

Gradient Descent (GD)

Mentor: PĐ Hải

Người trình bày: Nguyễn Minh Hiếu

PAYT CLUB PTIT

August 2025

Nội dung

- 1 Khái niệm, thuật toán
- 2 Các thuật toán tối ưu GD
- 3 Các biến thể của GD

1: Khái niệm, thuật toán

1.1: Đặt vấn đề

- Tìm GTNN của $f(x) = \frac{1}{2}(x - 1)^2 - 2$?
- Tìm GTNN của

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2N} \|\mathbf{y} - \bar{\mathbf{X}}\mathbf{w}\|_2^2$$

1.2: Khái niệm

- GD là một thuật toán tối ưu hóa **lặp đi lặp lại** được sử dụng để **cực tiểu hóa** một hàm mất mát bằng cách **tính toán gradient** (đạo hàm) của hàm số đó và cập nhật các tham số mô hình theo hướng **ngược với gradient** để giảm thiểu giá trị hàm mất mát qua từng vòng lặp.

1.3: Thuật toán

Đối với hàm mất mát chỉ có 1 biến

- Bước 1: Khởi tạo:

- ▶ Khởi tạo giá trị x_0
- ▶ Chọn learning rate $\alpha > 0$

- Bước 2: Lặp đến khi hội tụ:

- ▶ Tính gradient (đạo hàm) $f'(x_k)$.
- ▶ Cập nhật tham số: $x_{k+1} = x_k - \alpha \cdot f'(x_k)$
- ▶ Kiểm tra điều kiện dừng: $|x_{k+1} - x_k| < \epsilon$

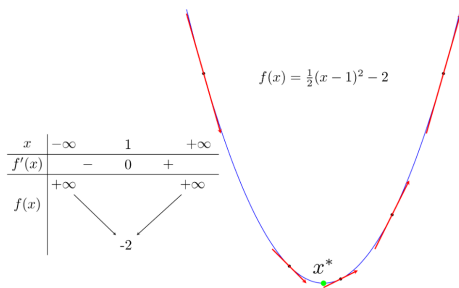
1.4: Ví dụ

$$x_0 = 0, \alpha = 0,1, f'(x) = x - 1$$

$$\begin{aligned} \blacktriangleright x_1 &= x_0 - 0,1 \cdot f'(x_0) \\ &= 0 - 0,1 \cdot (-1) \\ &= 0 + 0,1 = 0,1 \end{aligned}$$

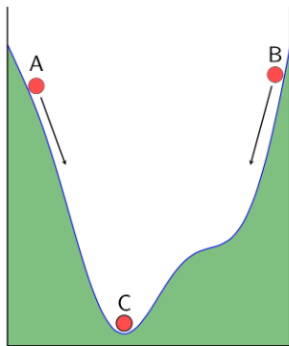
$$\begin{aligned} \blacktriangleright x_2 &= x_1 - 0,1 \cdot f'(x_1) \\ &= 0,1 - 0,1 \cdot (-0,9) \\ &= 0,1 + 0,09 = 0,19 \end{aligned}$$

$$\begin{aligned} \blacktriangleright x_3 &= x_2 - 0,1 \cdot f'(x_2) \\ &= 0,19 - 0,1 \cdot (-0,81) \\ &= 0,19 + 0,081 = 0,271 \end{aligned}$$

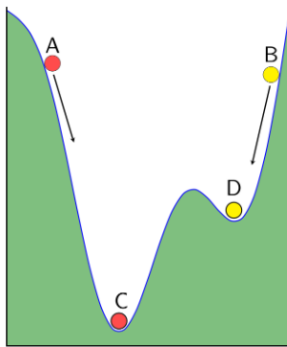


2: Các thuật toán tối ưu GD

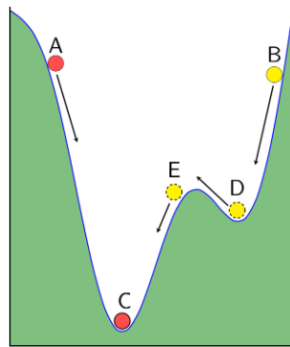
2.1: GD + momentum



a) GD



b) GD



c) GD with momentum

2.1: GD + momentum

Ý tưởng chính: Thay vì chỉ dựa vào **độ dốc** (gradient) tại điểm **hiện tại** để điều hướng, thuật toán sử dụng **"đà"** (momentum) được tích lũy từ các **bước di chuyển trước** đó để di chuyển nhanh hơn và ổn định hơn.

