TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**DỰ ÁN CUỐI KỲ MÔN NHẬP MÔN HỌC MÁY**

**TÌM HIỂU LÝ THUYẾT**

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

*Người thực hiện*: **AO THỤY NGỌC TRÂN - 52100937**

Lớp **: 21050301**

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**



**DỰ ÁN CUỐI KỲ MÔN NHẬP MÔN HỌC MÁY**

**TÌM HIỂU LÝ THUYẾT**

Người hướng dẫn: **TS LÊ ANH CƯỜNG**

Người thực hiện: **AO THỤY NGỌC TRÂN**

Lớp **: 21050301**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời cảm ơn chân thành đến các thầy cô giáo đã dành thời gian và công sức để chỉ dẫn chúng em. Đặc biệt là thầy Lê Anh Cường, giảng viên trực tiếp phụ trách giảng dạy bộ môn Nhập môn học máy. Chúng em đã nhận được rất nhiều sự quan tâm, giúp đỡ, hướng dẫn tận tình và đầy tâm huyết của thầy. Thầy đã giúp chúng em tích lũy thêm nhiều kiến thức bổ ích để có cái nhìn tổng quan và rõ ràng hơn về môn học. Từ những kiến thức mà thầy truyền tải, chúng em đã dần trả lời được rất nhiều dạng bài tập và câu hỏi mới, cũng như được thực hành thêm nhiều dạng bài thú vị. Và cũng nhờ đó góp phần giúp đỡ chúng em có nền tảng kiến thức vững vàng hơn, đa dạng hơn, cũng như tạo nên sức bật cho chúng em có thêm niềm vui và động lực để theo đuổi ước mơ của bản thân.

Kiến thức là vô cùng, vô tận, không có hạn định và sự tiếp nhận kiến thức của bản thân mỗi người luôn tồn tại những hạn chế nhất định. Do đó, trong quá trình hoàn thành bài báo cáo, chắc chắn sẽ không thể tránh khỏi những thiếu sót. Bản thân chúng em rất mong nhận những lời góp ý đến từ thầy để có thể sữa chữa, phát triển và học hỏi thêm càng nhiều từ những sai sót của bản thân và từ đó góp phần làm cho bài báo của chính mình càng trở nên toàn diện và hoàn thiện hơn nữa.

Lời cuối cùng, chúng em xin kính chúc thầy sức khỏe, hạnh phúc và càng ngày càng thành công trên con đường truyền tải tri thức cao quý.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi và được sự hướng dẫn của TS 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 21 tháng 12 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Trân*

*Ao Thụy Ngọc Trân*

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

Bài báo cáo trình bày về kết quả tìm hiểu được về các vấn đề:

* Tìm hiểu, so sánh các phương pháp Optimizer trong huấn luyện mô hình học máy;
* Tìm hiểu về Continual Learning và Test Production khi xây dựng một giải pháp học máy để giải quyết một bài toán

Bài báo cáo có 3 chương, trong đó:

* Chương 1: Phương pháp Optimizer
* Chương 2: Continual Learning
* Chương 3: Test Production

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc10312)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc3822)

[TÓM TẮT iv](#_Toc30008)

[MỤC LỤC 1](#_Toc13465)

[CHƯƠNG 1 - PHƯƠNG PHÁP OPTIMIZER 2](#_Toc7405)

[1.1 Optimizer là gì? 2](#_Toc14587)

[1.2 Các thuật toán tối ưu của Optimizer 2](#_Toc19555)

[1.2.1 Stochastic Gradient Descent (SGD) 2](#_Toc26467)

[1.2.2 Gradient Descent 3](#_Toc13032)

[1.2.3 Adagrad 5](#_Toc26173)

[1.2.4 RMSprop 5](#_Toc2115)

[1.2.5 Adam 6](#_Toc13585)

[1.2.6 Momentum 7](#_Toc24884)

[1.2.7 AMSGrad 8](#_Toc17302)

[CHƯƠNG 2 - CONTINUAL LEARNING 9](#_Toc11459)

[2.1 Đặc Điểm Cơ Bản của Continual Learning 9](#_Toc12848)

[2.2 Thách Thức Lớn 9](#_Toc18747)

[2.3 Phương Pháp Đối Phó với Thách Thức 9](#_Toc3990)

[2.4 Ứng Dụng của Continual Learning: 9](#_Toc21953)

