ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC

THUẬT TOÁN HƯỚNG GIẢM GRADIENT VÀ ỨNG DỤNG TRONG BÀI TOÁN HỒI QUY LOGISTIC

ĐỒ ÁN I

Chuyên ngành: HỆ THỐNG THÔNG TIN QUẢN LÝ

Chuyên sâu: Toán ứng dụng

Giảng viên hướng dẫn:	**TS. Phạm Thị Hoài**
Sinh viên thực hiện:	**Phạm Ngọc Thái**
Lớp:	**Hệ thống thông tin 02 – K65**

HÀ NỘI – 2023

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

1. **Mục tiêu và nội dung của đồ án**

   (a) Mục tiêu: Đề tài nghiên cứu về thuật toán hướng giảm gradient và ứng dụng trong bài toán hồi quy logistic;

   (b) Nội dung: Trình bày về các kiến thức toán học nền tảng và thuật toán hướng giảm gradient, thảo luận về các khía cạnh quan trọng của thuật toán cũng như hướng ứng dụng của thuật toán trong bài toán hồi quy logistic

2. **Kết quả đạt được**

   (a) Dịch, tổng hợp và trình bày lại kết quả trong các tài liệu nghiên cứu về quy hoạch toán học và khoa học dữ liệu;

   (b) Đưa ra phân tích đánh giá toàn diện về các khía cạnh của thuật toán hướng giảm gradient;

   (c) Ứng dụng các kiến thức lý thuyết vào việc phân tích dữ liệu và lập trình tính toán để xử lý vấn đề thực tiễn.

3. **Ý thức làm việc của sinh viên:**

   (a) Có ý thức, trách nhiệm cao trong quá trình học tập và làm đồ án;

   (b) Đam mê, ham học hỏi và tìm hiểu những kiến thức chuyên sâu liên quan đến đề tài đồ án;

   (c) Có khả năng tự nghiên cứu, dịch thuật tài liệu chuyên ngành và chế bản LaTeX tốt. Hoàn thành tốt đồ án theo đúng yêu cầu của giảng viên hướng dẫn.

Kết luận: Đồ án đã được hoàn thành rất tốt.

Hà Nội, ngày ... tháng ... năm 2023

Giảng viên hướng dẫn

TS. Phạm Thị Hoài

Lời cảm ơn

Em xin trân trọng cảm ơn TS. Phạm Thị Hoài đã tận tình hướng dẫn em hoàn thành đồ án này. Kính chúc cô thành công trong sự nghiệp giảng dạy và nghiên cứu.

Em xin chân thành cảm ơn!

Hà Nội, tháng 7 năm 2023

Tác giả đồ án

Phạm Ngọc Thái

Mục lục

Bảng ký hiệu

$\mathbb{R}$	tập các số thực
$\mathbb{R}^n$	không gian Euclid n chiều
$\langle x, y\rangle$	tích vô hướng của x và y
$\|x\|$	chuẩn Euclid của x
$[x_1, x_2]$	đoạn thẳng nối hai điểm x_1 và x_2
$epi(f)$	epigraph của hàm f
$hypo(f)$	hypograph của hàm f
$f'(x_0, d)$	đạo hàm theo hướng d của hàm f tại điểm x_0
$\nabla f(x)$	véctơ gradient của hàm f tại điểm x
$\nabla^2 f(x)$	ma trận Hesse của hàm f tại điểm x
$\sigma(x)$	hàm sigmoid theo biến x
A^T	ma trận chuyển vị của ma trận A

Mở đầu

Trong mọi lĩnh vực của đời sống, từ kinh tế đến kỹ thuật, việc lựa chọn ra phương án tốt nhất (theo tiêu chí xác định) từ tập hợp các phương án cho trước luôn là một vấn đề quan trọng và nhận được nhiều sự đầu tư. Trong toán học, vấn đề này được trình bày dưới dạng tìm giá trị tối đa hoặc tối thiểu của một hàm số. Đây chính là khái niệm quy hoạch toán học. Đã có rất nhiều phương pháp được nghiên cứu và phát triển để giải các bài toán quy hoạch. Một trong số các phương pháp nổi tiếng và hiệu quả nhất chính là thuật toán hướng giảm gradient.

Trong khuôn khổ Đồ án này, em xin trình bày một số tìm hiểu và nghiên cứu về thuật toán hướng giảm gradient cũng như ứng dụng của nó trong bài toán hồi quy logistic. Các kiến thức này có thể được xem là cơ sở và nền tảng cho những nghiên cứu sâu hơn về các lĩnh vực toán tối ưu và học máy.

Đầu tiên, các kiến thức cơ sở về giải tích tối ưu của Đồ án sẽ được trình bày trong Chương 1. Chương 2 sẽ đi sâu vào ý tưởng của thuật toán hướng giảm gradient và thảo luận về các khía cạnh của thuật toán. Để thể hiện tính ứng dụng của thuật toán, Chương 3 sẽ trình bày ví dụ thực tiễn của thuật toán trong bài toán hồi quy logistic.

Chương 1

Kiến thức chuẩn bị

Chương 1 sẽ trình bày các kiến thức toán học cơ bản của thuật toán hướng giảm gradient, bao gồm một số khái niệm giải tích trên hàm nhiều biến, giải tích lồi và điều kiện tối ưu của bài toán quy hoạch phi tuyến không ràng buộc. Nội dung của Chương 1 được viết dựa trên kiến thức từ cuốn sách [1] và [2].

1.1 Một số khái niệm giải tích trên hàm nhiều biến

1.1.1 Đạo hàm riêng

Định nghĩa 1.1 (đạo hàm riêng - xem [1]) Cho hàm số f xác định trên tập mở $S \subset \mathbb{R}^n$, $x^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})^T \in S$ và $h \in \mathbb{R}$ đủ nhỏ. Giới hạn

$$\lim_{h \to 0} \frac{f(x_1^{(0)}, \dots, x_{i-1}^{(0)}, x_i^{(0)} + h, x_{i+1}^{(0)}, \dots, x_n^{(0)}) - f(x_1^{(0)}, \dots, x_n^{(0)})}{h}$$

nếu tồn tại thì được gọi là *đạo hàm riêng* của f theo biến x_i tại điểm $x^{(0)}$, ký hiệu là $\dfrac{\partial f(x^{(0)})}{\partial x_i}$ hay $f'_{x_i}(x^{(0)})$.

Định nghĩa 1.2 (đạo hàm riêng cấp hai - xem [1]) Cho hàm số f xác định trên tập mở $S \subset \mathbb{R}^n$. Giả sử đạo hàm riêng $\dfrac{\partial f(x)}{\partial x_i}$ tồn tại $\forall x \in S$. Khi đó, tại $x^{(0)} \in S$, đạo hàm riêng theo biến x_j của hàm $\dfrac{\partial f}{\partial x_i}$, nếu tồn tại, thì được gọi là *đạo hàm riêng cấp hai* theo các biến x_i và x_j của hàm f tại x_0, ký hiệu là $\dfrac{\partial^2 f(x^{(0)})}{\partial x_i \partial x_j}$ hay $f''_{x_i x_j}(x^{(0)})$.

1.1.2 Gradient và ma trận Hesse

Định nghĩa 1.3 (gradient - xem [1]) Cho hàm số f xác định trên tập mở $S \subset \mathbb{R}^n$. Giả sử rằng tại $x^{(0)}$, các đạo hàm riêng của hàm f theo mọi biến đều tồn tại. Khi đó, véc tơ

$$\left(\frac{\partial f(x^{(0)}}{\partial x_1}, \frac{\partial f(x^{(0)}}{\partial x_2}, \ldots, \frac{\partial f(x^{(0)}}{\partial x_n}\right)^T$$

được gọi là *gradient* của f tại $x^{(0)}$, ký hiệu là $\nabla f(x^{(0)})$

Định nghĩa 1.4 (ma trận Hesse - xem [1]) Cho hàm số f xác định trên tập mở $S \subset \mathbb{R}^n$. Nếu các đạo hàm riêng cấp hai theo mọi biến của hàm f tại $x^{(0)}$ đều tồn tại thì ma trận

$$\begin{bmatrix} \frac{\partial^2 f(x^{(0)}}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x^{(0)}}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^{(0)}}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x^{(0)}}{\partial x_n \partial x_n} \end{bmatrix}$$

được gọi là *ma trận Hesse* của f tại $x^{(0)}$, ký hiệu là $\nabla^2 f(x^{(0)})$.

