VIETNAM GENERAL CONFEDERATION OF LABOR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL REPORT**

**MACHINE LEARNING**

*Instructor*: **LE VAN CUONG**

*Performer*: **DO HOANG DUY – 521H0395**

Class **: 21H50302**

Course  **: 25**

**HO CHI MINH CITY, YEAR 2023**

VIETNAM GENERAL CONFEDERATION OF LABOR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL REPORT**

**MACHINE LEARNING**

Instructor: **LE VAN CUONG**

Performer**: DO HOANG DUY****– 521H0395**

Class **: 21H50302**

Course  **: 25**

**HO CHI MINH CITY, YEAR 2023**

ACKNOWLEDGEMENTS

I'd like to take this time to thank my thesis advisor, Le Van Cuong, for his advice, support, and encouragement throughout this research effort. His insightful observations, critical input, and patience have been invaluable in advancing my grasp of Machine Learning.

I want to thank my family and friends for their continuous support and encouragement during our academic journey. Their love, support, and understanding have been the foundation of our success, and i am eternally grateful for having them in our life.

Finally, i would want to convey my gratitude to everyone who helped me finish this research project in some way. Your encouragement, support, and input have been invaluable, and i am grateful for your efforts.

**COMPLETED THESIS AT TON DUC THANG UNIVERSITY**

I hereby certify that this thesis is my own work and was guided by Le Van Cuong. The research contents and results in this topic are truthful and have not been published in any form before. The data in the tables used for analysis, comments, and evaluation were collected by the author from various sources, which are clearly stated in the reference section.

In addition, this thesis also uses some comments, evaluations, and data from other authors and organizations, all of which are cited and annotated with their original sources.

**If any academic dishonesty is discovered, I take full responsibility for the content of my thesis**. Ton Duc Thang University is not liable for any copyright infringement or violations caused by me during the process of completing this thesis (if any)..

*Ho Chi Minh City, December 20th, 2023*

*Author*

**

*Do Hoang Duy*

CONFIRMATION AND EVALUATION SECTION OF THE INSTRUCTOR

**Confirmation section of the instructor**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ho Chi Minh City, month day year

(Signature and printed name)

**Evaluation section of the grading instructor**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ho Chi Minh City, month day year

(Signature and printed name)

SUMMARY

The research problem addressed in this thesis is to understand and compare Optimizer approaches for training machine learning models, understand Continuous Learning and Test Production when developing a machine learning solution to a specific problem.

TABLE OF CONTENTS

[ACKNOWLEDGEMENTS i](#_Toc154223432)

[CONFIRMATION AND EVALUATION SECTION OF THE INSTRUCTOR iii](#_Toc154223433)

[SUMMARY iv](#_Toc154223434)

[TABLE OF CONTENTS 1](#_Toc154223435)

[CHAPTER 1 – LEARN AND COMPARE OPTIMIZER METHODS 2](#_Toc154223436)

[1.1 Stochastic Gradient Descent (SGD) 2](#_Toc154223437)

[1.2 Mini-Bacth Gradient Descent 3](#_Toc154223438)

[1.3 Momentum 3](#_Toc154223439)

[1.4 Adagrad 4](#_Toc154223440)

[1.5 Adam 4](#_Toc154223441)

[CHAPTER 2 – CONTINUOUS LEARNING AND TEST PRODUCTION 5](#_Toc154223442)

[2.1 Continuous Learning 5](#_Toc154223443)

[2.1.1 Definition 5](#_Toc154223444)

[2.1.2 Challenges 5](#_Toc154223445)

[2.1.3 Strategies and Methods 6](#_Toc154223446)

[2.1.4 Continuous Management 6](#_Toc154223447)

[2.2 Test Production 6](#_Toc154223448)

[2.2.1 Definition 6](#_Toc154223449)

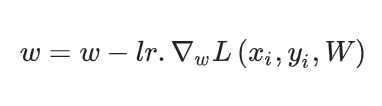
[2.2.2 Key Aspects 7](#_Toc154223450)

CHAPTER 1 – LEARN AND COMPARE OPTIMIZER METHODS

Optimization is a critical procedure in machine learning that involves adjusting the parameters of a model to minimize a loss function. To efficiently train machine learning models, many optimization techniques, often known as optimizers, have been created. Let's look at and contrast several popular optimizer methods:

* 1. Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) is a version of the Gradient Descent technique that is used to optimize machine learning models. It solves the computational inefficiencies of classic Gradient Descent algorithms when working with huge datasets in machine learning applications.



Advantages:

* SGD is faster than other Gradient Descent variations.
* Memory efficient and can handle large datasets
* Is capable of escaping local minima and converges to a global minimum

Disadvantages:

* SGD may require several iterations to reach the minimum
* Converge slowly
* May not reach the precise global minimum, resulting in a poor solution.

1.2 Mini-Bacth Gradient Descent

Mini-batch gradient descent is a gradient descent variation in which the training dataset is divided into small batches that are utilized to compute model error and update model coefficients.

Mini-batch gradient descent attempts a compromise between SGD and batch gradient descent

Advantages:

* The model update frequency is larger than in batch gradient descent
* Batched updates are a more efficient computational approach than SGD
* Is capable of escaping local minima and converges to a global minimum

Disadvantages:

* The learning algorithm must be configured with an additional "mini-batch size" hyperparameter.