[2.5 Thách Thức Đối Với Đạo Đức 10](#_Toc17525)

[2.6 Nghiên Cứu Liên Quan 10](#_Toc19947)

[2.7 Tương Lai và Hướng Phát Triển 10](#_Toc9539)

[CHƯƠNG 3 - TEST PRODUCTION 11](#_Toc9922)

[3.1 Ý Nghĩa Quan Trọng của Kiểm Thử Sản Xuất 11](#_Toc9124)

[3.2 Quy Trình Kiểm Thử Sản Xuất 11](#_Toc8461)

[3.3 Công Cụ và Phương Pháp Kiểm Thử 12](#_Toc9237)

[3.4 Ứng Dụng Thực Tế và Trong Quy Trình Xây Dựng Giải Pháp Học Máy: 12](#_Toc25316)

[3.5 Thách Thức và Giải Pháp Đối Với Kiểm Thử Sản Xuất 12](#_Toc25008)

[3.6 Tương Lai và Hướng Phát Triển 12](#_Toc15058)

[TỔNG KẾT 14](#_Toc22853)

CHƯƠNG 1 - PHƯƠNG PHÁP OPTIMIZER

* 1. Optimizer là gì?

Optimizer là một khái niệm trong lĩnh vực machine learning, đặc biệt là trong quá trình huấn luyện mô hình, để tối ưu hóa hàm mất mát và điều chỉnh các trọng số của mô hình. Nhiệm vụ chính của optimizer là tìm ra các giá trị trọng số cho mô hình sao cho hàm mất mát đạt giá trị nhỏ nhất, làm cho mô hình đạt được hiệu suất tốt nhất trên dữ liệu huấn luyện.

Optimizer thường xuyên làm việc bằng cách sử dụng độ dốc (gradient) của hàm mất mát đối với các trọng số và điều chỉnh chúng dựa trên các thuật toán tối ưu hóa. Các phương pháp optimizer thường sử dụng các siêu tham số như learning rate để điều chỉnh tốc độ học của mô hình.

Các optimizer phổ biến bao gồm Stochastic Gradient Descent (SGD) và các biến thể của nó như Momentum, Adagrad, RMSprop, Adam, và nhiều phương pháp khác. Mỗi optimizer có những đặc điểm riêng biệt và hiệu suất tốt trên một số loại mô hình và dữ liệu cụ thể. Sự lựa chọn của optimizer có thể ảnh hưởng đáng kể đến khả năng học và hiệu suất của mô hình trong quá trình huấn luyện.

* 1. Các thuật toán tối ưu của Optimizer
     1. Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) là một trong những phương pháp tối ưu hóa quan trọng trong quá trình huấn luyện mô hình học máy. SGD là một biến thể của thuật toán Gradient Descent (GD), được thiết kế để xử lý hiệu quả trên dữ liệu lớn và có thể có chiều cao số chiều cao. Dưới đây là một chi tiết hơn về SGD, bao gồm cách nó hoạt động, các ưu và nhược điểm, cũng như cách nó liên quan đến các biến thể khác của thuật toán tối ưu hóa.

1. Cách Hoạt Động của Stochastic Gradient Descent (SGD):

SGD hoạt động bằng cách cập nhật trọng số của mô hình dựa trên một lượng nhỏ mẫu (mini-batch) ngẫu nhiên từ tập dữ liệu huấn luyện thay vì toàn bộ dữ liệu. Quá trình này giúp giảm độ phức tạp tính toán và cho phép mô hình được cập nhật thường xuyên hơn, đặc biệt là khi làm việc với dữ liệu lớn.

1. Ưu Điểm của SGD:

* Hiệu Quả Trên Dữ Liệu Lớn: SGD hoạt động hiệu quả hơn trên dữ liệu lớn do sử dụng mini-batch thay vì toàn bộ dữ liệu.
* Cập Nhật Thường Xuyên: Tần suất cập nhật tham số giúp mô hình học nhanh chóng và thích ứng tốt với dữ liệu mới.

1. Nhược Điểm của SGD:

* Nhiễu: Vì sử dụng mini-batch ngẫu nhiên, SGD có thể mang lại nhiễu trong việc cập nhật trọng số, đặc biệt là khi dữ liệu là nhiễu.
* Hội Tụ Không Ổn Định: SGD có thể hội tụ không ổn định tùy thuộc vào lựa chọn learning rate và khởi tạo trọng số ban đầu.

