VIETNAM GENERAL CONFEDERATION OF LABOUR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL REPORT**

**MACHINE LEARNING**

Instructor: **Mr. LE ANH CUONG**

Student: **Huynh Anh Tu - 521H0325**

Class **: 21H50302**

Year  **: 23**

**HO CHI MINH CITY, 2023**

VIETNAM GENERAL CONFEDERATION OF LABOUR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL REPORT**

**MACHINE LEARNING**

Instructor: **Mr. LE ANH CUONG**

Student: **Huynh Anh Tu - 521H0325**

Class **: 21H50302**

Year  **: 23**

**HO CHI MINH CITY, 2023**

# ACKNOWLEDGEMENT

The first sincere thanks I want to give to Mr. Le Anh Cuong, who enthusiastically taught and worked tirelessly to give me enough tools and skills to complete this report. He played an important role in improving my mathematical logic and knowledge. The second thanks I would like to give to the teachers of the Department of Information Technology of Ton Duc Thang University for giving me the opportunity to do this report.

I am very open to receiving feedback from teachers so that I can improve my report writing skills.

Finally, I wish you good health and success in your noble career.

Ho Chi Minh city, 20th December, 2023

Author

(Sign and write full name)

Huynh Anh Tu

# THIS PROJECT WAS COMPLETED

# AT TON DUC THANG UNIVERSITY

I hereby declare that this thesis was carried out by myself under the guidance and supervision of Mr. Le Anh Cuong and that the work contained and the results in this thesis are true and have not been either submitted anywhere for any previous purpose or published in any other literature. The data and figures presented in this thesis are for analysis, comments, and evaluations from various resources by my own work and have been fully acknowledged in the reference part.

In addition, other comments, reviews and data from other authors, and organizations used in this thesis have been acknowledged, and explicitly cited.

I will take full responsibility for any fraud detected in my thesis. Ton Duc Thang University is unrelated to any copyright infringement caused on my work (if any).

Ho Chi Minh city, 20th December, 2023

Author

(Sign and write full name)

Huynh Anh Tu

# CONFIRMATION AND ASSESSMENT SECTION

**Instructor confirmation section**

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

Ho Chi Minh October, 2023

(Sign and write full name)

**Evaluation section for grading instructor**

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

Ho Chi Minh October, 2023

(Sign and write full name)

# SUMMARY

This is a midterm report on Machine Learning by the Faculty of Information Technology of Ton Duc Thang University. With the help of Mr. Le Anh Cuong. This report was finished at 11:35 PM on 20th December, 2023. If it weren’t for Mr. Le Anh Cuong lessons, I could not have finished this report. This article of mine has many errors. I am very open to receiving the constructive contributions of teachers and will use it as a lesson for the final articles.

# TABLE OF CONTENT

[**ACKNOWLEDGEMENT 3**](#_gjdgxs)

[**THIS PROJECT WAS COMPLETED 4**](#_30j0zll)

[**AT TON DUC THANG UNIVERSITY 4**](#_1fob9te)

[**CONFIRMATION AND ASSESSMENT SECTION 5**](#_3znysh7)

[**SUMMARY 6**](#_2et92p0)

[**TABLE OF CONTENT 7**](#_3dy6vkm)

[**CHAPTER 1 – INTRODUCTION 8**](#_1t3h5sf)

[**CHAPTER 2 - A Comparison of Optimization Techniques 9**](#_4d34og8)

[2.1 Stochastic Gradient Descent (SGD): 10](#_bxzj7yuhhfj7)

[2.2 Adam Optimization 11](#_x5swxb3c1ayi)

[2.3 RMSprop Optimization 12](#_iatrvs5whf5d)

[Summary: 13](#_battsuvfhx0b)

[**CHAPTER 3 - Continual Learning and Test Production 14**](#_y0oefg37jx9e)

[3.1 Continual Learning: 14](#_7zd6zl8l27il)

[3.2 Test Production: 15](#_l6f5tvhiw03v)

[Summary: 16](#_s288oriqsffm)

[**REFERENCE 17**](#_w0g1usbptsiv)

# CHAPTER 1 – INTRODUCTION

This study focuses on thoroughly examining several optimizer strategies used in training machine learning models. We aim to comprehensively analyze and contrast the efficacy of different optimization strategies, providing insights into their distinct advantages and constraints.

