Logistic Regression

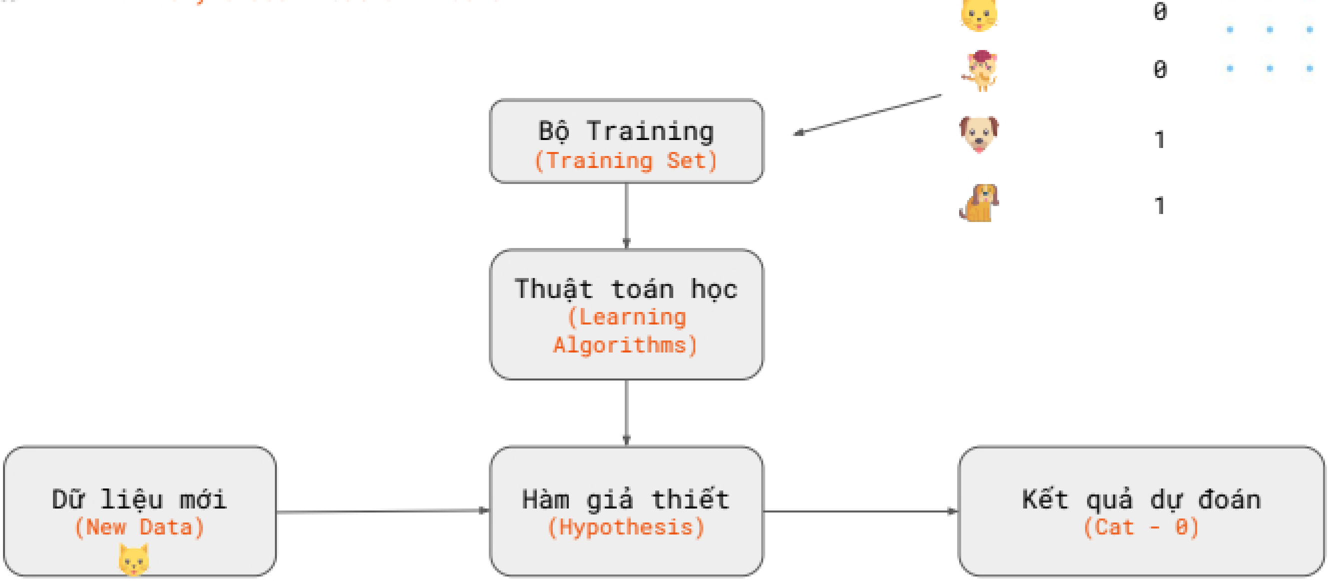
Hồi quy Logistic

PROTON X

PROTONX

Bài toán phân loại 2 lớp

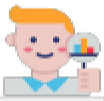
Binary Classification Problem



[Demo]

Hồi quy tuyến tính và hồi quy logistic

Linear Regression vs Logistic Regression

Số km	Thành tiền
2	13
7	35
9	41
3	19
10	45
6	28
1	10
8	55

Phân tích

- Đầu ra mô hình hồi quy tuyến tính là dạng **số vô hạn**
- Đầu ra mô hình hồi quy Logistic là dạng loại hữu hạn bao gồm 0 và 1

