TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**Nhập Môn Học Máy**

**DỰ ÁN CUỐI KỲ**

*Người hướng dẫn*: **PHS GS LÊ ANH CƯỜNG**

*Người thực hiện*: **LÊ NGUYỄN VIẾT HIỆP – 521H0398**

Lớp **: 21H50302**

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**



**Nhập Môn Học Máy**

**DỰ ÁN CUỐI KỲ**

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

*Người thực hiện*: **LÊ NGUYỄN VIẾT HIỆP – 521H0398**

Lớp **: 21H50302**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2014**

LỜI CẢM ƠN

Em xin chân thành cảm ơn PhS TS Lê Anh Cường - người trực tiếp giảng dạy những kiến thức cơ bản và cần thiết về Nhập môn Học Máy. Nhờ sự hướng dẫn, giúp đỡ và điều kiện thuận lợi, em đã hoàn thành được báo cáo cuối cùng.

Dù đã có sự đầu tư nhất định nhưng cũng khó tránh khỏi những sai sót trong quá trình thực hiện. Em trân trọng mong nhận được những ý kiến góp ý của thầy để báo cáo được hoàn thiện hơn.

Em 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 / chúng tôi và được sự hướng dẫn của PHS 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 7 tháng 12 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

**

*Lê Nguyễn Viết Hiệp*

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

Thực hiện các nghiên cứu về phương pháp Optimzer trong việc huấn luyện mô hình. So sánh các phương pháp và tìm ra điểm tối ưu của từng phương pháp. Tìm hiểu về Continual Learning và Test Production khi xây dựng giải pháp học máy cho một bài toán nào đó

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc153454403)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc153454404)

[TÓM TẮT iv](#_Toc153454405)

[MỤC LỤC 1](#_Toc153454406)

[DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT 2](#_Toc153454407)

[CHƯƠNG 1 – OPTIMIZER 3](#_Toc153454408)

[1.1 Optimizer 3](#_Toc153454409)

[1.1.1 Khái quát Optimizer 3](#_Toc153454410)

[1.1.2 Các phương pháp Optimizer 3](#_Toc153454411)

[1.1.3 So sánh các phương pháp Optimizer 7](#_Toc153454412)

[CHƯƠNG 2 – CONTINUAL LEARNING VÀ TEST PRODUCTION 9](#_Toc153454413)

[2.1 Continual Learning 9](#_Toc153454414)

[2.2 Test Production 10](#_Toc153454415)

[TÀI LIỆU THAM KHẢO 12](#_Toc153454416)

CHƯƠNG 1 – OPTIMIZER

1.1 Optimizer

1.1.1 Khái quát Optimizer

Thuật toán tối ưu hóa (Optimizer) giúp tìm ra các tham số tối ưu của mô hình bằng cách học các “features” của dữ liệu đầu vào, từ đó sử dụng 1 cặp “weights” và “bias” để chiều chỉnh các tham nhằm cải thiện độ chính xác của mô hình.

Một số thuật ngữ liên quan đến Optimizer:

+ Epoch: Số lần thuật toán chạy trên toàn bộ training data

+ Sample: Một hàng của tập dữ liệu

+ Batch: Số lượng mẫu được lấy để cập nhật các tham số cho mô hình

+ Learning rate: Tham số giúp cho mô hình có thể cập nhật được mức độ trọng số phù hợp

+ Cost Function/Loss Function: Hàm dùng để tính độ chênh lệch giữa giá trị dự đoán và giá trị thực tế

+ Weights: Trọng số, hay còn gọi là hệ số, là các giá trị mà mô hình học từ dữ liệu đầu vào. Chúng quyết định mức độ ảnh hưởng của mỗi đặc trưng đầu vào đến kết quả dự đoán

+ Bias: Sai số là một thành phần tuyến tính được áp dụng cho thông tin đầu vào. Thêm vào kết quả của phép nhân trọng số với đầu vào sẽ thành một hàm tuyến tính đi qua gốc tọa đô. Sử dụng “bias” giúp mạng nơ-ron có thể hàm dịch chuyển hàm tuyến tính một cách linh hoạt

1.1.2 Các phương pháp Optimizer

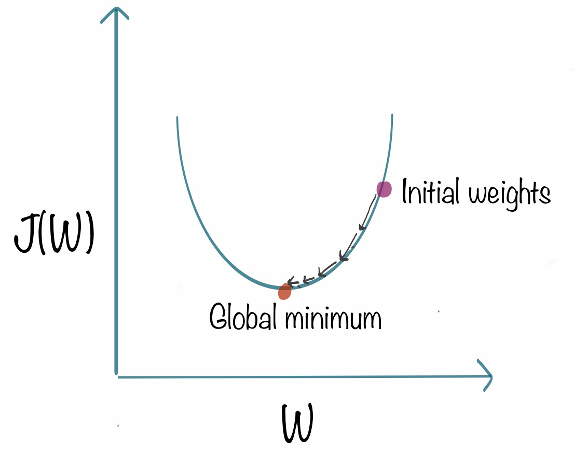
Gradient Descent (GD)

