

ĐẠI HỌC HUẾ

**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**





**BÁO CÁO ĐỒ ÁN**

**Học kỳ II, năm học 2022 - 2023**

**Học phần: Các thuật toán tối ưu cho phân tích dữ liệu**

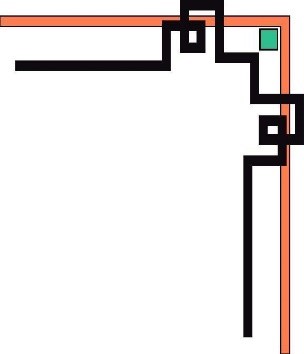
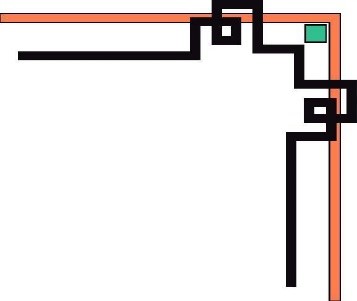
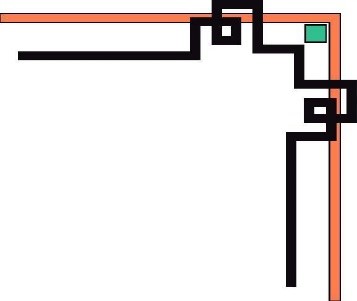
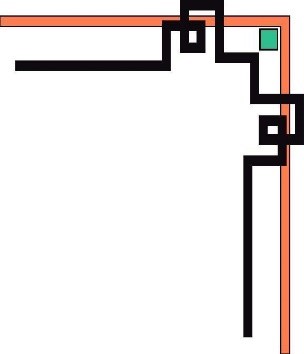
**Số phách**

*Do hội đồng chấm thi ghi*

*(*

*)*

**Thừa Thiên Huế, tháng….năm 2023**



ĐẠI HỌC HUẾ

**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**





**(**

**Bìa phụ**

**2)**

**BÁO CÁO ĐỒ ÁN**

**Học kỳ II, năm học 2022 - 2023**

**Học phần: Các thuật toán tối ưu cho phân tích dữ liệu**

**Giảng viên hướng dẫn: Hoàng Trọng Lợi**

**Lớp:KHDL & TTNT**

**Sinh viên thực hiện: Nguyễn Chính Phi Hùng**

*(*

*ký tên và ghi rõ họ tên*

*)*

**Số phách**

*(*

*)*

*Do hội đồng chấm thi ghi*

**Thừa Thiên Huế, tháng…..năm 2023**

**ĐẠI HỌC HUẾ**

**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**



**PHIẾU ĐÁNH GIÁ ĐỒ ÁN/TIỂU LUẬN/BÀI TẬP LỚN**

**Học kỳ II, năm học 2022 - 2023**

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1** | **Cán bộ chấm thi 2** |
| **Nhận xét:**  **Điểm đánh giá của CBCT1:**  Bằng số:  Bằng chữ: | **Nhận xét:**  **Điểm đánh giá của CBCT2:**  Bằng số:  Bằng chữ: |

Điểm kết luận:

Bằng số:

Bằng chữ:

*Thừa Thiên Huế, ngày tháng năm 2023*

**Cán bộ chấm thi 1 Cán bộ chấm thi 2**

*(Ký và ghi rõ họ và tên) (Ký và ghi rõ họ và tên)*

**Đề tài: Thuật toán Stochastis Gradient Descent và ứng dụng**

**Mục Lục**

[**Chương I. Giới thiệu về thuật toán Stochastic Gradient Descent 6**](#_Toc135318186)

[**1. Stochastic gradient descent là gì? 6**](#_Toc135318187)

[**2. Ưu điểm và nhược điểm của Stochastic Gradient Descent (SGD) 8**](#_Toc135318188)

[**Chương II. Ứng dựng của thuật toán Stochastic Gradient Descent (SGD) 8**](#_Toc135318189)

[**1. Hỗ trợ máy vector 8**](#_Toc135318190)

[**2. Hồi quy logistic 9**](#_Toc135318191)

[**3. Đảo ngược dạng sóng đầy đủ (FWI) 9**](#_Toc135318192)

