TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



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

**TÌM HIỂU VỀ OPTIMIZER, CONTINUAL LEARNING, TEST PRODUCTION VÀ SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER**

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

*Người thực hiện*: **HUỲNH HỒNG SƠN – 52000795**

Lớp **: 20050301**

Khoá  **: 24**

**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 KÌ MÔN NHẬP MÔN HỌC MÁY**

**TÌM HIỂU VỀ OPTIMIZER, CONTINUAL LEARNING, TEST PRODUCTION VÀ SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER**

Người hướng dẫn: **TS LÊ ANH CƯỜNG**

Người thực hiện: **HUỲNH HỒNG SƠN**

Lớp **: 20050301**

Khoá  **: 24**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Em xin chân thành cảm ơn thầy Lê Anh Cường đã giảng dạy, cung cấp kiến thức trên lớp và hỗ trợ em học tập để em có thể hoàn thành bài báo cáo này một cách hoàn thiện nhất.

Em xin chân thành cảm ơn.

**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 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 23 tháng 12 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Huỳnh Hồng Sơn*

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

Đây là báo cáo cuối kì của môn nhập môn học máy. Báo cáo này là tổng hợp các kiến thức mà em đã học trong quá trình học.

Chương I sẽ tiềm hiểu về về optimizer và so sánh các phương pháp

Chương II sẽ tiềm hiểu về Cotinual Learning và Test Production.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc154253507)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc154253508)

[TÓM TẮT iv](#_Toc154253509)

[MỤC LỤC 1](#_Toc154253510)

[DANH MỤC CÁC HÌNH VẼ 4](#_Toc154253511)

[CHƯƠNG 1 – TÌM HIỂU SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY 5](#_Toc154253512)

[1. Tìm hiểu, so sánh các phương pháp Optimizer trong huấn luyện mô hình học máy: 5](#_Toc154253513)

[1.1. Optimizer là gì? Nó hoạt động như thế nào? tại sao phải dùng nó? 5](#_Toc154253514)

[1.2. Các thuật toán tối ưu 6](#_Toc154253515)

[a) Gradient Descent (GD) 6](#_Toc154253516)

[b) Stochastic Gradient Descent (SGD) 7](#_Toc154253517)

[c) Momentum 9](#_Toc154253518)

[d) Adagrad 10](#_Toc154253519)

[e) RMSprop 11](#_Toc154253520)

[f) Adam 13](#_Toc154253521)

[CHƯƠNG 2 – TÌM HIỂU VỀ CONTINUAL LEARNING VÀ TEST PRODUCTION 15](#_Toc154253522)

[2.1. Continual Learning: 15](#_Toc154253523)

[a) Định nghĩa 15](#_Toc154253524)

[b) cách thức hoạt động chung của Continual Learning: 15](#_Toc154253525)

[c) Ưu điểm của Continual Learning: 16](#_Toc154253526)

[d) Nhược điểm của Continual Learning: 16](#_Toc154253527)

[2.2. Test Production 17](#_Toc154253528)

[a) Định nghĩa 17](#_Toc154253529)

[b) Cách thức hoạt động: 17](#_Toc154253530)

[c) Ưu điểm của Test Production: 18](#_Toc154253531)

[d) Nhược điểm của Test Production: 18](#_Toc154253532)

[TÀI LIỆU THAM KHẢO 19](#_Toc154253533)

**DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT**

**CÁC KÝ HIỆU**

**CÁC CHỮ VIẾT TẮT**

GD Gradient Descent

SGD Stochastic Gradient Descent

RSMprop Root Mean Square Propagation

Adam Adaptive Moment Estimation

DANH MỤC CÁC HÌNH VẼ

**DANH MỤC HÌNH**

**DANH MỤC BẢNG**

CHƯƠNG 1 – TÌM HIỂU SO SÁNH CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY

1. Tìm hiểu, so sánh các phương pháp Optimizer trong huấn luyện mô hình học máy:
   1. Optimizer là gì? Nó hoạt động như thế nào? tại sao phải dùng nó?

Optimizer là một thành phần quan trọng giúp cập nhật các trọng số của mô hình trong quá trình huấn luyện các mô hình học máy. Nó giúp giảm thiểu hàm loss và tối ưu hóa hiệu suất của các mô hình. Mục tiêu của việc huấn luyện mô hình là điều chỉnh các trọng số sao cho các mô hình có thể dự đoán chính xác nhất có thể trên các dữ liệu mới. Điều này sẽ được thực hiện bằng cách tối thiểu hàm loss, nó là một đô lường sự chệnh lệch giữa dự đoán mô hình và giá trị thực tế.