Là phương pháp tối ưu hóa phổ biến nhất trong học máy. Phương pháp chọn 1 nghiệm ngẫu nhiên sau mỗi epoch để giảm thiểu Loss Function

Công thức: **xnew = xold - learningrate.gradient(x)**

Cách hoạt động:

* Khởi tạo giá trị Weights và Bias: Bắt đầu từ một điểm ngẫu nhiên trong không gian tham số
* Tính Gradient (đạo hàm) của Loss Function tại điểm ngẫu nhiên vừa chọn
* Cập nhật tham số bằng cách di chuyển theo hướng ngược lại với gradient. Kích thước bước di chuyển (learning rate) quyết định tốc độ học của thuật toán.
* Lặp lại quá trình trên cho đến khi đạt được một điểm dừng (thường là gradient gần bằng 0 hoặc sau một số lượng lần lặp nhất định nào đó)



Ưu điểm: Thuật toán cơ bản, dễ hiểu

Nhược điểm:

* Vì thuật toán đơn giản nên còn nhiều hạn chế như phụ thuộc vào nghiệm khởi tạo ban đầu và learning rate
* Tốc độ học quá lớn sẽ khiến cho thuật toán không hội tụ, loanh quanh bên đích vì bước nhảy quá lớn, hoặc tốc độ học nhỏ ảnh hưởng đến tốc độ training

Stochastic Gradient Descent (SGD)

Là một biến thể của thuật toán Gradient Descent. Thuật toán này giải quyết sự không hiệu quả về mặt tính toán của các phương pháp Gradient Descent truyền thống khi xử lý với các tập dữ liệu lớn

Công thức:

Chú thích:

* : Điểm hiện tại
* : Hệ số học, quy định kích thước của bước di chuyển
* : Gradient (đạo hàm) của hàm tại điểm 
* : Điểm mới sau khi di chuyển

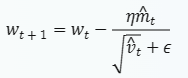
Cách hoạt động:

* Xáo trộn dữ liệu đào tạo
* Lặp lại qua từng ví dụ đào tạo (hoặc một batch) theo thứ tự đã xáo trộn
* Tính toán gradient của hàm chi phí liên quan đến các tham số mô hình bằng cách sử dụng ví dụ đào tạo hiện tại (hoặc batch).
* Cập nhật các tham số mô hình bằng cách thực hiện một bước theo hướng gradient âm, được co giãn bởi tỷ lệ học.
* Đánh giá tiêu chí hội tụ, chẳng hạn như sự khác biệt trong hàm chi phí giữa các lần lặp của gradient.

Ưu điểm: Hiệu suất tính toán cao, đặc biệt là khi xử lý các tập dữ liệu lớn.

Nhược điểm: Không hội tụ một cách trực tiếp do sự nhiễu

Adam

Tên đầy đủ là Adaptive Moment Estimation là một thuật toán tối ưu hóa lặp lại được sử dụng để giảm thiểu hàm mất mát trong quá trình huấn luyện mạng nơ-ron. Adam có thể được coi là sự kết hợp của RMSprop và Stochastic Gradient Descent với động lượng

Công thức:

Chú thích:

*  : Trọng số mô hình tại thời điểm t
* : Hệ số học, quy định kích thước của bước di chuyển
* : Giá trị được hiệu chỉnh của trung bình di chuyển của gradient và gradient bình phương
* : Một hằng số nhỏ để tránh chia cho 0

Cách hoạt động:

* Khởi tạo trung bình di chuyển của moment đầu tiên và thứ hai (m và v) bằng 0.
* Tính toán gradient của hàm mất mát đối với các tham số mô hình.
* Cập nhật các trung bình di chuyển bằng cách sử dụng trung bình giảm dần theo hàm mũ. Điều này liên quan đến việc tính toán mt​ và vt​ là trung bình trọng số của các moment trước đó và gradient hiện tại.
* Áp dụng hiệu chỉnh bias cho các trung bình di chuyển, đặc biệt là trong các vòng lặp đầu tiên

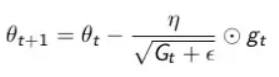
Ưu điểm:

* Hội tụ nhanh hơn Adagrad và RMSProp
* **Ít bị overfitting hơn:** Adam sử dụng một biến động độ dốc trung bình để làm mịn gradient. Điều này giúp mô hình tránh bị overfitting hơn so với các phương pháp tối ưu hóa khác, vốn có thể bị ảnh hưởng bởi nhiễu trong dữ liệu.