1. Mối Liên Quan với Các Phương Pháp Tối Ưu Hóa Khác:

* Batch Gradient Descent (GD): Là biến thể của SGD với mini-batch có kích thước bằng toàn bộ dữ liệu.
* Mini-Batch Gradient Descent: Một sự kết hợp giữa SGD và GD, sử dụng mini-batch có kích thước trung bình.
* Learning Rate Schedules: SGD thường kết hợp với các lịch trình learning rate để điều chỉnh tốt hơn.

*SGD là một phương pháp quan trọng trong quá trình huấn luyện mô hình học máy, và hiểu rõ về cách nó hoạt động cũng như ưu và nhược điểm sẽ giúp các nhà nghiên cứu và người làm machine learning đưa ra quyết định thông tin khi triển khai mô hình của mình.*

* + 1. Gradient Descent

Gradient Descent là một thuật toán tối ưu hóa phổ biến trong machine learning, được sử dụng để điều chỉnh các tham số của mô hình dựa trên độ dốc của hàm mất mát. Mục tiêu chính của Gradient Descent là tìm ra giá trị nhỏ nhất của hàm mất mát bằng cách di chuyển từ điểm khởi đầu theo hướng âm của độ dốc.

Dưới đây là một chi tiết về Gradient Descent:

1. Batch Gradient Descent:

* Mô Tả: Sử dụng toàn bộ tập dữ liệu huấn luyện để tính độ dốc và cập nhật trọng số.
* Ưu Điểm: Chắc chắn hội tụ đến điểm tối ưu toàn cục.
* Nhược Điểm: Đòi hỏi nhiều tài nguyên tính toán, chậm với dữ liệu lớn.

2. Stochastic Gradient Descent (SGD):

* Mô Tả: Sử dụng một mẫu ngẫu nhiên từ dữ liệu để tính độ dốc và cập nhật trọng số.
* Ưu Điểm: Cập nhật thường xuyên, phù hợp với dữ liệu lớn.
* Nhược Điểm: Nhiễu, hội tụ không ổn định.

3. Mini-Batch Gradient Descent:

* Mô Tả: Sử dụng một mini-batch (đoạn nhỏ) từ dữ liệu để tính độ dốc và cập nhật trọng số.
* Ưu Điểm: Kết hợp ưu điểm của cả Batch GD và SGD.
* Nhược Điểm: Cần lựa chọn kích thước mini-batch phù hợp.

4. Learning Rate:

* Mô Tả: Là hệ số quyết định kích thước của bước cập nhật trọng số.
* Ưu Điểm: Điều chỉnh tốc độ học của mô hình.
* Nhược Điểm: Learning rate không phù hợp có thể ảnh hưởng đến hiệu suất học.

5. Momentum:

* Mô Tả: Cải thiện SGD bằng cách tích tụ một lượng động để giảm độ dao động và tăng tốc quá trình học.
* Ưu Điểm: Giảm độ dao động, hội tụ nhanh hơn.
* Nhược Điểm: Cần điều chỉnh thêm một siêu tham số.

6. Convergence Criteria:

* Mô Tả: Điều kiện dừng quyết định khi nào dừng quá trình tối ưu hóa.
* Ưu Điểm: Ngăn chặn việc tiêu thụ tài nguyên tính toán khi đã đạt được độ chính xác mong muốn.
* Nhược Điểm: Cần lựa chọn đúng để tránh dừng quá sớm hoặc quá muộn.

7. Local Minimum vs. Global Minimum:

* Mô Tả: Thách thức của Gradient Descent là có thể hội tụ tới cực tiểu cục thay vì cực tiểu toàn cục.
* Ưu Điểm: Sử dụng các kỹ thuật như learning rate schedules để giúp tránh cực tiểu cục.
* Nhược Điểm: Không có đảm bảo hội tụ tới cực tiểu toàn cục.

*Gradient Descent là một công cụ quan trọng cho việc huấn luyện mô hình học máy và deep learning, và sự hiểu biết về cách nó hoạt động và các biến thể của nó là quan trọng để đạt được hiệu suất tốt nhất từ mô hình.*

* + 1. Adagrad

Adagrad là một phương pháp tối ưu hóa trong machine learning, đặc biệt được thiết kế để điều chỉnh learning rate của mỗi tham số của mô hình dựa trên lịch sử của độ dốc. Phương pháp này giúp ứng phó tốt với các tham số có độ dốc lớn và nhỏ, tự động điều chỉnh learning rate tương ứng với từng tham số.