Ví dụ

Nhãn chó và mèo được **số hóa** dưới giá trị 0 và 1

 0

 0

 1

 1

Bộ dữ liệu Titanic

Titanic Dataset

C: Cherbourg

Q: Queenstown

S: Southampton

Cảng lên tàu

Hạng ghế

1: Thương gia

2: Phổ thông

3: Phổ thông tiết kiệm

Số lượng vợ chồng/anh chị em ruột trên tàu

Số lượng cha mẹ/con trên tàu

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Nhãn: Khả năng sống sót

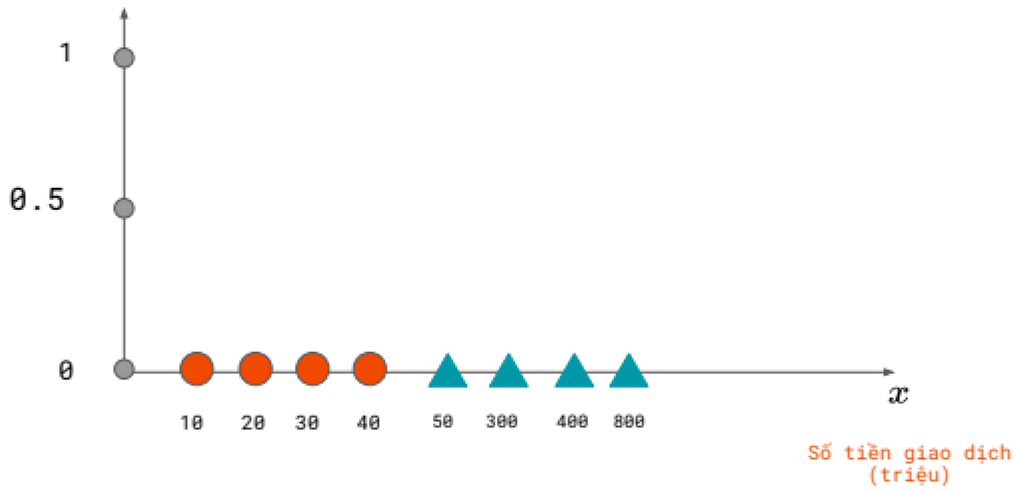
Chi tiết về bộ dữ liệu

Mã Vé

Chứa cả ký tự và số

Buồng

Phát hiện rủi ro
Fraud Detection

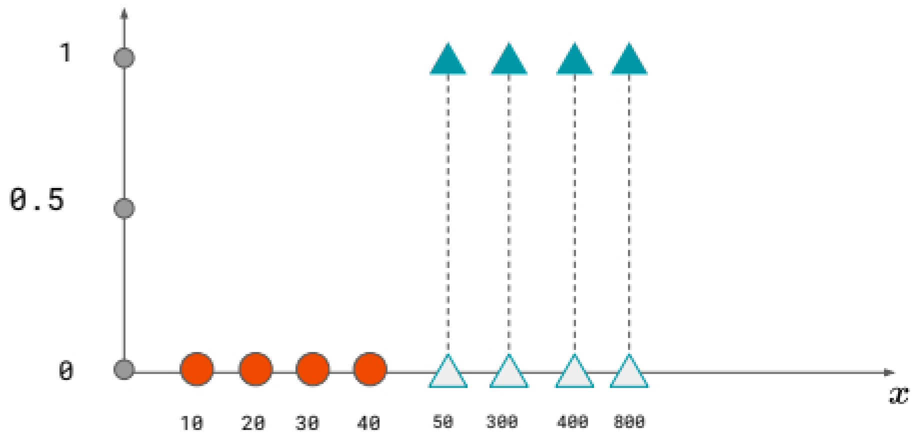


x	y
10	0
20	0
30	0
40	0
50	1
300	1
400	1
800	1

Áp dụng Linear Regression

Phát hiện rủi ro

Fraud Detection



$$y \in \{0, 1\}$$

Giao dịch bình thường

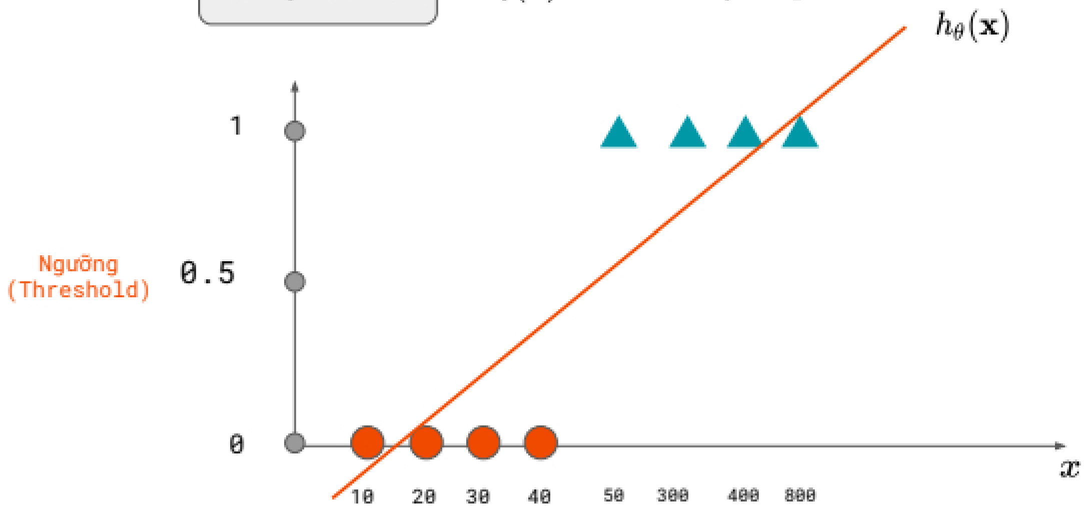
Giao dịch bất thường

Áp dụng Linear Regression

Phát hiện rủi ro

Fraud Detection

Hàm giả thiết $h_\theta(\mathbf{x}) = \theta^T \mathbf{x} = \theta_0 + \theta_1 x$

