Logistic regression

Tuan Nguyen - AI4E

Summary

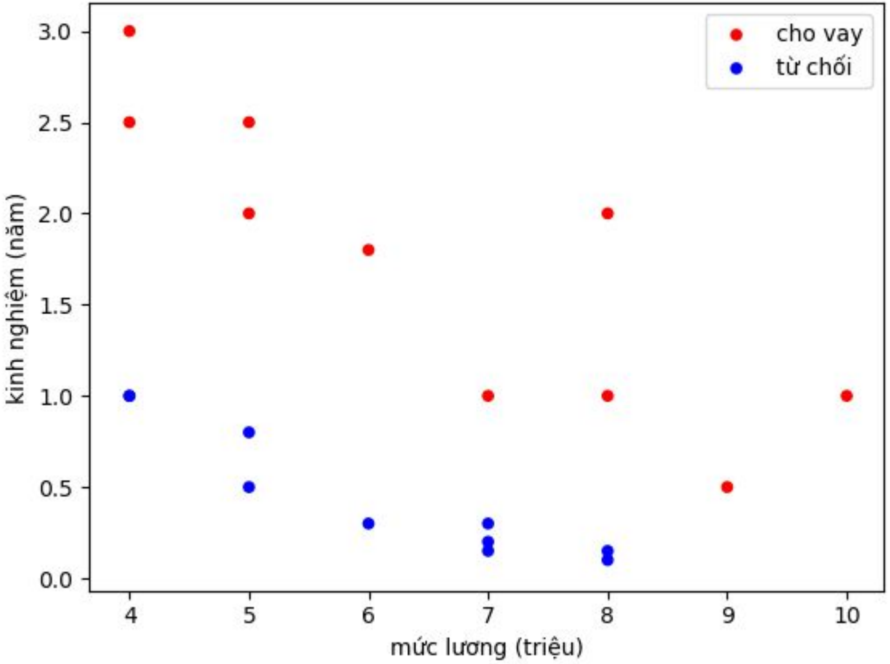
1. Define the problem
2. Visualize the data
3. Choose the model
4. Define Loss function
5. Minimize the loss function by gradient descent

Problem

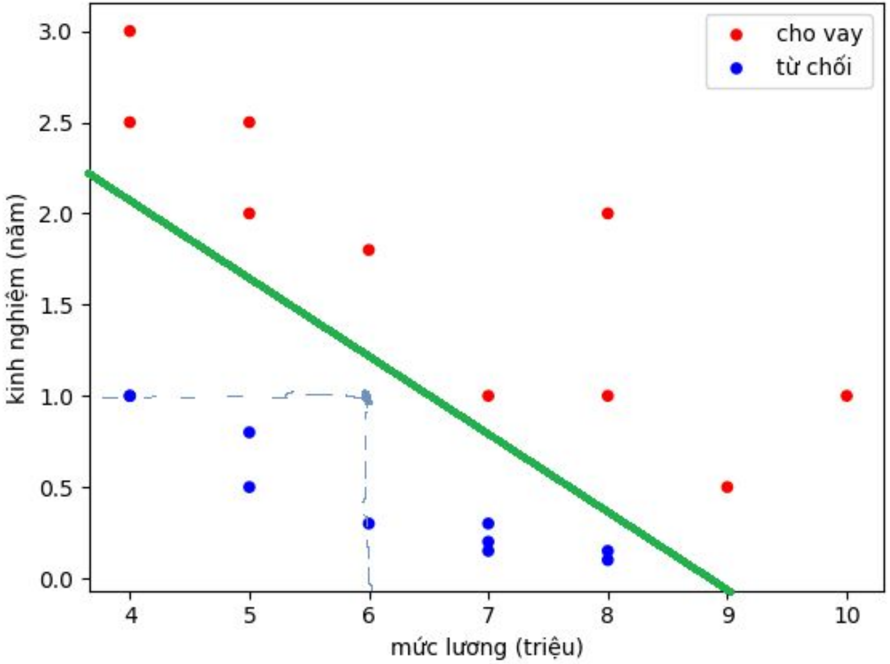
Dataset

Lương	Thời gian làm việc	Cho vay
10	1	1
5	2	1
...	...	1
8	0.1	0
7	0.15	0
...	...	0

Data visualization



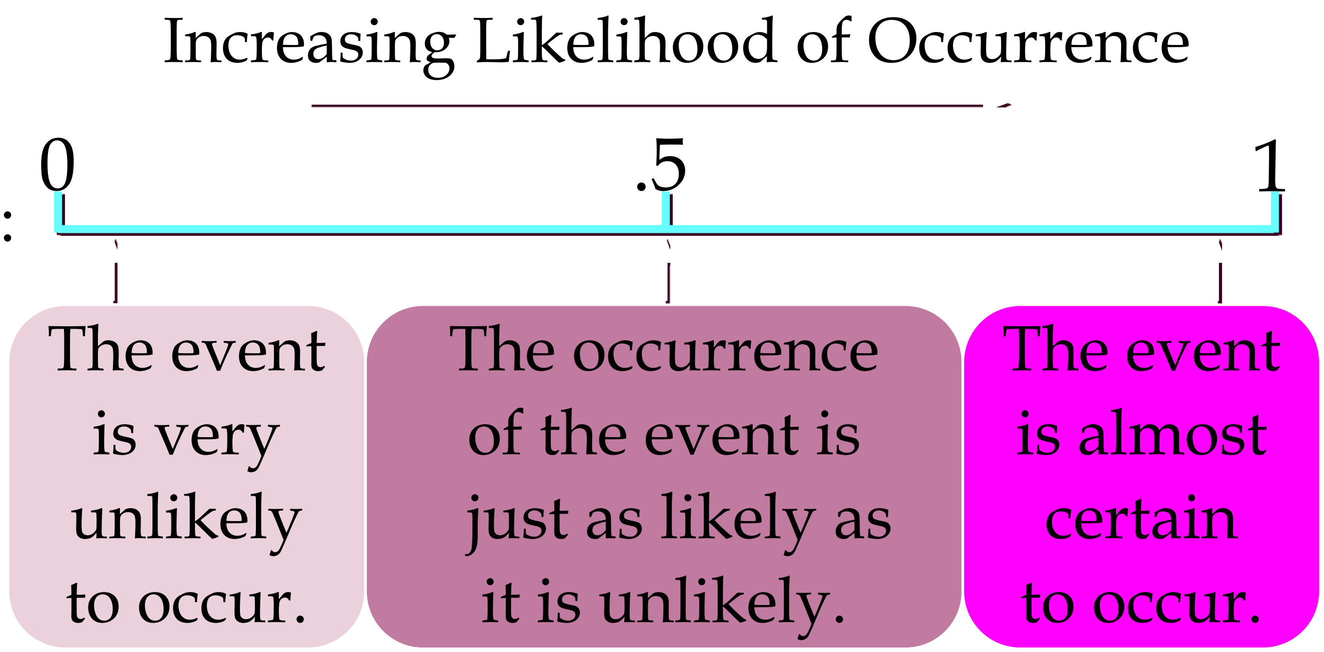
Prediction



cho vay
từ chối
kinh nghiệm (năm)
mức lương (triệu)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
4
5
6
7
8
9
10

Probability

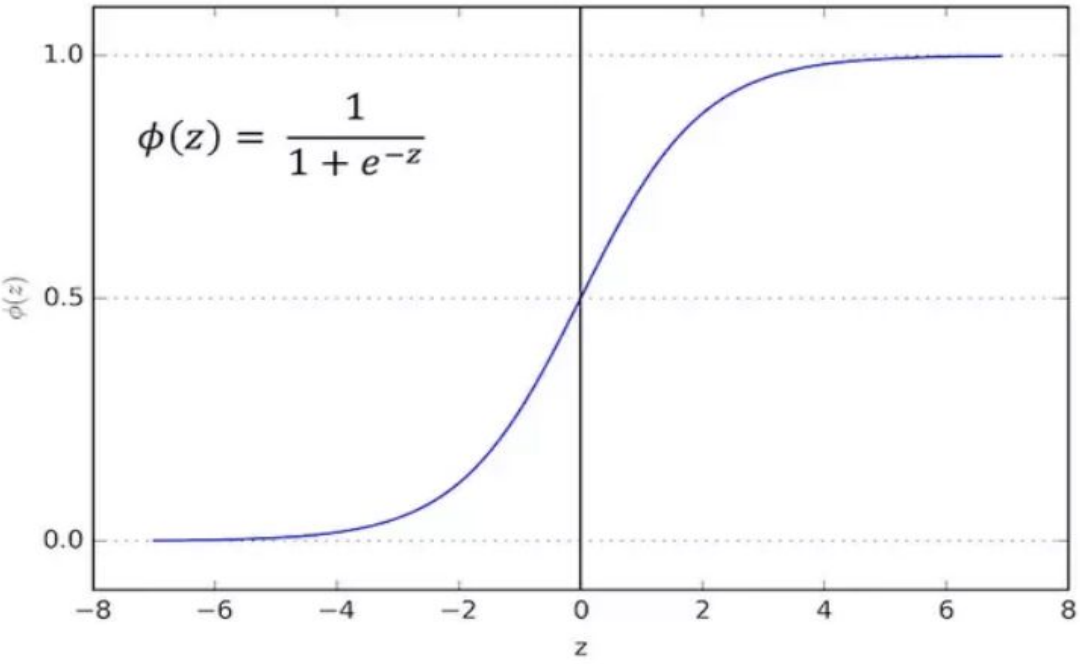
- Probability:



-

Sigmoid

- Hàm số liên tục, nhận giá trị thực trong khoảng (0,1).
- Hàm có đạo hàm tại mọi điểm (để áp dụng gradient descent)



Data presentation

Với dòng thứ i trong bảng dữ liệu, gọi $x_1^{(i)}$ là lương và $x_2^{(i)}$ là thời gian làm việc của hồ sơ thứ i .

p($x^{(i)} = 1) = \hat{y}_i$ là xác xuất mà model dự đoán hồ sơ thứ i được cho vay.

p($x^{(i)} = 0) = 1 - \hat{y}_i$ là xác xuất mà model dự đoán hồ sơ thứ i không được cho vay.

=> p($x^{(i)} = 1)$ + p($x^{(i)} = 0)$ = 1

Hàm sigmoid: $\sigma(x) = \frac{1}{1+e^{-x}}$.

Như bài trước công thức của linear regression là: $\hat{y}_i = w_0 + w_1 * x_i$ thì giờ công thức của logistic regression là:

$$\hat{y}_i = \sigma(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)}) = \frac{1}{1+e^{-(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)})}}$$

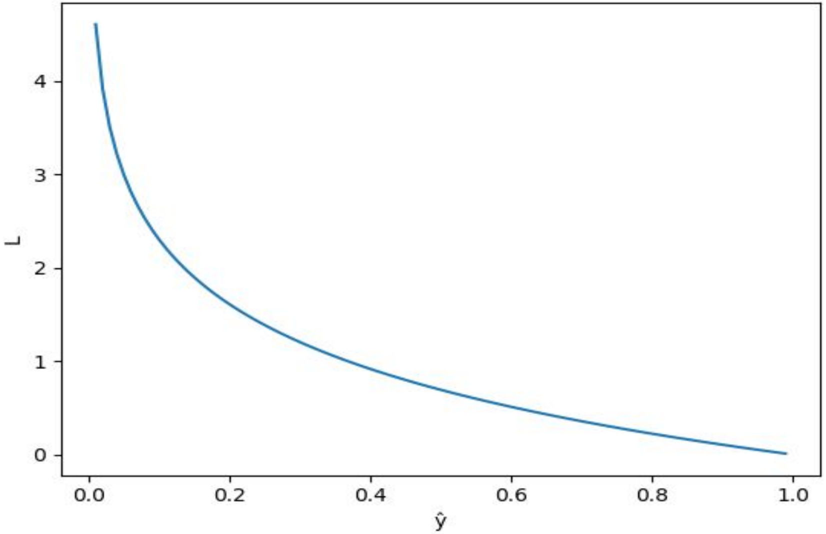
Loss function (binary_crossentropy)

$$L = -(y_i * log(\hat{y_i}) + (1-y_i) * log(1-\hat{y_i}))$$

- Nếu hồ sơ thứ i là cho vay (y=1) thì ta cũng mong muốn giá trị dự đoán càng gần 1 càng tốt hay model dự đoán xác xuất người thứ i được vay vốn càng cao càng tốt.
- Nếu hồ sơ thứ i không được vay (y=0) thì ta cũng mong muốn giá trị dự đoán càng gần 0 càng tốt hay model dự đoán xác xuất người thứ i được vay vốn càng thấp càng tốt

Loss function

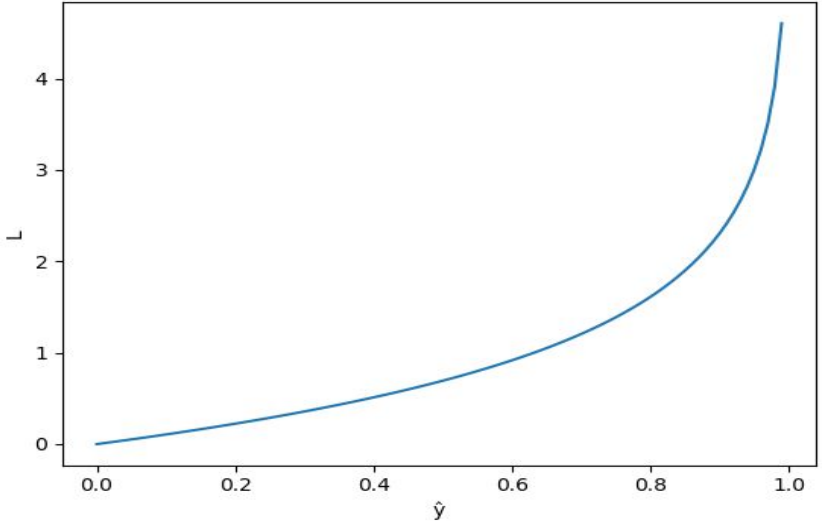
y=1



- Hàm L giảm dần từ 0 đến 1
- Khi model dự đoán $\hat{y_i}$ gần 1, tức giá trị dự đoán gần với giá trị thật y_i thì L nhỏ, xấp xỉ 0
- Khi model dự đoán $\hat{y_i}$ gần 0, tức giá trị dự đoán ngược lại giá trị thật y_i thì L rất lớn

Loss function

y=0



- Hàm L tăng dần từ 0 đến 1
- Khi model dự đoán $\hat{y_i}$ gần 0, tức giá trị dự đoán gần với giá trị thật y_i thì L nhỏ, xấp xỉ 0
- Khi model dự đoán $\hat{y_i}$ gần 1, tức giá trị dự đoán ngược lại giá trị thật y_i thì L rất lớn

Chain rule

$$z = f(y) \text{ và } y = g(x) \text{ hay } z = f(g(x)) \text{ thì } \frac{dz}{dx} = \frac{dz}{dy} * \frac{dy}{dx}$$

Ví dụ:

$z(x) = (2x+1)^2$, có thể thấy $z = f(g(x))$ trong đó $f(x) = x^2, g(x) = 2x + 1$

$$\frac{dz}{dx} = \frac{dz}{dy} * \frac{dy}{dx} = \frac{d(2x+1)^2}{d(2x+1)} * \frac{d(2x+1)}{dx} = 2 * (2x+1) * 2 = 4 * (2x+1)$$

$$\frac{d(\sigma(x))}{dx} = ???$$

Calculate gradient

L = $-(y_i * log(\hat{y_i}) + (1 - y_i) * log(1 - \hat{y_i}))$ trong đó $\hat{y_i} = \sigma(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)})$

$$\frac{dL}{dw_0} = \frac{dL}{d\hat{y_i}} * \frac{d\hat{y_i}}{dw_0}$$

$$\frac{dL}{d\hat{y_i}} = -\frac{d(y_i * log(\hat{y_i}) + (1 - y_i) * log(1 - \hat{y_i}))}{d\hat{y_i}} = -(\frac{y_i}{\hat{y_i}} - \frac{1 - y_i}{(1 - \hat{y})})$$

$$\frac{d\hat{y_i}}{dw_0} = \frac{\sigma(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)})}{dw_0} = \hat{y_i} * (1 - \hat{y_i})$$

$$\frac{dL}{dw_0} = \frac{dL}{d\hat{y_i}} * \frac{d\hat{y_i}}{dw_0} = -(\frac{y_i}{\hat{y_i}} - \frac{1 - y_i}{(1 - \hat{y_i})}) * \hat{y_i} * (1 - \hat{y_i}) = -(y_i * (1 - \hat{y_i}) - (1 - y_i) * \hat{y_i}) = \hat{y_i} - y_i$$

Matrix presentation

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)}) \\ 1 & x_1^{(2)} & x_2^{(2)}) \\ 1 & \ldots & \ldots \\ 1 & x_1^{(n)} & x_2^{(n)}) \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

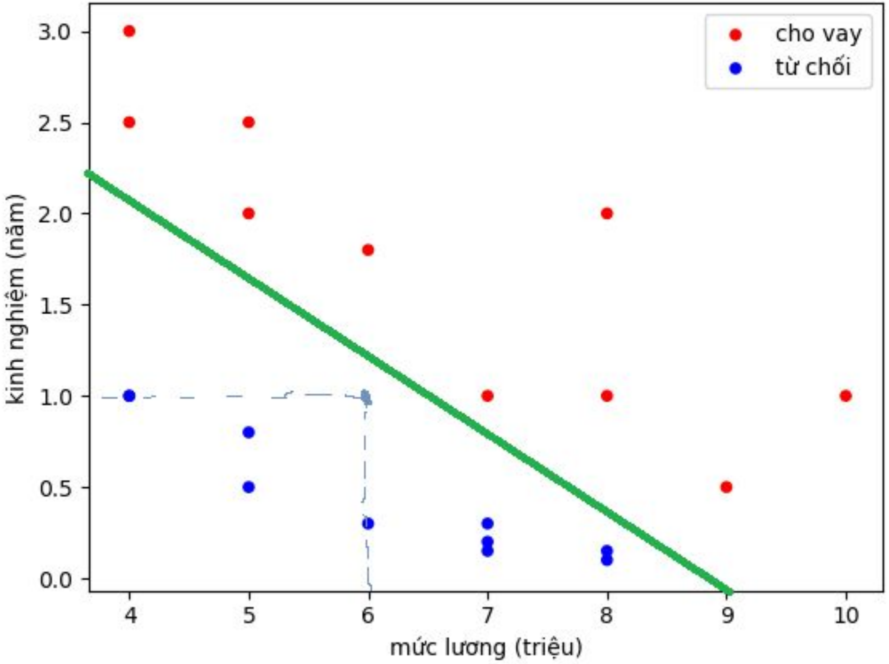
$$\hat{y} = \sigma(Xw)$$

$$J = -sum(y \otimes log(\hat{y}) + (1 - y) \otimes log(1 - \hat{y}))$$

$$\frac{dJ}{dw} = X^T * (\hat{y} - y), X^T = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1^{(1)} & x_1^{(2)} & \ldots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \ldots & x_2^{(n)}) \end{bmatrix}$$

Convert probability to line

$$p(x^{(i)} = 1) \Rightarrow$$

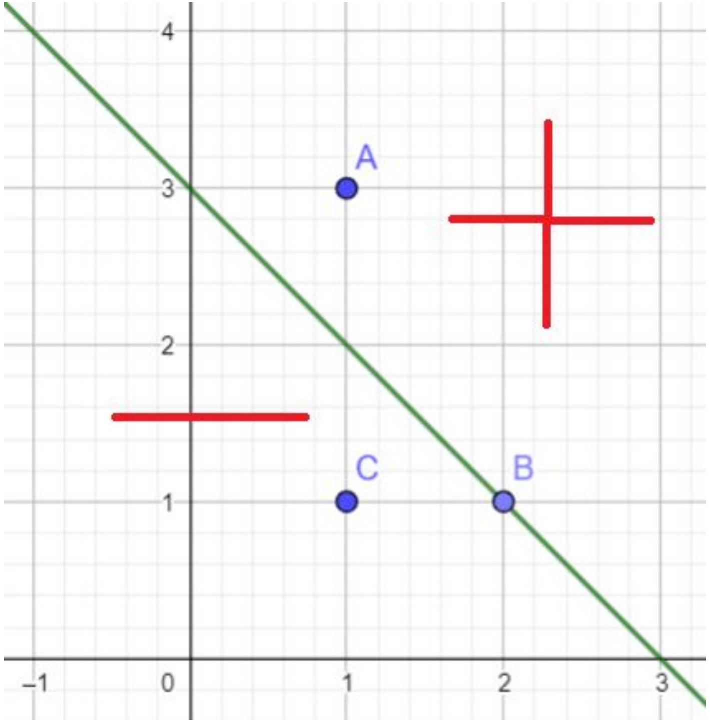


Choose a threshold

$$p(x^{(i)} = 1) >= 0.5 => x^{(i)}$$ Positive, nên cho vay

$$p(x^{(i)} = 1) < 0.5 => x^{(i)}$$ Negative, không cho vay

Positive/Negative



$$f = y - (ax+b)$$

Probability to line

$$\hat{y}_i >= 0.5 <=> \frac{1}{1 + e^{-(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)})}} >= 0.5$$

$$<=> 2 >= 1 + e^{-(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)})}$$

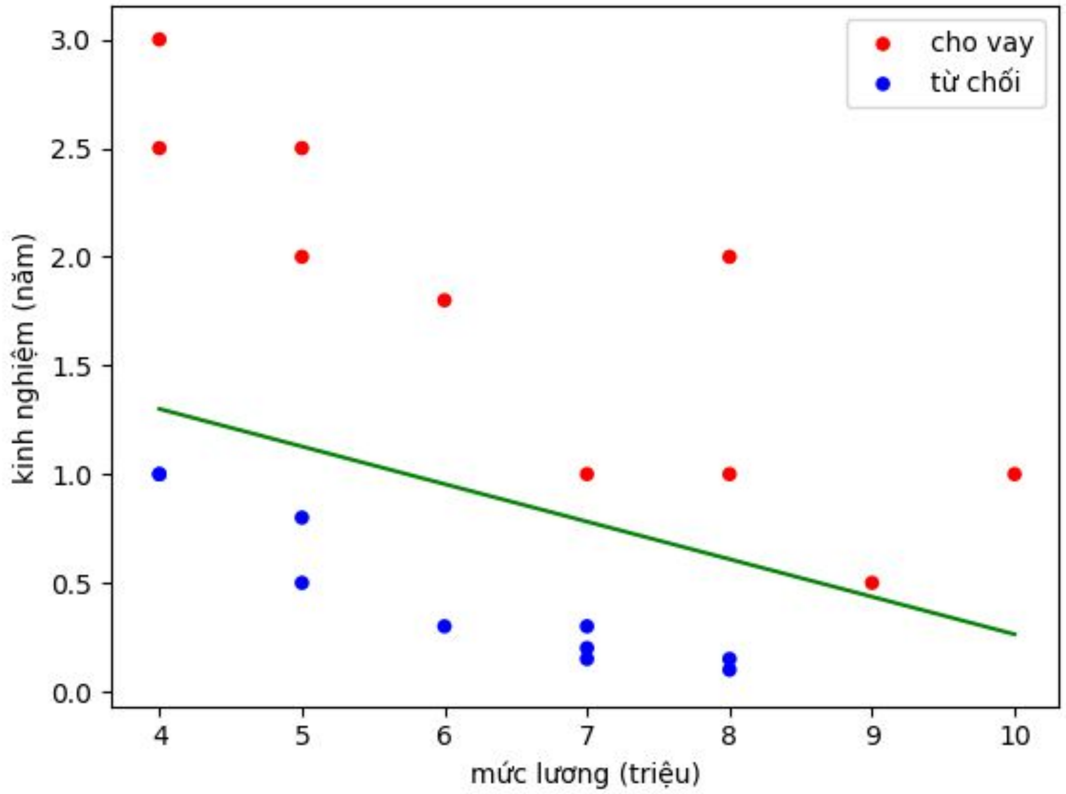
$$<=> e^{-(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)})} <= 1 = e^0$$

$$<=> -(w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)}) <= 0$$

$$<=> w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)} >= 0$$

Tương tự $\hat{y}_i < 0.5 <=> w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)} < 0$

Separated line

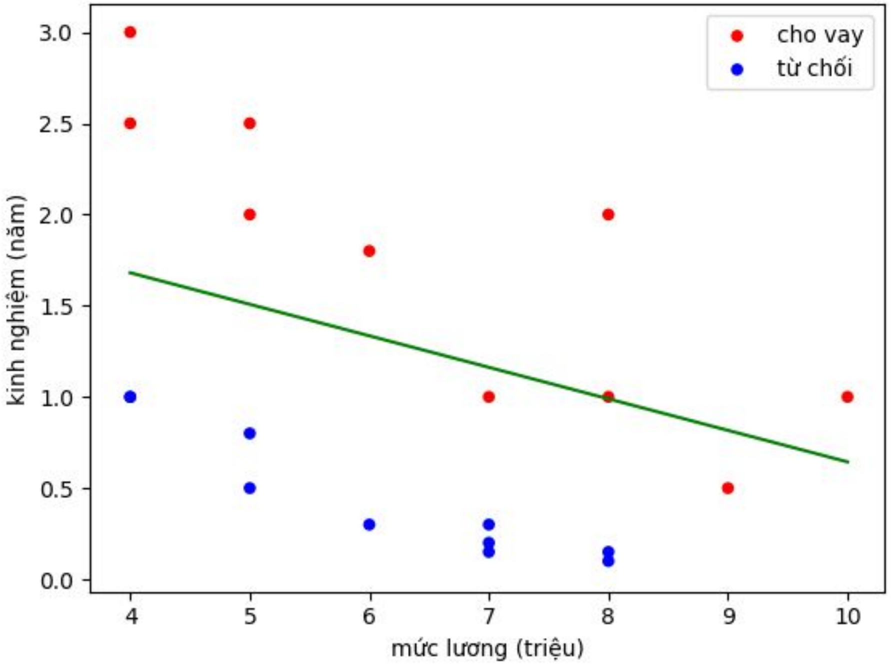


General formulation

Trong trường hợp tổng quát bạn lấy xác xuất lớn hơn t (0<t<1) thì mới cho vay tiền

$$\hat{y_i} > t <=> w_0 + w_1 * x_1^{(i)} + w_2 * x_2^{(i)} > -ln(\frac{1}{t} - 1)$$

Higher threshold (t = 0.8)



Q&A


Thank you.