Assignment 1: Difference between Gradient Descent, Stochastic Gradient Descent, and Mini-Batch Gradient Descent

Due date: 15/4/23

In this assignment, we will explore the differences between Gradient Descent, Stochastic Gradient Descent (SGD), and Mini-Batch Gradient Descent (MBGD). We will use a dataset in the form of a CSV file and evaluate the loss and value of each parameter 'a' and 'b' in the equation y = ax + b for each epoch, both in text and plots.

Gradient Descent is an optimization algorithm used to minimize the cost function of a model by adjusting its parameters iteratively. The cost function represents the difference between the predicted and actual values of the target variable. The algorithm starts with an initial guess for the values of the parameters and then iteratively updates them in the direction of the negative gradient of the cost function until it reaches the minimum point.

Stochastic Gradient Descent (SGD) is a variation of Gradient Descent that computes the gradient and updates the parameters for each training example. This process is repeated for all the training examples. SGD is much faster than Gradient Descent as it updates the parameters more frequently, but it may not converge to the global minimum due to the noise in the updates.

Mini-Batch Gradient Descent (MBGD) is a variation of Gradient Descent that updates the parameters using a subset of the training examples at each iteration. The size of the subset is called the batch size. MBGD combines the advantages of SGD and Gradient Descent by updating the parameters more frequently than Gradient Descent, but less frequently than SGD.

We will use the CSV file provided to evaluate the parameters 'a' and 'b' using Gradient Descent, SGD and MBGD. For each epoch, we will calculate the loss and the value of the parameters 'a' and 'b'. We will then plot the loss and the values of 'a' and 'b' for each epoch.

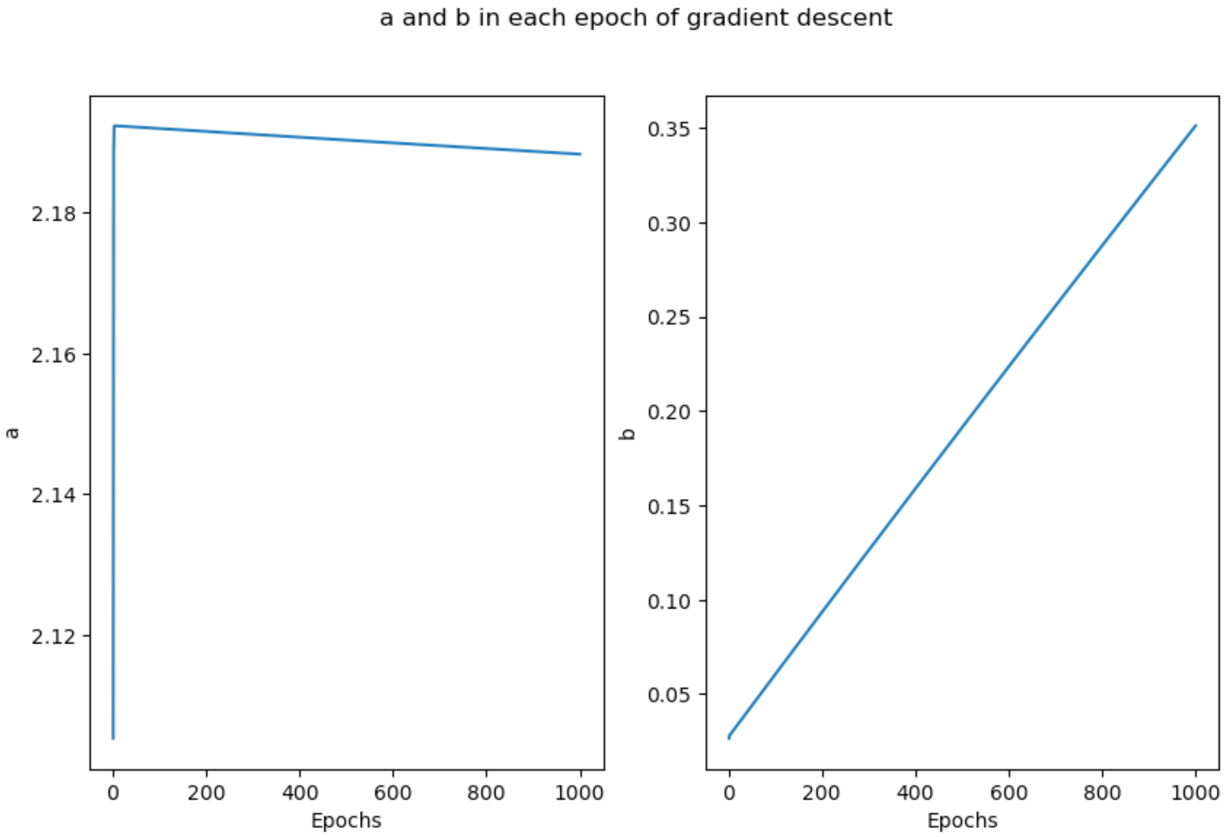
To complete this assignment, please follow these steps:

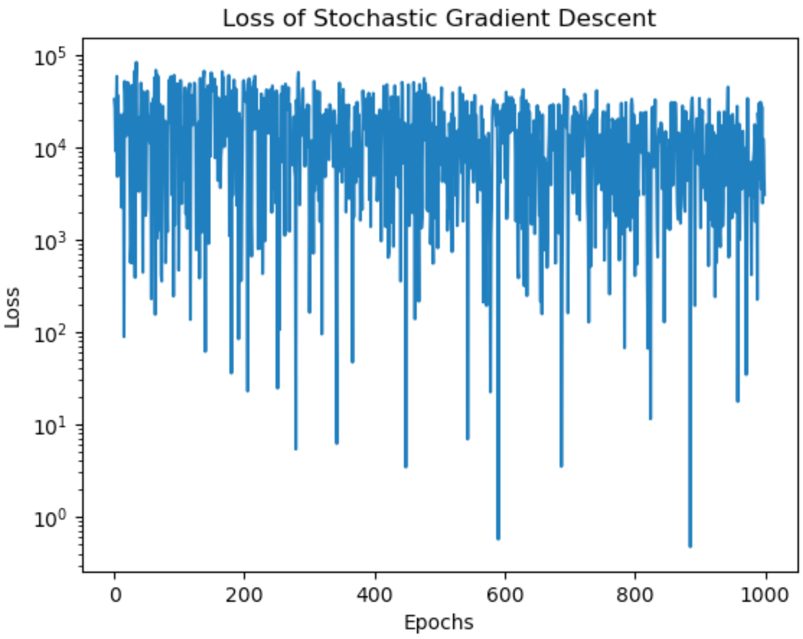
- Load the CSV file into your program.
- Initialize the values of 'a' and 'b' to zero.
- Choose the learning rate, batch size, and the number of epochs (1,000).
- Implement the Gradient Descent, SGD, and MBGD algorithms.
- For each epoch, calculate the loss and the value of the parameters 'a' and 'b' using each algorithm.
- Plot the loss and the values of 'a' and 'b' for each epoch using a suitable library such as Matplotlib.

Complete the assignment for the following learning rates: 0.0001 and 0.1.

Please submit your code and plots, along with a brief explanation of your findings. Good luck!

Your plots should look similar to those ones (6 plots of loss and 6 plots of the values):





I hope you found the assignment on Gradient Descent and its variants challenging and informative. As you may have noticed, one of the key aspects of building effective machine learning models is selecting the optimal hyperparameters for each algorithm, such as the learning rate, batch size, and the number of epochs.

Learning is a continuous process, and the more you practice, the better you will become. To further enhance your skills and build a strong portfolio for future work interviews, I encourage you to continue this assignment on your own and experiment with different hyperparameters. This will not only improve your understanding of the algorithms but also demonstrate your ability to optimize them, which is a valuable skill in the industry. You can then upload your code to your GitHub profile as part of your portfolio, which will showcase your skills and expertise to potential employers.