

**TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN**



**Đỗ Nhật Huy – mssv 521h0237**

**BÀI BÁO CÁO CUỐI KỲ**

**NHẬP MÔN HỌC MÁY**

**KHOA HỌC MÁY TÍNH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

**TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN**



**TRỊNH MINH TÍCH THIỆN – 521H0305**

**BÀI BÁO CÁO CUỐI KỲ**

**NHẬP MÔN HỌC MÁY**

**KHOA HỌC MÁY TÍNH**

Người hướng dẫn  
**TS. Lê Anh Cường**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

## LỜI CẢM ƠN

Em xin gửi lời cảm ơn sâu sắc đến thầy Lê Anh Cường về những kiến thức quý báu và sự hướng dẫn tận tâm mà thầy đã chia sẻ trong môn học Nhập Môn Học Máy. Thầy không chỉ giúp chúng em hiểu rõ về những khái niệm cơ bản trong lĩnh vực Học Máy mà còn truyền đạt với tinh thần sự say mê và đam mê nghề nghiệp. Những bài giảng của thầy không chỉ là việc truyền đạt thông tin mà còn là nguồn động viên lớn để chúng em không ngừng nỗ lực và phấn đấu trong hành trình học tập của mình. Đặc biệt, chúng em rất trân trọng những hỗ trợ của thầy để giúp chúng em giải quyết những khó khăn trong quá trình thực hiện bài báo cáo. Nhờ có sự hướng dẫn của Thầy, chúng em đã có được cái nhìn tổng quan và sâu sắc hơn về lĩnh vực này. Chúng em mong rằng sau khóa học này, sẽ có nhiều cơ hội được học thêm từ thầy và tiếp tục nhận được sự hỗ trợ quý báu từ thầy trên con đường nghề nghiệp của mình. Một lần nữa, chân thành cảm ơn thầy Lê Anh Cường và hy vọng sẽ tiếp tục được học hỏi từ sự am hiểu sâu sắc của thầy về lĩnh vực Học Máy.

*TP. Hồ Chí Minh, ngày 20 tháng 12 năm 2023*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Đỗ Nhật Huy*

**PHIẾU ĐÁNH GIÁ CỦA GIẢNG VIÊN HƯỚNG DẪN**  
(BOLD, size 16)

Tên giảng viên hướng dẫn:.....

Ý kiến nhận xét:.....

.....

.....

.....

Điểm tổng theo phiếu đánh giá rubrik:.....

*TP. Hồ Chí Minh, ngày      tháng      năm 20*  
*Giảng viên hướng dẫn*  
*(Ký tên và ghi rõ họ tên)*

## **CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Nhóm chúng em xin cam đoan đây là công trình nghiên cứu của riêng chúng em và được sự hướng dẫn khoa học của thầy Lê Anh Cường. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong báo cáo còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào nhóm chúng em xin hoàn toàn chịu trách nhiệm về nội dung Báo cáo Dự án CNTT 2 của mình.** Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do chúng em gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày    tháng    năm*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Đỗ Nhật Huy*

## 1) Learn and compare Optimizer methods in training machine learning models;

During the process of training a machine learning model, optimization is an important factor to improve model performance. There are many different optimizer methods developed to help optimize the loss function and update the model weights. Below, we will learn and compare some popular optimizer methods.

### **Gradient Descent (GD):**

Gradient Descent is the most basic optimization method in machine learning.

How it works: Update the weights by moving in the direction opposite to the derivative of the loss function.

Advantages: Easy to deploy, works well on convex problems.

Disadvantages: Easy to fall into the local optimum, the update direction may not be effective on nonlinear problems.

### **Stochastic Gradient Descent (SGD):**

- SGD is a variation of GD.

How it works: Updates the weights based on a random

sample from the data set. Advantages: Suitable for large data sets, can escape local optimization.

Disadvantages: Ability to fluctuate strongly, unstable in convergence.

### **Mini-Batch Gradient Descent:**

Mini-Batch GD is a combination of GD and SGD.

How it works: Updates weights based on a small sample of data (mini-batch) selected from the data set.

Advantages: Combines the benefits of both GD and SGD, suitable for machine learning on large data sets.

Disadvantages: Need to set appropriate mini-batch size.

