Einav Diar 319010807**

**Introduction:**

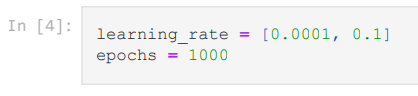
First, I upload the data and created two variables x and y.



I showed a plot of the data.



As you can see, the plot shows a linear relationship between x and y variables.



I defined the learning rate and the epochs.

Afterwards, I tried to implement the gradient descent, the stochastic gradient descent, and the mini-batch gradient descent algorithms.

**Gradient descent:**

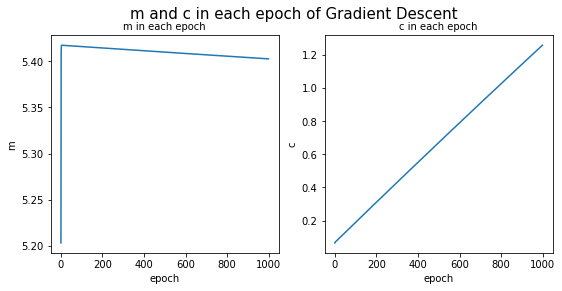
The gradient descent is an optimization algorithm that tries to find the optimal values of parameters a and b (y=ax+b) and to minimize the loss function.

The plots of a/b in each epoch of gradient descent will display the changes in the values over iterations.

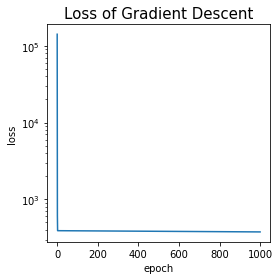
Ideally, the plot should show a smooth convergence of the values of a/b towards the optimal value, and eventually stabilize at the optimal value.

The loss function values should generally decrease, indicating that the algorithm is moving closer to the optimal parameters.

**Learning rate = 0.0001:**



At the beginning of the optimization process, the m (a) and c (b) values start from 0, and the plot start to show a trend towards the optimal value (m=5.4, c= 1.256).



The loss plot shows a high initial value of 141,350, and after several iterations the model reached a low number of 390.

As the algorithm updates the values of a and b (that are based on the gradient of the loss function), the plot shows a trend of decreasing loss function values.

The loss final value is 376.77.

**Learning rate = 0.1:**

Shape, square

Description automatically generated

Shape, square

Description automatically generated

The plot of the loss function is increasing during the iterations, it indicates that the optimization process is not progressing as excepted.

The data contain inf (infinity) values, and therefore, the final values we received is Nan.

I assume that the reason that the plot is increasing is because the learning rate is too high.

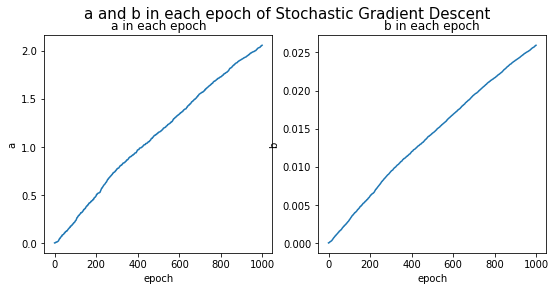
**Stochastic Gradient Descent:**

The plots for SGD may show more noise compared to regular gradient descent, due to the more frequent updates after each data point.

Since SGD updates the parameters after each individual data point, the plot of a/b may show more noise compared to regular gradient descent.

the plot of loss may also exhibit more fluctuations compared to regular gradient descent, as the updates are more frequent and can result in rapid changes in the model's predictions.

**Learning rate = 0.0001:**



At the beginning of the optimization process, the a and b values start from 0, and the plot start to show a trend towards the optimal value (a=2.06, b= 0.02).

