

StudywithT – Study

Gradient Descent

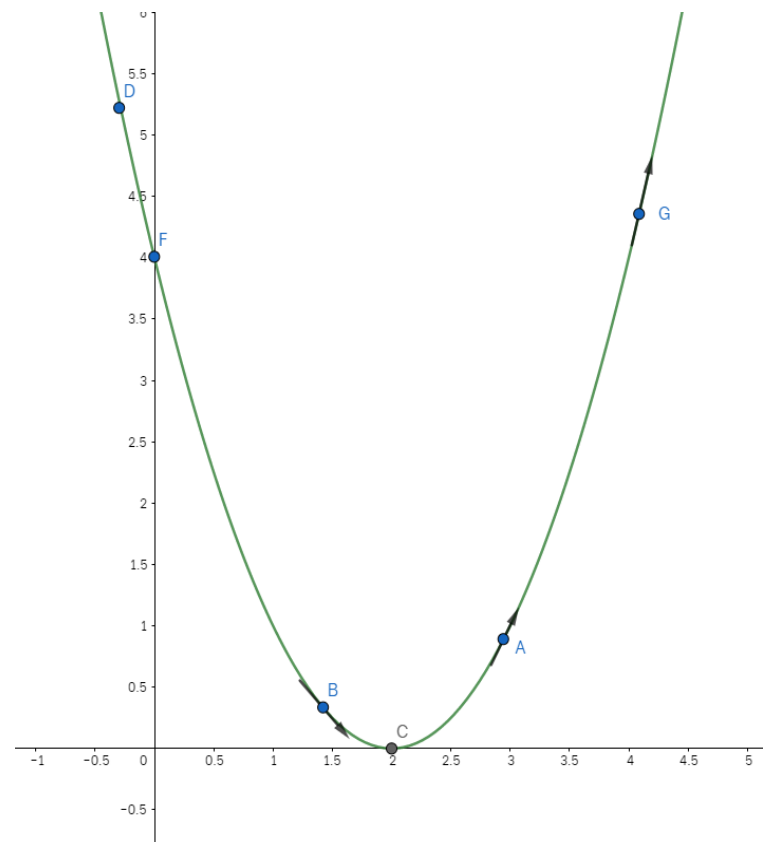
Ngô Doãn Thịnh



1 Cực trị của hàm số

x	$-\infty$	2	∞
$f'(x)$	$-$	0	$+$
$f(x)$	$-\infty$	4	∞

- Cực trị của hàm số là điểm có **giá trị lớn nhất** so với xung quanh và **giá trị nhỏ nhất** so với xung quanh mà hàm số có thể đạt được.
- Đường tiếp tuyến với đồ thị hàm số đó tại 1 điểm bất kỳ có hệ số góc chính bằng đạo hàm của hàm số tại điểm đó.



2

Gradient Descent

Gradient Descent cho 1 biến

- Nhìn chung, trong các bài toán Machine Learning, chúng ta thường sẽ cố gắng tìm những điểm cực tiểu (local minimum).

⇒ Các điểm local minimum là nghiệm của phương trình đạo hàm của hàm số.

- Gradient Descent cho hàm 1 biến

- Tại x_t : $f'(x_t) > 0$ thì x_t nằm ở bên phải so với điểm C. Vậy để tiến gần điểm C, thì cần di chuyển x_t về bên trái.

$$x_{t+1} = x_t + \Delta$$

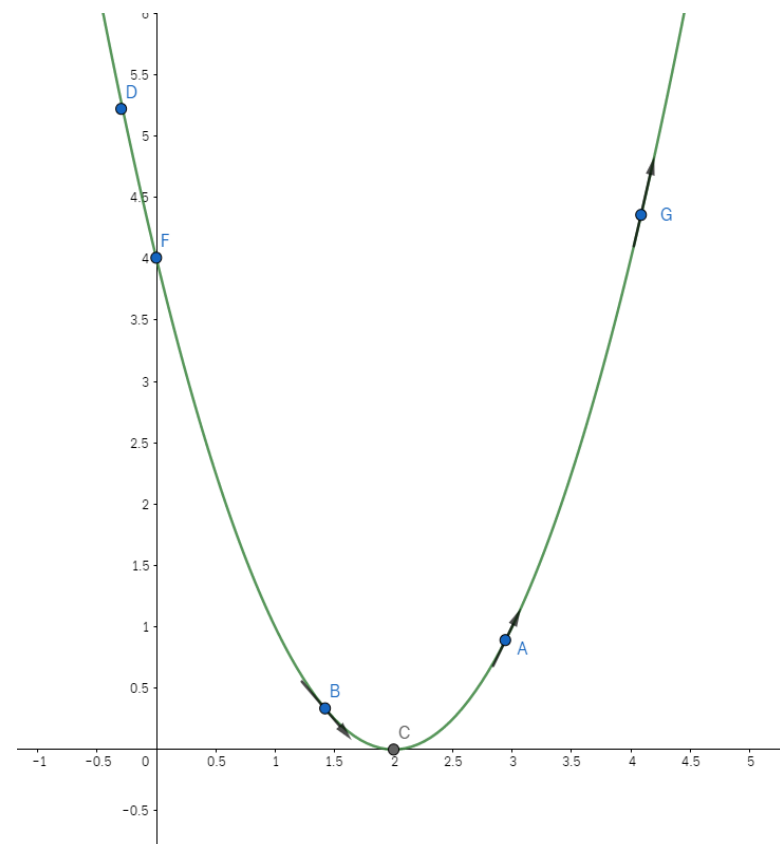
Trong đó, Δ là đại lượng nghịch dấu với $f'(x)$

- Tại x_t : $f'(x_t) < 0$ thì x_t nằm ở bên trái so với điểm C. Vậy để tiến gần điểm C, thì cần di chuyển x_t về bên phải.