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

Adam is a recently developed self-tuning optimizer method.

How it works: Uses information about the first and second derivatives to update the weights. Advantages: Effective on many types of problems, automatically adjusts learning rate.

Disadvantage: Can sometimes be unstable on some problems.

## **RMSProp (Root Mean Square Propagation):**

RMSProp is another optimizer method that uses adaptive learning rate.

How it works: Adjust the learning rate for each parameter based on the magnitude of its gradient.

Advantages: Effective on many problems, more stable than SGD.

Disadvantage: Inconsistent update capabilities for parameters

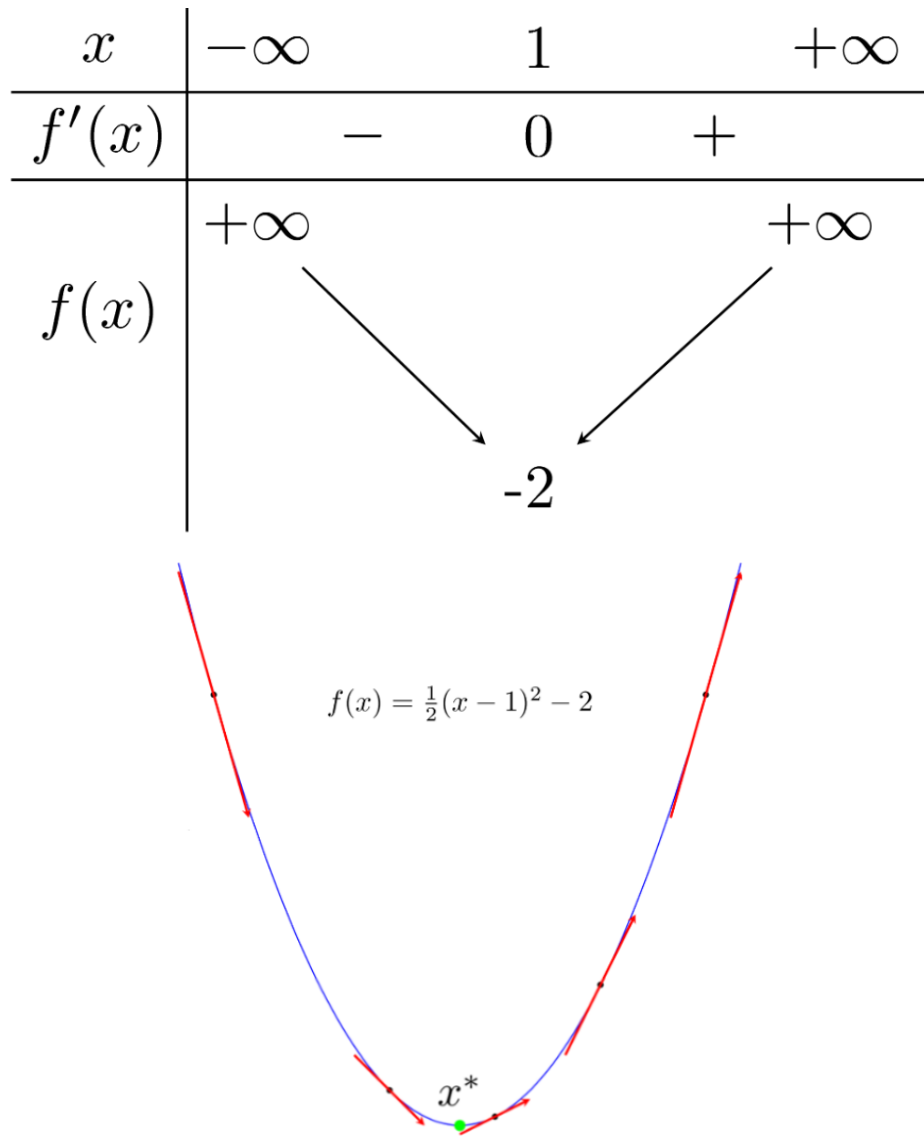
### **Conclude:**

Optimizer methods have their own advantages and disadvantages, and the choice of optimizer depends on the specific problem and data set. To determine the optimal method for your model, you need to test and evaluate each method on your specific data set and problem to find the most optimal method.

clear

## **1.1 Gradient descent for a function of 1 variable**





Green is the point that has both the minimum value and the local minimum or minimum point. From here on, I will use local minimum to mark the minimum point and global minimum to mark the point of minimum value. International Minimum is an example of it.

