VIETNAME GENERAL CONFEDERATION OF LABOR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL PROJECT**

**LEARN ABOUT MACHINE LEARNING**

*Instructor*: **MR. LE ANH CUONG**

*Performed by*: **PHAM DO TRONG DAI - 519H0148**

Course**: 23-25**

**HO CHI MINH CITY, YEAR 2023**

VIETNAME GENERAL CONFEDERATION OF LABOR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**MIDTERM PROJECT**

**LEARN ABOUT MACHINE LEARNING**

*Instructor*: **MR. LE ANH CUONG**

*Performed by*: **PHAM DO TRONG DAI - 519H0148**

Course**: 23-25**

**HO CHI MINH CITY, YEAR 2023**

ACKNOWLEDGEMENT

I would like to extremly thank Ton Duc Thang university choose to be a student of school and has come up with introduce to machine learning program my teaching process and I would like to sincerely thank teacher Le Anh Cuong is dedicated to teaching and give me the best knowledge for the lesson.

I EXTREMLY THANK YOU!

**THE PROJECT WAS COMPLETED**

**AT TON DUC THANG UNIVERSITY**

We hereby declare that this is the product our own project and under the guidance of Mr. Le Anh Cuong. The research content, results in this topic are honest and not published in any form before. The data in the tables for analysis, comments and evaluation are collected by the author himself from different sources, which are clearly stated in the reference section.

**If any fraud is detected, I would like to take full responsibility for the content of my project**. Ton Duc Thang University is not related to the rights and copyright crimes caused by me during the implementation process (if any).

*Ho Chi Minh City, 22 October, 2023*

*Author*

CONFIRMATION AND EVALUATION BY THE INSTRUCTO

**Confirmation section of the instructor’s guidance**

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

Ho Chi Minh City, date month year

(sign and write full name)

**Evaluation section of the teacher’s grading**

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

Ho Chi Minh City, date month year

(sign and write full name)

CHAPTER 1 – Learn and compare Optimizer methods in training machine learning models

**1.1** What is the Optimizer method?  
- Optimizer in the context of machine learning and machine learning is an algorithm used to adjust the parameters of a machine learning model to minimize the value of the loss function. The goal of optimization is to find the minimum value of the loss function, and update the model's weights so that the model has good prediction ability on new data. Optimization helps adjust model parameters in a way that minimizes the bias between predictions and actual values.

**1.2** Lists optimization methods in training machine learning models and outlines their strengths and weaknesses

**1.2.1** Gradient Descent:

* Gradient Descent is an optimization method used during machine learning model training to adjust model parameters based on the gradient of the loss function.
* Formula to update parameters in GD:

note:

* is the parameter vector at the iteration .
* is the slope of the loss function at the point
* is the learning rate, which is a hyperparameter that determines the size of the update step.
* Advantages and disadvantages:
* Advantages:
* Easy to implement and understand.
* Suitable for problems with small data sets.
* Disadvantages:
* Can be absorbed into local minimums.
* Slow convergence in problems with complex parameter spaces.

**1.2.2** Stochastic Gradient Descent:

* Stochastic Gradient Descent is an optimization method used in the process of training machine learning models, especially in machine learning and neural network problems. Different from traditional Gradient Descent, SGD performs model parameter updates based on a random data point from the training set instead of based on the entire data set.
* SGD has the following process:
* Choose a random data point from the training set.
* Calculate the derivative of the loss function at that time.
* Update the model parameters in the opposite direction to the gradient and to the learning rate.
* Formula to update parameters in SGD:

Note:

* is the parameter vector at the iteration .
* is the slope of the loss function at the point . based on a specific data point
* is the learning rate, which is a hyperparameter that determines the size of the update step.
* Advantages and disadvantages
* Advantages:
  + Suitable for big data.
  + Reduced risk of minimal local absorption.
* Disadvantages:
  + High fluctuations due to random noise.
  + Need to set appropriate mini-batch size.

**1.2.3** Mini-batch Gradient Descent:

* Mini-batch Gradient Descent is a variation of the Gradient Descent (GD) optimization method used in machine learning model training. In Mini-batch Gradient Descent, the training data is not completely used during each parameter update, but instead, only a small portion of the data is used. Specifically, the data is divided into mini-batches (small batches), and each update uses only one mini-batch to calculate the gradient and update the parameters.
* Formula to update parameters in Mini-batch Gradient Descent:

Note:

* is the parameter vector at the iteration .
* is the slope of the loss function calculated on a mini-batch at the pointt .
* is the learning rate, which is a hyperparameter that determines the size of the update step.
* Advantages and disadvantages
  + Advantages:
    - Combine the benefits of GD and SGD.
    - Speed ​​up the convergence process.
  + Disadvantages:
    - Need to adjust mini-batch size.

**1.2.4** Momentum Optimization:

* Momentum Optimization is an optimization method used during machine learning model training to reduce oscillations and speed up convergence. This method is designed to solve the problem of Gradient Descent, especially when the path of the optimization process has many local minima or when the loss function is long and curved.
* Formula to update parameters in Momentum Optimization:

Note:

* + is the vector of model parameters.
  + is the slope of the loss function J at.
  + is the learning rate, which determines the size of the update step.
  + is the momentum coefficient, which determines the degree to which the algorithm "remembers" the previous updated direction.
* Advantages and disadvantages
  + Advantages:
    - Reduce oscillation and increase convergence speed.
  + Disadvantages:
    - It is possible to overshoot the target if the momentum is too great.