Dưới đây là chi tiết về Adagrad:

1. Cách Hoạt Động:

Adagrad sử dụng learning rate khác nhau cho mỗi tham số của mô hình tùy thuộc vào độ dốc của tham số đó.

2. Ưu Điểm:

* Tự Điều Chỉnh Learning Rate: Adagrad tự động điều chỉnh learning rate cho từng tham số dựa trên lịch sử của độ dốc. Các tham số có độ dốc lớn sẽ có learning rate giảm mạnh hơn, và ngược lại.
* Phù Hợp cho Dữ Liệu Thưa: Phương pháp này đặc biệt hiệu quả khi làm việc với dữ liệu thưa, nơi một số đặc trưng có độ quan trọng cao hơn.

3. Nhược Điểm:

Accumulation of Squared Gradients: Adagrad tích tụ bình phương của độ dốc qua thời gian, có thể dẫn đến việc learning rate giảm quá nhanh. Điều này có thể làm cho learning rate trở nên rất nhỏ, đặc biệt là với các tham số đã được cập nhật nhiều lần.

4. Lưu Ý:

RMSprop: Để giải quyết vấn đề giảm quá nhanh của learning rate trong Adagrad, phương pháp RMSprop đã được đề xuất. RMSprop thay thế tổng tích lũy bình phương độ dốc bằng một phần độ dốc trung bình động để làm giảm giá trị tích lũy theo thời gian.

*Adagrad là một công cụ quan trọng trong hội tụ của thuật toán tối ưu hóa và thường được sử dụng trong các bài toán machine learning khi cần tự động điều chỉnh learning rate cho từng tham số. Tuy nhiên, cũng cần lưu ý đến nhược điểm của nó và xem xét các biến thể như RMSprop để cải thiện hiệu suất.*

* + 1. RMSprop

RMSprop (Root Mean Square Propagation) là một phương pháp tối ưu hóa được thiết kế để giải quyết vấn đề của learning rate giảm quá nhanh trong phương pháp Adagrad. RMSprop giữ lại ưu điểm của Adagrad như tự động điều chỉnh learning rate cho từng tham số, nhưng tránh tình trạng learning rate giảm quá nhanh bằng cách sử dụng trung bình trọng số bình phương của độ dốc.

Dưới đây là chi tiết về lý thuyết của RMSprop:

1. Cách Hoạt Động:

Giống như Adagrad, RMSprop sử dụng learning rate khác nhau cho từng tham số của mô hình dựa trên lịch sử của độ dốc.

2. Ưu Điểm:

* Điều Chỉnh Learning Rate Tự Động: RMSprop tự động điều chỉnh learning rate cho từng tham số dựa trên trung bình trọng số bình phương của độ dốc.
* Phù Hợp cho Dữ Liệu Thưa: Như Adagrad, RMSprop thích hợp với việc làm việc với dữ liệu thưa.

3. Nhược Điểm:

Cần Lựa Chọn Siêu Tham Số: Siêu tham số �β cần được lựa chọn sao cho phù hợp với dữ liệu và mô hình cụ thể.

4. Lưu Ý:

Adadelta: Adadelta là một biến thể của RMSprop giảm sự phụ thuộc vào learning rate toàn cục bằng cách thay thế nó bằng một trung bình trọng số bình phương độ dốc động.

*RMSprop là một trong những phương pháp tối ưu hóa phổ biến trong deep learning, đặc biệt là khi làm việc với các mô hình sâu và dữ liệu lớn. Tuy nhiên, như mọi phương pháp tối ưu hóa, cần lưu ý đến việc lựa chọn siêu tham số để đảm bảo hiệu suất tốt.*

* + 1. Adam

Adam (Adaptive Moment Estimation) là một phương pháp tối ưu hóa kết hợp các ý tưởng từ RMSprop và Momentum. Adam giữ lại ưu điểm của cả hai phương pháp này, tự động điều chỉnh learning rate cho từng tham số và tích tụ động để tăng tốc quá trình học. Dưới đây là chi tiết về lý thuyết của Adam:

