TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ MÔN HỌC MÁY**

**BÁO CÁO CUỐI KÌ**

*Người hướng dẫn*: **TS LÊ ANH CƯỜNG**

*Người thực hiện*: **NGUYỄN HUỲNH ĐỨC - 52100613**

Khoá  **: 25**

**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**



**ĐỒ ÁN CUỐI KÌ MÔN HỌC MÁY**

**BÁO CÁO CUỐI KÌ**

*Người hướng dẫn*: **TS LÊ ANH CƯỜNG**

*Người thực hiện*: **NGUYỄN HUỲNH ĐỨC - 52100613**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

**LỜI CẢM ƠN**

Nhóm chúng em xin gửi lời cảm ơn chân thành đến Khoa Công Nghệ Thông Tin, Trường Đại học Tôn Đức Thắng đã tạo điều kiện thuận lợi cho nhóm em học tập và hoàn thành bài báo cáo này. Đặc biệt, chúng em xin bày tỏ lòng biết ơn sâu sắc đến thầy Lê Anh Phong, giảng viên bộ Học máy đã truyền đạt kiến thức và hướng dẫn cho chúng em trong quá trình học tập và làm bài.

Nhóm đã cố gắng hết sức trong quá trình thực hiện bài báo cáo nhưng không thể không tránh khỏi những hạn chế và sai sót không đáng có. Rất mong nhận được những góp ý của thầy trong quá trình chấm bài để chúng em tiếp tục học hỏi và cải thiện cho các báo cáo lần sau.

Kính chúc quý thầy cô, giảng viên bộ môn thuận lợi và thành công trong công việc. Chúc thầy cô dồi dào sức khỏe và hạnh phúc bên người thân gia đình.

Chúng em xin chân thành cảm ơn!

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Chúng em xin cam đoan đây là sản phẩm đồ án của chúng em và được sự hướng dẫn của ThS 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 đồ án 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 chúng em xin hoàn toàn chịu trách nhiệm về nội dung đồ án 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)*

*Đức*

*Nguyễn Huỳnh Đức*

**PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN**

**Phần xác nhận của GV hướng dẫn**

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

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

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

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**MỤC LỤC**

[Câu 1: Tìm hiểu, so sánh các phương pháp Optimizer trong huấn luyện mô hình học máy; 2](#_Toc154219106)

[1. Tìm hiểu về Optimizer là gì ? 2](#_Toc154219107)

[2. Các phương pháp Optimizer trong huấn luyện mô hình học máy 2](#_Toc154219108)

[2.1 Gradient Descent (GD) 2](#_Toc154219109)

[2.2 Stochastic Gradient Descent (SGD) 2](#_Toc154219110)

[2.3 Mini-batch Gradient Descent 2](#_Toc154219111)

[2.4 Momentum 3](#_Toc154219112)

[2.5 Adagrad 3](#_Toc154219113)

[2.6 RMSprop (Root Mean Square Propagation) 3](#_Toc154219114)

[2.7 Adam (Adaptive Moment Estimation) 4](#_Toc154219115)

[2.8 Adadelta 4](#_Toc154219116)

[2.9 Nadam 4](#_Toc154219117)

[2.10 FTRL-Proximal 4](#_Toc154219118)

[Câu 2: Tìm hiểu về Continual Learning và Test Production khi xây dựng một giải pháp học máy để giải quyết một bài toán nào đó. 5](#_Toc154219119)

[1. Continual Learning 5](#_Toc154219120)

[2. Test Production 5](#_Toc154219121)

[TÀI LIỆU THAM KHẢO 7](#_Toc154219122)

# Câu 1: Tìm hiểu, so sánh các phương pháp Optimizer trong huấn luyện mô hình học máy;

## Tìm hiểu về Optimizer là gì ?

* Optimizer là một thành phần quan trọng trong quá trình huấn luyện mô hình máy học và học sâu. Nhiệm vụ của optimizer là cập nhật trọng số của mô hình sao cho hàm mất mát (loss function) giảm dần và mô hình đạt được hiệu suất tốt hơn. Dưới đây là một số phương pháp optimizer phổ biến:

## Các phương pháp Optimizer trong huấn luyện mô hình học máy

### 2.1 Gradient Descent (GD)

- Dựa trên đạo hàm của hàm mất mát để di chuyển trọng số theo hướng ngược với độ dốc của hàm mất mát.

Cập nhật trọng số: *θ*=*θ*−*α*∇*J*(*θ*) trong đó *α* là learning rate, ∇*J*(*θ*) là đạo hàm của hàm mất mát theo *θ*.