Furthermore, the research explores the complexities of Continual Learning and Test Production in the context of developing machine learning solutions to tackle specific challenges. Our objective is to carefully analyze and evaluate these techniques in order to understand its intricate characteristics and how they affect the construction of models.

Furthermore, the research explores the complexities of Continual Learning and Test Production in the context of developing machine learning solutions to tackle specific challenges. Our objective is to carefully analyze and evaluate these techniques in order to understand its intricate characteristics and how they affect the construction of models.

# CHAPTER 2 - A Comparison of Optimization Techniques

Chapter 2 introduces a thorough examination of different optimization approaches used in training machine learning models. Optimization is essential for improving model performance, and it is important to have a deep understanding of various optimization strategies to train models effectively.

This chapter will explore the complexities of well-known optimization techniques, including Stochastic Gradient Descent (SGD), Adam, and RMSprop. By conducting a thorough and systematic comparison, our objective is to reveal the distinct attributes, benefits, and constraints of each technique. This knowledge is crucial for practitioners who want to refine their models for optimal outcomes.

## 2.1 Stochastic Gradient Descent (SGD):

**Definition:**Stochastic Gradient Descent is an iterative optimization approach used to train machine learning models. The model parameters are updated by utilizing the gradients of the loss function, which are computed on randomly chosen portions of the training data called mini-batches.

**Requirement:** Stochastic Gradient Descent (SGD) is crucial for optimizing models on large datasets, as it strikes a balance between computing efficiency and accurate parameter updates.

**Formula:**

**Explanation:**

* Updated model parameters at time t+1
* : Current model parameters at time t
* Learning rate, controls the step size for parameter updates
* : Gradient of the loss function J with respect to the model parameters, calculated on a mini-batch (

**Working Principle:**

* The model parameters are updated iteratively by taking steps in the direction opposite to the gradient of the loss function. These steps are tiny in order to minimize the loss.

## 2.2 Adam Optimization

**Definition**: Adam is an optimization approach that integrates the advantages of both momentum and RMSprop. It is short for Adaptive Moment Estimation. The system calculates the moving averages of the gradients and their squares, and adjusts the learning rates for each parameter by taking into account past data.

**Requirement:** Adam surpasses the constraints of conventional optimization techniques to attain quicker convergence, especially in situations involving intricate and high-dimensional data.

**Formulas:**

**Explanation:**

* : First moment estimate (mean) of the gradients.
* : Second moment estimate (uncentered variance) of the gradients.
* , : Exponential decay rates for the moment estimates.
* : Learning rate.
* : Small constant to avoid division by zero

**Working Principle:**

* Maintains moving averages of gradients and their squares.
* Adapts learning rates for each parameter based on historical information.

## 2.3 RMSprop Optimization

**Definition**:RMSprop is an optimization algorithm designed to tackle the issue of diminishing learning rate in Stochastic Gradient Descent. It adjusts the learning rates for each parameter by using the root mean square of previous gradients.

**Need**: RMSprop is essential for overcoming the difficulties presented by uneven terrain in the loss landscape. It enables more stable and adaptable learning rates during training.

**Formula**:

**Explanation**:

* : Exponential moving average of squared gradients
* : Decay rate for the moving average.
* ϵ: Smoothing term to prevent division by zero.

**Working Principle:**

* Addresses the diminishing learning rate problem of SGD by adapting learning rates for each parameter.
* Controls the step size based on the root mean square of past gradients.

## Summary:

In conclusion, the optimization techniques discussed—Stochastic Gradient Descent (SGD), Adam, and RMSprop—are pivotal for enhancing model optimization in machine learning. SGD efficiently handles large datasets, Adam blends momentum and RMSprop for quicker convergence, and RMSprop addresses diminishing learning rates. These methods collectively contribute to increased efficiency and effectiveness in various machine learning applications.

**1. Stochastic Gradient Descent (SGD):**

* **Purpose**: Efficient optimization, especially with large datasets.
* **Advantages**: Addresses computational inefficiencies using random subsets.
* **Limitations**: May exhibit slow convergence and noise susceptibility.