Chart, histogram

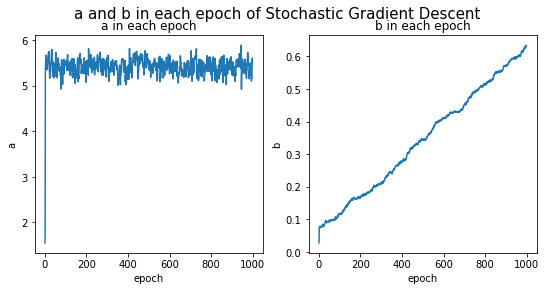
Description automatically generated

The loss function plot shows a significate decrease, with a lot of noise.

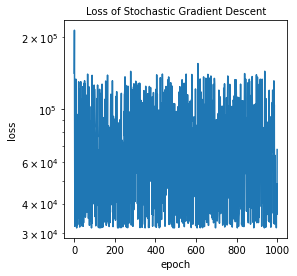
At the beginning of the optimization process, the loss value is very high, and as the algorithm updates the values of a and b, the plot shows a trend of decreasing loss function values.

The loss value in the last iteration is still very high (loss= 139,262), but we can see a decreasing trend in the loss function values, indicating that the optimization process is improving the model's performance.

**Learning rate = 0.1:**



At the beginning of the optimization process, the a and b values start from 0, and the plot start to show a trend towards the optimal value (a=4.925, b= 0.6).



The loss final value is 114,961, the plot doesn't show a trend of decreasing in the loss values and instability can be seen.

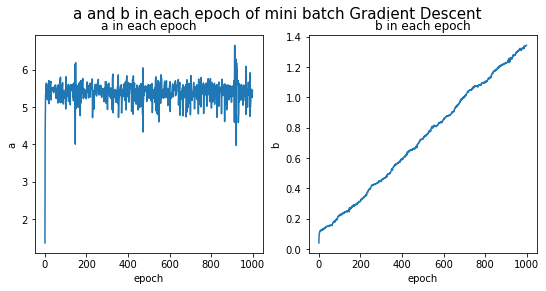
**Mini-Batch Gradient Descent:**

The mini batch gradient descent plots may show some noise, as the updates are based on a random group of data at each epoch.

over time, the plots of a/b are expected to show a trend towards the optimal value, that minimizes the loss function.

The loss function plot may also exhibit some noise, but over time, the plot is expected to show a trend of decreasing loss function values, indicating that the optimization process is progressing towards finding the optimal parameters.

**Learning rate = 0.0001:**



At the beginning of the optimization process, the a and b values start from 0, and the plot start to show a trend towards the optimal value (a=5.42, b= 1.34).

Chart

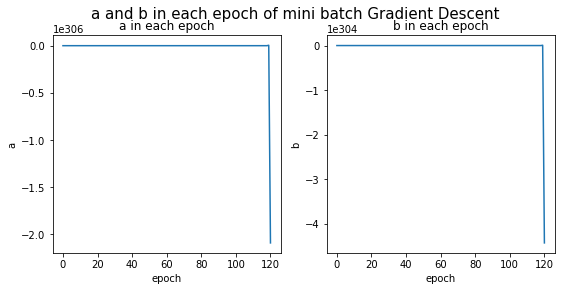
Description automatically generated

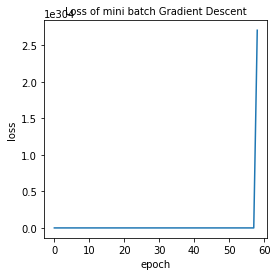
The loss plot shows a high initial value of 473, and after several iterations the model reached a low number of 0.56.

The loss function plot shows a decreased trend, and the final value of the loss is 1.94.

The low value indicates that the values we found of a and b fit the data and produce the smallest possible loss value.

**Learning rate = 0. 1:**





The plot of the loss function is increasing during the iterations, it indicates that the optimization process is not progressing as excepted.

The data contain inf (infinity) values, and therefore, the final values we received is Nan.

I assume that the reason that the plot is increasing is because the learning rate is too high.