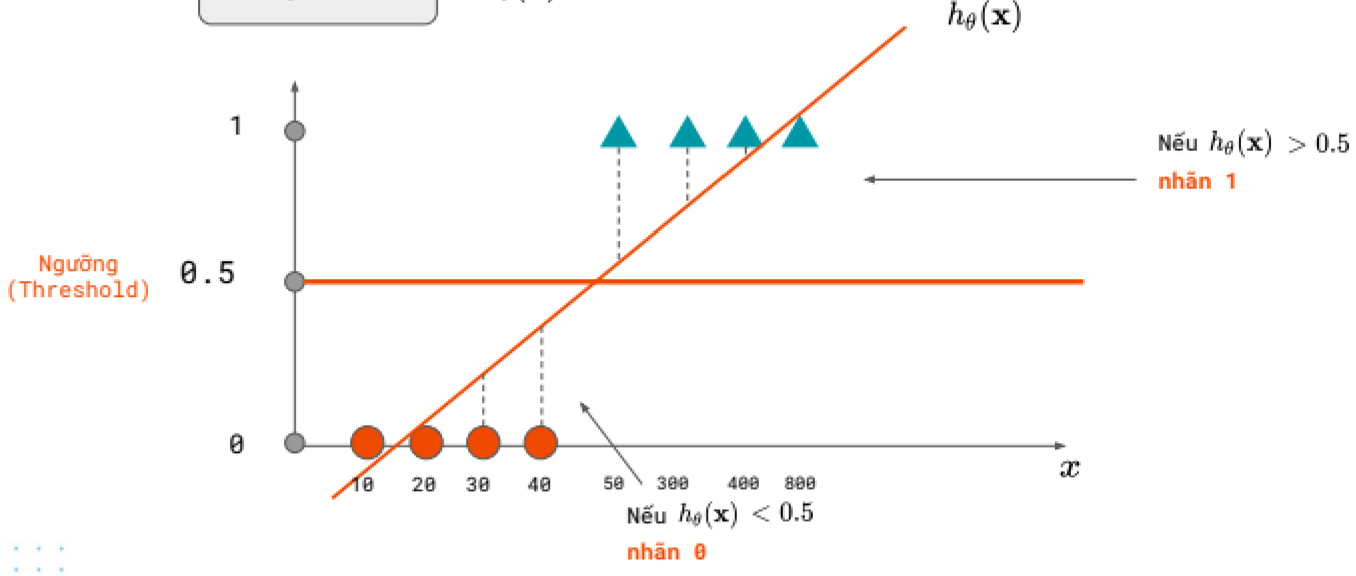


Áp dụng Linear Regression

Phát hiện rủi ro

Fraud Detection

Hàm giả thiết $h_\theta(\mathbf{x}) = \theta^T \mathbf{x} = \theta_0 + \theta_1 x$