**2. Adam (Adaptive Moment Estimation):**

* **Purpose**: Combines momentum and RMSprop for faster convergence.
* **Advantages**: Adapts learning rates effectively, balancing momentum and gradients.
* **Limitations**: Requires hyperparameter tuning and sensitivity to certain data types.

**3. RMSprop (Root Mean Square Propagation):**

* **Purpose**: Mitigates diminishing learning rates in SGD.
* **Advantages**: Crucial for stabilizing learning rates, especially in uneven terrains.
* **Limitations**: Requires careful tuning and performance influenced by dataset characteristics.

# CHAPTER 3 - Continual Learning and Test Production

Welcome to Chapter 3, where we delve into the dynamic realms of Continual Learning and Test Production in the development of machine learning solutions. This chapter explores the significance of lifelong learning in adapting to evolving data distributions and the critical role of effective testing procedures.

## 3.1 Continual Learning:

**Definition**:Continual Learning, also known as lifetime learning, pertains to the capacity of a machine learning model to dynamically adjust and assimilate novel information during its existence. The process entails training models on sequential tasks while ensuring that previously acquired knowledge is not lost.

**Challenges and Solutions:**

* Catastrophic Forgetting: Models exhibit a tendency to lose or disregard information from previous tasks when they are trained on new tasks. Techniques such as Elastic Weight Consolidation (EWC) and Synaptic Intelligence address this problem by imposing penalties on modifications made to significant parameters.
* Task Interference: Acquiring knowledge of a new task can disrupt the execution of previously learned activities. Approaches like Task-agnostic Networks and Progressive Neural Networks tackle this issue by breaking down the learning process into separate modules for different tasks.

**Applications**: Continuous learning is essential in situations where models must adjust to new data distributions, such as evolving user preferences, changing environments, or emerging trends. It is utilized in online learning systems, adaptive robotics, and personalized recommendation systems.

## 3.2 Test Production:

**Definition**: Test Production encompasses the methodical creation and implementation of test cases to assess the efficacy and resilience of machine learning solutions. It has a crucial function in guaranteeing that models have a broad applicability to various contexts.

**Key Aspects:**

* Data Quality and Diversity: Test Production entails generating test datasets that are both varied and representative, accurately simulating real-world settings. It guarantees a thorough evaluation of models and their ability to handle various input changes.
* Adversarial Testing: involves deliberately introducing difficult scenarios or adversarial inputs during the testing process to detect vulnerabilities and weaknesses in models. Adversarial testing improves the resilience of machine learning solutions.

**Importance**: Efficient test generation is crucial for implementing dependable machine learning solutions. It aids in identifying possible problems, guarantees that models have broad applicability, and instills trust in the model's performance under different circumstances.

## Summary:

Continual Learning and Test Production are indispensable elements in the creation of resilient and effective machine learning solutions. Continual Learning empowers models to adeptly respond to dynamic environments and evolving tasks, fostering adaptability and long-term efficacy. Simultaneously, Test Production plays a critical role in subjecting models to comprehensive evaluation and validation across diverse conditions, ensuring robust performance.

**Continual Learning:**

* **Purpose**: Enables models to adapt to changing environments and tasks over time.
* **Advantages**: Enhances adaptability, allowing models to evolve and improve continuously.
* **Applications**: Particularly useful in scenarios where data distributions or task requirements evolve.

**Test Production:**

* **Purpose**: Ensures thorough evaluation and validation of models under diverse conditions.
* **Advantages**: Validates model performance across various scenarios, improving reliability.
* **Applications**: Essential in guaranteeing the robustness and generalization capabilities of machine learning systems.

# REFERENCE

1. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
2. Hinton, G. E., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
3. Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*.
4. McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychological Review, 96*(2), 285-297.
5. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences, 114*(13), 3521-3526.
6. Chen, Y., Liu, S., Kira, Z., & Wang, Y. (2018). A closer look at memorization in deep networks. *arXiv preprint arXiv:1706.05394*.
7. Bishop, C. M. (1995). Neural networks for pattern recognition. *Oxford university press*.
8. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). *MIT press Cambridge*.
9. Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 427-436.
10. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision, 115*(3), 211-252.