Comparative Analysis: Stochastic Gradient Descend and Adam Optimizer for Neural Networks Training

Nattapoom Dumronglaohapun

April 2023

1 Introduction

In recent years, neural networks have achieved remarkable success in various machine learning tasks such as image classification, natural language processing, and speech recognition. However, training neural networks can be computationally expensive, especially for large datasets and complex architectures. To address this issue, many optimization algorithms have been proposed, including stochastic gradient descent (SGD) and Adam optimizer. While these algorithms have shown promising results, their performance can vary significantly depending on the dataset, network architecture, and hyperparameters.

Therefore, the aim of this report aims to compare and contrast the performance of SGD and Adam optimizer for neural network training across different types of problems such as binary classification, image multi-class classification, and natural language processing. The project will explore how the performance of these algorithms varies with different network architectures and hyperparameters, such as learning rate or batch size.

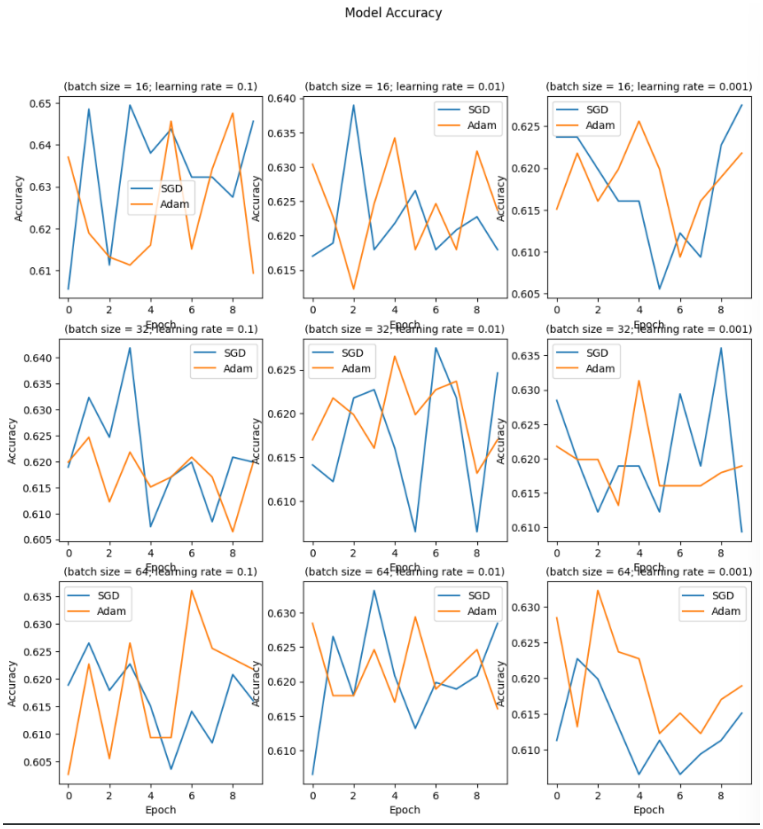
2 Methodology

The objective of this research project is to evaluate the performance of stochastic gradient descent (SGD) and Adam optimizer in training neural networks for various machine learning tasks. To achieve this, we will collect and preprocess three datasets from the TensorFlow Datasets module: Titanic, MNIST, and IMDB Reviews.

The datasets will be used to train and test neural networks with different architectures and hyperparameters, such as learning rate and batch size, to evaluate the performance of SGD and Adam in terms of convergence speed and final accuracy. The experiments will be conducted on different types of problems, including binary classification, multi-class image classification, and sentimental analysis.

We will compare the performance of both optimization algorithms across all three datasets and different hyperparameters. To present the results, we will use graphs and tables to show the performance of SGD and Adam at learning rates of 0.1, 0.01, and 0.001, and at batch sizes of 16, 32, and 64. Finally, we will analyze and interpret the results to provide insights into the strengths and weaknesses of each optimization algorithm and their suitability for different types of problems.

3 Results



4 Conclusion

The results of the study indicate that for the binary classification problem using a feed-forward neural network, SGD outperforms Adam for learning rates of 0.1 and 0.01, while Adam performs better for a learning rate of 0.001. However, no clear trend was observed in terms of batch size.

References

[1] Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS) (pp. 1-2). Ieee.

[2] Desai, C. (2020). Comparative analysis of optimizers in deep neural networks. International Journal of Innovative Science and Research Technology, 5(10), 959-962.

[3] Vani, S., Rao, T. M. (2019, April). An experimental approach towards the performance assessment of various optimizers on convolutional neural network. In 2019 3rd international conference on trends in electronics and informatics (ICOEI) (pp. 331-336). IEEE.