

MSSV: 19520954

Họ tên: Lê Thị Thanh Thanh

THỊ GIÁC MÁY TÍNH NÂNG CAO

BÀI TẬP BUỔI 2: CÂU 1

Câu 1: Tại sao không dùng 2 hàm Loss theo công thức như (1) và (2) mà lại dùng công thức (3)?

1. $L(\theta) = \hat{y} - y = \theta_1 x + \theta_0 - y$

2. $L(\theta) = |\hat{y} - y| = |\theta_1 x + \theta_0 - y|$

3. $L(\theta) = (\hat{y} - y)^2 = (\theta_1 x + \theta_0 - y)^2$

Trả lời:

Mục đích của quá trình training là tìm ra bộ số để độ lớn hàm Loss đạt giá trị cực tiểu. Hàm Loss để chỉ khoảng cách/chênh lệch giữa giá trị thực và giá trị dự đoán của mục tiêu, trong quá trình training cần rất nhiều mẫu trong tập dữ liệu, có thể $y^{(1)}$ lớn hơn $y^{(1)}$, và có thể $y^{(2)}$ bé hơn $y^{(2)}$, vậy khi muốn lấy tổng loss thì $L^{(1)}$ và $L^{(2)}$ có thể triệt tiêu nhau. Vậy nên Hàm Loss nên là một hàm có giá trị không âm, không thể là hàm số (1).

Khi thêm trị tuyệt đối như hàm số (2), hàm Loss đã mang giá trị không âm. Tuy nhiên trong quá trình tìm ra bộ số để hàm Loss đạt giá trị cực tiểu, chúng ta cần phải đạo hàm. Và việc tìm bộ số để đạo hàm bằng 0 sẽ rất khó khăn khi có giá trị tuyệt đối như trong công thức (2).

Công thức số (3) vừa thỏa mãn là một giá trị không âm, vừa có thể dễ tính đạo hàm hơn, bên cạnh đó các giá trị dự đoán quá xa so với giá trị thực sẽ lấy loss nặng hơn so với các giá trị ở gần giá trị thực (vì bình phương lên).