* Ưu điểm: Đơn giản, dễ hiểu.
* Nhược điểm: Chậm, đặc biệt trên dữ liệu lớn.

### 2.2 Stochastic Gradient Descent (SGD)

- Thực hiện cập nhật trọng số sau mỗi mẫu dữ liệu, làm giảm động lượng của quá trình học.

Cập nhật trọng số: *θ*=*θ*−*α*∇*J*(*θ*,*xi*​,*yi*​), với *xi*​ là mẫu dữ liệu, *yi* là giá trị đích tương ứng.

* Ưu điểm: Hiệu suất cao trên dữ liệu lớn.
* Nhược điểm: Dao động lớn, không ổn định.

### 2.3 Mini-batch Gradient Descent

- Kết hợp ưu điểm của GD và SGD bằng cách cập nhật trọng số sau mỗi mini-batch dữ liệu.

Cập nhật trọng số: *θ*=*θ - α*∇*J*(*θ*,*X*mini-batch​,*Y*mini-batch​), với *X*mini-batch​ và *Y*mini-batch là mini-batch dữ liệu.

* Ưu điểm: Kết hợp lợi ích của GD và SGD, phù hợp cho nhiều loại dữ liệu.
* Nhược điểm: Cần lựa chọn kích thước mini-batch phù hợp.

### 2.4 Momentum

- Sử dụng động lượng để giảm đà của quá trình học và tránh rơi vào các điểm cực tiểu địa phương.

Cập nhật trọng số: *v*=*βv - α*∇*J*(*θ*), *θ*=*θ* +*v với β* là hệ số động lượng

* Ưu điểm: Giảm dao động, giúp vượt qua các điểm địa phương.
* Nhược điểm: Cần điều chỉnh hệ số momentum.

### 2.5 Adagrad

- Điều chỉnh learning rate cho từng tham số dựa trên tần suất xuất hiện của nó.

Cập nhật trọng số: *G*=*G*+ (∇*J*(*θ*))2, *θ*=*θ* −*G*+*ϵ*​*α*​∇*J*(*θ*), trong đó G là ma trận đường chéo chứa tổng bình phương đạo hàm theo từng tham số, ϵ là một số nhỏ để tránh chia cho 0.

* Ưu điểm: Hiệu quả với các biến đổi thưa thớt.
* Nhược điểm: Learning rate giảm quá nhanh, không phù hợp cho mô hình deep learning.

### 2.6 RMSprop (Root Mean Square Propagation)

- Cải thiện Adagrad bằng cách duy trì trọng số thơi gian.

Cập nhật trọng số: *G*= *βG* +(1−*β*) (∇*J*(*θ*))2, *θ*=*θ* −*G*+*ϵ*​*α*​∇*J*(*θ*), với *β* là hệ số giảm động lượng

* Ưu điểm: Hiệu quả và ổn định hơn so với Adagrad.
* Nhược điểm: Cần điều chỉnh learning rate.

### 2.7 Adam (Adaptive Moment Estimation)

- Kết hợp động lượng và RMSprop, sử dụng động lượng và thơi gian để cập nhật learning rate.

Cập nhật trọng số: *m = β*1​*m +* (1−*β*1​) ∇*J*(*θ*), *v*=*β*2​*v* +(1−*β*2​) (∇*J*(*θ*))2, *θ*=*θ* −*v*​+*ϵα*​*m với m và v* là moment của bậc nhất và bậc hai,*β*1​ và *β*2​ là hệ số giảm động lượng.

* Ưu điểm: Hiệu suất tốt trên nhiều loại mô hình.
* Nhược điểm: Cần điều chỉnh nhiều siêu tham số.

### 2.8 Adadelta

- Loại bỏ nhu cầu điều chỉnh learning rate, sử dụng tỉ lệ giảm trọng số thay vì learning rate cố định.

Cập nhật trọng số: Δ*θ*= −*G*+*ϵ*​Δ*θ*+*ϵ*​​∇*J*(*θ*), *θ*=*θ* +Δ*θ*, với Δ*θ* là ma trận trọng số thay đổi, G là ma trận đường chéo chứa tổng bình phương đạo hàm.

* Ưu điểm: Stabilizes learning rate, ít phụ thuộc vào siêu tham số.
* Nhược điểm: Cần nhiều bộ nhớ hơn trong quá trình huấn luyện.

### 2.9 Nadam

- Kết hợp Adam và Nesterov Momentum.

Cập nhật trọng số: Tương tự như Adam, nhưng thêm đạo hàm thơi gian với động lượng.