Nhược điểm:

* Có thể bị quá đà
* **Yêu cầu điều chỉnh tham số:** Adam có hai tham số, beta1 và beta2. Các tham số này cần được điều chỉnh để đạt được hiệu quả tốt nhất.

Adagrad

Công thức:

Chú thích:

* n: hằng số
* gt: gradient tại thời điểm t
* ϵ : hệ số tránh lỗi ( chia cho mẫu bằng 0)
* G : là ma trận chéo mà mỗi phần tử trên đường chéo (i,i) là bình phương của đạo hàm vectơ tham số tại thời điểm t.

Ưu điểm: Tránh việc điều chỉnh learning rate bằng tay, chỉ cần để tốc độ học default là 0.01 thì thuật toán sẽ tự động điều chỉnh

Nhược điểm: Tổng bình phương biến thiên sẽ lớn dần theo thời gian cho đến khi nó làm tốc độ học cực kì nhỏ, làm việc training trở nên đóng băng

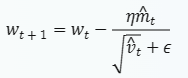
1.1.3 So sánh các phương pháp Optimizer

Gradient Descent

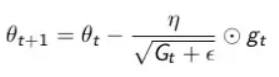
* Công thức: **xnew = xold - learningrate.gradient(x)**
* Ưu điểm: Chính xác vì nó sử dụng gradient trung bình được tính toán trên toàn bộ tập dữ liệu
* Nhược điểm: Cần phải xử lý toàn bộ tập dữ liệu cho mỗi lần lặp, điều này có thể gây ra vấn đề về hiệu suất khi xử lý với các tập dữ liệu lớn

**Stochastic Gradient Descent

* Công thức:
* Ưu điểm: Hiệu suất tính toán tốt hơn khi xử lý với các tập dữ liệu lớn
* Nhược điểm: Ít chính xác hơn Gradient Descent, vì nó chỉ tính toán gradient trên một ví dụ duy nhất, có thể không đại diện chính xác cho toàn bộ tập dữ liệu

**Adam

* Công thức:
* Ưu điểm: Hiệu suất tính toán tốt khi xử lý với các tập dữ liệu lớn. Adam cũng kết hợp các ưu điểm của RMSprop và SGD với động lượng.
* Nhược điểm: Có thể không hội tụ một cách trực tiếp do sự nhiễu

**Adagrad

* Công thức:
* Ưu điểm: Không cần phải điều chỉnh thủ công tốc độ học. Adagrad cải thiện sự hội tụ bằng cách điều chỉnh tốc độ học cho mỗi tham số một cách riêng biệt
* Nhược điểm: Một nhược điểm chính của Adagrad là alpha (t) có thể trở nên lớn khi số lần lặp tăng lên và do đó sẽ giảm ở tỷ lệ lớn hơn. Điều này sẽ làm cho trọng số cũ gần như bằng với trọng số mới, có thể dẫn đến sự hội tụ chậm

CHƯƠNG 2 – CONTINUAL LEARNING VÀ TEST PRODUCTION

2.1 Continual Learning

Continual Learning là một lĩnh vực học máy tập trung vào việc xây dựng các mô hình có thể học hỏi và thích nghi với dữ liệu mới mà không làm mất kiến thức đã học trước đó

Khi xây dựng một giải pháp học máy để giải quyết bài toán nào đó, cần lưu ý đến khả năng học liên tục của mô hình. Nếu mô hình không thể học liên tục, có thể dẫn đến các vấn đề sau:

* Overfitting: Mô hình có thể bị quá đà với dữ liệu cũ và không thể học hỏi từ dữ liệu mới.
* Performance degradation: Mô hình có thể giảm hiệu suất khi tiếp xúc với dữ liệu mới.

Có nhiều kỹ thuật có thể được sử dụng để cải thiện khả năng học liên tục của mô hình:

* Data augmentation: Tạo thêm dữ liệu dừ dữ liệu hiện có để giúp mô hình học hỏi nhiều hơn
* Regulaziration: Sử dụng các kỹ thuật như L1 regularization hoặc L2 regularization để giúp mô hình tránh bị quá đà
* Memory-based approaches: Sử dụng bộ nhớ để lữu trữ kiến thức đã học trước đó và sử dụng kiến thức này để học hỏi từ dữ liệu mới
* Incremental learning approaches: Sử dụng các kỹ thuật để cập nhật mô hình một cách dần dần khi tiếp xúc với dữ liệu mới.