1.3 Momentum

Momentum is a gradient descent optimization technique extension that creates inertia in a search direction to avoid local minima and noisy gradient oscillation.

Advantages:

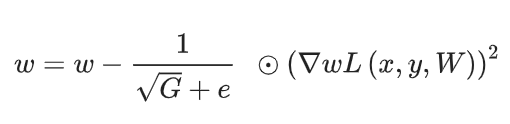
* May assist the optimization process in fast escaping from local minima and converge to the global minimum

Disadvantages:

* It may exceed the global minimum and instead converge to a local minimum.
* Requires tuning of momentum hyperparameter

1.4 Adagrad

Adagrad is an optimizer with parameter-specific learning rates that are adjusted based on how frequently a parameter is updated during training. The more updates a parameter receives, the smaller the updates.



Advantages:

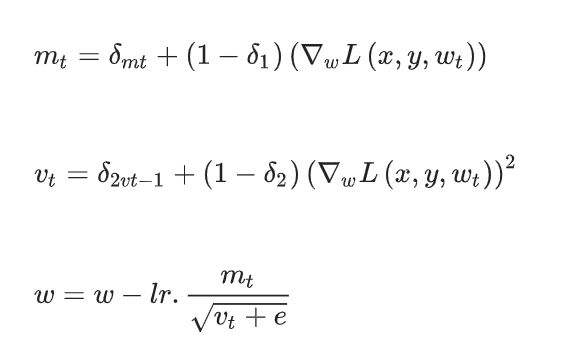
* Automatic learning rate adjustment

Disadvantages:

* May have a diminishing learning rate problem

1.5 Adam

Adam is a machine learning optimization algorithm that iteratively updates network weights depending on training data. Adam is a stochastic gradient descent extension that has lately gained traction in deep learning applications such as computer vision and natural language processing.



Advantages:

* Adaptive learning rates, effective on a wide range of problems.

Disadvantages:

* Sensitive to hyperparameter choices

When selecting an optimizer, it is critical to test several algorithms and hyperparameter settings on your specific dataset and model architecture. There is no one-size-fits-all option, and your optimizer selection might have a considerable impact on the training performance of your machine learning model.

CHAPTER 2 – CONTINUOUS LEARNING AND TEST PRODUCTION

2.1 Continuous Learning

2.1.1 Definition

In the context of machine learning, continuous learning refers to the continuing process of updating and enhancing a machine learning model over time. This is critical because the distribution of real-world data may vary, and models may become less successful if not modified to new trends.

2.1.2 Challenges

Continual Learning faces some significant challenges:

**Catastrophic Forgetting**: The model may forget old knowledge when learning new information, especially when the new data doesn't adequately represent the old data.

**Managing Complexity**: For complex models, retaining old knowledge and learning new knowledge simultaneously can be a challenge.

2.1.3 Strategies and Methods

**Regularization**: Use techniques like Elastic Weight Consolidation (EWC) to retain important information from previous tasks and prevent catastrophic forgetting.

**Buffer Replay**: Store a buffer of data samples from previous tasks and use them to train the model on new tasks, minimizing catastrophic forgetting.

**Network Expansion**: Expand the model architecture when faced with a new task, helping the model increase its learning capacity and retain old information.

2.1.4 Continuous Management

**Monitoring and Controlling Data Changes**: Continuously monitor and control changes in data to ensure the model can adapt to the dynamics of the data.

**Periodic Retraining**: Conduct periodic retraining with new data to update the model and maintain performance.

Continual Learning is crucial when building machine learning solutions for problems where data changes over time, and the model needs to continuously update to maintain predictive capabilities.

2.2 Test Production

2.2.1 Definition

Test production entails introducing a machine learning model to a production environment for validation and testing prior to full deployment. It allows for the evaluation of the model's behavior in a real-world situation and assures that it meets performance, reliability, and safety criteria.

2.2.2 Key Aspects

**A/B Testing**: Comparing the performance of different versions (A and B) of the application in a live environment to assess the impact on user experience and other key metrics.

**Canary Releases**: Gradual deployment of new features or versions to a subset of users before a full release, helping to detect and mitigate potential issues early.

**Monitoring and Logging**: Implementing extensive monitoring and logging systems to track the behavior of the application in real-time, detect anomalies, and gather insights.

**Rollback Strategies**: Planning strategies for quickly rolling back to a previous version in case issues arise after a deployment.

**Scalability Testing**: Evaluating how well the system performs under various loads and conditions to ensure scalability in the production environment.

Deploying a machine learning model in the production environment is a crucial part of the machine learning solution development process, and this Test Production process ensures that the model operates effectively and reliably in real-world conditions

REFERENCES

|  |  |
| --- | --- |
| [1] | Ayush Gupta, "A Comprehensive Guide on Optimizers in Deep Learning". Available : https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/ |
| [2] | Springer Link, "The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study". Available: https://link.springer.com/article/10.1007/s11042-022-13820-0. |
| [3] | upGrad, " Types of Optimizers in Deep Learning Every AI Engineer Should Know". Available: https://www.upgrad.com/blog/types-of-optimizers-in-deep-learning/ |
| [4] | Oreilly, " Continual Learning and Test in Production". Available: https://www.oreilly.com/library/view/designing-machine learning/9781098107956/ch09.html. |
| [5] | Omdena, " A Guide to Continuous Training of Machine Learning Models in Production". Available: https://www.omdena.com/blog/continuous-training-machine-learning-models. |