1.1.3 Tính khả vi

Định nghĩa 1.5 (khả vi - xem [1]) Cho hàm số f xác định trên tập mở $S \subset \mathbb{R}^n$. Hàm f được gọi là *khả vi* tại $x_0 \in \mathbb{R}^n$ nếu $\forall x_0 \neq d \in \mathbb{R}^n$, $\|d\|$ đủ nhỏ:

$$f(x_0 + d) = f(x_0) + \langle \nabla f(x_0), d \rangle + o(\|d\|)$$

Biểu thức tương đương:

$$\lim_{\|d\| \to 0} \frac{f(x_0 + d) - f(x_0) - \langle \nabla f(x_0), d \rangle}{\|d\|} = 0$$

Mệnh đề 1.1 (xem [1]) *Cho hàm số f xác định trên tập mở $S \subset \mathbb{R}^n$:*

(i) *Nếu f khả vi tại $x_0 \in S$ thì f liên tục và có các đạo hàm riêng tại x_0;*

(ii) *Nếu f có các đạo hàm riêng liên tục trong lân cận của $x_0 \in S$ thì f khả vi tại x_0.*

Định nghĩa 1.6 (khả vi hai lần - xem [1]) Cho hàm f xác định trên tập mở $S \subset \mathbb{R}^n$. Hàm f được gọi là *khả vi hai lần* tại $x_0 \in S$ nếu $\forall x_0 \neq d \in \mathbb{R}^n$, $\|d\|$ đủ nhỏ:

$$f(x_0 + d) = f(x_0) + \langle \nabla f(x_0), d \rangle + \frac{d^T \nabla^2 f(x_0) d}{2} + o(\|d\|^2)$$

1.1.4 Đạo hàm theo hướng

Định nghĩa 1.7 (đạo hàm theo hướng - xem [1]) Cho hàm f xác định trên $\mathbb{R}^n$ và véc tơ $d \in \mathbb{R}^n \setminus \{0\}$. Giới hạn

$$\lim_{t \to 0^+} \frac{f(x_0 + td) - f(x_0)}{t}$$

nếu tồn tại (hữu hạn hoặc vô cùng), được gọi là *đạo hàm theo hướng* d của hàm f tại điểm $x_0 \in \mathbb{R}^n$, ký hiệu là $f'(x_0, d)$.

Mệnh đề 1.2 (xem [1]) *Cho hàm f xác định trên $\mathbb{R}^n$. Nếu f khả vi tại $x_0 \in \mathbb{R}^n$ thì $\forall d \in \mathbb{R}^n \setminus \{0\}$:*

$$f'(x_0, d) = \langle \nabla f(x_0), d \rangle.$$

Mệnh đề 1.3 (xem [1]) *Đạo hàm theo hướng d của hàm f tại điểm x_0 là lớn nhất khi hướng $d = \dfrac{\nabla f(x_0)}{||\nabla f(x_0)||}$ và nhỏ nhất khi hướng $d = -\dfrac{\nabla f(x_0)}{||\nabla f(x_0)||}$ trong tất cả các hướng có $||d|| = 1$. Nói cách khác, giá trị hàm tăng nhanh nhất theo hướng gradient và giảm nhanh nhất theo hướng ngược với gradient.*

1.1.5 Hướng giảm

Định nghĩa 1.8 (hướng giảm - xem [1]) Cho $x_0 \in \mathbb{R}^n$. Véc tơ $d \in \mathbb{R}^n$ được gọi là *hướng giảm* của hàm f tại x_0 nếu $\exists \varepsilon > 0$ sao cho $\forall t \in (0; \varepsilon)$:

$$f(x_0 + td) < f(x_0).$$

Mệnh đề 1.4 (xem [1]) *Cho hàm f khả vi trên $\mathbb{R}^n$, điểm $x_0 \in \mathbb{R}^n$ và hướng $d \in \mathbb{R}^n$. Nếu:*

$$\langle f(x_0), d \rangle < 0$$

thì d là hướng giảm của f tại x_0.

1.2 Giải tích lồi

1.2.1 Các định nghĩa

Định nghĩa 1.9 (tập lồi - xem [1]) Tập $S \subset \mathbb{R}^n$ được gọi là *tập lồi* nếu $\forall x_1, x_2 \in S; \lambda \in [0,1]$:

$$\lambda x_1 + (1-\lambda)x_2 \in S,$$

tức là tập S chứa trọn đoạn thẳng nối hai điểm bất kỳ thuộc nó.

Định nghĩa 1.10 (hàm lồi - xem [1]) Hàm f được gọi là *hàm lồi* xác định trên tập lồi $S \subset \mathbb{R}^n$ nếu $\forall x_1, x_2 \in S; \lambda \in [0,1]$:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2),$$

tức là đoạn thẳng nối hai điểm bất kỳ trên đồ thị hàm f không nằm dưới đồ thị đó.

Định nghĩa 1.11 (hàm lồi chặt - xem [1]) Hàm f được gọi là *hàm lồi chặt* xác định trên tập lồi $S \subset \mathbb{R}^n$ nếu $\forall x_1 \neq x_2 \in S; \lambda \in (0,1)$:

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2).$$

Định nghĩa 1.12 (hàm lõm - xem [1]) Hàm f được gọi là *hàm lõm (chặt)* trên S nếu hàm $-f$ lồi (chặt) trên S.

Mệnh đề 1.5 (xem [1]) *Cho hàm f xác định trên tập lồi $S \subset \mathbb{R}^n$. Khi đó:*

(i) *Hàm f lồi trên S khi và chỉ khi tập epigraph của f:*

$$epi(f) := \{(x,\mu) : x \in S, \mu \geq f(x) \in \mathbb{R}\}$$

là tập lồi;

(ii) *Hàm f lõm trên S khi và chỉ khi tập hypograph của f:*

$$hypo(f) := \{(x,\mu) : x \in S, \mu \leq f(x) \in \mathbb{R}\}$$

là tập lồi.

Mệnh đề 1.6 (xem [1]) *Cho hàm f xác định trên tập lồi $S \subset \mathbb{R}^n$. Khi đó, $\forall \alpha \in \mathbb{R}$:*

(i) *Nếu f là hàm lồi thì tập mức dưới:*

$$L_\alpha(f) := \{x \in S : f(x) \leq \alpha\}$$

là tập lồi;

(ii) *Nếu* f *là hàm lõm thì tập mức trên:*

$$L^{\alpha}(f) := \{x \in S : f(x) \geq \alpha\}$$

là tập lồi.

1.2.2 Tiêu chuẩn nhận biết hàm lồi khả vi

Định lý 1.1 (xem [1]) *Cho hàm* f *khả vi trên tập lồi mở* $S \subset \mathbb{R}^n$. *Khi đó:*

(i) *Hàm* f *là hàm lồi trên* S *khi và chỉ khi* $\forall x_1, x_2 \in S$:

$$f(x_2) - f(x_1) \geq \langle \nabla f(x_1), x_2 - x_1 \rangle;$$

(ii) *Hàm* f *là hàm lõm trên* S *khi và chỉ khi* $\forall x_1, x_2 \in S$:

$$f(x_2) - f(x_1) \leq \langle \nabla f(x_1), x_2 - x_1 \rangle;$$

Định lý 1.2 (xem [1]) *Cho hàm* f *khả vi hai lần trên tập lồi mở* $S \subset \mathbb{R}^n$. *Khi đó:*

(i) *Hàm* f *là hàm lồi trên* S *khi và chỉ khi* $\nabla^2 f(x)$ *nửa xác định dương trên* S, *tức là* $\forall x \in S, y \in \mathbb{R}^n$:

$$y^T \nabla^2 f(x) y \geq 0;$$

(ii) *Hàm* f *là hàm lồi chặt trên* S *khi và chỉ khi* $\nabla^2 f(x)$ *xác định dương trên* S, *tức là* $\forall x \in S, y \in \mathbb{R}^n \setminus \{0\}$:

$$y^T \nabla^2 f(x) y > 0;$$

(iii) *Hàm* f *là hàm lõm trên* S *khi và chỉ khi* $\nabla^2 f(x)$ *nửa xác định âm trên* S, *tức là* $\forall x \in S, y \in \mathbb{R}^n$:

$$y^T \nabla^2 f(x) y \leq 0;$$

(iv) *Hàm* f *là hàm lõm chặt trên* S *khi và chỉ khi* $\nabla^2 f(x)$ *xác định âm trên* S, *tức là* $\forall x \in S, y \in \mathbb{R}^n \setminus \{0\}$:

$$y^T \nabla^2 f(x) y < 0.$$

1.2.3 Hướng giảm trên hàm lồi

Mệnh đề 1.7 (xem [1]) *Cho hàm lồi* f *khả vi trên* $\mathbb{R}^n$, *điểm* $x_0 \in \mathbb{R}^n$ *và hướng* $d \in \mathbb{R}^n$. *Khi đó:*

$$\langle \nabla f(x_0), d \rangle < 0$$

khi và chỉ khi d là hướng giảm của f tại x_0.

