

**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:* **Nguyễn Văn Anh - 51800009**

*Lớp:* **18050202**

*Khoá:* **22**

**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:* **Nguyễn Văn Anh - 51800009**

*Lớp:* **18050202**

*Khoá:* **22**

**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)*

*Nguyễn Văn Anh*

## **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 .....	1
1.2    Các phương pháp Optimizer .....	1
1.3    Kết Luận.....	2
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 .....	5
2.3    Kết luận.....	<b>Error! Bookmark not defined.</b>
Chương 3 - TÀI LIỆU THAM KHẢO .....	<b>Error! Bookmark not defined.</b>

## 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 Tổng quan Optimizer

Optimizer là một thuật toán được sử dụng để tìm các giá trị tối ưu của các tham số trong mô hình học máy. Mục tiêu của optimizer là giảm thiểu hàm mất mát (loss function) của mô hình. Optimizer đóng một vai trò quan trọng trong việc xác định hiệu quả của mô hình, đặc biệt là đối với các mô hình phức tạp.

## 1.2 Các phương pháp Optimizer

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:

**Gradient Descent (GD):** Đây là phương pháp optimizer đơn giản nhất. GD hoạt động bằng cách tính toán đạo hàm của hàm mất mát theo từng tham số và sau đó cập nhật các tham số theo hướng làm giảm hàm mất mát.

**Stochastic gradient descent (SGD):** SGD là một biến thể của GD sử dụng chỉ một mẫu dữ liệu từ tập dữ liệu huấn luyện để cập nhật các giá trị tham số của mô hình trong mỗi bước. SGD có thể hiệu quả hơn GD trong các trường hợp tập dữ liệu lớn. Tuy nhiên, trong một số trường hợp, SGD có thể không phải là lựa chọn tốt nhất. Ví dụ, nếu dữ liệu có nhiễu hoặc nếu mô hình dễ bị quá khớp (overfitting), thì các phương pháp optimizer khác như Momentum hoặc Adam có thể là lựa chọn tốt hơn.

**Momentum:** Momentum là một phương pháp Optimizer nâng cao dựa trên gradient descent. Phương pháp này sử dụng một biến gọi là momentum để tăng tốc độ hội tụ của mô hình, giúp mô hình tránh bị mắc kẹt tại các điểm cực tiểu cục bộ, nhưng có thể làm cho mô hình không ổn định.

**Adagrad:** Adagrad là một phương pháp Optimizer tự động điều chỉnh tốc độ cập nhật các tham số của mô hình dựa trên độ lớn của gradient.

**RMSprop:** RMSprop là một phương pháp Optimizer tương tự như Adagrad, nhưng có độ ổn định cao hơn.

Adam: Adam là một phương pháp Optimizer kết hợp Momentum và RMSprop. Phương pháp này được coi là một trong những phương pháp Optimizer hiệu quả nhất hiện nay.

Phương pháp	Ưu điểm	Nhược điểm
Gradient descent	Đơn giản, dễ hiểu	Không hiệu quả với các mô hình phức tạp
Stochastic gradient descent	Hiệu quả hơn gradient descent	Có thể làm cho mô hình không ổn định
Momentum	Tăng tốc độ hội tụ	Có thể làm cho mô hình không ổn định
Adagrad	Tăng tốc độ hội tụ, đặc biệt là đối với các mô hình có các tham số có giá trị khác nhau	Có thể làm cho mô hình không ổn định
RMSprop	Tăng tốc độ hội tụ, đặc biệt là đối với các mô hình có các tham số có giá trị thay đổi nhanh	Có thể làm chậm quá trình hội tụ
Adam	Hiệu quả với các mô hình phức tạp	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, có thể cân nhắc sử dụng các phương pháp Optimizer có tốc độ hội tụ nhanh.

### 1.3 Kết Luận

Optimizer là một yếu tố quan trọng trong huấn luyện mô hình học máy. Việc lựa chọn phương pháp Optimizer phù hợp có thể giúp mô hình học máy đạt được hiệu suất tốt hơn.