[**Chương III. Thực hành kiểm thử thuật toán Stochastic Gradient Descent 9**](#_Toc135318193)

[**Chương IV. Tài liệu tham khảo 13**](#_Toc135318194)

# **Chương I. Giới thiệu về thuật toán Stochastic Gradient Descent.**

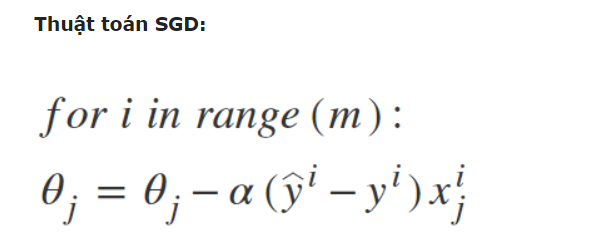
## **1. Stochastic gradient descent là gì?**

- Stochastic gradient descent (SGD) là một thuật toán tối ưu hóa thường được sử dụng trong các ứng dụng học máy để tìm các tham số mô hình tương ứng với sự phù hợp nhất giữa đầu ra được dự đoán và thực tế. Đó là một kỹ thuật không chính xác nhưng mạnh mẽ.

- SGD được cho là thuật toán quan trọng nhất khi đào tạo mạng lưới thần kinh sâu. Mặc dù hóa thân ban đầu của SGD đã được giới thiệu nhiều năm trước, nó vẫn là công cụ cho phép chúng ta đào tạo các mạng lớn để tìm hiểu các mẫu từ các điểm dữ liệu.

- Giả sử, bạn có một triệu mẫu trong tập dữ liệu của mình, vì vậy nếu bạn sử dụng kỹ thuật tối ưu hóa Gradient Descent điển hình, bạn sẽ phải sử dụng tất cả một triệu mẫu để hoàn thành một lần lặp trong khi thực hiện Gradient Descent, và nó phải được thực hiện cho mỗi lần lặp cho đến khi đạt đến cực tiểu. Do đó, nó trở nên rất tốn kém về mặt tính toán để thực hiện.

- Vấn đề này được giải quyết bằng Stochastic Gradient Descent. Trong SGD, nó chỉ sử dụng một mẫu duy nhất, tức là kích thước lô của một mẫu, để thực hiện mỗi lần lặp. Mẫu được xáo trộn ngẫu nhiên và được chọn để thực hiện lặp lại.

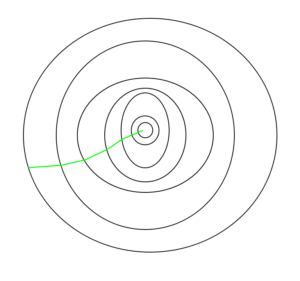


Vì vậy, trong SGD, chúng ta tìm ra gradient của chi phí hàm của một ví dụ duy nhất tại mỗi lần lặp thay vì tổng gradient của hàm chi phí của tất cả các ví dụ.

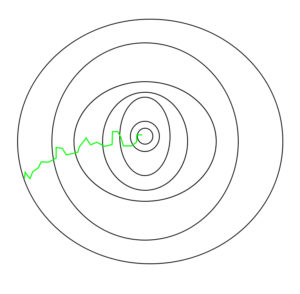
-->

Trong SGD, vì chỉ có một mẫu từ tập dữ liệu được chọn ngẫu nhiên cho mỗi lần lặp, nên đường dẫn mà thuật toán thực hiện để đến cực tiểu thường ồn ào hơn so với thuật toán Gradient Descent điển hình của bạn. Nhưng điều đó không quan trọng lắm vì đường đi của thuật toán không quan trọng, miễn là chúng ta đạt đến cực tiểu và với thời gian đào tạo ngắn hơn đáng kể.

Đường dẫn được thực hiện bởi Batch Gradient Descent –



Đường dẫn được thực hiện bởi Stochastic Gradient Descent –



Một điều cần lưu ý là, vì SGD nói chung ồn hơn so với Gradient Descent điển hình, nên nó thường mất một số lần lặp lại cao hơn để đạt đến cực tiểu, vì tính ngẫu nhiên trong dữ liệu của nó. Mặc dù nó yêu cầu số lần lặp lại cao hơn để đạt đến cực tiểu so với Gradient Descent truyền thống, nhưng về mặt tính toán nó vẫn ít tốn kém hơn nhiều so với Gradient Descent truyền thống. Do đó, trong hầu hết các tình huống, SGD được ưu tiên hơn Batch Gradient Descent để tối ưu hóa thuật toán học tập.