- Bước 1: Khởi tạo:
 - ▶ Khởi tạo giá trị x_0
 - ▶ Chọn learning rate $\alpha > 0, \gamma > 0$
- Bước 2: Lặp đến khi hội tụ:
 - ▶ Tính gradient (đạo hàm) $f'(x_k)$.
 - ▶ Cập nhật tham số: $x_{k+1} = x_k - v_k$, với $v_k = \gamma v_{k-1} + \eta f'(x_k)$
 - ▶ Kiểm tra điều kiện dừng: $|f'(x_{k+1})| < \epsilon$

2.2: GD + Nesterov Accelerated Gradient (NAG)

Ý tưởng chính: Nếu Momentum là một quả bóng lăn xuống dốc, thì NAG là một quả bóng có **khả năng dự đoán** tương lai. Nó không chỉ dựa vào đà hiện có mà còn 'thò đầu' nhìn về phía trước để điều chỉnh hướng đi một cách thông minh và chủ động hơn

- Bước 1: Khởi tạo:

- ▶ Khởi tạo giá trị x_0
- ▶ Chọn learning rate $\alpha > 0, \gamma > 0$

- Bước 2: Lặp đến khi hội tụ:

- ▶ Tính gradient (đạo hàm) $f'(x_k)$.
- ▶ Cập nhật tham số: $x_{k+1} = x_k - v_k$, với $v_k = \gamma v_{k-1} + \eta f'(x_k - \gamma v_{k-1})$
- ▶ Kiểm tra điều kiện dừng: $|f'(x_{k+1})| < \epsilon$

3: Các biến thể của GD

3.1: Các biến thể của Gradient Descent

- 1. Batch Gradient Descent:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t)$$

- 2. Stochastic Gradient Descent (SGD):

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta; \mathbf{x}_i; \mathbf{y}_i)$$

- 3. Mini-batch Gradient Descent:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta; \mathbf{x}_{i:i+n}; \mathbf{y}_{i:i+n})$$

3.2: So sánh các phương pháp Gradient Descent

Tiêu chí	Batch Gradient Descent (GD)	Stochastic Gradient Descent (SGD)	Mini-batch Gradient Descent
Lượng dữ liệu/lần cập nhật	Toàn bộ tập huấn luyện	1 mẫu dữ liệu duy nhất	Một nhóm nhỏ (mini-batch) dữ liệu
Tốc độ mỗi lần cập nhật	Rất chậm	Rất nhanh	Nhanh
Độ ổn định của quá trình hội tụ	Mượt, ổn định, tiến thẳng đến cực tiểu	"Nhiều", dao động mạnh	Dao động ít hơn SGD, tương đối mượt
Yêu cầu bộ nhớ	Rất cao (phải chứa toàn bộ dữ liệu)	Rất thấp	Trung bình, linh hoạt theo kích thước batch
Hiệu quả tính toán	Thấp (tính toán trên toàn bộ dữ liệu)	Cao (tính toán đơn giản)	Rất cao (tận dụng vector hóa)
Khả năng thoát khỏi cực tiểu địa phương	Kém, dễ bị kẹt	Tốt, nhờ sự "nhiều"	Tốt, tốt hơn Batch GD
Phù hợp với	Tập dữ liệu nhỏ, hàm loss lồi	Tập dữ liệu rất lớn, học trực tuyến (online learning)	Hầu hết các bài toán học sâu hiện đại