- (1) In case the derivative of the function at  $x_t$  is  $f'(x_t) > 0$ , then  $x_t$  is to the right of  $x^*$  (and vice versa). We need to move  $x_t$  to the left, that is, to the negative side, so that the next point  $x_{t+1}$  is closer to  $x^*$ . In other words, we have to go against the derivative

$$X_{t+1} = x_t + \Delta$$

In which  $\Delta$  is a quantity with opposite sign to the derivative  $f'(x_t)$ .

- (2) The farther  $x^*$  is to the right, the larger  $f'(x_t)$  is than 0 (and vice versa). So, the amount of movement  $\Delta$ , most intuitively, is proportional to  $-f'(x_t)$

The two comments above give us a simple update:

$$X_{t+1} = x_t - \eta f'(x_t)$$

**2) Learn about Continuous Learning and Test Production when building a machine learning solution to solve a certain problem.**

When building a machine learning solution to solve a specific problem, two important aspects to consider are Continuous Learning and Test Production.

## 1. Continuous Learning:

Continuous Learning is a machine learning method designed to allow models to learn from new data without forgetting previously learned knowledge. This is especially important in applications that require the model to continuously update and adapt to new data, such as machine learning on streaming data or in changing environmental situations. .

Some popular Continuous Learning methods include:

Elastic Weight Consolidation (EWC): EWC is a method to preserve previous model knowledge by setting weights on important parameters and not allowing them to change too quickly when learning from new data.

Progressive Neural Networks (PNN): PNN uses a passive network to maintain old knowledge and add new layers to learn from new data.

LwF (Learning without Forgetting): LwF solves the

problem of learning without forgetting by using old data to evaluate the model's performance on new data and using an auxiliary loss function to ensure knowledge old is not lost.

## 2. Test Production:

Test Production involves creating tests to evaluate the performance of a machine learning model. When building a machine learning solution, creating effective and reliable testing is important to ensure that the model performs well on real-world data.

Important steps in Test Production include:

Choose test data: You must choose a test data set that represents the real situation the model will face.

Define evaluation metrics: Define metrics or indices to evaluate model performance, such as accuracy, precision, recall, F1-score, RMSE, MAE, etc.

Design the test pipeline: Determine how the model will be evaluated on the test data, including how the test data will be preprocessed and how the model's predictions will be evaluated.

Conclude:

Continuous Learning and Test Production are two important aspects in building machine learning solutions. Continuous Learning helps the model adapt to new data without forgetting old knowledge, while Test Production ensures that the model's performance can be reliably evaluated on real data.

### So sánh các Optimizer

Optimizer	Advantage	Defect
SGD	<ul style="list-style-type: none"><li>- Easy to implement and effective calculation.</li><li>- Effective for large data sets with high-dimensional feature space.</li></ul>	<ul style="list-style-type: none"><li>SGD can get stuck in local minima.</li><li>- High sensitivity with initial learning rate</li></ul>

Adam	<ul style="list-style-type: none"> <li>- Effective and easy to perform.</li> <li>- Applicable to large data sets and multidimensional models.</li> <li>- Good generalization ability.</li> </ul>	<ul style="list-style-type: none"> <li>- Careful adjustment of hyperparameters is required.</li> </ul>
RMSProp	<ul style="list-style-type: none"> <li>- Adaptive learning rate per parameter helps limit gradient accumulation.</li> <li>- Effective against non-fixed targets.</li> </ul>	<ul style="list-style-type: none"> <li>There may be a slow convergence rate in some cases.</li> </ul>
Adagrad	<ul style="list-style-type: none"> <li>Adaptive learning rate for each parameter.</li> <li>- Effective for sparse data.</li> </ul>	<ul style="list-style-type: none"> <li>- Việc tích lũy gradient bình phương trong mẫu số có thể khiến tốc độ học giảm xuống quá nhanh.</li> <li>- Có thể dừng</li> </ul>

		việc học quá sớm.
--	--	-------------------