

Họ và tên: Đặng Anh Tiến
MSSV: 20520800
Lớp: CS431.N11

Bài tập LT01: Hàm loss function cho Linear Regression

Problem:

Linear Regression dựa trên giả định rằng các feature của tập dữ liệu có mối quan hệ tuyến tính. Huấn luyện model để tạo đường thẳng hoặc mặt phẳng sao cho “fit” với dữ liệu nhất. Fit với dữ liệu có nghĩa là sự khác biệt giữa kết quả model dự đoán, với kết quả thực tế là thấp nhất. Vấn đề bài toán đặt ra là làm sao để biết được mô hình đã được huấn luyện có thực sự “fit” với dữ liệu?

Why:

Để đánh giá chất lượng mô hình dự đoán tốt trên các tập dữ liệu, đánh giá dựa trên kết quả dự đoán với kết quả thực tế. Cần tính “độ lỗi” này để thuật toán Gradient Descent sử dụng để tối ưu các parameter sao cho độ lỗi là nhỏ nhất. Model fit với dữ liệu là model có độ lỗi nhỏ nhất.

Solution:

Trong bài toán Linear Regression, Mean Squared Error (MSE) là metric thường được chọn bởi nó cho biết đường thẳng hay mặt phẳng đó gần với tập dữ liệu như thế nào. MSE được tính bằng tổng bình phương khoảng cách của các điểm dữ liệu tới đường thẳng hay mặt phẳng đó, sau đó lấy giá trị trung bình. MSE càng thấp cho thấy model càng fit với dữ liệu.