1.3 Bài toán quy hoạch phi tuyến không ràng buộc và điều kiện tối ưu

1.3.1 Bài toán quy hoạch phi tuyến không ràng buộc

Định nghĩa 1.13 (quy hoạch toán học - xem [1]) Cho hàm f xác định trên $S \subset \mathbb{R}^n$. *Bài toán tối ưu* tổng quát có dạng:

$$\min_{x \in S} f(x),$$

trong đó, S được gọi là *tập nghiệm chấp nhận được (tập ràng buộc)*, mỗi điểm $x \in S$ được gọi là một *nghiệm chấp nhân được (phương án chấp nhận được)* và f được gọi là *hàm mục tiêu.*

Định nghĩa 1.14 (tối ưu toàn cục - xem [2]) Cho hàm f xác định trên $S \subset \mathbb{R}^n$ và điểm $x^* \in S$. Khi đó:

(i) Điểm $x^* \in S$ được gọi là *nghiệm cực tiểu toàn cục* của f trên S nếu:

$$f(x^*) \leq f(x) \quad \forall x \in S \setminus \{x^*\};$$

(ii) Điểm $x^* \in S$ được gọi là *nghiệm cực tiểu toàn cục chặt* của f trên S nếu:

$$f(x^*) < f(x) \quad \forall x \in S \setminus \{x^*\};$$

(iii) Điểm $x^* \in S$ được gọi là *nghiệm cực đại toàn cục* của f trên S nếu:

$$f(x^*) \geq f(x) \quad \forall x \in S \setminus \{x^*\};$$

(iv) Điểm $x^* \in S$ được gọi là *nghiệm cực đại toàn cục chặt* của f trên S nếu:

$$f(x^*) > f(x) \quad \forall x \in S \setminus \{x^*\}.$$

Định nghĩa 1.15 (tối ưu địa phương - xem [2]) Cho hàm f xác định trên $S \subset \mathbb{R}^n$ và điểm $x^* \in S$. Khi đó, nếu tồn tại lân cận $U(x^*)$ thì:

(i) x^* được gọi là *nghiệm cực tiểu địa phương* của f trên S nếu:

$$f(x^*) \leq f(x) \quad \forall x \in S \cap U(x^*);$$

(ii) x^* được gọi là *nghiệm cực tiểu địa phương chặt* của f trên S nếu:

$$f(x^*) < f(x) \quad \forall x \in S \cap U(x^*);$$

(iii) x^* được gọi là *nghiệm cực đại địa phương* của f trên S nếu:

$$f(x^*) \geq f(x) \quad \forall x \in S \cap U(x^*);$$

(iv) x^* được gọi là *nghiệm cực đại địa phương chặt* của f trên S nếu:

$$f(x^*) > f(x) \quad \forall x \in S \cap U(x^*).$$

Định nghĩa 1.16 (quy hoạch phi tuyến không ràng buộc - xem [1]) Cho hàm phi tuyến f xác định trên $S \subset \mathbb{R}^n$. *Bài toán quy hoạch phi tuyến không ràng buộc* có dạng:

$$\min_{x \in \mathbb{R}^n} f(x).$$

1.3.2 Điều kiện tối ưu bậc nhất

Định lý 1.3 (xem [2]) *Cho hàm f khả vi trên $\mathbb{R}^n$. Nếu $x^* \in \mathbb{R}^n$ là nghiệm cực tiểu (hoặc cực đại) địa phương của hàm f trên $\mathbb{R}^n$ thì:*

$$\nabla f(x^*) = 0.$$

Định nghĩa 1.17 (điểm yên ngựa - xem [2]) Cho hàm f khả vi trên $S \subset \mathbb{R}^n$ và điểm $x^* \in S$ thỏa mãn:

$$\nabla f(x^*) = 0.$$

Nếu x^* không là nghiệm cực tiểu (hoặc cực đại) địa phương của f trên S, x^* được gọi là *điểm yên ngựa.*

Định lý 1.4 (xem [1]) *Giả sử f là hàm lồi khả vi trên $\mathbb{R}^n$. Khi đó, $x^* \in \mathbb{R}^n$ là nghiệm cực tiểu toàn cục của hàm f trên $\mathbb{R}^n$ khi và chỉ khi:*

$$\nabla f(x^*) = 0.$$

1.3.3 Điều kiện tối ưu bậc hai

Định lý 1.5 (xem [1]) *Giả sử hàm f khả vi liên tục hai lần trên $\mathbb{R}^n$. Khi đó:*

(i) *Nếu* $x^* \in \mathbb{R}^n$ *là điểm cực tiểu địa phương của* f *trên* $\mathbb{R}^n$ *thì:*

$$\begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ nửa xác định dương.} \end{cases}$$

(ii) *Ngược lại, nếu:*

$$\begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ xác định dương} \end{cases}$$

thì x^* *là điểm cực tiểu địa phương chặt của* f *trên* $\mathbb{R}^n$.

Mệnh đề 1.8 (xem [1]) *Cho hàm* f *khả vi liên tục hai lần trên* $\mathbb{R}^n$. *Khi đó:*

(i) *Nếu* $x^* \in \mathbb{R}^n$ *là điểm cực đại địa phương của* f *trên* $\mathbb{R}^n$ *thì:*

$$\begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ nửa xác định âm.} \end{cases}$$

(ii) *Ngược lại, nếu:*

$$\begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ xác định âm} \end{cases}$$

thì x^* *là điểm cực đại địa phương chặt của* f *trên* $\mathbb{R}^n$.

Chương 2

Thuật toán hướng giảm gradient

Chương 2 sẽ đi sâu vào ý tưởng của thuật toán hướng giảm gradient và thảo luận về các khía cạnh của thuật toán, bao gồm hướng giảm, cỡ bước và sự hội tụ. Cũng như Chương 1, nội dung của Chương 2 được viết dựa trên tham khảo từ cuốn sách [1] và [2].

2.1 Ý tưởng của thuật toán

Với $f : \mathbb{R}^n \to \mathbb{R}$ là hàm phi tuyến, khả vi trên $\mathbb{R}^n$, xét bài toán quy hoạch phi tuyến không ràng buộc:

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Theo Định lý 1.3, nếu f là hàm khả vi trên $\mathbb{R}^n$ thì điều kiện cần của nghiệm cực tiểu địa phương x^* là:

  $$\nabla f(x^*) = 0$$

  Đây là một hệ phương trình có n ẩn, n phương trình. Tuy nhiên, ngoại trừ một số trường hợp đơn giản, việc giải trực tiếp hệ này khó thực hiện được.
- Để giải quyết vấn đề này, thay vì giải trực tiếp, một thuật toán lặp được đề xuất có dạng:

  $$x^{(k+1)} = x^{(k)} + t_k * d_k \quad (k = 0, 1, 2, \dots)$$

  trong đó tại bước thứ k, t_k được gọi là cỡ bước và d_k được gọi là hướng giảm của f. Dãy $\{x^{(k)}\}$ cần hội tụ đến điểm dừng $x^* \in \mathbb{R}^n$ của hàm f, tức $\nabla f(x^*) = 0$. Đây chính là ý tưởng cơ bản của thuật toán hướng giảm.

- Thuật toán gradient thuộc lớp thuật toán hướng giảm. Với phương pháp này, hướng giảm d_k được chọn của hàm f tại điểm $x^{(k)}$ sẽ là $-\nabla f(x^{(k)})$. Theo Mệnh đề 1.3, giá trị của hàm sẽ giảm nhanh nhất theo hướng ngược với gradient, vậy nên thuật toán này còn được gọi là phương pháp hướng giảm nhanh nhất.
- Có ba cách xác định cỡ bước trong thuật toán hướng giảm gradient, tương ứng với ba thủ tục tiến hành thuật toán: thủ tục cỡ bước cố định, thủ tục tìm chính xác theo tia và thủ tục quay lui.

2.2 Thủ tục cỡ bước cố định

2.2.1 Thuật toán

Thuật toán 2.1 (Thuật toán hướng giảm gradient với thủ tục cỡ bước cố định - xem [2])

Bước khởi tạo:

– Chọn cỡ bước $t > 0$;

– Chọn sai số $\varepsilon > 0$ đủ nhỏ;

– Chọn điểm tùy ý $x^{(0)} \in \mathbb{R}^n$ sao cho $\nabla f(x^{(0)}) \neq 0$;

– Đặt $k = 0$.