**1.2.5** Adam (Adaptive Moment Estimation):

* Momentum Optimization is an optimization method used during machine learning model training to reduce oscillations and speed up convergence. This method is designed to solve the problem of Gradient Descent, especially when the path of the optimization process has many local minima or when the loss function is long and curved.
* Adam generally performs well on a wide range of problems and has become one of the popular optimization methods in the machine learning community. However, the choice of optimization method still depends on the specific characteristics of the data and model.
* Advantages and disadvantages
  + Advantages:
    - Effective on many types of problems.
    - Self-adjust learning rate and weight reduction.
  + Disadvantages:
    - Can be sensitive to hyperparameters.

**1.2.6** Adagrad:

* Adagrad (Adaptive Gradient Algorithm) is an optimization method in machine learning designed to automatically adjust the learning rate for each parameter of the model based on how often they appear during training.
* Adagrad's idea is to update the learning rate of each parameter so that the parameter with a larger gradient will have a lower learning rate and vice versa. This helps the learning rate automatically adapt to each parameter based on previous learning experience.
* Formula to update parameters in Adagrad for each parameter *θi* at a time t:

Note:

* + is the initial learning rate.
  + is the gradient of parameter i at time t.
  + is the sum of squares of the previous gradients of parameter i, calculated from time 1 to t.
  + is a small constant (usually ) to avoid division by 0.
* Advantages and disadvantages
  + Advantages:
    - Integrate the learning rate for each parameter.
  + Disadvantages:
    - Accumulate the square of the gradient, leading to a reduction in the overlearning rate.

**1.2.7** RMSprop:

* RMSprop (Root Mean Square Propagation) is an optimization method used in the process of training machine learning models. This method is designed to reduce the problem of Adagrad, another optimization method.
* RMSprop has the following operating principles: In RMSprop, the learning rate of each parameter is adjusted based on the degree of change in the gradient of that parameter. The goal is to reduce the learning rate for parameters with large gradients and increase the learning rate for parameters with small gradients. This stabilizes the optimization process and reduces the risk of overlearning on unimportant parameters.
* Advantages and disadvantages
  + Advantages:
    - Reduce Adagrad's problem by reducing the weight of too steep gradients.
  + Disadvantages:
    - It is necessary to select the weight reduction parameter.

**1.2.8 *Nesterov Accelerated Gradient (NAG)***:

* Nesterov Accelerated Gradient (NAG) là một biến thể của phương pháp tối ưu hóa Momentum Optimization trong quá trình huấn luyện mô hình máy học.
* NAG improves on the idea of ​​Momentum Optimization by using lookahead when updating the weights. Instead of updating directly in the gradient direction, the NAG first predicts where the weights will move if momentum is used, and then it calculates the gradient at the predicted location.
* Formula to update parameters in NAG:

Note:

* + is the model parameter vector.
  + is the slope of the loss function at the current location.
  + is the learning rate.
  + is the momentum coefficient.
  + is the momentum vector at time t.
* Advantages and disadvantages
  + Advantages:
    - Reduces oscillations and helps converge faster.
  + Disadvantages:
    - Need to adjust hyperparameters.

CHAPTER 2 - Learn about Continuous Learning and Test Production

**2.1** Continual Learning:

* Continuous Learning (CL) is an important field in machine learning and artificial intelligence that focuses on the ability of machine learning models to retain knowledge over time when dealing with multiple tasks or data new. Continuous Learning helps the model adapt and not forget previously learned knowledge.
* Some important challenges in Continuous Learning:
  + Catastrophic Forgetting: The model can forget previously learned knowledge when learning a new task.
  + Interference: Knowledge of old tasks can affect the model's ability to learn on new tasks.
  + Scalability: Model complexity and ability to handle multiple tasks.
* Continuous Learning methods include:
  + Replay: Store and reuse old data to help the model recall knowledge.
  + Regularization: Use methods like Elastic Weight Consolidation (EWC) to minimize the influence of new learning on old task importance weights.
  + Dynamic Architectures: Model structure can change to match the complexity of the task.

**2.1** Test Production:

* Test Production is the process of creating test data to evaluate the performance of a machine learning model. Building a machine learning solution often involves establishing a quality test data set to ensure that the model generalizes well to new data.
* Important methods in the Test Production process:
  + Prepare Test Data: Identify and collect test data that represents the data distribution that the model is likely to encounter in a real-world environment.
  + Create Test Data That Covers a Variety of Scenarios: Includes both large and small cases, boundary situations, and unusual data.
  + Diversified Testing: Make sure the test data is diverse enough to evaluate the model's ability to generalize.
  + Continuous Testing: Update and expand test data as new data becomes available or as the model is updated.

**2.1** Overview:

* Both of these aspects are important to build and maintain a high-performance and flexible machine learning solution in real-world environments. Continuous Learning helps the model not forget knowledge when faced with multiple tasks, while Test Production ensures that the model is properly evaluated on test data representative of any real machine learning field.