

BÀI TOÁN LOGISTIC REGRESSION

1. Tư Tưởng

- Mặc dù nó có tên regression nhưng thuật toán dùng để giải quyết cho bài toán classification.

- Tư tưởng bài toán về cơ bản giống như bài toán Linear regression nhưng kết quả được đi qua hàm sigmoid và giá trị hàm này nằm trong khoảng 0,1 cho ta phân loại kết quả.

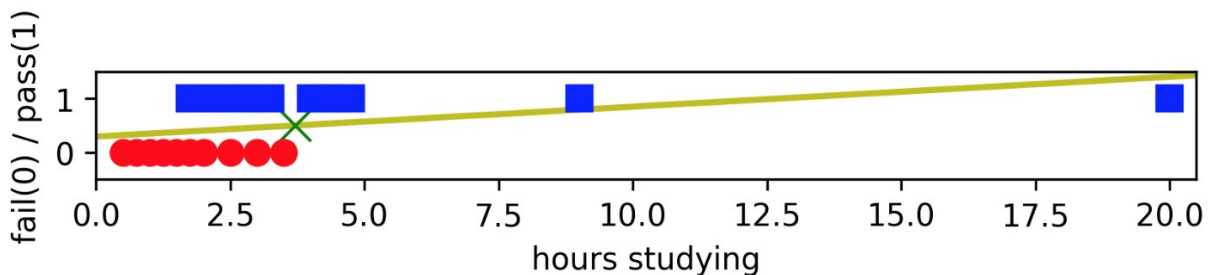
- VD: Dựa vào thời gian học và điểm thi giữa kì của sinh viên hãy dự đoán sinh viên có được giấy khen hay không....

2. Ý tưởng

- Đối với bài toán Regression ta đã biết:

$$Y = w_1.x_1 + w_2.x_2 + w_3.x_3 + \dots + w_n.x_n + w_0$$

- Đối với đầu ra là hàm tuyến tính giá trị liên tục ta không dễ để phân loại được 1 sinh viên liệu có được giấy khen hay không, email có spam hay không..



Hình 3: Tại sao Linear Regression không phù hợp?

- Chính vì vậy ta cần quy giá trị output và 1 ngưỡng mà ta có thể xác định được chẳng hạn về 0->1. Từ đó, nếu ta lấy ngưỡng là 0.5 thì giá trị output lớn hơn 0.5 ta cho nó là 1, nhỏ hơn 0.5 thì là 0. Cái ngưỡng này tùy chúng ta điều chỉnh.

3 Hàm sigmoid

$$q(z) = \frac{1}{1 + e^{-z}}$$

-Ta nhận thấy giá trị sigmoid này luôn nằm trong khoảng 0,1 z càng lớn thì output càng tiến về 1, ngược lại thì tiến về 0.

