

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



ĐỒ ÁN NHẬP MÔN HỌC MÁY

# CÁC PHƯƠNG PHÁP OPTIMIZE CONTINUAL LEARNING VÀ TEST PRODUCTION

*Người hướng dẫn:* GV. Lê Anh Cường

*Người thực hiện:* Hoàng Quốc Bảo - 51603024

*Lớp:* 18050200

*Khoá:* 20

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 NHẬP MÔN HỌC MÁY**

# **CÁC PHƯƠNG PHÁP OPTIMIZE CONTINUAL LEARNING VÀ TEST PRODUCTION**

*Người hướng dẫn:* **GV. Lê Anh Cường**

*Người thực hiện:* **Hoàng Quốc Bảo - 51603024**

*Lớp:* **18050200**

*Khoá:* **20**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

## LỜI CẢM ƠN

Để hoàn thành bài báo cáo này trước tiên nhóm chúng em xin gửi đến các quý thầy, cô giảng viên trường Đại học Tôn Đức Thắng lời cảm ơn chân thành và sâu sắc nhất. Đặc biệt, nhóm chúng em xin gửi đến thầy Lê Anh Cường – người đã tận tình hướng dẫn, giúp đỡ chúng em hoàn thành tiểu luận này lời cảm ơn sâu sắc nhất. Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế của một sinh viên, bài báo cáo này không thể tránh được những thiếu sót. Chúng em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các quý thầy cô để chúng em có điều kiện bổ sung, nâng cao ý thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Chúng em xin chân thành cảm ơn!.

## **BÁO CÁO ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Chúng tôi xin cam đoan đây là sản phẩm đồ án của riêng chúng tôi và được sự hướng dẫn 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 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 tôi 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 tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 20 tháng 07 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Hoàng Quốc Bảo*

## **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)

## TÓM TẮT

## MỤC LỤC

LỜI CẢM ƠN .....	i
BÁO CÁO ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG ...i	
PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN .....	iii
TÓM TẮT .....	iv
MỤC LỤC.....	v
DANH MỤC HÌNH .....	vi
DANH MỤC BẢNG.....	vii
Chương 1 - TÌM HIỂU, SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY .....	1
1.1    Tổng quan Optimizer .....	<b>Error! Bookmark not defined.</b>
1.2    Các phương pháp Optimizer .....	1
1.3    Kết Luận.....	<b>Error! Bookmark not defined.</b>
Chương 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. ....	4
2.1    Continual Learning .....	4
2.2    Test Production .....	4
2.3    Kết luận .....	<b>Error! Bookmark not defined.</b>
Chương 3 - TÀI LIỆU THAM KHẢO .....	5

## DANH MỤC HÌNH



## **DANH MỤC BẢNG**

# Chương 1 - TÌM HIỂU, SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY

## 1.1 Các phương pháp Optimizer

Trong học máy, phương pháp tối ưu hóa (optimizer) là một thuật toán được sử dụng để tìm giá trị tối ưu của một hàm mục tiêu. Trong bối cảnh huấn luyện mô hình học máy, hàm mục tiêu thường là hàm mất mát, đại diện cho mức độ sai số giữa dự đoán của mô hình và dữ liệu thực tế.

Một số phương pháp gradient phổ biến bao gồm:

- Gradient descent (tìm đường xuống dốc): Phương pháp này đơn giản là cập nhật các tham số của mô hình theo hướng ngược với đạo hàm của hàm mục tiêu.
- Momentum (sức quán tính): Phương pháp này thêm một thành phần quán tính vào hướng cập nhật của các tham số. Điều này giúp mô hình vượt qua các vùng địa phương tối ưu.
- Adagrad (tìm đường xuống dốc với tốc độ học tập giảm dần): Phương pháp này điều chỉnh tốc độ học tập của mô hình dựa trên độ lớn của đạo hàm. Điều này giúp mô hình tránh bị mắc kẹt trong các vùng địa phương tối ưu.
- RMSProp (tìm đường xuống dốc với tốc độ học tập giảm dần dựa trên RMS): Phương pháp này tương tự như Adagrad nhưng sử dụng RMS của đạo hàm để điều chỉnh tốc độ học tập. Điều này giúp mô hình ổn định hơn.
- Adam (tìm đường xuống dốc với tốc độ học tập giảm dần dựa trên tích phân đạo hàm): Phương pháp này kết hợp các ý tưởng của Momentum và RMSProp để tạo ra một phương pháp hiệu quả và ổn định.