Bước lặp $k = 0, 1, 2, \ldots$:

(k_1) Tính $x^{(k+1)} = x^{(k)} - t * \nabla f(x^{(k)})$;

(k_2) Tính $\nabla f(x^{(k+1)})$;

(k_3) IF $\|\nabla f(x^{(k+1)})\| < \varepsilon$ THEN lấy $x^* := x^{(k+1)}$, dừng thuật toán ELSE $k := k + 1$, quay lại Bước lặp k.

2.2.2 Đánh giá thuật toán

- Thủ tục này có ưu điểm là đơn giản, dễ cài đặt;
- Tuy nhiên, việc chọn ra hằng số làm cỡ bước là một vấn đề phức tạp. Một hằng số lớn có thể khiến thuật toán không thể hội tụ. Ngược lại, một hằng số nhỏ có thể khiến thuật toán hội tụ với tốc độ rất chậm.

2.3 Thủ tục tìm chính xác theo tia

2.3.1 Thuật toán

Thuật toán 2.2 (Thuật toán hướng giảm gradient với thủ tục tìm chính xác theo tia - xem [1])

Bước khởi tạo:

– Chọn sai số $\varepsilon > 0$ đủ nhỏ;

– Chọn điểm tùy ý $x^{(0)} \in \mathbb{R}^n$ sao cho $\nabla f(x^{(0)}) \neq 0$;

– Đặt $k = 0$.

Bước lặp $k = 0, 1, 2, \ldots$**:**

(k_1) Đặt $\varphi_k(t) := f(x^{(k)} - t_k * \nabla f(x^{(k)}))$

Tính $t_k = argmin\{\varphi_k(t), t_k > 0\}$;

(k_2) Tính $x^{(k+1)} = x^{(k)} - t_k * \nabla f(x^{(k)})$;

(k_3) Tính $\nabla f(x^{(k+1)})$;

(k_4) IF $\|\nabla f(x^{(k+1)})\| < \varepsilon$ THEN lấy $x^* := x^{(k+1)}$, dừng thuật toán

ELSE $k := k + 1$, quay lại Bước lặp k.

2.3.2 Sự hội tụ

Định lý 2.1 (xem [1]) *Cho $x^{(0)} \in \mathbb{R}^n$, hàm f khả vi liên tục trên $\mathbb{R}^n$ và có tập mức dưới $\{x \in \mathbb{R}^n | f(x) \leq f(x^{(0)})\}$ bị chặn. Khi đó, điểm tụ x^* của dãy $\{x^{(k)}\}$ được chọn như trong Thuật toán 2.2 thỏa mãn:*

$$\nabla f(x^*) = 0.$$

2.3.3 Đánh giá thuật toán

- Do t_k được lựa chọn chính xác sao cho $f(x^{(k+1)})$ nhỏ nhất, tốc độ hội tụ của thuật toán là nhanh nhất có thể;
- Tuy nhiên, việc giải bài toán cực tiểu để xác định cỡ bước theo thủ tục tìm chính xác theo tia cũng không dễ dàng, chi phí tính toán cao.

2.4 Thủ tục quay lui

2.4.1 Thuật toán

Thuật toán 2.3 (Thuật toán hướng giảm gradient với thủ tục quay lui - xem [1])

Bước khởi tạo:

– Chọn $\alpha, \beta \in (0,1)$;

– Chọn sai số $\varepsilon > 0$ đủ nhỏ;

– Chọn điểm tùy ý $x^{(0)} \in \mathbb{R}^n$ sao cho $\nabla f(x^{(0)}) \neq 0$;

– Đặt $k = 0$.

Bước lặp $k = 0, 1, 2, \ldots$:

(k_1) Đặt $t_k := 1$;

(k_2) Tính $x^{(k+1)} = x^{(k)} - t_k * \nabla f(x^{(k)})$
Tính $f(x^{(k+1)})$;

(k_3) IF $f(x^{(k+1)}) - f(x^{(k)}) \leq -\alpha t_k \|\nabla f(x^{(k)})\|^2$ THEN đến bước (k_4)
ELSE $t_k := \beta t_k$, quay về bước (k_2);

(k_4) Tính $\nabla f(x^{(k+1)})$;

(k_5) IF $\|\nabla f(x^{(k+1)})\| < \varepsilon$ THEN lấy $x^* := x^{(k+1)}$, dừng thuật toán
ELSE $k := k + 1$, quay lại Bước lặp k.

2.4.2 Sự hội tụ

Định lý 2.2 (xem [1]) *Giả sử hàm $f(x)$ bị chặn dưới, gradient $\nabla f(x)$ thỏa mãn điều kiện Lipschitz, tức tồn tại $L > 0$ sao cho $\forall x, y \in \mathbb{R}^n$:*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

Khi đó, với bất kỳ điểm xuất phát $x^{(0)}$, dãy $\{x^{(k)}\}$ được chọn như trong Thuật toán 2.3 có tính chất:

$$\|\nabla f(x^{(k)})\| \to 0 \textit{ khi } k \to \infty.$$

2.4.3 Đánh giá thuật toán

Thủ tục quay lui có thể được coi là sự kết hợp và cân bằng các ưu nhược điểm của hai thủ tục cỡ bước cố định và tìm chính xác theo tia:

- Thủ tục quay lui không cố tìm chính xác một cỡ bước "hoàn hảo" mà là một cỡ bước đủ tốt để thuật toán có tốc độ hội tụ nhanh, hơn hẳn thủ tục cỡ bước cố định;
- Bên cạnh đó, việc xác định cỡ bước trong thủ tục quay lui có chi phí thấp hơn hẳn so với việc giải bài toán cực tiểu trong thủ tục tìm chính xác theo tia;
- Tất nhiên, ở chiều ngược lại, thủ tục quay lui không thể xác định được cỡ bước "hoàn hảo" như thủ tục tìm chính xác theo tia và có chi phí xác định cỡ bước cao hơn thủ tục cỡ bước cố định;
- Trong thực tế tính toán, thủ tục quay lui thường được sử dụng.

2.5 Sự hội tụ của thuật toán hướng giảm gradient

- Đối với hàm khả vi, dù xuất phát ở điểm khởi đầu nào, thuật toán hướng giảm gradient, nếu hội tụ, sẽ hội tụ đến điểm có chuẩn của gradient bằng 0, tức là điểm cực tiểu địa phương hoặc điểm yên ngựa (theo Định lý 1.3 và Định nghĩa 1.17);
- Để thuật toán có thể hội tụ đến điểm cực tiểu toàn cục, một số kỹ thuật có thể được áp dụng như xuất phát từ nhiều điểm khởi đầu khác nhau, áp dụng quán tính vào từng bước cập nhật thuật toán, tạo thêm bước "nhảy cóc" đủ lớn khi đã đạt đến một điểm dừng,...
- Đối với hàm lồi khả vi, theo Định lý 1.4, một điểm là nghiệm cực tiểu toàn cục khi và chỉ khi điểm đó có chuẩn của gradient bằng 0. Do đó, điểm hội tụ của thuật toán hướng giảm gradient trên hàm lồi chính là nghiệm cực tiểu toàn cục.

Chương 3

Ứng dụng trong bài toán hồi quy logistic

Chương 3 sẽ trình bày ứng dụng của thuật toán hướng giảm gradient trong bài toán hồi quy logistic, thông qua ví dụ dự đoán bệnh tim mạch vành từ tập dữ liệu [3]. Kiến thức về hồi quy logistic được tham khảo từ nghiên cứu [4] trong khi kiến thức về suy giảm trọng số được tham khảo từ nghiên cứu [5].

3.1 Bài toán phân tích dữ liệu

3.1.1 Vấn đề dự đoán bệnh tim mạch vành

Bệnh tim mạch vành (CHD) là một bệnh lý rất phổ biến, xuất hiện khi các động mạch vành (động mạch lớn cung cấp máu đến cơ tim) bị bít kín hoặc hẹp lại do sự tích tụ của mảng bám (từ mỡ, canxi, tế bào bạch cầu và các loại tế bào khác) trong thành động mạch. Căn bệnh này được xem là một vấn đề sức khỏe nghiêm trọng trên toàn cầu, gây ra thiệt hại lớn về an sinh xã hội. Một giải pháp quan trọng để giải quyết tình trạng trên là dự đoán sớm nguy cơ mắc bệnh để có các biện pháp ứng phó kịp thời và hiệu quả.