1. Cách Hoạt Động:

Adam tính toán hai trung bình động:

* Trung bình động bình phương của độ dốc (squared gradient):



* Trung bình động của độ dốc (gradient):



Sau đó, Adam tính toán các giá trị được sử dụng để cập nhật trọng số:

* Điều chỉnh trung bình động bình phương:



* Điều chỉnh trung bình động:

​

* Cập nhật trọng số:



2. Ưu Điểm:

* Tích Hợp Momentum và RMSprop: Adam kết hợp cả Momentum (để tăng tốc quá trình học) và RMSprop (để điều chỉnh learning rate cho từng tham số).
* Phù Hợp cho Dữ Liệu và Mô Hình Đa Dạng: Adam thường hoạt động hiệu quả trên nhiều loại dữ liệu và cấu trúc mô hình.

3. Nhược Điểm: Phụ Thuộc vào Siêu Tham Số

*Adam là một trong những phương pháp tối ưu hóa được sử dụng rộng rãi trong deep learning và machine learning. Sự linh hoạt và hiệu suất của nó làm cho nó trở thành một lựa chọn phổ biến khi huấn luyện mô hình đa dạng. Tuy nhiên, cũng như các phương pháp tối ưu hóa khác, việc lựa chọn siêu tham số là một phần quan trọng để đạt được hiệu suất tốt.*

* + 1. Momentum

Momentum là một trong những phương pháp tối ưu hóa phổ biến trong machine learning, được thiết kế để giảm độ dao động và tăng tốc quá trình học của mô hình. Ý tưởng cơ bản của Momentum là tích tụ một lượng động dựa trên độ dốc của các bước trước đó, giúp mô hình vượt qua các "bumps" và "dips" trên đường đào tạo. Dưới đây là một cái nhìn chi tiết về lý thuyết của Momentum:

1. Ưu Điểm:

Giảm Độ Dao Động: Momentum giúp giảm độ dao động của quá trình học, đặc biệt là trong trường hợp địa hình có nhiều độ dốc và cong vòng.

1. Nhược Điểm:

Cần Lựa Chọn Siêu Tham Số: Việc lựa chọn hệ số momentum cần được thực hiện cẩn thận để phù hợp với bài toán cụ thể.

* + 1. AMSGrad

AMSGrad là một biến thể của phương pháp Adam, được thiết kế để giải quyết vấn đề của Adam khi không đảm bảo sự giảm của learning rate theo thời gian. AMSGrad giữ lại cấu trúc chính của Adam, nhưng thay đổi cách cập nhật một trong các trung bình động. Dưới đây là chi tiết về lý thuyết của AMSGrad:

1. Vấn Đề Của Adam:

Adam có thể không đảm bảo sự giảm của learning rate theo thời gian do sự tăng lên không ngừng của trung bình động bình phương của độ dốc.

1. Ưu Điểm:

Giữ Nguyên Learning Rate: AMSGrad giữ nguyên learning rate theo thời gian hơn so với Adam.

1. Nhược Điểm:

Không Phù Hợp cho Mọi Trường Hợp: AMSGrad có thể không hoạt động tốt trong mọi trường hợp và phụ thuộc nhiều vào đặc điểm cụ thể của dữ liệu.

CHƯƠNG 2 - CONTINUAL LEARNING

Continual Learning (CL), hay còn được biết đến với tên gọi khác như Lifelong Learning hoặc Incremental Learning, là một lĩnh vực quan trọng trong machine learning mà giải quyết thách thức của việc học từ dữ liệu mới liên tục mà không làm mất đi kiến thức đã học từ trước. Đây không chỉ là một khía cạnh quan trọng của học máy mà còn đặt ra những thách thức độc đáo. Dưới đây là một cái nhìn tổng quan và lý thuyết sâu sắc về Continual Learning

2.1 Đặc Điểm Cơ Bản của Continual Learning

Liên Tục và Tiếp Tục: Trong bối cảnh của học máy truyền thống, mô hình thường được đào tạo và đánh giá trên một tập dữ liệu cố định. Ngược lại, Continual Learning đặt ra yêu cầu học liên tục và tiếp tục từ dữ liệu mới khi nó xuất hiện.

2.2 Thách Thức Lớn

Quên Đi (Catastrophic Forgetting): Một trong những thách thức lớn nhất của Continual Learning là đối mặt với hiện tượng quên đi. Khi mô hình học từ dữ liệu mới, có thể xảy ra việc mất kiến thức đã học từ dữ liệu trước đó.