Optimizer là một công cụ được sử dụng để điều chỉnh các trọng số của mô hình dựa trên gradient của hàm loss. Bằng cách sử dụng Gradient, optimizer sẽ cập nhật các trọng số của mô hình thei hướng giảm độ dốc để tìm ra giá trị tối ưu của hàm loss.

Trong quá trình huấn luyện các mô hình, mô hình cần điều chỉnh các trọng số một cách tự động dựa trên các dữ liệu huấn luyện để học từ dữ liệu và cải thiện hiệu suất. Optimizer sẽ giúp định rõ cách cập nhật các trọng số của các mô hình một cách tự động và hiệu quả. Nếu không sử dụng optimizer thì việc tinh chỉnh các trọng số có thể trở nên rất khó khăn và không hiệu quả.

* Có nhiều loại optimizer khác nhau như là SGD, Adam, RMSprop và nhiều biển thể khác. Mỗi loại sẽ có cách thức cập nhật các trọng số cho mô hình khác nhau và có thể tốt hơn một loại mô hình hoặc bài toán cụ thể. Việc lựa chọn optimizer sẽ phụ thuộc và các yếu tố như là kích thước của mô hình, loại dữ liệu và yêu cầu về tốc độ học để có thể đạt được kết quả tốt nhất trong quá trình huấn luyện các mô hình.
  1. Các thuật toán tối ưu

1. Gradient Descent (GD)

Gradient Descent (GD) là một phương pháp tối ưu hóa được sử dụng rộng rãi trong machine learning và optimization để tối thiểu hóa hàm mất mát (loss function) của một mô hình bằng cách điều chỉnh các tham số (ví dụ: trọng số) của mô hình dựa trên gradient của hàm mất mát đó.

**Cách hoạt động của Gradient Descent:**

1. Gradient (Đạo hàm): Gradient là hướng và độ lớn tăng lên nhanh nhất của hàm số tại một điểm cụ thể. Nó cho biết hướng mà hàm số tăng nhanh nhất. Trong Gradient Descent, gradient của hàm mất mát được tính để xác định hướng và độ lớn mà ta cần đi theo để giảm thiểu hàm mất mát.
2. Cập nhật trọng số: Gradient Descent cập nhật các trọng số (ví dụ: các trọng số của một mô hình học máy) dựa trên gradient của hàm mất mát. Nó thực hiện việc di chuyển các trọng số theo hướng ngược lại với gradient để giảm thiểu hàm mất mát.
3. Learning Rate (Tốc độ học): Là tham số quan trọng trong Gradient Descent, quyết định kích thước của bước di chuyển. Learning rate quyết định tốc độ học của mô hình, nếu learning rate quá lớn, có thể dẫn đến việc vượt qua điểm cực tiểu, trong khi nếu quá nhỏ, nó có thể làm cho quá trình học chậm và mất thời gian.

**Ưu điểm của Gradient Descent:**

* Đơn giản và dễ triển khai: GD là một phương pháp tối ưu đơn giản và dễ hiểu, có thể triển khai một cách tương đối dễ dàng trong các mô hình học máy cơ bản.
* Tính toán hiệu quả: Nó có thể hoạt động tốt với dữ liệu lớn, vì việc tính toán gradient chỉ yêu cầu việc tính toán đạo hàm của hàm mất mát.

**Nhược điểm của Gradient Descent:**

* Có thể rơi vào cực tiểu cục bộ: Gradient Descent không đảm bảo tìm ra cực tiểu toàn cục của hàm mất mát. Điều này có thể dẫn đến việc hội tụ ở cực tiểu cục bộ.
* Yêu cầu chọn đúng learning rate: Việc chọn learning rate thích hợp là rất quan trọng. Nếu learning rate quá lớn hoặc quá nhỏ, có thể gây ra vấn đề không mong muốn như không hội tụ hoặc hội tụ quá chậm.
* Khả năng bị kẹt ở điểm tối thiểu: Nếu hàm mất mát có định dạng không phẳng (non-convex) hoặc có nhiều điểm cực tiểu, GD có thể bị kẹt ở điểm tối thiểu không mong muốn.

