Vietnam General Confederation of Labor
TON DUC THONG UNIVERSITY
FACULTY OF INFORMATION TECHNOLOGY

FINAL PROJECT

MACHINE LEARNING

REPORT OF FINAL PROJECT

Instructor : **Dr.LÊ ANH CƯỜNG**

Executor : **ĐẶNG NHẬT KHANG – 520H0371**

Grade **: 20H50204**

Course **: 24**

HO CHI MINH CITY, 2023

Vietnam General Confederation of Labor
TON DUC THONG UNIVERSITY
FACULTY OF INFORMATION TECHNOLOGY

FINAL PROJECT

MACHINE LEARNING

REPORT OF FINAL PROJECT

Instructor : **Dr.LÊ ANH CƯỜNG**
Executor : **PHAN NGỌC HOÀNG ANH - 520H0511**
Grade **: 20H50204**
Course **: 24**

HO CHI MINH CITY, 2023

ACKNOWLEDGEMENT

We are really grateful for the instruction of Dr.Lê Anh Cường because you helped and explained the problem of the final project of Introduction of Machine Learning to us to complete perfectly as soon as possible.

PROJECT COMPLETED
AT TON DUC THONG UNIVERSITY

I hereby declare that this is our own project and is under the guidance of Dr. Lê Anh Cường. The research contents and results in this topic are honest and have not been published in any form before. The data in the tables for analysis, comments and evaluation are collected by the author himself from different sources, clearly stated in the reference section.

In addition, the project also uses a number of comments, assessments as well as data from other authors, other agencies and organizations, with citations and source annotations.

If I find any fraud, I will take full responsibility for the content of my project. Ton Duc Thang University is not related to copyright and copyright violations caused by me during the implementation process (if any).

City. Ho Chi Minh, April 23 2023

Author

(sign and write full name)

Anh

Phan Ngọc Hoàng Anh

Khang

Đặng Nhật Khang

TEACHER'S CONFIRMATION AND ASSESSMENT SECTION

The confirmation part of the instructor

_____

_____

_____

_____

_____

_____

_____

_____

City. Ho Chi Minh, May Day
(signature and full name)

The evaluation part of the teacher marks the test

_____

_____

_____

_____

_____

_____

_____

_____

City. Ho Chi Minh, May Day
(signature and full name)

SUMMARY

Chapter 1 Will illustrate Task 1 of the final project.

TABLE OF CONTENTS

CHAPTER 1: THEORY

1.1 Investigate and compare different Optimizer methods in machine learning model training.

1. **Stochastic Gradient Descent (SGD)** excels in handling large datasets due to its use of random data batches, reducing computational costs compared to full-batch methods.

Advantages: Efficient for large datasets, less memory usage.

Disadvantages: Noisy path, slower convergence, sensitive to feature scaling.

2. **SGD with Momentum** improves upon standard SGD by accelerating convergence, though it requires careful adjustment of momentum and learning rate to avoid overshooting the optimal solution.

Advantages: Faster convergence than SGD, less oscillation.

Disadvantages: Risk of overshooting the minimum, hyperparameter tuning required.

3. **Mini Batch Gradient Descent** strikes a balance between computational efficiency and accuracy by working with data subsets, necessitating fine-tuning of the batch size.

Advantages: Faster than full-batch methods, better for large datasets.

Disadvantages: Hyperparameter (batch size) tuning needed, potential for suboptimal accuracy.

4. **Adagrad** adapts the learning rate for each parameter, benefiting from sparse data features but potentially decreasing the learning rate too quickly.

Advantages: Adaptive learning rates, good for sparse data.

Disadvantages: Learning rate might decrease too quickly, not ideal for non-convex optimization.

5. **RMSProp** addresses the rapid decrease in learning rate seen in Adagrad, yet it still requires manual tuning of the learning rate.

Advantages: Overcomes learning rate diminishing issue in Adagrad, good for non-stationary objectives.

Disadvantages: Requires manual setting of learning rate.

6. **AdaDelta** builds on RMSProp and Adagrad, using state variables to provide a more stable and robust learning rate adjustment.

Advantages: Addresses decaying learning rate, robust to hyperparameter choices.

Disadvantages: Potentially complex implementation, performance may vary across tasks.

7. **Adam Optimizer** combines advantages of both AdaGrad and RMSProp, adjusting learning rates based on the gradients' history and their second moments, enhancing efficiency in many scenarios.

Advantages: Efficient for large datasets and high-dimensional spaces, adaptive learning rate.

Disadvantages: Requires tuning of learning rate and first/second moment decay rates.

REFERENCES.