TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CUỐI KỲ MÔN NHẬP MÔN NHẬP MÔN HỌC MÁY**

**BÁO CÁO CUỐI KỲ**

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

*Người thực hiện*: **ĐỖ TRẦN ANH VĨNH – 52100948**

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**



**BÁO CÁO CUỐI KỲ MÔN NHẬP MÔN NHẬP MÔN HỌC MÁY**

**BÁO CÁO GIỮA KỲ**

Người hướng dẫn: **TS LÊ ANH CƯỜNG**

Người thực hiện: **ĐỖ TRẦN ANH VĨNH**

Lớp **: 21050301**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Em xin chân thành cảm ơn thầy Lê Anh Cường – giảng viên bộ môn Nhập môn học máy đã hỗ trợ, giúp đỡ trong quá trình học tập, nghiên cứu để có thể thực hiện được bài báo cáo.

Xin chân thành cảm ơn khoa Công Nghệ Thông Tin trường Đại học Tôn Đức Thắng đã tạo điều kiên để em có thể học tập, nghiên cứu, hoàn thiện kiến thức của bản thân để có thể thực hiện bài báo cáo. Do hạn chế về kiến thức và khả năng nên bài báo cáo còn nhiều thiếu xót, mong rằng quý thầy cô có thể đóng góp, ý kiến để em có thể học hỏi và hoàn thiên các bài báo cáo khác trong tương lai. 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 / chúng tôi và được sự hướng dẫn của TS Trịnh Hùng 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)*

*Vĩnh*

*Đỗ Trần Anh Vĩnh*

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

Báo cáo sẽ trình bày về các nghiên cứu tìm hiểu về các Optimizer trong huấn luyện các mô hình học máy, cụ thể là cách hoạt động, công thức, cách thực hiện và so sánh giữa chúng. Có tổng cộng 4 Optimizer được tìm hiểu là Gradient Descent, Adaptive Gradient Descent, Root Mean Square Propagation và Adaptive Moment Estimation. Tiếp theo sẽ là về Continual Learning và Test Production khi xây dựng một giải pháp học máy để giải quyết một bài toán nào đó.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc154092687)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc154092688)

[TÓM TẮT iv](#_Toc154092689)

[MỤC LỤC 1](#_Toc154092690)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 3](#_Toc154092691)

[CHƯƠNG 1 – OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY 4](#_Toc154092692)

[1.1 Các Optimizer 4](#_Toc154092693)

[1.1.1 Gradient Descent 4](#_Toc154092694)

[1.1.2 Adaptive Gradient Algortithm (AdaGrad) 5](#_Toc154092695)

[1.1.3 Root Mean Squared Propagation (RMSProp) 6](#_Toc154092696)

[1.1.4 Adaptive Moment Estimation (Adam) 7](#_Toc154092697)

[1.2 So sánh các Optimizer 8](#_Toc154092698)

**DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT**

**CÁC KÝ HIỆU**

**CÁC CHỮ VIẾT TẮT**

DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

**DANH MỤC HÌNH**

**DANH MỤC BẢNG**

CHƯƠNG 1 – OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY

Optimizer là các thuật toán, phương pháp giúp mô hình cập nhật các tham số của nó dựa trên kết quả huấn luyện để cải thiện hiệu quả, độ chính xác của mô hình. Các Optimizer sẽ hoạt động dựa trên loss function của bài toán, cố gắng cập nhất các trọng số, bias để làm tối thiểu loss function.

Việc lựa chọn Optimizer sẽ phụ thuộc vào loss function hoặc activation function trong mạng neural network và cũng sẽ ảnh hưởng rất nhiều đến mô hình. Một số Optimizer phổ biến như Gradient Descent, AdaGrad, RMSProp, Adam,…

1.1 Các Optimizer

1.1.1 Gradient Descent

Gradient Descent là một optimizer phổ biến và đơn giản, hoạt động bằng cách tối thiểu loss function L của bài toán để tìm giá trị tốt nhất cho tham số trong bài. Vì mục tiêu là tối thiểu loss function L nên thông qua đạo hàm của L theo tham số wi bất kỳ ta có thể biết được hướng biến thiên của loss function theo wi, khi đó ta sẽ thay đổi giá trị wi một khoảng bằng , dấu trừ sẽ đảo chiều của đạo hàm để di chuyển dần đến điểm cực tiểu của hàm, learning rate γ sẽ là một số rất nhỏ, giá trị này sẽ quyết định độ lớn trong giá trị của mỗi lần cập nhật tham số. Thuật toán sẽ dừng lại khi tham số không thể cải thiện được nữa hoặc khi hết số vòng lặp.

Công thức:

wi: tham số cần cập nhật

γ: learning rate

: đạo hàm của loss function (L) theo wi

Các bước thực hiện:

Khởi tạo giá trị ban đầu bất kỳ cho các tham số w

Tính đạo hàm của Loss Function theo từng w*­*i

Cập nhật từng wi theo công thức

1.1.2 Adaptive Gradient Algortithm (AdaGrad)

Là thuật toán mở rộng từ Gradient Descent, thay vì cố định một learning rate xuyên suốt cho tất cả các tham số thì ta sẽ có từng learning rate cho từng tham số và sẽ thay đổi xuyên suốt thuật toán. Learning rate sẽ được tính bằng cách chia một tham số step\_size cố định cho tổng bình phương của các gradient theo wi của cost function trong từng vòng lặp.