Framingham Heart Study (FHS) là một dự án nghiên cứu lớn và dài hạn được thực hiện tại thành phố Framingham, tiểu bang Massachusetts, Hoa Kỳ. Nghiên cứu bắt đầu từ năm 1948 bởi Viện Tim mạch, Phổi và

Huyết học Quốc gia (NHLBI) của Hoa Kỳ và vẫn kéo dài cho đến ngày nay, góp phần quan trọng trong việc hiểu rõ hơn về bệnh tim mạch và các yếu tố nguy cơ liên quan (xem [6]). Đơn vị nghiên cứu đã tiến hành theo dõi, chẩn đoán và thu thập dữ liệu bệnh lý của năm nghìn cư dân thành phố trong nhiều thập kỷ. Sử dụng tập dữ liệu này, đồ án sẽ tiến hành xây dựng mô hình dự đoán nguy cơ mắc bệnh tim mạch vành.

3.1.2 Tập dữ liệu

Tập dữ liệu được sử dụng trong đồ án là phiên bản đã trải qua tiền xử lý bởi Özlem Niksch vào năm 2021 (xem [3]). Tập dữ liệu bao gồm 16 thuộc tính và 4133 bản ghi, các giá trị dữ liệu đều ở dạng số thực. 16 thuộc tính bao gồm 15 yếu tố nguy cơ và 1 biến tình trạng bệnh:

- Nhóm yếu tố nhân khẩu học:
  - male: giới tính (1 là "nam", 0 là "nữ");
  - age: độ tuổi;
  - education: học vấn (0 là "giáo dục phổ thông trở xuống", 1 là "giáo dục đại học trở lên");
- Nhóm yếu tố hành vi:
  - currentSmoker: tình trạng hút thuốc (1 là "có", 0 là "không");
  - cigsPerDay: số điếu thuốc/ngày;
- Nhóm yếu tố tiền sử bệnh:
  - BPMeds: tiền sử dùng thuốc huyết áp (1 là "có", 0 là "không");
  - prevalentStroke: tiền sử đột quỵ (1 là "có", 0 là "không");
  - prevalentHyp: tiền sử cao huyết áp (1 là "có", 0 là "không");
  - diabetes: tiền sử bệnh tiểu đường (1 là "có", 0 là "không");
- Nhóm yếu tố thể trạng:
  - totChol: tổng mức cholesterol;
  - sysBP: huyết áp tâm thu;
  - diaBP: huyết áp tâm trương;
  - BMI: chỉ số khối cơ thể;
  - heartRate: nhịp tim;
  - glucose: nồng độ đường huyết;

- Tình trạng bệnh:
  - TenYearCHD: nguy cơ mắc bệnh tim mạch vành trong vòng 10 năm (1 là "có", 0 là "không").

3.2 Hồi quy logistic

Hồi quy logistic là một kỹ thuật hồi quy trong lĩnh vực học máy. Trong mô hình này, đầu ra có thể được thể hiện ở dạng xác suất. Mặc dù trong tên có từ "hồi quy", hồi quy logistic thường được sử dụng nhiều hơn cho các bài toán phân lớp (xem [4]).

3.2.1 Hàm sigmoid

- Xác suất dự đoán một người bị bệnh của bài toán luôn nằm trong khoảng $(0, 1)$, nên cần một hàm có giá trị luôn nằm trong khoảng này, đó là hàm sigmoid:
$$\sigma(s) := \frac{1}{1 + \mathrm{e}^{-s}}$$
- Tính chất của hàm sigmoid:
  - Hàm sigmoid bị chặn trong khoảng $(0, 1)$, rất phù hợp trong việc tìm xác suất;
  - Hàm sigmoid có tính đơn điệu tăng, cho phép mô hình hồi quy logistic dễ dàng học các mối quan hệ tuyến tính và phi tuyến giữa giá trị đầu vào và xác suất dự đoán;
  - Hàm sigmoid có tính khả vi với gradient là:
$$\nabla\sigma(s) = \sigma(s) * (1 - \sigma(s)) \tag{3.1}$$
    thuận tiện trong việc áp dụng thuật toán hướng giảm gradient.

3.2.2 Mô hình bài toán

- Với bản ghi thứ i trong tập dữ liệu, gọi:
  - $x_n^{(i)}$ là giá trị của biến đầu vào ở thuộc tính thứ n;
  - y_i là giá trị thuộc tính Outcome về tình trạng bệnh;
  - $\widehat{y_i}$ là xác suất mà mô hình dự đoán nguy cơ mắc bệnh.

- Để thuận tiện trong tính toán, đặt:
  – $x^{(i)} = (1, x_1^{(i)}, \ldots, x_n^{(i)})^T$ là vec tơ các giá trị thuộc tính;
  – $w = (w_0, w_1, \ldots, w_n)^T$ là vec tơ trọng số của các thuộc tính.
- Khi đó, ta có công thức của mô hình hồi quy logistic:

$$\widehat{y_i} = \sigma(w^T x^{(i)}) \tag{3.2}$$

- Mục tiêu của bài toán là tìm ra vec tơ trọng số w sao cho giá trị dự đoán $\widehat{y_i}$ càng gần giá trị thực tế y_i càng tốt.

3.2.3 Hàm mất mát

- Ta có: $\widehat{y_i}$ là xác suất mà mô hình dự đoán nguy cơ mắc bệnh, ngược lại, $1 - \widehat{y_i}$ là xác suất không bị bệnh. Hai xác suất này có thể viết chung dưới dạng:

$$P := \widehat{y_i}^{y_i}(1 - \widehat{y_i})^{1-y_i}$$

  vì với tình trạng mắc bệnh ($y_i = 1$) thì biểu thức trở thành:

$$P = \widehat{y_i}^{y_i}$$

  và với tình trạng không bệnh ($y_i = 0$) thì:

$$P = (1 - \widehat{y_i})^{1-y_i}.$$

- Giả sử các thuộc tính là độc lập với nhau, xét trên toàn tập dữ liệu:

$$P = \prod_{i=1}^{N} \widehat{y_i}^{y_i}(1 - \widehat{y_i})^{1-y_i} \quad \text{với } N \text{ là số lượng bản ghi.}$$

- Lấy logarit tự nhiên, đổi dấu, lấy trung bình, ta thu được hàm số:

$$L(w) := -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log \widehat{y_i} + (1 - y_i)\log(1 - \widehat{y_i})\right]$$

- $L(w)$ được gọi là hàm mất mát của hàm số. Mục tiêu bài toán là tìm w sao cho $L(w)$ nhỏ nhất. Đây chính là bài toán quy hoạch phi tuyến không ràng buộc.

3.2.4 Tối ưu hàm mất mát

- Xét gradient của hàm $L(w)$:

$$\nabla L(w) = \frac{1}{N}\sum_{i=1}^{N} \frac{\widehat{y_i} - y_i}{\widehat{y_i}(1-\widehat{y_i})}\nabla \widehat{y_i} \tag{3.3}$$

Theo công thức tính gradient của hàm sigmoid (3.1), gradient của $\widehat{y_i} = \sigma(w^T x^{(i)})$ là:

$$\nabla \widehat{y_i} = \widehat{y_i}(1-\widehat{y_i})x^{(i)} \tag{3.4}$$

Từ (3.3) và (3.4), suy ra gradient của hàm $L(w)$:

$$\nabla L(w) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y_i} - y_i)x^{(i)}$$

- Từ gradient của hàm $L(w)$, xét ma trận Hesse bằng cách tiếp tục đạo hàm:

$$\nabla^2 L(w) = \frac{1}{N}\sum_{i=1}^{N}\nabla \widehat{y_i} x^{(i)}$$

Kết hợp với (3.4), suy ra ma trận Hesse của hàm $L(w)$:

$$\nabla^2 L(w) = \frac{1}{N}\sum_{i=1}^{N}\widehat{y_i}(1-\widehat{y_i}) * x^{(i)}x^{(i)T}$$

- Xét tính lồi của hàm $L(w)$:

Nhận xét:

$$\begin{cases} \widehat{y_i} \textit{ là hàm sigmoid nên } \widehat{y_i}(1-\widehat{y_i}) \textit{ bị chặn trong khoảng } (0,1); \\ x^{(i)}x^{(i)T} \textit{ là dạng toàn phương nên là ma trận xác định dương.} \end{cases}$$

Suy ra, $\widehat{y_i}(1-\widehat{y_i}) * x^{(i)}x^{(i)T}$ và $\nabla^2 L(w)$ là ma trận xác định dương.