2.3 Phương Pháp Đối Phó với Thách Thức

Replay: Một trong những phương pháp chính để đối phó với thách thức quên là sử dụng Replay. Điều này đòi hỏi lưu giữ và sử dụng lại mẫu dữ liệu từ quá khứ để mô hình có thể giữ lại kiến thức cũ trong quá trình học từ dữ liệu mới.

Regularization Techniques: Các kỹ thuật chống quên như Elastic Weight Consolidation (EWC) và Learning without Forgetting (LwF) giúp giảm thiểu ảnh hưởng của việc học mới đối với kiến thức cũ. EWC tập trung vào việc giữ lại trọng số quan trọng cho các nhiệm vụ trước đó, trong khi LwF cố gắng đồng thời tối thiểu hóa độ chệch giữa các đầu ra của mô hình trước và sau khi học từ dữ liệu mới.

Architecture Design: Thiết kế kiến trúc mô hình để hỗ trợ Continual Learning là một phương pháp quan trọng. Mạng nơ-ron có khả năng mở rộng, tức là có thể thêm các tầng mới khi cần thiết, có thể giúp mô hình mở rộng để học từ dữ liệu mới mà không ảnh hưởng nhiều đến kiến thức cũ.

2.4 Ứng Dụng của Continual Learning:

Học Máy trong Robotics: Trong ngữ cảnh của robotics, Continual Learning là quan trọng để robot có thể liên tục học từ môi trường mới và thích ứng với các tình huống mới mà không cần phải được đào tạo lại từ đầu.

Dịch Ngôn Ngữ Liên Tục: Trong lĩnh vực dịch ngôn ngữ, mô hình có thể được cập nhật để xử lý các từ ngữ mới và xu hướng ngôn ngữ mới theo thời gian.

Ứng Dụng Trong Thiết Bị Di Động: Continual Learning có thể được áp dụng để mô hình trên các thiết bị di động có thể liên tục học từ các tương tác mới và cải thiện trải nghiệm người dùng.

2.5 Thách Thức Đối Với Đạo Đức

Quyền Riêng Tư: Vấn đề về quyền riêng tư là một trong những thách thức đạo đức quan trọng khi mô hình liên tục học từ dữ liệu người dùng mới.

2.6 Nghiên Cứu Liên Quan

Meta-Learning: Nghiên cứu về Meta-Learning, nơi mô hình học cách học để có khả năng nhanh chóng thích ứng với nhiều nhiệm vụ, cũng liên quan chặt chẽ với Continual Learning.

2.7 Tương Lai và Hướng Phát Triển

Kết Hợp Continual Learning với Transfer Learning: Kết hợp Continual Learning với Transfer Learning có thể là hướng đi để tận dụng thông tin từ các nhiệm vụ trước đó và giảm thiểu thời gian đào tạo.

Mở Rộng Đối Với Học Máy Đa Nhiệm: Mở rộng Continual Learning để xem xét cả các mô hình học máy đa nhiệm, nơi có nhiều nhiệm vụ cần thực hiện đồng thời.

Trong khi Continual Learning mở ra cơ hội cho việc áp dụng machine learning trong các bối cảnh đa dạng và thay đổi, thì đối mặt với thách thức lớn về việc quản lý kiến thức và thích ứng liên tục với môi trường mới. Sự tiếp tục nghiên cứu trong lĩnh vực này sẽ giúp tạo ra các giải pháp học máy linh hoạt và có khả năng thích ứng cao hơn trong thế giới thực.

CHƯƠNG 3 - TEST PRODUCTION

Kiểm thử sản xuất (Test Production) đóng vai trò quan trọng trong quy trình xây dựng giải pháp học máy, là giai đoạn cuối cùng và quyết định đối với việc triển khai mô hình vào môi trường thực tế. Trong phần này, chúng ta sẽ tìm hiểu chi tiết và mở rộng về ý nghĩa, quy trình, công cụ, và ứng dụng của kiểm thử sản xuất để đảm bảo mô hình là ổn định, hiệu suất cao, và an toàn khi hoạt động trong môi trường sản xuất đa dạng và thay đổi.