Một số mẹo sử dụng optimizer

Dưới đây là một số mẹo sử dụng optimizer hiệu quả:

- Khởi tạo tham số: Khởi tạo tham số của mô hình một cách hợp lý có thể giúp cải thiện hiệu quả của optimizer.
- Chọn tham số: Các phương pháp optimizer thường có một số tham số cần được điều chỉnh. Việc điều chỉnh các tham số này có thể giúp cải thiện hiệu quả của optimizer.
- Theo dõi kết quả: Theo dõi kết quả huấn luyện của mô hình có thể giúp bạn xác định xem optimizer đang hoạt động hiệu quả hay không. Nếu optimizer không hoạt động hiệu quả, bạn có thể cần thử một phương pháp optimizer khác hoặc điều chỉnh các tham số của optimizer.

Trong các trường hợp tập dữ liệu lớn, SGD hoặc Adam là những lựa chọn tốt. Trong các trường hợp tập dữ liệu nhỏ, GD có thể là lựa chọn tốt hơn. Trong các trường hợp hàm mất mát có nhiều điểm dừng, Adagrad hoặc RMSprop có thể không hiệu quả.

## **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 cách khác nhau để áp dụng continuous learning. Một cách phổ biến là sử dụng hệ thống giám sát dữ liệu. Hệ thống giám sát dữ liệu có thể được sử dụng để thu thập dữ liệu mới và gửi dữ liệu này đến mô hình để cập nhật.

Một cách khác để áp dụng continuous learning là sử dụng các kỹ thuật học tự động. Các kỹ thuật học tự động có thể được sử dụng để tự động cập nhật mô hình dựa trên dữ liệu mới.

Bài toán dự đoán giá cổ phiếu

Trong lĩnh vực tài chính, các nhà giao dịch có thể sử dụng continuous learning để cập nhật mô hình dự đoán giá cổ phiếu. Mô hình có thể được cập nhật dựa trên dữ liệu mới, chẳng hạn như dữ liệu giá cổ phiếu, dữ liệu kinh tế và dữ liệu tin tức.

Ví dụ, một mô hình dự đoán giá cổ phiếu có thể được huấn luyện trên tập dữ liệu gồm giá cổ phiếu trong quá khứ, dữ liệu kinh tế và dữ liệu tin tức. Sau khi mô hình được huấn luyện, nó có thể được sử dụng để dự đoán giá cổ phiếu trong tương lai.

Tuy nhiên, dữ liệu thực tế luôn thay đổi. Giá cổ phiếu có thể biến động do nhiều yếu tố, chẳng hạn như tình hình kinh tế, tin tức và các sự kiện bất ngờ. Do đó,

mô hình dự đoán giá cổ phiếu cần được cập nhật thường xuyên để đảm bảo độ chính xác.

Continuous learning có thể được sử dụng để cập nhật mô hình dự đoán giá cổ phiếu. Hệ thống giám sát dữ liệu có thể được sử dụng để thu thập dữ liệu mới, chẳng hạn như giá cổ phiếu hiện tại, dữ liệu kinh tế mới nhất và dữ liệu tin tức mới nhất. Dữ liệu mới này có thể được sử dụng để cập nhật mô hình dự đoán giá cổ phiếu.

Việc cập nhật mô hình dự đoán giá cổ phiếu thường xuyên có thể giúp cải thiện độ chính xác của mô hình. Điều này có thể giúp các nhà giao dịch đưa ra quyết định giao dịch tốt hơn.

Continuous learning là một kỹ thuật quan trọng cần được áp dụng khi xây dựng một giải pháp học máy để giải quyết một bài toán nào đó. Continuous learning giúp mô hình học hỏi và thích ứng với những thay đổi của dữ liệu thực tế, từ đó cải thiện độ chính xác của mô hình.

## 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.

Lợi ích của Continual Learning và Test Production

Continual Learning và Test Production mang lại nhiều lợi ích cho các giải pháp học máy, bao gồm:

- Hiệu suất cao hơn: Continual Learning giúp các giải pháp học máy luôn học hỏi và thích ứng với dữ liệu mới, dẫn đến hiệu suất cao hơn.
- Tính ổn định cao hơn: Test Production giúp đảm bảo rằng các giải pháp học máy đang hoạt động hiệu quả và có thể đáp ứng các yêu cầu của doanh nghiệp.
- Tính linh hoạt cao hơn: Continual Learning giúp các giải pháp học máy có thể thích ứng với các thay đổi của môi trường.

## Tài Liệu Tham Khảo

### Tiếng Việt

1. <https://www.stdio.vn/ai-ml/bai-toan-phan-lop-trong-machine-learning-classification-in-machine-learning-515OIh.> {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 ]