## **2. Ưu điểm và nhược điểm của Stochastic Gradient Descent (SGD)**

- Stochastic Gradient Descent (SGD) là một biến thể của thuật toán Gradient Descent được sử dụng để tối ưu hóa các mô hình máy học. Trong biến thể này, chỉ một ví dụ đào tạo ngẫu nhiên được sử dụng để tính toán độ dốc và cập nhật các tham số ở mỗi lần lặp. Dưới đây là một số ưu điểm và nhược điểm của việc sử dụng SGD:

**\* Ưu điểm** :

**- Tốc độ:** SGD nhanh hơn các biến thể khác của Gradient Descent như Batch Gradient Descent và Mini-Batch Gradient Descent vì nó chỉ sử dụng một ví dụ để cập nhật các tham số.

**- Hiệu quả bộ nhớ:** Vì SGD cập nhật từng tham số cho từng ví dụ đào tạo nên nó tiết kiệm bộ nhớ và có thể xử lý các tập dữ liệu lớn không thể vừa với bộ nhớ.

**- Tránh cực tiểu cục bộ:** Do các bản cập nhật ồn ào trong SGD, nó có khả năng thoát khỏi cực tiểu cục bộ và hội tụ về cực tiểu toàn cục.

**\* Nhược điểm:**

**- Cập nhật ồn ào:** Các cập nhật trong SGD ồn ào và có phương sai cao, điều này có thể làm cho quá trình tối ưu hóa kém ổn định hơn và dẫn đến dao động xung quanh mức tối thiểu.

**- Hội tụ chậm:** SGD có thể yêu cầu nhiều lần lặp lại để hội tụ ở mức tối thiểu vì nó cập nhật từng tham số cho từng ví dụ đào tạo tại một thời điểm.

**- Độ nhạy với tốc độ học:** Việc lựa chọn tốc độ học có thể rất quan trọng trong SGD vì sử dụng tốc độ học cao có thể khiến thuật toán vượt quá mức tối thiểu, trong khi tốc độ học thấp có thể khiến thuật toán hội tụ chậm.

**- Ít chính xác hơn:** Do các bản cập nhật ồn ào, SGD có thể không hội tụ đến mức tối thiểu toàn cầu chính xác và có thể dẫn đến một giải pháp dưới mức tối ưu. Điều này có thể được giảm thiểu bằng cách sử dụng các kỹ thuật như lập lịch tỷ lệ học tập và cập nhật dựa trên động lượng

# **Chương II. Ứng dựng của thuật toán Stochastic Gradient Descent (SGD).**

SGD, thường được gọi là nền tảng cho học sâu, là một thuật toán để đào tạo một loạt các mô hình trong học máy. [Học sâu](https://en.wikipedia.org/wiki/Deep_learning" \o "wikipedia:Deep learning) là một kỹ thuật học máy dạy máy tính làm những gì tự nhiên đến với con người. Trong học sâu, một mô hình máy tính học cách thực hiện các tác vụ phân loại trực tiếp từ hình ảnh, văn bản hoặc âm thanh. Các mô hình được đào tạo bằng cách sử dụng một tập hợp lớn dữ liệu được gắn nhãn và kiến trúc mạng nơ-ron chứa nhiều lớp. Mạng lưới thần kinh tạo nên xương sống của các thuật toán học sâu. Một mạng lưới thần kinh bao gồm nhiều hơn ba lớp sẽ bao gồm các đầu vào và đầu ra có thể được coi là một thuật toán học sâu. Do hiệu quả của SGD trong việc xử lý các bộ dữ liệu quy mô lớn, đây là phương pháp phổ biến nhất để đào tạo [các mạng nơ-ron sâu](https://en.wikipedia.org/wiki/Deep_learning" \l "Deep_neural_networks). Hơn nữa, SGD đã nhận được sự chú ý đáng kể và được áp dụng cho phân loại văn bản và [xử lý ngôn ngữ tự nhiên](https://en.wikipedia.org/wiki/Natural_language_processing). Nó phù hợp nhất cho các vấn đề tối ưu hóa không bị hạn chế và là cách chính để đào tạo các mô hình tuyến tính lớn trên các tập dữ liệu rất lớn. Việc thực hiện giảm độ dốc ngẫu nhiên bao gồm các khu vực trong hồi quy [sườn núi](https://en.wikipedia.org/wiki/Tikhonov_regularization) và [hồi quy logistic](https://en.wikipedia.org/wiki/Logistic_regression) thường xuyên. Các vấn đề khác, chẳng hạn như Lasso và các máy vectơ hỗ trợ có thể được giải quyết bằng cách giảm độ dốc ngẫu nhiên.