Tóm lại, Gradient Descent là một phương pháp tối ưu cơ bản và quan trọng, nhưng nó cũng có nhược điểm cần được cân nhắc và điều chỉnh để đạt được hiệu suất tốt nhất trong quá trình huấn luyện mô hình.

1. Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) là một biến thể của phương pháp Gradient Descent (GD) được sử dụng rộng rãi trong machine learning và optimization để tối ưu hóa hàm mất mát của mô hình học máy. Điểm khác biệt chính giữa SGD và GD là cách mà chúng cập nhật trọng số của mô hình.

**Cách hoạt động của Stochastic Gradient Descent:**

1. Tương tác với từng điểm dữ liệu: Trong SGD, việc cập nhật trọng số của mô hình được thực hiện trên từng điểm dữ liệu một cách ngẫu nhiên hoặc theo thứ tự của dữ liệu huấn luyện. Điều này khác với GD, nơi việc tính gradient và cập nhật trọng số được thực hiện dựa trên toàn bộ tập dữ liệu.
2. Cập nhật trọng số nhanh chóng: Do việc chỉ tính toán gradient trên một điểm dữ liệu duy nhất, SGD thường có tốc độ cập nhật nhanh hơn so với GD. Điều này đặc biệt hữu ích khi làm việc với dữ liệu lớn.

**Ưu điểm của Stochastic Gradient Descent:**

* Tính toán hiệu quả: SGD thường nhanh hơn so với GD vì nó chỉ sử dụng một điểm dữ liệu hoặc một số điểm dữ liệu nhỏ để tính toán gradient và cập nhật trọng số.
* Đặc biệt tốt với dữ liệu lớn: Khi làm việc với dữ liệu lớn, việc cập nhật trọng số trên từng điểm dữ liệu giúp SGD hội tụ nhanh hơn so với việc cập nhật trên toàn bộ tập dữ liệu.

**Nhược điểm của Stochastic Gradient Descent:**

* Khả năng dao động: Do tính ngẫu nhiên trong việc chọn điểm dữ liệu, SGD có thể dao động quanh điểm cực tiểu hoặc không hội tụ một cách ổn định, đặc biệt là trong các hàm mất mát không phẳng (non-convex).
* Yêu cầu lựa chọn learning rate cẩn thận: Việc chọn learning rate phù hợp là rất quan trọng trong SGD. Nếu learning rate quá lớn hoặc quá nhỏ, có thể ảnh hưởng đến khả năng hội tụ của mô hình.
* Khả năng không ổn định: Do tính ngẫu nhiên, SGD có thể không ổn định hơn so với GD trong việc hội tụ tới điểm tối ưu.

Tóm lại, Stochastic Gradient Descent là một phương pháp tối ưu hóa hiệu quả, đặc biệt là với dữ liệu lớn, nhưng cũng có nhược điểm cần được cân nhắc và điều chỉnh một cách cẩn thận để đạt được hiệu suất tốt nhất trong quá trình huấn luyện mô hình.

1. Momentum

Momentum là một kỹ thuật được sử dụng trong quá trình tối ưu hóa mô hình học máy, đặc biệt là trong việc cập nhật trọng số của mô hình khi sử dụng các phương pháp tối ưu như Gradient Descent (GD) hoặc Stochastic Gradient Descent (SGD).

**Cách hoạt động của Momentum:**

Momentum giúp tăng tốc độ hội tụ của thuật toán tối ưu hóa bằng cách tích lũy một lượng đại diện cho vận tốc trong quá trình cập nhật trọng số. Thay vì chỉ dựa vào gradient tại điểm hiện tại, nó giữ lại thông tin về hướng di chuyển trước đó của trọng số và sử dụng nó để định hướng cho việc cập nhật tiếp theo.

Cụ thể, khi cập nhật trọng số tại mỗi bước, Momentum tính toán một vector vận tốc (velocity) dựa trên gradient hiện tại và vận tốc trước đó. Vector vận tốc này được sử dụng để điều chỉnh và cập nhật trọng số của mô hình.

**Ưu điểm của Momentum:**

* Hỗ trợ việc hội tụ nhanh hơn: Momentum giúp mô hình hội tụ nhanh hơn bằng cách tận dụng thông tin về hướng di chuyển trước đó của trọng số.
* Giảm dao động và tránh rơi vào các cực tiểu cục bộ: Điều chỉnh trọng số dựa trên momentum có thể giúp mô hình vượt qua các điểm cực tiểu cục bộ và tránh được việc dao động quanh điểm cực tiểu.