⇒ Vậy để di chuyển x_t , thì ta nên dùng đại lượng tỉ lệ thuận với $f'(x_t)$

Ta có công thức tổng quát:

$$x_{t+1} = x_t - \mu f'(x_t)$$

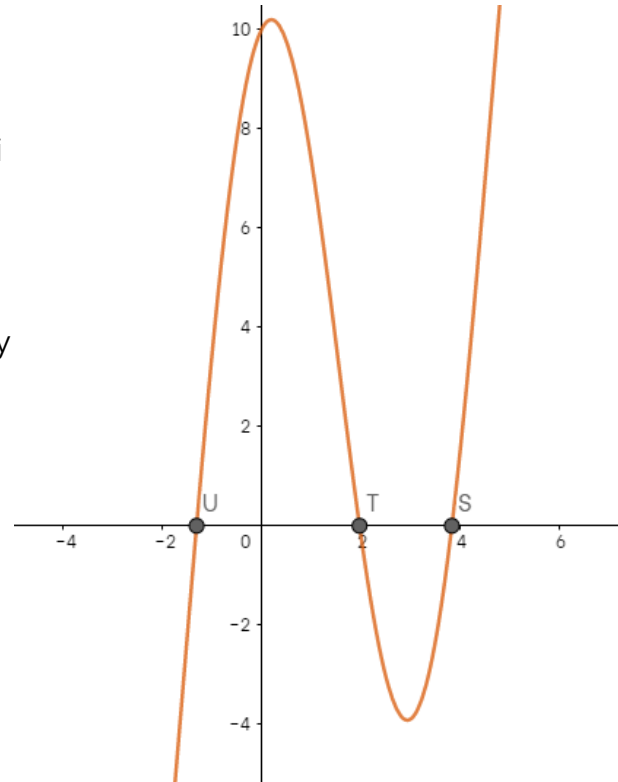


2

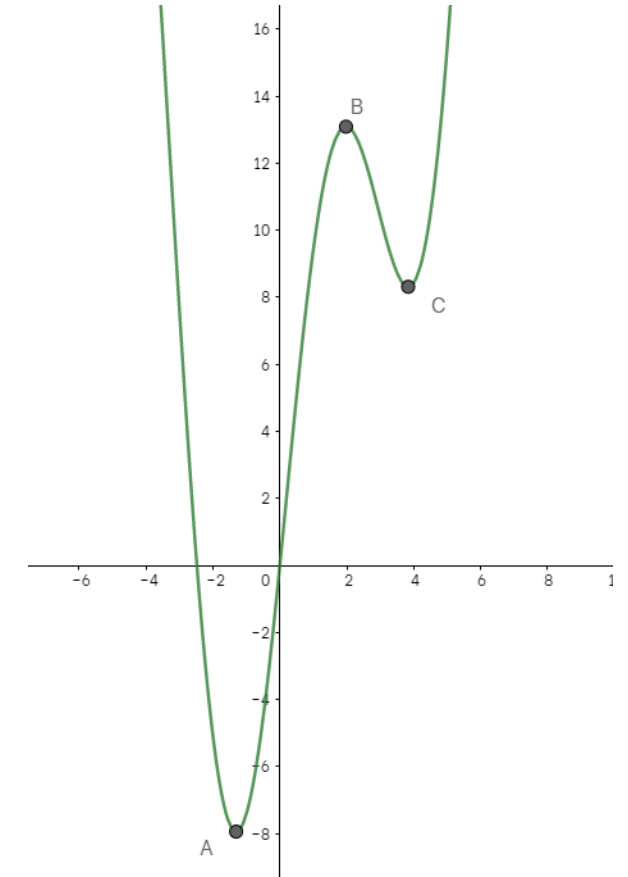
Gradient Descent

Tối ưu Gradient Descent

- Bài toán đang xét là PT bậc 2 nên sẽ chỉ có 1 điểm cực trị. Vậy bài toán có nhiều hơn 1 điểm cực trị thì Gradient Descent sẽ gặp phải vấn đề gì?
- Xét hàm số sau: $f(x) = x^2 + 10\sin(x)$.
 - Hàm số có đạo hàm: $f'(x) = 2x + 10\cos(x)$
 - Đạo hàm của hàm số có 3 cực trị là U, T, S tương ứng với 3 điểm A, B, C
 - Theo công thức tổng quát trên, nếu chúng ta xét điểm nằm bên phải điểm S thì chúng ta chỉ dịch chuyển điểm đấy tới điểm S (tương đương điểm C trên đồ thị). Thực tế, điểm A mới là kết quả mong muốn của chúng ta



Đồ thị đạo hàm $f'(x)$



Đồ thị đạo hàm $f(x)$

2

Gradient Descent

Tối ưu Gradient Descent

3

Gradient Descent

Gradient Descent cho nhiều biến