## **Hỗ trợ máy vector**

SGD là một cách tiếp cận đơn giản nhưng rất hiệu quả để phù hợp với các bộ phân loại tuyến tính và hồi quy dưới các chức năng lồi như [Máy vector hỗ trợ](https://en.wikipedia.org/wiki/Support_vector_machine) (SVM) (tuyến tính). Máy vector hỗ trợ là một mô hình học máy được giám sát sử dụng các thuật toán phân loại cho các vấn đề phân loại hai nhóm. SVM tìm thấy cái được gọi là siêu mặt phẳng tách biệt: một siêu mặt phẳng (một đường, trong trường hợp hai chiều) ngăn cách hai lớp điểm với nhau. Đây là một thuật toán phân loại nhanh và đáng tin cậy, hoạt động rất tốt với lượng dữ liệu hạn chế để phân tích. Tuy nhiên, vì SVM tốn kém về mặt tính toán, các ứng dụng phần mềm thường không cung cấp đủ hiệu suất để đáp ứng yêu cầu về thời gian cho lượng lớn dữ liệu. Để cải thiện khả năng mở rộng SVM liên quan đến kích thước của tập dữ liệu, các thuật toán SGD được sử dụng như một quy trình đơn giản hóa để đánh giá độ dốc của hàm.

## **Hồi quy logistic**

Hồi quy logistic mô hình hóa xác [suất](https://en.wikipedia.org/wiki/Probability) cho các vấn đề phân loại với hai kết quả có thể xảy ra. Đó là một phần mở rộng của mô hình hồi quy tuyến tính cho các vấn đề phân loại. Nó là một kỹ thuật thống kê với các biến đầu vào là các biến liên tục và biến đầu ra là biến nhị phân. Nó là một lớp hồi quy trong đó biến độc lập được sử dụng để dự đoán biến phụ thuộc. Mục tiêu của việc đào tạo một mô hình học máy là giảm thiểu sự mất mát hoặc sai sót giữa các sự thật cơ bản và dự đoán bằng cách thay đổi các tham số có thể đào tạo. Hồi quy logistic có hai giai đoạn: đào tạo và thử nghiệm. Hệ thống, cụ thể là trọng số w và b, được đào tạo bằng cách sử dụng giảm độ dốc ngẫu nhiên và tổn thất entropy chéo.

## **Đảo ngược dạng sóng đầy đủ (FWI)**

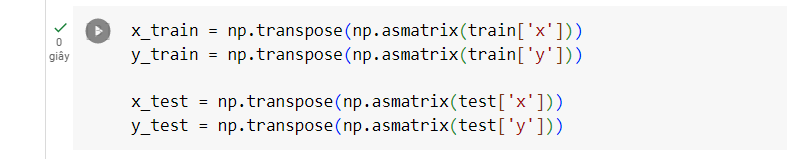
Đảo ngược dạng sóng đầy đủ (FWI) là một quá trình [hình ảnh địa chấn](https://en.wikipedia.org/wiki/Geophysical_imaging) bằng cách lấy thông tin từ các thông số vật lý của mẫu. Các công ty sử dụng quy trình này để tạo ra các mô tả tốc độ cao có độ phân giải cao về các hoạt động dưới bề mặt. SGD hỗ trợ quá trình này vì nó có thể xác định mức tối thiểu và mức tối thiểu toàn cầu tổng thể trong thời gian ngắn hơn vì có nhiều mức tối thiểu cục bộ.

