Experiment 1

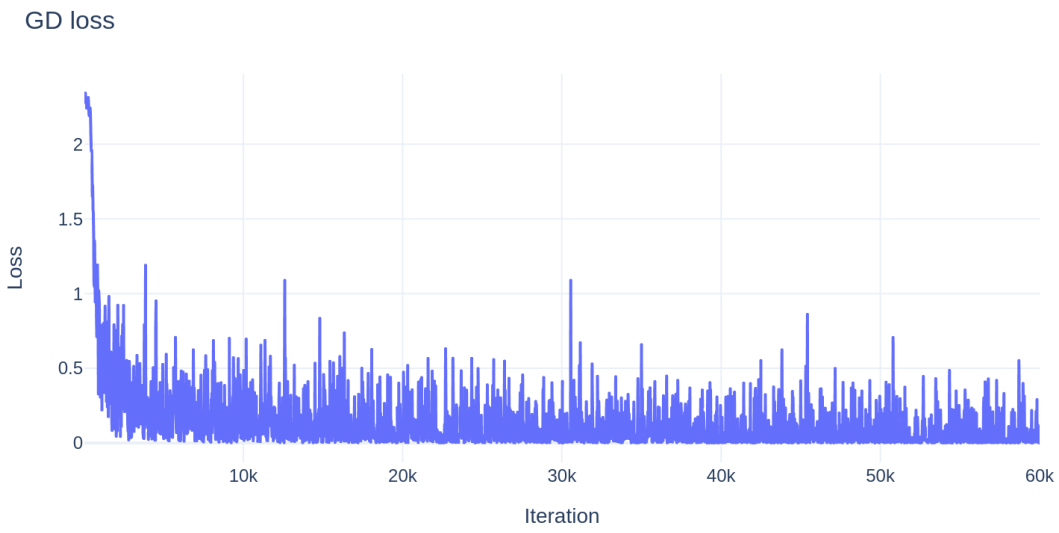
Overview

In this task, I have employed three distinct optimizers: the regular Gradient Descent, the Stochastic Gradient Descent, and the ADAM Optimizer.

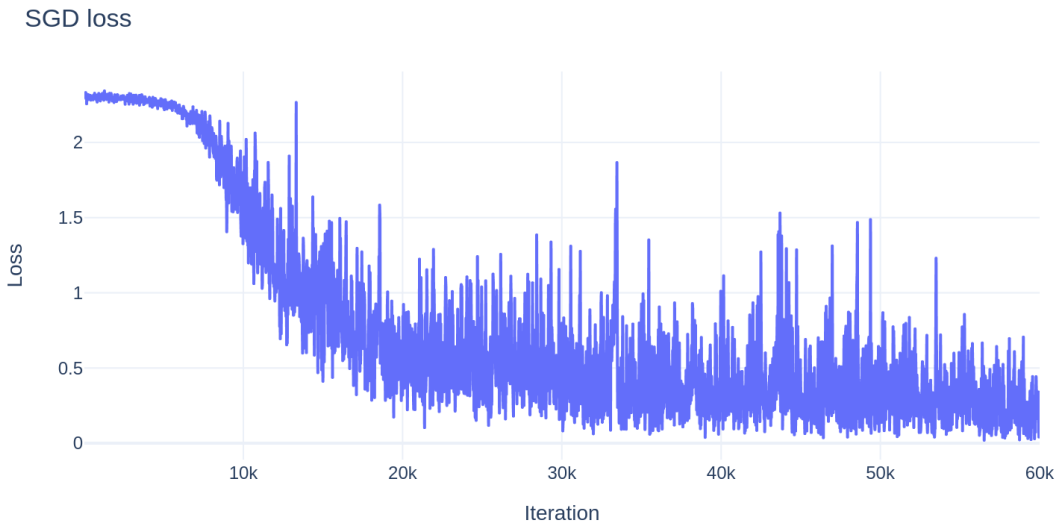
Subsequently, I have preserved and graphed the values of loss, carrying out a analysis of the results.

Results

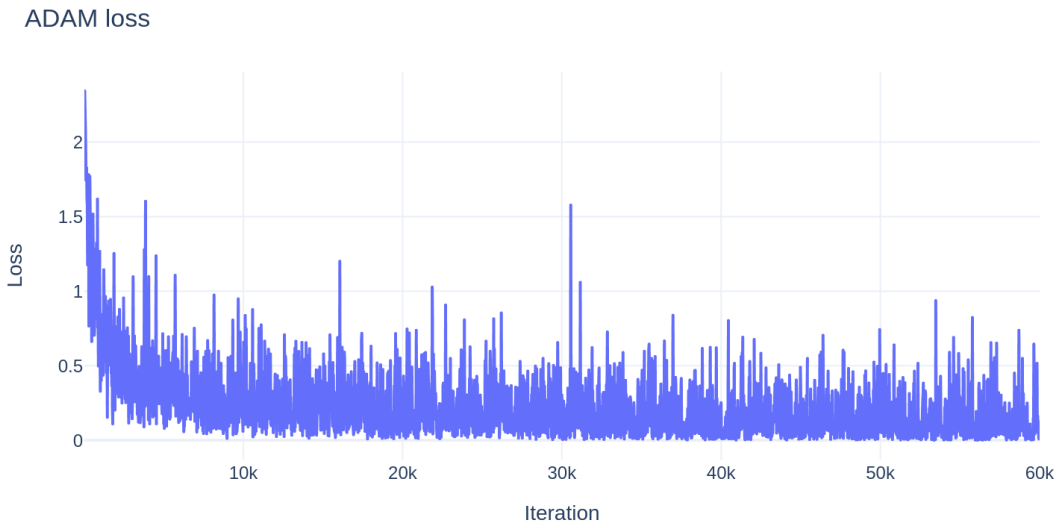
Gradient Descent: Accuracy = 98.44% (lr=0.01)



SGD: Accuracy = 91.36% (lr=0.01)



ADAM: Accuracy = 96.5% (lr=0.01, beta1=0.8, beta2=0.999)



Analysis

It is clear that regular GD takes less iterations to converge, but it is known that each iteration is more computationally difficult. ADAM optimizer convergence depends on values of $beta_1$ and $beta_2$. For high values of $beta_1 (beta_1 = 0.9)$ it is not always converge. For the plot I have used $beta_1 = 0.8$. ADAM convergence speed seems similar to regular GD. All optimizers have similar accuracy value, while GD have the highest and SGD the lowest, but depends on hyperparameters and chosen seed.