* Ưu điểm: Kết hợp lợi ích của cả Adam và Nesterov.
* Nhược điểm: Có thể không phù hợp cho mọi loại mô hình.

### 2.10 FTRL-Proximal

- Dùng cho các mô hình với số lượng lớn các đặc trưng, sử dụng chiến lược tỷ lệ giảm trọng số.

Cập nhật trọng số: Sử dụng tỷ lệ giảm trọng số dựa trên đạo hàm của hàm mất mát và trọng số hiện tại.

* Ưu điểm: Hiệu quả với các dữ liệu có nhiều đặc trưng.
* Nhược điểm: Cần điều chỉnh nhiều tham số.

# Câu 2: Tìm hiểu về Continual Learning và Test Production khi xây dựng một giải pháp học máy để giải quyết một bài toán nào đó.

## Continual Learning

Continual Learning (hay còn gọi là Lifelong Learning) là một lĩnh vực trong học máy nghiên cứu về khả năng của mô hình học máy để liên tục học và thích ứng với dữ liệu mới mà không quên đi kiến thức đã học trước đó. Trong môi trường thực tế, dữ liệu có thể thay đổi theo thời gian, và mô hình cần phải cập nhật kỹ thuật học của mình mà không gặp vấn đề quên đi kiến thức cũ.

Một số thách thức chính của Continual Learning bao gồm:

**Quên quá mức (Catastrophic Forgetting):** Mô hình có thể quên kiến thức đã học khi đối mặt với dữ liệu mới.

**Chuyển giao thông tin (Transfer of Knowledge):** Làm sao chia sẻ và sử dụng kiến thức đã học từ một nhiệm vụ cho các nhiệm vụ mới.

**Quản lý bộ nhớ (Memory Management):** Làm thế nào để hiệu quả lưu trữ và quản lý kiến thức cũ trong mô hình.

## Test Production

Test Production là một quá trình quan trọng trong xây dựng giải pháp học máy. Sau khi mô hình đã được huấn luyện trên dữ liệu, quá trình kiểm thử (test) được thực hiện để đánh giá hiệu suất của mô hình trên dữ liệu chưa nhìn thấy trước đó. Quá trình này giúp đảm bảo rằng mô hình có khả năng tổng quát hóa và không chỉ làm việc tốt trên dữ liệu huấn luyện.

Các bước chính trong quá trình Test Production bao gồm:

**Chuẩn bị Dữ liệu Kiểm thử**: Lựa chọn một tập dữ liệu kiểm thử đại diện để đánh giá hiệu suất của mô hình.

**Thực Hiện Kiểm thử:** Áp dụng mô hình đã được huấn luyện lên tập dữ liệu kiểm thử và đánh giá kết quả dự đoán so với nhãn thực tế.

**Đánh Giá và Hiệu Chuẩn Mô hình:** Dựa trên kết quả kiểm thử, đánh giá hiệu suất của mô hình và có thể điều chỉnh các siêu tham số hoặc cải thiện mô hình nếu cần.

**Ghi Chú và Báo Cáo:** Ghi lại kết quả kiểm thử và tạo báo cáo về hiệu suất của mô hình, bao gồm cả các khía cạnh như độ chính xác, độ đồng đều, và các độ đo khác tùy thuộc vào bài toán cụ thể.

* Cả hai khía cạnh trên đều quan trọng trong quá trình xây dựng và duy trì một hệ thống học máy thực tế và hiệu quả. Continual Learning giúp mô hình làm việc hiệu quả trên dữ liệu thay đổi liên tục, trong khi Test Production đảm bảo rằng mô hình hoạt động chính xác và hiệu quả trong môi trường sản xuất.

# TÀI LIỆU THAM KHẢO

**1.** D. P. Kingma và J. Ba, ["Adam: A Method for Stochastic Optimization"](https://arxiv.org/abs/1412.6980): Bài báo giới thiệu về phương pháp Adam.

**2.** J. Reddi, S. Sra, B. Póczos, và A. Smola, "Stochastic Variance Reduction for Nonconvex Optimization": Bài báo giới thiệu về phương pháp SVRG, một trong những ý tưởng cơ bản của các optimizer hiện đại.

**3.** Sebastian Ruder, ["An overview of gradient descent optimization algorithms"](https://books.google.com/books/about/An_Overview_of_Gradient_Descent_Optimiz.html?id=2lTzswEACAAJ): Sách này nắm vững các thuật toán tối ưu hóa gradient descent và cung cấp một cái nhìn chi tiết về chúng.