# **Chương III. Thực hành kiểm thử thuật toán Stochastic Gradient Descent.**

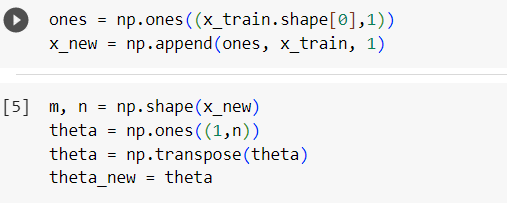
* Nhập thư viện và đọc dataset



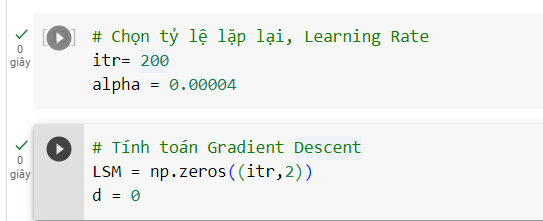
* Thay đổi tập dữ liệu thành ma trận và chuyển đổi chúng



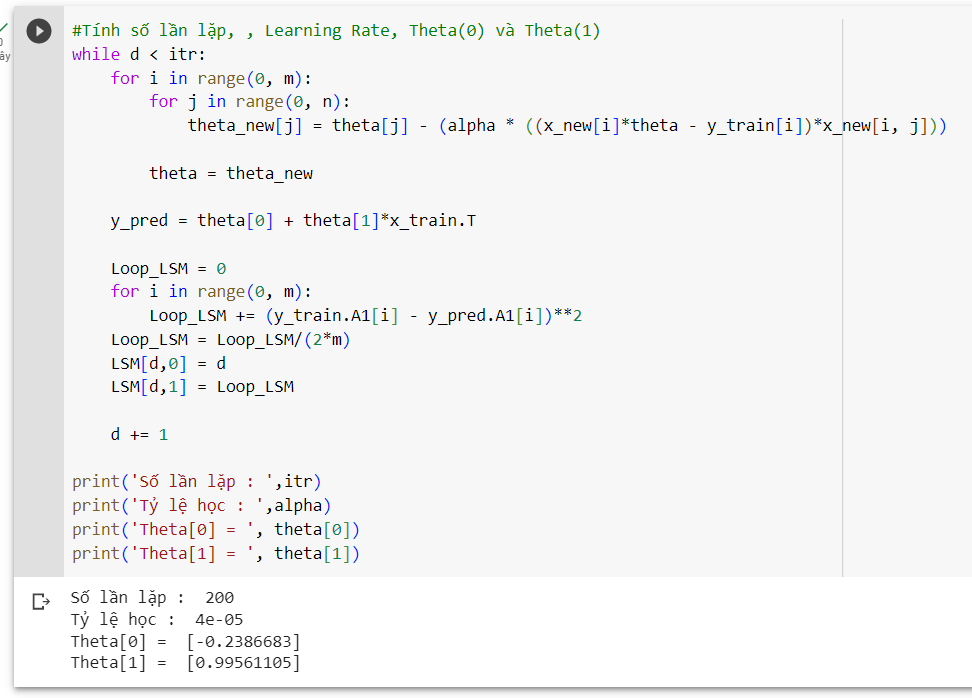
* Thêm X0 vào dữ liệu, tạo m , n



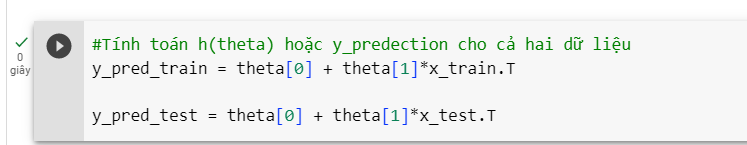
* Chọn tỷ lệ lặp lại, Learning Rate và tính toán Gradient Descent