  - Theo Định lý 1.2, vì $\nabla^2 L(w)$ là ma trận xác định dương nên $L(w)$ là hàm lồi.
  - Theo Định lý 1.4, vì $L(w)$ là hàm lồi, điểm hội tụ của thuật toán hướng giảm gradient chính là nghiệm tối ưu toàn cục của bài toán.

- Kết luận, công thức cập nhật nghiệm của bài toán theo thuật toán hướng giảm gradient là:

$$w := w - t_k * \frac{1}{N}\sum_{i=1}^{N}(\widehat{y_i} - y_i)x^{(i)}$$

với t_k là cỡ bước ở vòng lặp thứ k.

3.2.5 Chính quy hóa

Quá khớp là hiện tượng mô hình tìm được có độ khớp quá cao với dữ liệu huấn luyện, dẫn đến việc không thực sự mô tả tốt dữ liệu ngoài tập huấn luyện. Một kỹ thuật chính quy hóa giúp tránh quá khớp là sử dụng suy giảm trọng số, trong đó một đại lượng tỉ lệ với bình phương chuẩn Euclid của vector trọng số được cộng vào hàm mất mát để hạn chế độ lớn của các trọng số (xem [5]).

- Kết hợp hệ số chính quy hóa λ, hàm mất mát trở thành:

$$L(w) := \frac{1}{N}\sum_{i=1}^{N}\Big[-y_i\log\widehat{y_i} - (1-y_i)\log(1-\widehat{y_i}) + \frac{\lambda}{2}\|w\|^2\Big]$$

- Đạo hàm của hàm mất mát:

$$\nabla L(w) = \frac{1}{N}\sum_{i=1}^{N}\big[(\widehat{y_i} - y_i)x^{(i)} + \lambda w\big]$$

- Công thức cập nhật nghiệm của bài toán:

$$w := w - t_k * \frac{1}{N}\sum_{i=1}^{N}\big[(\widehat{y_i} - y_i)x^{(i)} + \lambda w\big]$$

với t_k là cỡ bước ở vòng lặp thứ k.

3.3 Triển khai thuật toán bằng ngôn ngữ lập trình Python

3.3.1 Khởi tạo chương trình

- Công cụ:
  - Ngôn ngữ lập trình: Python (miễn phí và thuận tiện trong khoa học tính toán);
  - Thiết bị chạy chương trình: laptop Dell Latitude E6540 Intel Core i7-4800MQ CPU @ 2.70GHz 16GB RAM.
- Quy trình phân tích dữ liệu:

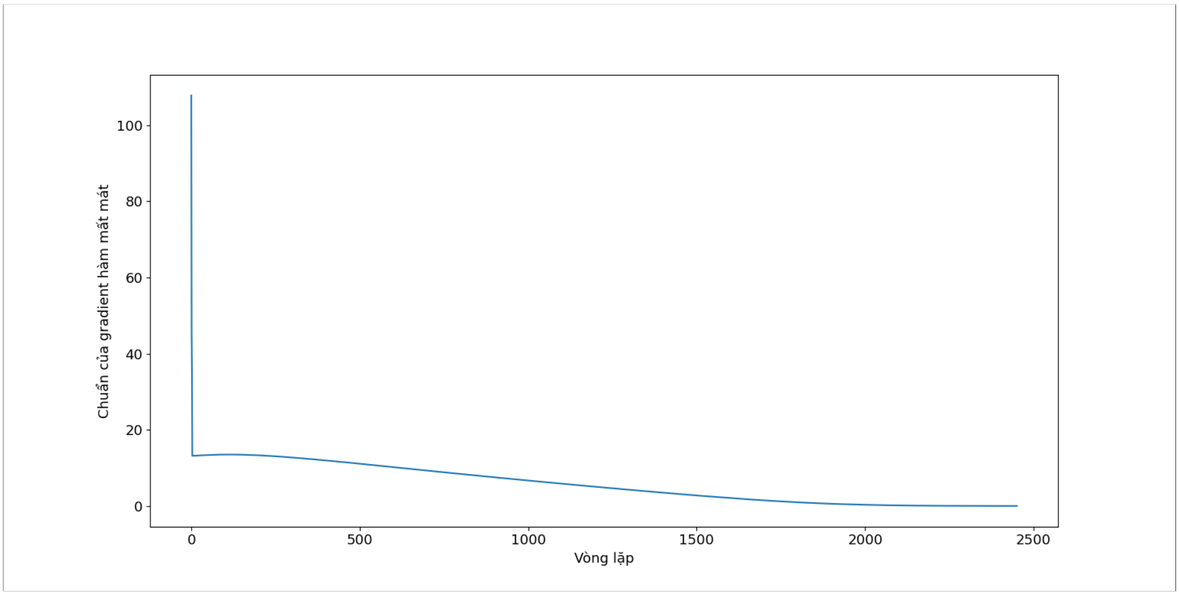
  - B1: Chia dữ liệu thành tập huấn luyện và tập kiểm tra một cách ngẫu nhiên theo tỷ lệ 70 - 30;
  - B2: Từ dữ liệu trong tập huấn luyện, xây dựng mô hình hồi quy logistic dự đoán bệnh tim mạch vành, thực hiện thuật toán hướng giảm gradient để tối ưu mô hình;
  - B3: Áp dụng mô hình xây dựng ở B2 vào tập kiểm tra, đưa ra dự đoán về nguy cơ mắc bệnh cho các đối tượng trong tập kiểm tra;
  - B4: So sánh giá trị dự đoán ở B3 với giá trị thực về tình trạng bệnh của các đối tượng trong tập kiểm tra, từ đó đánh giá độ chính xác của mô hình và hiệu chỉnh nếu cần thiết.
- Thuật toán hướng giảm gradient tối ưu mô hình:
  - Thủ tục: Trong bài toán này, việc áp dụng thủ tục tìm chính xác theo tia sẽ có chi phí tìm cỡ bước rất cao và khó khăn trong việc lập trình tính toán. Vì vậy, chương trình tính toán trong đồ án này sẽ chỉ xét:
    - Thủ tục cỡ bước cố định
    - Thủ tục quay lui
  - Tham số:
    - Chọn điểm khởi đầu: $w^{(0)} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)^T$
    - Chọn hệ số chính quy hóa: $\lambda = 0,01$
    - Chọn sai số: $\varepsilon = 0,05$

3.3.2 Thủ tục cỡ bước cố định

- Phương án chọn cỡ bước: đầu tiên chọn một cỡ bước thật nhỏ để thuật toán có thể hội tụ, sau đó thử nghiệm tiếp bằng cách tăng dần cỡ bước lên.
- Xét lần lượt 9 cỡ bước cố định, chạy chương trình thu được kết quả về số vòng lặp, thời gian thực thi và độ chính xác như trong Bảng 3.1.
- Nhận xét:
  - Chi phí tính toán (số vòng lặp và thời gian thực thi) giảm dần khi cỡ bước tăng từ 0,000001 lên 0,000167. Sau đó chi phí tính toán đảo chiều, tăng nhanh trong khi cỡ bước chỉ tăng từ 0,000167 lên 0,000167952. Đến khi cỡ bước tăng lên 0,0001679525, sau 1 triệu

Trường hợp	Cỡ bước cố định t	Số vòng lặp	Thời gian (giây)	Thời gian/vòng lặp	Độ chính xác
A1	$t = 0,000001$	392601	400,230	$1,02 \times 10^{-3}$	83,31%
A2	$t = 0,00001$	39260	37,725	$0,96 \times 10^{-3}$	83,31%
A3	$t = 0,0001$	3926	3,924	$1,00 \times 10^{-3}$	83,31%
A4	$t = 0,00016$	2454	2,440	$0,99 \times 10^{-3}$	83,31%
A5	$t = 0,000167$	2451	2,388	$0,97 \times 10^{-3}$	83,39%
A6	$t = 0,0001679$	7336	8,661	$1,18 \times 10^{-3}$	83,23%
A7	$t = 0,00016795$	12324	12,085	$0,98 \times 10^{-3}$	83,31%
A8	$t = 0,000167952$	13478	13,286	$0,99 \times 10^{-3}$	83,31%
A9	$t = 0,0001679525$	> 1000000	$> 1197,223$	$1,20 \times 10^{-3}$	null