3.1 Ý Nghĩa Quan Trọng của Kiểm Thử Sản Xuất

Kiểm thử sản xuất không chỉ là bước kiểm thử cuối cùng mà còn là bước quyết định cho sự thành công của mô hình khi đối mặt với thách thức của môi trường sản xuất. Những điểm chính về ý nghĩa của kiểm thử sản xuất bao gồm:

* Đảm Bảo Độ Tin Cậy: Kiểm thử sản xuất được thực hiện để đảm bảo rằng mô hình hoạt động đúng đắn và tin cậy trong mọi tình huống có thể xảy ra trong môi trường sản xuất thực tế.
* Đánh Giá Hiệu Suất: Kiểm thử hiệu suất được thực hiện để đánh giá hiệu suất của mô hình dưới nhiều điều kiện và đảm bảo rằng nó đáp ứng mong đợi trong mọi tình huống.
* Kiểm Thử Tính Bảo Mật: Đảm bảo rằng mô hình là an toàn và có khả năng chống lại các loại tấn công trong môi trường sản xuất, đặc biệt là khi đối mặt với dữ liệu thực tế.

3.2 Quy Trình Kiểm Thử Sản Xuất

Quy trình kiểm thử sản xuất bao gồm một loạt các bước cụ thể để đảm bảo mô hình đã được chuẩn bị và đánh giá một cách toàn diện. Những bước quan trọng trong quy trình này bao gồm:

* Chuẩn Bị Dữ Liệu: Dữ liệu đầu vào cho kiểm thử phải phản ánh đầy đủ các điều kiện thực tế và đa dạng đủ để kiểm tra tất cả các khả năng của mô hình.
* Triển Khai Mô Hình: Mô hình được đặt vào môi trường sản xuất, và quá trình này được kiểm soát để đảm bảo rằng mô hình được tích hợp một cách chính xác và hoàn toàn.
* Kiểm Thử Tích Hợp: Kiểm thử tích hợp được thực hiện để đảm bảo rằng mô hình tương tác chính xác với các phần mềm và hệ thống khác trong môi trường sản xuất.
* Kiểm Thử Hiệu Suất: Đánh giá hiệu suất của mô hình dưới nhiều tình huống, bao gồm cả những điều kiện đặc biệt và tải cao.
* Kiểm Thử Tính Bảo Mật: Đảm bảo rằng mô hình có khả năng chống lại các loại tấn công thông dụng và rằng dữ liệu nhạy cảm được bảo vệ đúng cách.
* Kiểm Thử Sự Linh Hoạt: Kiểm thử sự linh hoạt của mô hình, đảm bảo rằng nó có khả năng thích ứng với sự thay đổi trong môi trường và dữ liệu mới

3.3 Công Cụ và Phương Pháp Kiểm Thử

* Automation Testing: Sử dụng các công cụ tự động hóa kiểm thử để giảm thiểu sai sót con người và đảm bảo tính liên tục của quy trình kiểm thử.
* A/B Testing: Áp dụng phương pháp A/B Testing để so sánh hiệu suất giữa các phiên bản của mô hình trong môi trường thực tế. Điều này giúp xác định phiên bản nào hoạt động tốt hơn dựa trên dữ liệu thực tế.

3.4 Ứng Dụng Thực Tế và Trong Quy Trình Xây Dựng Giải Pháp Học Máy:

* Đảm Bảo Độ Ổn Định và Tính Tin Cậy: Kiểm thử sản xuất giúp đảm bảo rằng mô hình là ổn định và tin cậy trong điều kiện thực tế, không gặp vấn đề lớn khi triển khai.
* Cập Nhật Liên Tục: Test Production có thể liên tục được áp dụng để đảm bảo rằng mô hình không gặp phải những vấn đề lớn sau khi triển khai và có khả năng thích ứng với sự thay đổi trong môi trường.

3.5 Thách Thức và Giải Pháp Đối Với Kiểm Thử Sản Xuất

* Thách Thức: Môi trường thực tế thường phức tạp và đa dạng, khó khăn trong việc đặc trưng và mô phỏng đầy đủ.
* Giải Pháp: Sử dụng dữ liệu thực tế và các kịch bản kiểm thử phức tạp để đảm bảo rằng mô hình sẽ phản ứng đúng đắn trong mọi tình huống.