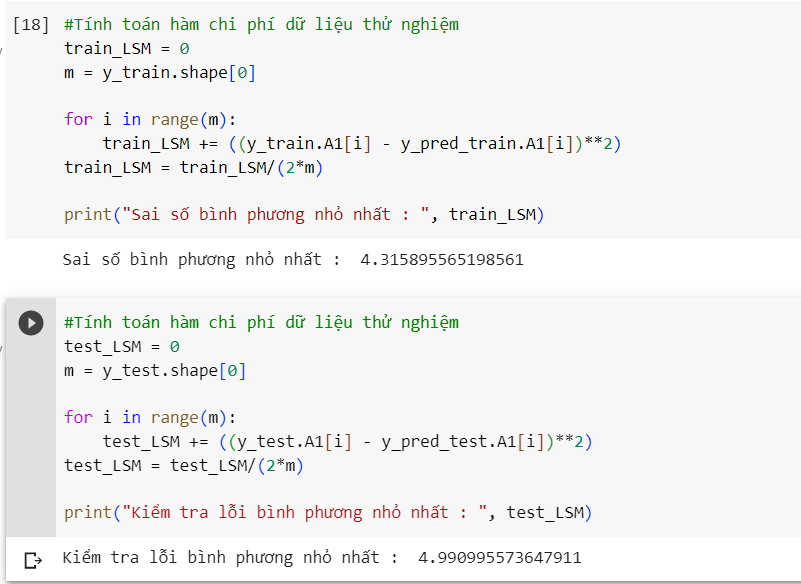
* Tính số lần lặp, , Learning Rate, Theta(0) và Theta(1)



* Tính toán h(theta) hoặc y\_predection cho cả hai dữ liệu



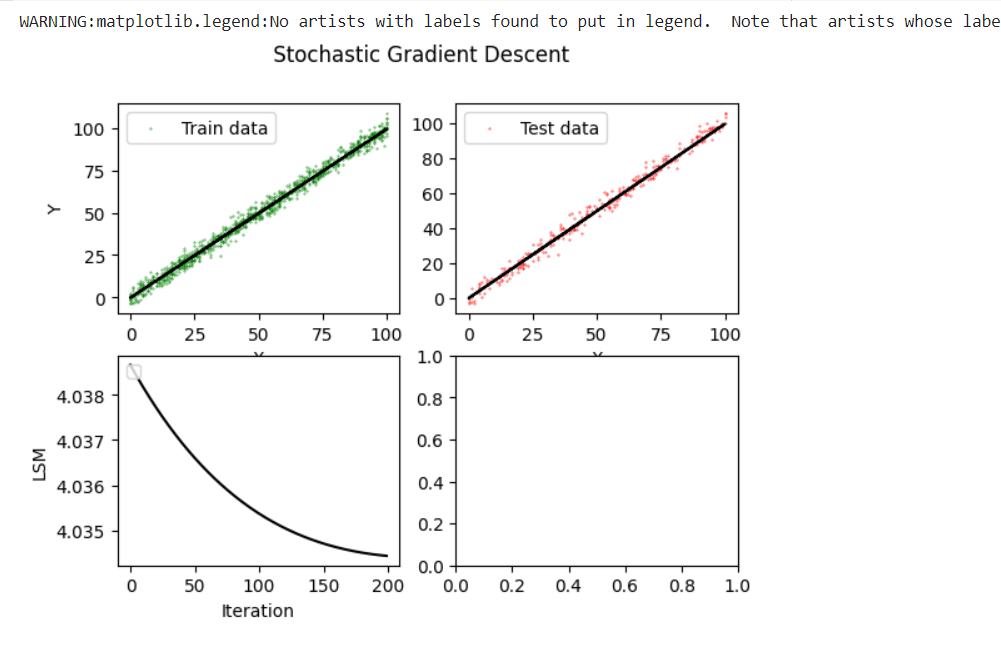
* Tính toán hàm chi phí dữ liệu thử nghiệm



* Vẽ dữ liệu và Gradient



Kết quả:



**Chương IV. Tài liệu tham khảo**

(1)

<https://optimization.cbe.cornell.edu/index.php?title=Stochastic_gradient_descent&fbclid=IwAR1BgZZ-68cjkTHQ8Am-uBarxmduXp7NFMiC1-yw-qPUOg9R4hD73kVB__c>

(2)

<https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/>