Bảng 3.1: Xét 9 trường hợp theo thủ tục cỡ bước cố định

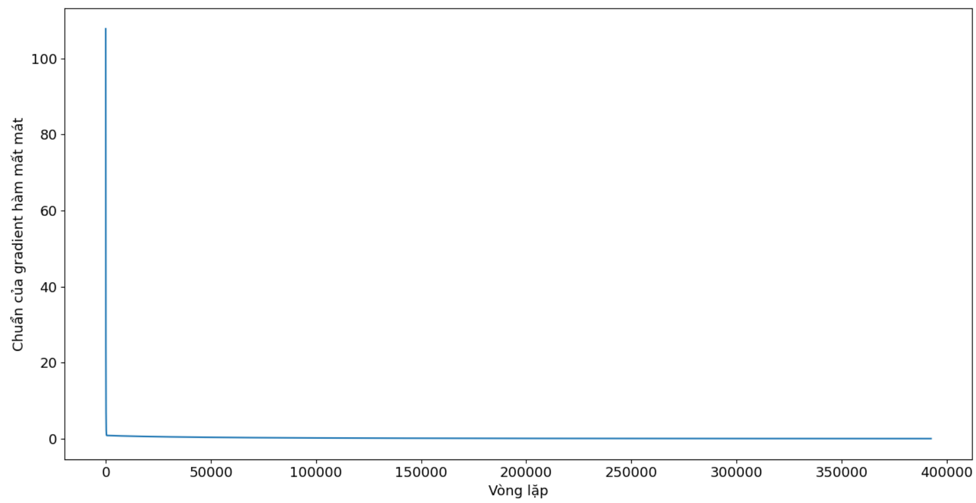


Hình 3.1: Chuẩn của gradient hàm mất mát - thủ tục cỡ bước cố định $t = 0,000167$

bước lặp, chương trình vẫn không dừng lại;

- Thời gian trung bình trên mỗi vòng lặp của 9 trường hợp tương đối đồng đều nhau, thể hiện tính chất lý thuyết của thủ tục quay lui;
- Độ chính xác của 8 trường hợp đầu cũng gần như nhau do có cùng ngưỡng sai số.

- Kết quả đại diện của trường hợp tốt nhất và xấu nhất (trong số 8 trường hợp đạt đến điểm dừng xét ở mục này):
  - Chuẩn của gradient hàm mất mát theo mỗi vòng lặp:
    - Trường hợp tốt nhất ($t = 0,000167$): Hình 3.1
    - Trường hợp xấu nhất ($t = 0,000001$): Hình 3.2



Hình 3.2: Chuẩn của gradient hàm mất mát - thủ tục cỡ bước cố định $t = 0,000001$

– Kết quả tìm được - vec tơ trọng số:

$$w^*_{t=0,000167} = \begin{bmatrix} -0,0033546 \\ 0,002789 \\ 0,026203 \\ -0,001969 \\ -0,001833 \\ 0,018551 \\ 0,001500 \\ -0,000132 \\ 0,009052 \\ 0,0015112 \\ -0,003755 \\ 0,023506 \\ -0,029015 \\ -0,029891 \\ -0,030868 \\ 0,000032 \end{bmatrix} \qquad w^*_{t=0,000001} = \begin{bmatrix} -0,003234 \\ 0,002676 \\ 0,025976 \\ -0,001902 \\ -0,001744 \\ 0,018488 \\ 0,001445 \\ -0,000126 \\ 0,008719 \\ 0,001457 \\ -0,003756 \\ 0,023428 \\ -0,028885 \\ -0,029264 \\ -0,030911 \\ 0,000022 \end{bmatrix}$$

– Nhận xét: kết quả tìm được ở hai trường hợp rất tương đồng nhau.

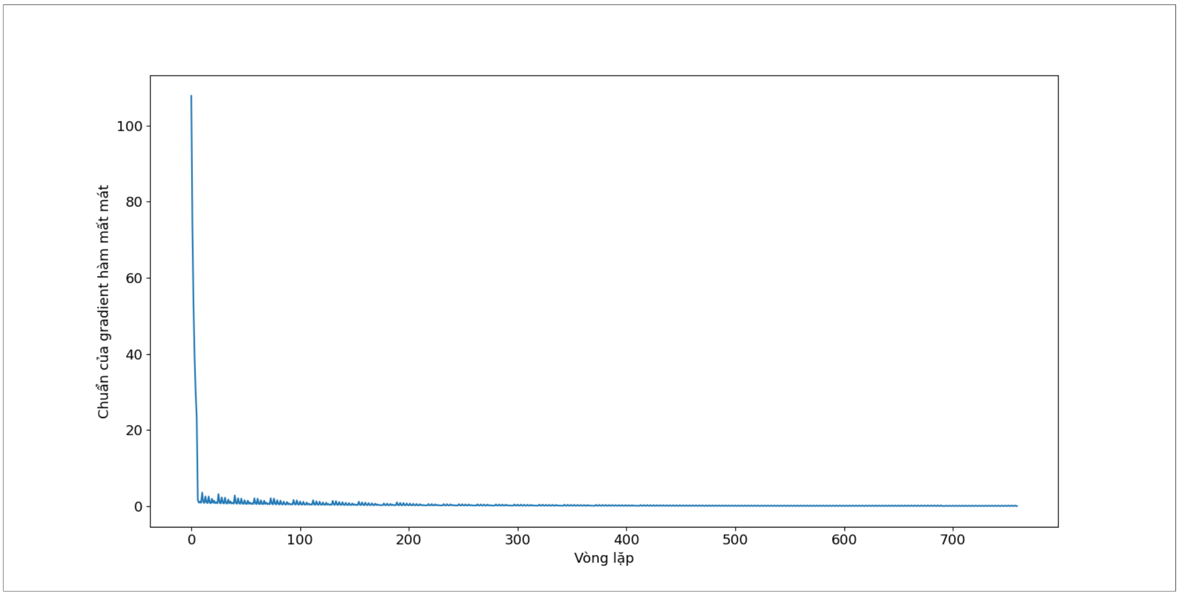
3.3.3 Thủ tục quay lui

- Phương án chọn hệ số α và β: xét lần lượt các tổ hợp của hai hệ số α và β với các giá trị được chọn là 0,2; 0,5 và 0,7.
- Xét lần lượt 9 trường hợp, chạy chương trình thu được kết quả về số vòng lặp, thời gian thực thi và độ chính xác như trong Bảng 3.2.

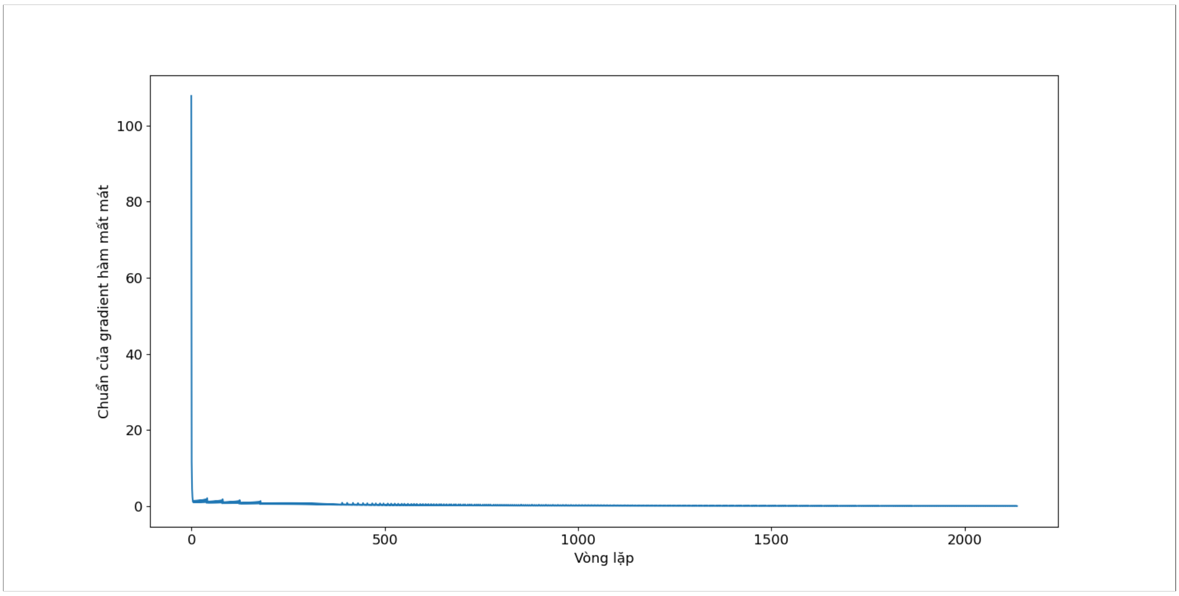
Trường hợp	Hệ số α	Hệ số β	Số vòng lặp	Thời gian (giây)	Thời gian/vòng lặp	Độ chính xác
B1	$\alpha = 0,2$	$\beta = 0,2$	826	2,894	$3,50 \times 10^{-3}$	83,31%
B2	$\alpha = 0,2$	$\beta = 0,5$	2135	13,527	$6,33 \times 10^{-3}$	93,38%
B3	$\alpha = 0,2$	$\beta = 0,7$	3597	31,14	$8,66 \times 10^{-3}$	83,31%
B4	$\alpha = 0,5$	$\beta = 0,2$	759	2,748	$3,62 \times 10^{-3}$	83,31%
B5	$\alpha = 0,5$	$\beta = 0,5$	2132	13,244	$6,21 \times 10^{-3}$	84,39%
B6	$\alpha = 0,5$	$\beta = 0,7$	880	9,261	$10,52 \times 10^{-3}$	83,30%
B7	$\alpha = 0,7$	$\beta = 0,2$	1278	4,696	$3,67 \times 10^{-3}$	83,31%
B8	$\alpha = 0,7$	$\beta = 0,5$	1939	12,219	$6,30 \times 10^{-3}$	83,39%
B9	$\alpha = 0,7$	$\beta = 0,7$	635	7,017	$11,05 \times 10^{-3}$	83,31%