Có rất nhiều phương pháp Optimizer khác nhau, mỗi phương pháp có ưu và nhược điểm riêng. Một số phương pháp Optimizer phổ biến bao gồm:

Phương pháp	Ưu điểm	Nhược điểm
Gradient Descent	Đơn giản, dễ hiểu	Tốc độ hội tụ chậm
Phương pháp	Ưu điểm	Có thể bị vượt biên
Gradient Descent	Đơn giản, dễ hiểu	Có thể bị lắc lư
RMSprop	Độ ổn định cao hơn Adagrad	Tốc độ hội tụ chậm hơn Adagrad
Adam	Hiệu quả cao	Có thể phức tạp hơn các phương pháp khác

Lựa chọn phương pháp Optimizer phù hợp phụ thuộc vào nhiều yếu tố, bao gồm:

- Loại mô hình: Một số phương pháp Optimizer phù hợp hơn với một số loại mô hình nhất định. Ví dụ, Gradient Descent thường được sử dụng cho các mô hình đơn giản, trong khi Adam thường được sử dụng cho các mô hình phức tạp.
- Kích thước tập dữ liệu: Đối với các tập dữ liệu lớn, các phương pháp Optimizer có tốc độ hội tụ nhanh như Adam thường được ưu tiên.
- Thời gian huấn luyện: Nếu bạn có hạn chế về thời gian huấn luyện, bạn có thể cân nhắc sử dụng các phương pháp Optimizer có tốc độ hội tụ nhanh.

### Phương pháp metaheuristics

Phương pháp metaheuristics không sử dụng đạo hàm của hàm mục tiêu và thường là các thuật toán tìm kiếm dựa trên bầy đàn hoặc dựa trên sự tiến hóa. Các phương pháp này có thể hiệu quả trong các trường hợp mà hàm mục tiêu phức tạp hoặc có nhiều cực trị địa phương.

Một số phương pháp metaheuristics phổ biến bao gồm:

- Thuật toán di truyền (genetic algorithm): Phương pháp này mô phỏng quá trình tiến hóa tự nhiên để tìm giải pháp tối ưu.
- Thuật toán tối ưu đàn kiến (ant colony optimization): Phương pháp này mô phỏng hành vi của đàn kiến để tìm đường đi tối ưu.
- Thuật toán tối ưu đàn bướm (bee colony optimization): Phương pháp này mô phỏng hành vi của đàn bướm để tìm giải pháp tối ưu.

Nhìn chung, phương pháp gradient là lựa chọn tốt cho hầu hết các trường hợp. Tuy nhiên, phương pháp metaheuristics có thể là lựa chọn tốt hơn trong các trường hợp sau:

- Hàm mục tiêu phức tạp hoặc có nhiều cực trị địa phương.
- Tập dữ liệu lớn.
- Thời gian và tài nguyên hạn chế.

## **Chương 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.**

### **2.1 Continual Learning**

Continual Learning là một lĩnh vực của học máy tập trung vào việc học các mô hình có thể học hỏi và thích ứng với dữ liệu mới mà không bị mất hiệu suất trên dữ liệu cũ. Điều này là cần thiết trong nhiều ứng dụng thực tế, nơi dữ liệu luôn thay đổi theo thời gian.

Có nhiều phương pháp khác nhau để thực hiện Continual Learning. Một số phương pháp phổ biến bao gồm:

- Entropy regularization: Phương pháp này thêm một thuật toán entropy vào hàm mất mát của mô hình. Điều này giúp mô hình tránh tập trung quá nhiều vào các tập dữ liệu cũ.
- Data augmentation: Phương pháp này tạo ra các dữ liệu mới từ dữ liệu cũ. Điều này giúp mô hình học được các đặc trưng mới mà không cần phải học lại các đặc trưng cũ.
- Incremental learning: Phương pháp này huấn luyện mô hình mới trên dữ liệu mới. Điều này giúp mô hình học được các đặc trưng mới mà không cần phải thay đổi mô hình cũ.

### **2.2 Test Production**

Test Production là một quá trình trong đó một giải pháp học máy được triển khai trên môi trường sản xuất và được theo dõi hiệu suất. Điều này giúp đảm bảo rằng giải pháp học máy đang hoạt động hiệu quả và có thể đáp ứng các yêu cầu của doanh nghiệp.

Có nhiều công cụ và phương pháp khác nhau để thực hiện Test Production. Một số công cụ phổ biến bao gồm:

- Đánh giá hiệu suất: Các phương pháp đánh giá hiệu suất được sử dụng để xác định xem giải pháp học máy đang hoạt động hiệu quả hay không.

- Theo dõi hiệu suất: Các phương pháp theo dõi hiệu suất được sử dụng để giám sát hiệu suất của giải pháp học máy theo thời gian.
- Khắc phục sự cố: Các phương pháp khắc phục sự cố được sử dụng để giải quyết các vấn đề phát sinh với giải pháp học máy.