3.6 Tương Lai và Hướng Phát Triển

* Kiểm Thử Học Máy Tự Động: Phát triển các phương pháp và công cụ kiểm thử tự động chuyên sâu dành cho mô hình học máy để giảm độ phức tạp của quy trình kiểm thử.
* Kiểm Thử Tự Động Động Hóa: Kết hợp kiểm thử sản xuất với các hệ thống tự động động hóa để có thể phản ứng nhanh chóng và hiệu quả với sự thay đổi trong môi trường.

Kiểm thử sản xuất không chỉ là một bước cuối cùng quan trọng mà còn là một phần không thể thiếu để đảm bảo tính ổn định và hiệu suất của mô hình học máy trong môi trường thực tế. Sự kết hợp giữa kiểm thử sản xuất và các bước khác trong quy trình xây dựng giải pháp học máy đảm bảo rằng mô hình không chỉ đạt được kết quả tốt trong điều kiện kiểm thử mà còn có khả năng hoạt động hiệu quả trong môi trường sản xuất đa dạng và thay đổi.

TỔNG KẾT

Trong khi khám phá về các phương pháp Optimizer trong huấn luyện mô hình học máy và hai khía cạnh quan trọng là Continual Learning và Test Production, chúng ta đã đi sâu vào nhiều khái niệm, lý thuyết, và ứng dụng quan trọng. Dưới đây là sự tổng kết của chúng ta:

**1. Phương Pháp Optimizer:**

Các phương pháp như Stochastic Gradient Descent (SGD), Gradient Descent, Adagrad, RMSprop, Adam, Momentum, và AMSGrad cung cấp các kỹ thuật tinh tế để cập nhật trọng số của mô hình trong quá trình huấn luyện.

Mỗi phương pháp có những ưu điểm và nhược điểm riêng, phụ thuộc vào đặc tính của dữ liệu và đặc điểm của mô hình.

**2. Continual Learning:**

Continual Learning giải quyết thách thức của việc học từ dữ liệu mới mà không làm mất đi kiến thức cũ.

Phương pháp như Replay, Regularization Techniques, và Architecture Design đóng vai trò quan trọng trong việc duy trì và cập nhật kiến thức của mô hình liên tục.

**3. Test Production trong Giải Pháp Học Máy:**

Kiểm thử sản xuất là bước quyết định cuối cùng để đảm bảo mô hình có thể hoạt động hiệu quả trong môi trường thực tế.

Quy trình kiểm thử bao gồm chuẩn bị dữ liệu, triển khai mô hình, kiểm thử tích hợp, kiểm thử hiệu suất, kiểm thử tính bảo mật, và kiểm thử sự linh hoạt.

**4. Ứng Dụng và Tương Lai:**

Cả hai lĩnh vực, Continual Learning và Test Production, đều có ứng dụng rộng rãi trong các ngành công nghiệp như robotics, dịch ngôn ngữ, thiết bị di động, và nhiều lĩnh vực khác.

Tương lai của chúng phụ thuộc vào sự kết hợp linh hoạt và thông minh để giải quyết những thách thức ngày càng phức tạp.

**5. Hệ Quả và Khả Năng Phát Triển:**

Sự hiểu biết sâu sắc về các phương pháp optimizer, kỹ thuật continual learning, và quy trình kiểm thử sản xuất là quan trọng để xây dựng những mô hình học máy linh hoạt và có tính ổn định.

Sự kết hợp thông minh giữa các lĩnh vực này có thể đưa ra những phương pháp mới và hiệu quả hơn.

Nhìn chung, sự liên kết giữa quá trình huấn luyện mô hình, khả năng liên tục học, và khả năng thích ứng với môi trường thực tế là chìa khóa để xây dựng những giải pháp học máy mạnh mẽ và có sức mạnh linh hoạt. Sự tiếp tục nghiên cứu và đổi mới trong các lĩnh vực này sẽ không chỉ định hình tương lai của học máy mà còn giúp chúng ta đối mặt với những thách thức lớn trong thế giới số đang ngày càng phát triển.

**TÀI LIỆU THAM KHẢO**

**Tiếng Việt**

1. Optimizer - hiểu sâu về các thuật toán tối ưu

<https://viblo.asia/p/optimizer-hieu-sau-ve-cac-thuat-toan-toi-uu-gdsgdadam-Qbq5QQ9E5D8>

**Tiếng Anh**

1. https://www.datacamp.com/blog/what-is-continuous-learning