Bảng 3.2: Xét 9 trường hợp với thủ tục quay lui

- Nhận xét:
  - Chi phí tính toán có sự biến thiên nhưng không theo quy luật và cũng không có sự chênh lệch quá lớn giữa các trường hợp;
  - Thời gian trung bình trên các vòng lặp của mỗi trường hợp có sự chênh lệch do chi phí tìm cỡ bước ở mỗi vòng lặp là khác nhau;
  - Độ chính xác của 9 trường hợp cũng gần như nhau do có cùng ngưỡng sai số.
- Kết quả đại diện của trường hợp tốt nhất và xấu nhất (trong số 9 trường hợp đang xét ở mục này):
  - Chuẩn của gradient hàm mất mát theo mỗi vòng lặp:
    * Trường hợp tốt nhất ($\alpha = 0,5; \beta = 0,2$): Hình 3.3
    * Trường hợp xấu nhất ($\alpha = 0,2; \beta = 0,5$): Hình 3.4
  - Kết quả tìm được - vec tơ trọng số:



Hình 3.3: Chuẩn của gradient hàm mất mát - thủ tục quay lui với $\alpha = 0,5; \beta = 0,2$



Hình 3.4: Chuẩn của gradient hàm mất mát - thủ tục quay lui với $\alpha = 0,2; \beta = 0,5$

$$w^*_{\alpha=0,5;\beta=0,2} = \begin{bmatrix} -0,003313 \\ -0,002744 \\ 0,026113 \\ -0,001947 \\ -0,001801 \\ 0,018528 \\ 0,001480 \\ -0,000130 \\ 0,008935 \\ 0,001492 \\ -0,003755 \\ 0,023487 \\ -0,028981 \\ -0,029702 \\ -0,030877 \\ 0,000030 \end{bmatrix} \qquad w^*_{\alpha=0,2;\beta=0,5} = \begin{bmatrix} -0,003843 \\ -0,003203 \\ 0,026717 \\ -0,002244 \\ -0,002197 \\ 0,018670 \\ 0,001715 \\ -0,000151 \\ 0,010387 \\ 0,001731 \\ -0,003738 \\ 0,023707 \\ -0,029235 \\ -0,032255 \\ -0,030664 \\ 0,000076 \end{bmatrix}$$

– Nhận xét: kết quả tìm được ở hai trường hợp rất tương đồng nhau.

3.3.4 So sánh và đánh giá

- Do có cùng ngưỡng sai số, thủ tục cỡ bước cố định và thủ tục quay lui có kết quả tìm được và độ chính xác của mô hình rất tương đồng nhau;
- Thủ tục cỡ bước cố định phải tiến hành nhiều bước lặp hơn hẳn thủ tục quay lui vì thủ tục quay lui có cơ chế để tìm ra cỡ bước tương đối phù hợp cho từng bước lặp;
- Tuy nhiên, các bước lặp của thủ tục quay lui có chi phí tính toán cao hơn thủ tục cỡ bước cố định, nguyên nhân là vì thủ tục quay lui tốn chi phí tìm cỡ bước cho từng bước lặp;
- Cân nhắc cả hai khía cạnh trên, đánh giá tổng quan trên nhiều trường hợp vẫn có thể thấy được rằng thủ tục quay lui có chi phí tính toán thấp hơn thủ tục cỡ bước cố định (1).
- Việc xác định cỡ bước cố định phù hợp trong thủ tục cỡ bước cố định là rất khó khăn, nếu cỡ bước quá nhỏ thuật toán sẽ hội tụ rất chậm trong khi nếu cỡ bước quá lớn sẽ khiến thuật toán không thể hội tụ.

Như thực hành ở mục 3.3.2, chỉ một thay đổi rất nhỏ ở cỡ bước cố định cũng làm chi phí tính toán tăng lên rất nhiều.

- Ngược lại, việc thay đổi hệ số α và β ở thủ tục quay lui không có ảnh hưởng quá lớn đến tốc độ hội tụ của thuật toán do đã được xử lý bởi một vòng lặp con nằm bên trong mỗi bước lặp để tìm cỡ bước phù hợp (2).
- Đánh giá chủ quan: việc cài đặt thủ tục quay lui cũng rất đơn giản, không quá khó khăn so với thủ tục cỡ bước cố định (3).

$\Rightarrow$ Từ các đánh giá (1), (2) và (3) có thể rút ra kết luận: thủ tục quay lui hiệu quả hơn thủ tục cỡ bước cố định do có chi phí tính toán tổng thể thấp hơn, không gặp khó khăn khi lựa chọn các tham số cho thuật toán và cũng rất dễ dàng cài đặt. Trong thực tế tính toán, nên sử dụng thủ tục quay lui.

Kết luận

1 Kết luận

Đồ án đã nghiên cứu tìm hiểu về thuật toán hướng giảm gradient và ứng dụng trong bài toán hồi quy logistic.

Kết quả đạt được

- Trình bày các kiến thức về thuật toán hướng giảm gradient và thảo luận về các khía cạnh quan trọng của thuật toán như cơ sở lý thuyết, ưu nhược điểm của các thủ tục tìm cỡ bước và sự hội tụ của thuật toán;
- Trình bày về ứng dụng của thuật toán hướng giảm gradient trong bài toán hồi quy logistic qua ví dụ dự đoán bệnh tim mạch vành, tiến hành phân tích toán học và lập trình tính toán.

Kỹ năng đạt được

- Bước đầu biết tìm kiếm, đọc, dịch tài liệu chuyên ngành liên quan đến nội dung đồ án;
- Học cách tổng hợp các kiến thức tìm hiểu để áp dụng vào nội dung đồ án;
- Biết trình bày đồ án một cách logic, chặt chẽ theo chuẩn khoa học;
- Học được cách chế bản đồ án bằng LaTeX;
- Nâng cao kỹ năng lập trình trong việc xử lý bài toán thực tế.

2 Hướng phát triển của đồ án trong tương lai

- Nghiên cứu sâu hơn các khía cạnh của thuật toán hướng giảm gradient đồng thời phát triển các phương pháp cải tiến từ thuật toán;
- Nghiên cứu hướng ứng dụng của thuật toán vào các bài toán khác trong thực tế của lĩnh vực toán tối ưu và học máy.

Tài liệu tham khảo

Tiếng Việt

[1] Nguyễn Thị Bạch Kim (2008), *Giáo trình Các Phương pháp Tối ưu: Lý thuyết và Thuật toán*, NXB Bách Khoa Hà Nội.

Tiếng Anh

[2] A. Beck (1984), *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*, Society for Industrial and Applied Mathematics, Philadelphia.

[3] Ö. Niksch (2021), "Framingham CHD Preprocessed Data". [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database (Accessed: Jul. 16, 2023).

[4] J. Peng, K. Lee, and G. Ingersoll (2002), "An Introduction to Logistic Regression Analysis and Reporting", The Journal of Educational Research, vol. 96, pp. 3–14. Doi: 10.1080/00220670209598786.

[5] A. Krogh and J. Hertz (1991), "A Simple Weight Decay Can Improve Generalization," in *Advances in Neural Information Processing Systems*, Morgan-Kaufmann, Burlington, Massachusetts.

[6] S. Mahmood, D. Levy, R. Vasan, and T. Wang (2013), "The Framingham Heart Study and the Epidemiology of Cardiovascular Diseases: A Historical Perspective," Lancet, vol. 383. Doi: 10.1016/S0140-6736(13)61752-3.