### Ứng dụng trong xây dựng giải pháp học máy

Continuous learning và test production có thể được áp dụng trong việc xây dựng giải pháp học máy để giải quyết một bài toán nào đó.

Continuous learning có thể giúp cải thiện hiệu suất của mô hình học máy theo thời gian. Điều này là do dữ liệu mới có thể cung cấp thông tin mới về thế giới thực, giúp mô hình học hỏi và cải thiện khả năng dự đoán.

Test production có thể giúp đảm bảo rằng mô hình học máy hoạt động tốt trong môi trường sản xuất. Điều này là do dữ liệu thực tế có thể khác với dữ liệu đào tạo, do đó cần phải kiểm tra hiệu suất của mô hình trong môi trường sản xuất để đảm bảo rằng mô hình vẫn hoạt động tốt.

### Ví dụ

Ví dụ, một công ty sản xuất có thể sử dụng continuous learning để cải thiện hiệu suất của mô hình dự đoán lỗi sản xuất. Công ty có thể thu thập dữ liệu mới về lỗi sản xuất theo thời gian và huấn luyện lại mô hình trên dữ liệu mới. Điều này có thể giúp mô hình học cách dự đoán lỗi sản xuất chính xác hơn.

Hoặc, một công ty tài chính có thể sử dụng test production để đảm bảo rằng mô hình định giá tài sản của họ hoạt động tốt trong môi trường thực tế. Công ty có thể triển khai mô hình trong môi trường sản xuất và thu thập dữ liệu thực tế về giá tài sản. Điều này có thể giúp công ty xác định bất kỳ vấn đề nào với mô hình và thực hiện các điều chỉnh cần thiết.

### Lợi ích

Continuous learning và test production có thể mang lại một số lợi ích cho việc xây dựng giải pháp học máy, bao gồm:

- Cải thiện hiệu suất: Continuous learning có thể giúp cải thiện hiệu suất của mô hình học máy theo thời gian.
- Đảm bảo chất lượng: Test production có thể giúp đảm bảo rằng mô hình học máy hoạt động tốt trong môi trường sản xuất.
- Tăng tính linh hoạt: Continuous learning và test production có thể giúp giải pháp học máy thích ứng với sự thay đổi của dữ liệu và môi trường.

### Thử thách

Continuous learning và test production cũng có một số thách thức, bao gồm:

- Yêu cầu dữ liệu: Continuous learning yêu cầu dữ liệu mới để cập nhật mô hình.
- Yêu cầu kỹ thuật: Continuous learning và test production yêu cầu kiến thức và kỹ năng về học máy và kỹ thuật phần mềm.

### Kết luận

Continuous learning và test production là những kỹ thuật quan trọng có thể được sử dụng để cải thiện hiệu suất và đảm bảo chất lượng của giải pháp học máy. Tuy nhiên, các kỹ thuật này cũng có một số thách thức, do đó cần được cân nhắc kỹ lưỡng trước khi triển khai.

## Chương 3 - TÀI LIỆU THAM KHẢO

### Tiếng Việt

1. <https://www.stdio.vn/ai-ml/bai-toan-phan-lop-trong-machine-learning-classification-in-machine-learning-5150Ih.> {Đã truy cập 10/12/2023 ]
2. <http://hoctructuyen123.net/tong-quan-ve-thuat-toan-phan-lop-naive-bayes-classification-nbc/>. {Đã truy cập 10/12/2023 ]
3. <https://machinelearningcoban.com/2017/08/08/nbc/>. {Đã truy cập 10/12/2023 ]
4. <https://whitehat.vn/threads/thuat-toan-phan-loai-naive-bayes-va-ung-dung.13775/> {Đã truy cập 10/12/2023 ]
5. [https://www.academia.edu/7482756/Decision\\_Tree](https://www.academia.edu/7482756/Decision_Tree) {Đã truy cập 10/12/2023 ]
6. <https://machinelearningcoban.com/2018/01/14/id3/> {Đã truy cập 10/12/2023 ]
7. <https://vietnambiz.vn/thuat-toan-k-lang-gieng-gan-nhat-k-nearest-neighbor-knn-la-gi-2020022911113334.htm> {Đã truy cập 10/12/2023 ]
8. <https://neralnetwork.wordpress.com/2018/05/11/thuat-toan-support-vector-machine-svm/> {Đã truy cập 10/12/2023 ]
9. <https://machinelearningcoban.com/2017/04/09/smv/> {Đã truy cập 10/12/2023 ]
10. <https://towardsdatascience.com/implementing-sgd- from-scratch-d425db18a72c>. {Đã truy cập 10/12/2023 ]
11. [https://d2l.ai/chapter\\_optimization/adam.html](https://d2l.ai/chapter_optimization/adam.html). {Đã truy cập 10/12/2023 ]