Lựa chọn mô hình Continual Learning để xây dựng một giải pháp học máy phụ thuốc vào nhiều yếu tố như loại bài toán, kích thước của tập dữ liệu hoặc khả năng tính toán.

2.2 Test Production

Test Production là quá trình kiếm tra và đánh giá mô hình học máy trong môi trường sản xuất thực tế. Mục đích của việc này đảm bảo rằng mô hình học máy có thể hoạt động hiệu quả trong môi trường sản xuất thực tế. Quá trình này giúp phát hiện các vấn đề tiềm ẩn hoặc không lường trước của mô hình

Các bước của Test Production:

* Thu thập dữ liệu từ môi trường sản xuất:  Dữ liệu cần được thu thập từ môi trường sản xuất thực tế để đảm bảo rằng mô hình có thể hoạt động tốt trong môi trường đó. Dữ liệu cần được thu thập một cách đại diện để mô hình có thể học hỏi từ các trường hợp khác nhau.
* Phân tách dữ liệu: Dữ liệu cần được phân tách thành hai tập: tập huấn luyện và tập kiểm tra. Tập huấn luyện được sử dụng để đào tạo mô hình, tập kiểm tra được sử dụng để đánh giá hiệu suất của mô hình
* Đào tạo mô hình: Mô hình được đào tạo trên tập huấn luyện
* Đánh giá mô hình: Mô hình được đánh giá trên tập kiểm tra
* Điều chỉnh mô hình: Nếu hiệu suất của mô hình không đạt yêu cầu, mô hình cần được điều chỉnh. Điều chỉnh mô hình có thể bao gồm việc thay đổi các tham số của mô hình, thay đổi kiến trúc của mô hình, hoặc thu thập thêm dữ liệu.

Có nhiều cách khác nhau để đánh giá Test Production:

* Đánh giá định lượng: Đánh giá định lượng sử dụng các số liệu thống kê để đánh giá hiệu suất của mô hình. Các số liệu thống kê phổ biến bao gồm:
* Accuracy: Tỷ lệ các dự đoán chính xác.
* Precision: Tỷ lệ các dự đoán chính xác trong số tất cả các dự đoán dương.
* Recall: Tỷ lệ các dự đoán dương được dự đoán chính xác.
* F1-score: Hỗn hợp giữa precision và recall.
* Đánh giá định tính: Đánh giá định tính sử dụng các mô tả bằng văn bản để đánh giá hiệu suất của mô hình
* Mức độ tin cậy của dự đoán: Mô hình có thể tin tưởng vào dự đoán của mình đến mức nào?
* Khả năng giải thích của dự đoán: Mô hình có thể giải thích lý do tại sao nó đưa ra dự đoán đó không?
* Sự phù hợp với mục tiêu: Mô hình có đáp ứng các yêu cầu của mục tiêu không?
* Các chỉ số đánh giá khác: Ngoài các chỉ số trên, có thể sử dụng các chỉ số đánh giá khác để đánh giá Test Production:
* Thời gian đào tạo: Thời gian cần thiết để đào tạo mô hình.
* Thời gian dự đoán: Thời gian cần thiết để mô hình đưa ra dự đoán.
* Sử dụng tài nguyên: Số lượng tài nguyên tính toán cần thiết để chạy mô hình.

Lặp lại quá trình Test Production: Quá trình Test Production có thể được lặp lại nhiều lần để cải thiện hiệu suất của mô hình. Khi mô hình được cải thiện, dữ liệu cần được thu thập lại để đảm bảo rằng mô hình vẫn hoạt động tốt trong môi trường sản xuất thực tế.

# TÀI LIỆU THAM KHẢO

**Tiếng Việt**

|  |  |
| --- | --- |
| [1] | N. T. Đạt, "Optimizer- Hiểu sâu về các thuật toán tối ưu ( GD,SGD,Adam,..)," 20 12 2021. [Online]. Available: https://viblo.asia/p/optimizer-hieu-sau-ve-cac-thuat-toan-toi-uu-gdsgdadam-Qbq5QQ9E5D8. |
| [2] | N. T. Đạt, "Tại sao bạn nên kiểm thử trên môi trường Production?," 20 07 2023. [Online]. Available: https://viblo.asia/p/tai-sao-ban-nen-kiem-thu-tren-moi-truong-production-WAyK8oOm5xX. |

**Tiếng Anh**

|  |  |
| --- | --- |
|  |  |
| [3] | A. Vidhya, "Optimizers in Deep Learning: A Comprehensive Guide," 25 01 2021. [Online]. Available: https://www.scribbr.com/apa-style/format/. |
| [4] | Optimizely, "What is 'testing in production'?," 20 07 2022. [Online]. Available: https://www.optimizely.com/optimization-glossary/testing-in-production/. |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |