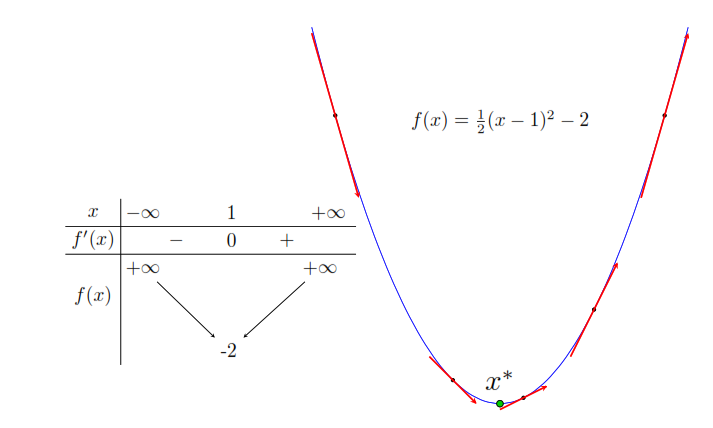
1. Giới Thiệu



* Điểm local minimum x\* của hàm số có đạo hàm f’(x\*) = 0.
  + Đạo hàm của các điểm phía bên trái x\* không dương
  + Đạo hàm của các điểm phía bên phải x\* không âm
* Đường tiếp tuyến với đồ thị f(x) tại bất kì 1 điểm nào có hệ số góc bằng chính đạo hàm tại điểm đó.

Trong Machine Learning chúng ta thường xuyên phải tìm các giá trị lớn nhất hoặc nhỏ nhất. Nếu chỉ xét riêng các hàm khả vi liên tục, việc giải phương trình đạo hàm bằng không thường rất phức tạp hoặc có thể ra vô số nghiệm. Thay vào đó người ta thường cố gắng tìm các điểm local minimum, và ở một mức độ nào đó, coi đó là nghiệm của bài toán.

Các điểm local minimum là nghiệm của phương trình đạo hàm bằng không. Nếu bằng một cách nào đó có thể tìm được toàn bộ (hữu hạn) các điểm cực tiểu, ta chỉ cần thay từng điểm local minimum đó vào hàm số rồi tìm điểm làm cho hàm số rồi tìm điểm làm cho hàm có giá trị nhỏ nhất.

Hướng tiếp cận phổ biến nhất để giải quyết các bài toán tối ưu là xuất phát từ một điểm được coi là gần với nghiệm của bài toán, sau đó dùng một phép toán lặp để tiến dần đến điểm cần tìm, tức đến khi đạo hàm gấn với không. Gradient descent và các biến thể của nó là một trong những phương pháp được dùng nhiều nhất.

1. GD cho hàm một biến

Note:

- Độ dốc càng cao thì tốc độ càng nhanh, ngược lại.

- Việc chọn learning rate rất quan trọng. Việc này phụ thuộc vào từng bài toán và phải làm một vài thí nghiệm để chọn ra giá trị tốt nhất.