Công thức:

wi: tham số cần cập nhật

N: số thuộc tính của bộ dữ liệu

S­: tổng bình phương gradient theo từng w của L

: đạo hàm của loss function (L) theo wi

γi: learning rate của wi

step\_size: một giá trị cố định để tính learning rate

1e-8: một số cực kỳ nhỏ để tránh trường hợp S = 0

Các bước thực hiện:

Khởi tạo giá trị ban đầu bất kỳ cho các tham số wi

Tính gradient của L theo tất cả w

Tính tổng bình phương tất cả gradient S

Tính learning rate γi

Cập nhật từng wi theo công thức

1.1.3 Root Mean Squared Propagation (RMSProp)

RMSProp là phiên bản mở rộng từ AdaGrad nhằm khắc phục tốc độ học chậm của nó. Một tham số mới rho được gọi là decay rate sẽ được thêm vào để tính một giá trị momentum S mới để cập nhật learning rate. Momentum sẽ khiến cho trọng số của các gradient cũ giảm dần đi, giúp thuật toán điều chỉnh chủ yếu dựa trên các gradient gần đây.

Công thức:

wi: tham số cần cập nhật

rho: decay rate

Si: momentum

: đạo hàm của loss function (L) theo wi

γi: learning rate của wi

step\_size: một giá trị cố định để tính learning rate

1e-8: một số cực kỳ nhỏ để tránh trường hợp Si = 0

Cách thực hiện:

Khởi tạo giá trị ban đầu bất kỳ cho các tham số wi

Tính gradient của L theo wi

Tính Si

Cập nhật learning rate γi

Cập nhật từng wi theo công thức

1.1.4 Adaptive Moment Estimation (Adam)

Là thuật toán tiếp tục được mở rộng từ AdaGrad và RMSProp, ta sẽ có 2 momentum m, v để điều chỉnh các learning rate thay vì 1 như 2 thuật toán trên. Momentum thứ 2 sẽ giúp normalize step\_size, khi step\_size quá lớn v sẽ giảm giá trị của nó và ngược lại, khi step\_size nhỏ thì v sẽ tăng giá trị của nó lên.

Công thức:

wi: tham số cần cập nhật

m: momentum thứ nhất

β1: decay rate của m

: đã được điều chỉnh bias

v: momentum thứ hai

β2: decay rate của v

: đã được điều chỉnh bias

: đạo hàm của loss function (L) theo wi

γi: learning rate của wi

step\_size: một giá trị cố định để tính learning rate

1e-8: một số cực kỳ nhỏ để tránh trường hợp chia cho 0

Cách thực hiện:

Khởi tạo giá trị ban đầu bất kỳ cho các tham số wi

Tính gradient của L theo wi

Tính m và v

Điều chỉnh m và v thành và

Cập nhật learning rate γi

Cập nhật từng wi theo công thức

1.2 So sánh các Optimizer

|  |  |  |
| --- | --- | --- |
| Thuật toán | Ưu điểm | Nhược Điểm |
| Gradient Descent | * Đơn giản, dễ thực hiện * Có thể áp dụng cho các hàm nào có thể đạo hàm * Tính toán hiệu quả với các bài toán kích thước nhỏ. * Phổ biến | * Có thể bị kẹt ở các điểm local minimal. * Dễ bị ảnh hưởng bởi nhiễu. * Bị ảnh hưởng bởi việc chọn learning rate. * Tốc độ chậm. |
| AdaGrad | * Ít ảnh hưởng bởi giá trị khởi tạo. * Dễ tìm được điểm cực tiểu. * Tự động điều chỉnh learning rate. * Tính toán hiệu quả hơn Gradient Descent nhờ sử dụng các learning rate khác nhau. | * Learning rate có thể giảm xuống giá trị rất nhỏ, dẫn tới dừng thuật toán sớm. * Learning rate tự điều chỉnh có thể giúp thoát khỏi local minima. |
| RMSProp | * Ổn định * Hội tụ nhanh * Khắc phục nhược điểm trong learning rate của AdaGrad * Khả năng mắc kẹt ở local minima được giảm thiểu. | * Dễ bị ảnh hưởng bởi các tham số. * Cần phải điều chỉnh một số tham số. |
| Adam | * Có các ưu điểm của AdaGrad và RMSProp * Tự động normalize learning rate, giúp tính toán hiệu quả hơn. * Kết hợp từ AdaGrad và RMSProp giúp thuật toán có khả năng thoát khỏi local minima. | * Dễ bị ảnh hưởng bởi các tham số. * Có khả năng dẫn tới overfitting. * Hội tụ kết quả tệ trong một số trường hợp. |

**TÀI LIỆU THAM KHẢO**

**Tiếng Việt**

**Tiếng Anh**

1. [Gradient Descent For Machine Learning - MachineLearningMastery.com](https://machinelearningmastery.com/gradient-descent-for-machine-learning/)
2. [Gradient Descent With AdaGrad From Scratch - MachineLearningMastery.com](https://machinelearningmastery.com/gradient-descent-with-adagrad-from-scratch/)
3. [Gradient Descent With RMSProp from Scratch - MachineLearningMastery.com](https://machinelearningmastery.com/gradient-descent-with-rmsprop-from-scratch/)
4. [Code Adam Optimization Algorithm From Scratch - MachineLearningMastery.com](https://machinelearningmastery.com/adam-optimization-from-scratch/)

**PHỤ LỤC**