**Nhược điểm của Momentum:**

* + - * Có thể dẫn đến việc vượt qua điểm cực tiểu: Trong một số trường hợp, momentum có thể khiến mô hình vượt qua điểm cực tiểu mà chưa hội tụ tới điểm tối ưu thực sự.
      * Yêu cầu thiết lập thêm siêu tham số: Việc điều chỉnh siêu tham số (như hệ số momentum) trong Momentum cũng là một thách thức, và cần phải được điều chỉnh một cách thích hợp để đạt được hiệu suất tốt nhất.

Tóm lại, Momentum là một kỹ thuật hữu ích giúp tăng tốc độ hội tụ của thuật toán tối ưu hóa trong quá trình huấn luyện mô hình. Tuy nhiên, cũng như các kỹ thuật tối ưu hóa khác, Momentum cũng có nhược điểm cần được xem xét khi áp dụng vào huấn luyện mô hình.

1. Adagrad

Adagrad là một phương pháp tối ưu hóa được sử dụng trong quá trình huấn luyện mô hình học máy, đặc biệt là trong việc điều chỉnh tỷ lệ học tập (learning rate) cho từng tham số của mô hình.

**Cách hoạt động của Adagrad:**

Adagrad được thiết kế để điều chỉnh tỷ lệ học tập tự động cho từng tham số của mô hình dựa trên lịch sử của gradient đã tính toán. Cụ thể:

1. Điều chỉnh learning rate: Adagrad điều chỉnh tỷ lệ học tập của mỗi tham số bằng cách chia learning rate cho căn bậc hai của tổng bình phương các gradient đã xuất hiện cho tham số đó.
2. Tích lũy các gradient: Kỹ thuật này tích lũy bình phương của tất cả các gradient đã xuất hiện cho mỗi tham số trong quá trình huấn luyện.
3. Tối ưu hóa tỷ lệ học tập: Bằng cách thực hiện việc chia tỷ lệ học tập cho căn bậc hai của tổng bình phương các gradient, Adagrad giảm tỷ lệ học tập cho các tham số đã có gradient lớn và tăng tỷ lệ học tập cho các tham số có gradient nhỏ.

**Ưu điểm của Adagrad:**

* Tính cá nhân hóa tỷ lệ học tập: Adagrad cá nhân hóa tỷ lệ học tập cho từng tham số, giúp các tham số với gradient lớn được cập nhật với tỷ lệ nhỏ hơn và ngược lại. Điều này hữu ích khi huấn luyện mô hình trên dữ liệu có các đặc trưng không đồng nhất.
* Khả năng điều chỉnh tỷ lệ học tập tự động: Adagrad tự động điều chỉnh tỷ lệ học tập dựa trên lịch sử của gradient đã tính toán, giúp tiết kiệm thời gian cho việc tinh chỉnh siêu tham số.

**Nhược điểm của Adagrad:**

* Sự giảm tỷ lệ học tập quá nhanh: Trong quá trình huấn luyện, tỷ lệ học tập có thể giảm quá nhanh với việc tích lũy bình phương của gradient. Điều này có thể dẫn đến việc learning rate trở nên rất nhỏ, khiến quá trình học chậm lại.
* Không cập nhật learning rate được điều chỉnh: Adagrad không cập nhật learning rate theo cách được điều chỉnh hoặc tối ưu trong quá trình huấn luyện, dẫn đến việc learning rate giảm đột ngột.

Tóm lại, Adagrad là một phương pháp tối ưu hóa tỷ lệ học tập tự động dựa trên lịch sử gradient của các tham số, nhưng cũng có nhược điểm cần được cân nhắc để đạt được hiệu suất tốt nhất trong quá trình huấn luyện mô hình.

1. RMSprop

RMSprop là một phương pháp tối ưu hóa trong quá trình huấn luyện mô hình học máy, thường được sử dụng để điều chỉnh tỷ lệ học tập (learning rate) của các tham số mô hình dựa trên lịch sử của gradient.

**Cách hoạt động của RMSprop:**

RMSprop được thiết kế để giảm tốc độ giảm learning rate nhanh chóng và không cho phép nó giảm đến gần với 0. Cụ thể:

1. Tính toán giá trị trung bình bình phương của gradient: RMSprop tính toán trung bình bình phương của gradient của mỗi tham số dựa trên một trọng số (thường là một hằng số beta) của gradient cũ và gradient mới.
2. Điều chỉnh learning rate: RMSprop sử dụng giá trị trung bình bình phương của gradient để điều chỉnh learning rate cho mỗi tham số. Nó chia learning rate cho căn bậc hai của giá trị trung bình bình phương này, từ đó giảm learning rate cho các tham số có gradient lớn và tăng learning rate cho các tham số có gradient nhỏ.

**Ưu điểm của RMSprop:**

* Hiệu quả với các dữ liệu không đồng nhất: RMSprop giúp điều chỉnh tỷ lệ học tập tương ứng với từng tham số, giúp tối ưu hóa hiệu quả hơn trên các dữ liệu có các đặc trưng không đồng nhất.
* Khả năng chống lại giảm learning rate quá nhanh: RMSprop giúp ngăn chặn việc giảm learning rate quá nhanh, giúp quá trình huấn luyện vẫn tiếp tục được thực hiện với learning rate có giá trị khá lớn.

**Nhược điểm của RMSprop:**

* Yêu cầu lựa chọn siêu tham số: Cũng như các phương pháp tối ưu hóa khác, RMSprop cũng cần lựa chọn các siêu tham số (ví dụ: hằng số beta) phù hợp để đạt được hiệu suất tốt nhất.
* Có thể mất tính ổn định: Trong một số trường hợp, RMSprop có thể mất tính ổn định hoặc không hiệu quả, đặc biệt khi sử dụng trên các kiến trúc mô hình phức tạp hoặc trong các tình huống cụ thể.

Tóm lại, RMSprop là một phương pháp tối ưu hóa tỷ lệ học tập dựa trên lịch sử của gradient, giúp cải thiện hiệu suất huấn luyện mô hình. Tuy nhiên, cũng như các kỹ thuật tối ưu hóa khác, RMSprop cũng có nhược điểm cần được xem xét khi áp dụng vào huấn luyện mô hình.

1. Adam

Adam (Adaptive Moment Estimation) là một phương pháp tối ưu hóa trong quá trình huấn luyện mô hình học máy, kết hợp cả momentum và RMSprop để điều chỉnh tỷ lệ học tập (learning rate) cho từng tham số của mô hình.

**Cách hoạt động của Adam:**

Adam kết hợp hai khái niệm chính: momentum và RMSprop, để điều chỉnh learning rate tự động cho từng tham số. Cụ thể:

1. Momentum: Tương tự như kỹ thuật Momentum, Adam sử dụng động lượng (momentum) để lưu trữ thông tin về hướng di chuyển trước đó của trọng số. Điều này giúp tăng tốc độ hội tụ của thuật toán.
2. RMSprop: Adam tính toán trung bình bình phương của gradient như RMSprop để điều chỉnh tỷ lệ học tập cho từng tham số.
3. Kết hợp momentum và RMSprop: Adam kết hợp thông tin từ cả momentum và RMSprop để điều chỉnh learning rate cho từng tham số. Nó tính toán learning rate được điều chỉnh dựa trên cả gradient hiện tại và độ lớn của gradient trước đó.
4. Biased correction: Adam thực hiện sửa đổi lượng bias cho các giá trị của gradient ước tính và trung bình bình phương của gradient để điều chỉnh sai số ban đầu khi bắt đầu huấn luyện.

**Ưu điểm của Adam:**

* Hiệu suất tốt và ổn định: Adam thường có hiệu suất tốt và ổn định khi huấn luyện mô hình, đặc biệt trong các bài toán học sâu (deep learning).
* Tự điều chỉnh learning rate: Adam tự động điều chỉnh learning rate cho từng tham số dựa trên thông tin từ cả động lượng và RMSprop, giúp tối ưu hóa hiệu quả hơn.

**Nhược điểm của Adam:**

* Yêu cầu thiết lập thêm siêu tham số: Adam cũng cần phải thiết lập các siêu tham số (ví dụ: hằng số beta) để điều chỉnh và có thể ảnh hưởng đến hiệu suất của thuật toán.
* Tốn tài nguyên tính toán: Adam có thể tốn tài nguyên tính toán hơn so với các phương pháp tối ưu hóa đơn giản khác trong một số trường hợp.

Tóm lại, Adam là một phương pháp tối ưu hóa tỷ lệ học tập được kết hợp từ momentum và RMSprop, có khả năng cải thiện hiệu suất và ổn định trong quá trình huấn luyện mô hình học máy. Tuy nhiên, cũng cần xem xét và điều chỉnh các siêu tham số để đạt được hiệu suất tốt nhất.

CHƯƠNG 2 – TÌM HIỂU VỀ CONTINUAL LEARNING VÀ TEST PRODUCTION

* 1. Continual Learning:

1. Định nghĩa

Continual Learning (CL), còn được gọi là lifelong learning hoặc incremental learning, là một lĩnh vực quan trọng trong machine learning tập trung vào khả năng của mô hình học máy để liên tục học và thích nghi với dữ liệu mới mà không quên hoặc quá mức thích ứng với dữ liệu cũ. Điều này thường xảy ra trong một số tình huống thực tế khi dữ liệu liên tục được cập nhật, không gian đặc trưng thay đổi theo thời gian hoặc khi có nhiều tác vụ cần được học liên tục mà không cần phải huấn luyện lại mô hình từ đầu.

1. cách thức hoạt động chung của Continual Learning:
2. Huấn luyện ban đầu: Mô hình được huấn luyện ban đầu trên tập dữ liệu ban đầu để học cách giải quyết một nhiệm vụ cụ thể.
3. Tiếp nhận dữ liệu mới: Dữ liệu mới được đưa vào mô hình để học từ các điểm dữ liệu mới mà mô hình chưa từng gặp.
4. Quản lý quên (Catastrophic Forgetting): Một trong những thách thức lớn nhất của CL là quản lý tri thức đã học từ dữ liệu cũ khi học dữ liệu mới. Điều này gọi là "catastrophic forgetting". Các kỹ thuật như replay, regularization, hoặc distillation được sử dụng để giảm thiểu vấn đề này. Ví dụ, mô hình có thể được huấn luyện lại với một phần dữ liệu cũ hoặc dữ liệu mẫu được lưu trữ để cung cấp lại thông tin từ quá khứ.
5. Phân tách kiến thức: Mô hình cần phải biết phân biệt và tách các kiến thức đã học từ các tác vụ hoặc dữ liệu khác nhau. Điều này có thể được thực hiện thông qua các kỹ thuật như modularization, task boundaries hoặc các cơ chế như attention mechanisms.
6. Tối ưu hóa cho khả năng học liên tục: Các thuật toán và phương pháp CL như transfer learning, meta learning, online learning, lifelong learning được sử dụng để cải thiện khả năng học liên tục của mô hình.
7. Đánh giá và kiểm thử: Mô hình cần được đánh giá thường xuyên trên dữ liệu mới để đảm bảo rằng nó vẫn hiệu quả và không quên tri thức đã học từ dữ liệu cũ.
8. Ưu điểm của Continual Learning:

* Khả năng học liên tục: CL cho phép mô hình tiếp tục học từ dữ liệu mới mà không cần phải huấn luyện lại từ đầu, tiết kiệm thời gian và tài nguyên tính toán.
* Thích nghi với môi trường thay đổi: Mô hình học liên tục có khả năng thích nghi với sự thay đổi trong môi trường, dữ liệu mới, và yêu cầu của bài toán, phù hợp với các tình huống thực tế có dữ liệu đa dạng và thay đổi theo thời gian.
* Tiến triển kiến thức không gian lớn: CL cho phép mô hình tích lũy kiến thức từ nhiều nguồn dữ liệu khác nhau và từ nhiều bài toán, từ đó cải thiện khả năng tổng hợp thông tin và áp dụng kiến thức đã học vào các tác vụ mới.

1. Nhược điểm của Continual Learning:

* Catastrophic Forgetting: Một trong những thách thức lớn nhất của CL là sự quên tri thức từ dữ liệu cũ khi học dữ liệu mới. Điều này gọi là "catastrophic forgetting" và yêu cầu các phương pháp như replay, regularization để giảm thiểu vấn đề này.
* Dữ liệu không cân đối: Mô hình CL có thể gặp phải vấn đề khi dữ liệu mới không cân đối so với dữ liệu cũ, dẫn đến việc mất cân bằng và ảnh hưởng đến khả năng học của mô hình.
* Yêu cầu tính toán phức tạp: Mô hình CL thường yêu cầu tính toán phức tạp hơn so với mô hình học cổ điển vì nó phải xử lý và học từ dữ liệu liên tục.

Tóm lại, Continual Learning mang lại nhiều ưu điểm quan trọng trong việc xử lý dữ liệu đa dạng và thay đổi theo thời gian, nhưng cũng đặt ra nhiều thách thức đối với việc quản lý tri thức cũ và mới, và yêu cầu sự cân nhắc trong việc xử lý sự quên và bất cân đối trong dữ liệu.

* 1. Test Production

1. Định nghĩa

Test Production là quá trình kiểm thử và sản xuất các bài kiểm tra hoặc thử nghiệm để đánh giá hiệu suất và tính ổn định của một sản phẩm hoặc dịch vụ trong môi trường thực tế. Trong lĩnh vực học máy và phát triển phần mềm, Test Production thường được áp dụng để kiểm tra hiệu suất của mô hình học máy trước khi triển khai vào môi trường thực tế.

1. Cách thức hoạt động:
2. Tạo bộ dữ liệu kiểm thử: Một bộ dữ liệu kiểm thử được tạo ra, bao gồm các dữ liệu mà mô hình chưa từng thấy trước đó. Đây có thể là dữ liệu thực tế từ môi trường hoạt động thực tế hoặc dữ liệu được tạo ra mô phỏng.
3. Áp dụng mô hình vào dữ liệu kiểm thử: Mô hình học máy sẽ được thử nghiệm bằng cách áp dụng vào bộ dữ liệu kiểm thử để đánh giá hiệu suất và dự đoán trên dữ liệu mới.
4. Đánh giá hiệu suất: Mô hình được đánh giá dựa trên các metric như accuracy, precision, recall, F1-score hoặc các metric phù hợp với bài toán cụ thể. Điều này giúp đánh giá khả năng dự đoán và hiệu suất của mô hình trên dữ liệu mới.
5. Kiểm tra tính ổn định và tin cậy: Test Production cũng giúp kiểm tra tính ổn định của mô hình trong các điều kiện khác nhau và đánh giá mức độ tin cậy của dự đoán trong môi trường thực tế.
6. Ưu điểm của Test Production:

* Đánh giá chất lượng sản phẩm: Quá trình này cho phép đánh giá chất lượng và hiệu suất của mô hình hoặc sản phẩm trước khi triển khai vào môi trường thực tế.
* Phát hiện lỗi và cải thiện: Nó giúp phát hiện lỗi, điều chỉnh và cải thiện mô hình trước khi đưa vào sử dụng thực tế, giúp tránh các vấn đề và sự cố xảy ra sau này.

1. Nhược điểm của Test Production:

* Khó khăn trong việc tạo dữ liệu kiểm thử đa dạng và phản ánh thực tế: Việc tạo ra một bộ dữ liệu kiểm thử chính xác và đủ đa dạng để phản ánh mọi trường hợp có thể xảy ra trong môi trường thực tế có thể khó khăn.
* Chi phí và thời gian: Test Production yêu cầu chi phí và thời gian để tạo ra và thực hiện các bài kiểm tra, đặc biệt là khi đòi hỏi môi trường và điều kiện phức tạp.

Tóm lại, Test Production là quá trình đánh giá hiệu suất và tính ổn định của mô hình hoặc sản phẩm trong môi trường thực tế, giúp đảm bảo chất lượng và tin cậy trước khi triển khai vào sử dụng thực tế. Tuy nhiên, nó cũng đặt ra những thách thức về độ đa dạng của dữ liệu kiểm thử và chi phí thời gian để thực hiện.

TÀI LIỆU THAM KHẢO

(1)

Slide bài giảng nhập môn học máy của trường Đại học Tôn Đức Thắng

(2)

Trần Trung Trực. *Optimizer- Hiểu sâu về các thuật toán tối ưu ( GD,SGD,Adam,..)*. Viblo.

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

(3)

Sen Tây Hồ. *Optimizer- Hiểu sâu về các thuật toán tối ưu ( GD,SGD,Adam,.)*. Sen Tây Hồ.

<https://sentayho.com.vn/optimizer-la-gi.html>.

(4)

Ferreiro, Y. *What is Continuous Learning? Revolutionizing Machine Learning & Adaptability*. Datacamp.com.

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