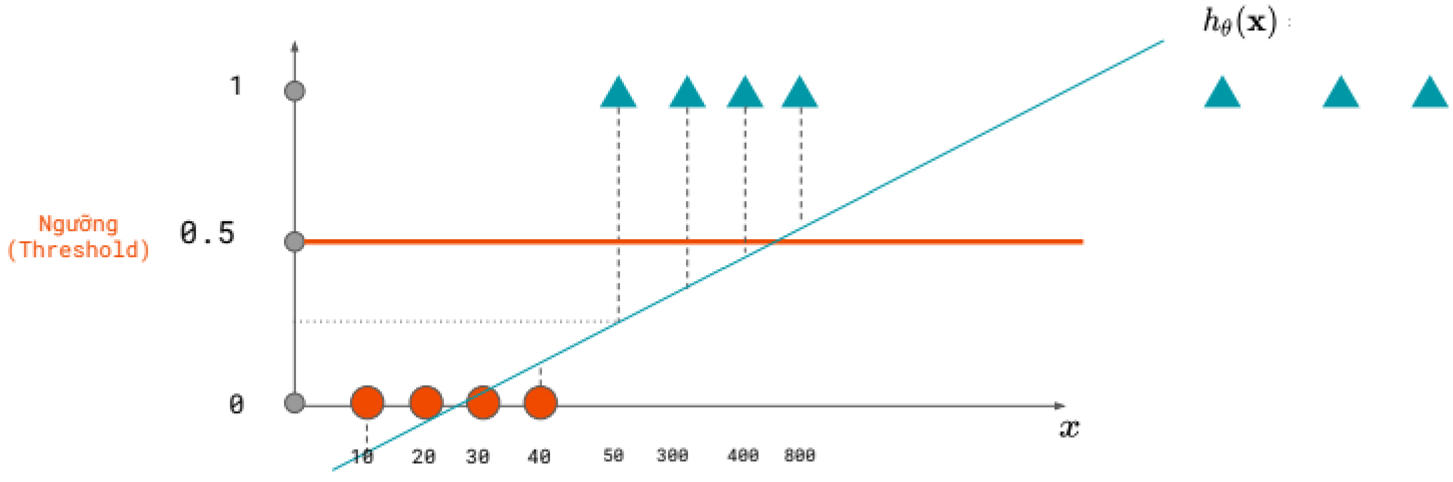
Áp dụng Linear Regression

Phát hiện rủi ro

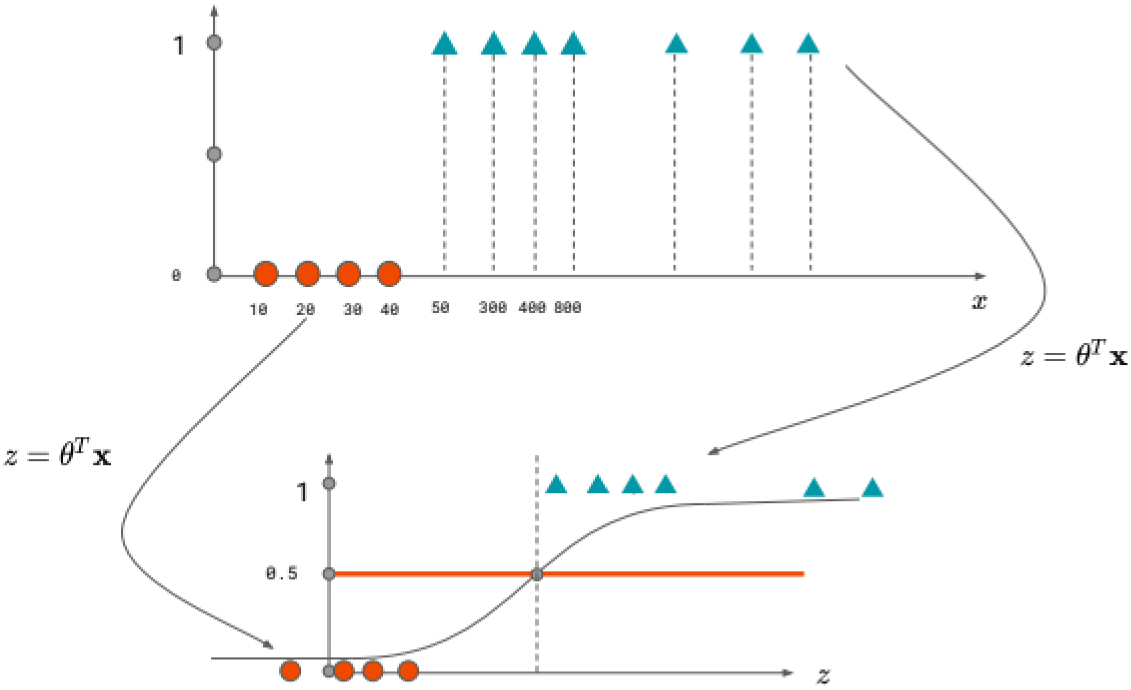
Fraud Detection

Hàm giả thiết $h_\theta(\mathbf{x}) = \theta^T \mathbf{x}$

Khi thêm một điểm dữ liệu mới làm cho mô hình tuyến tính **dự đoán sai** những điểm lúc trước



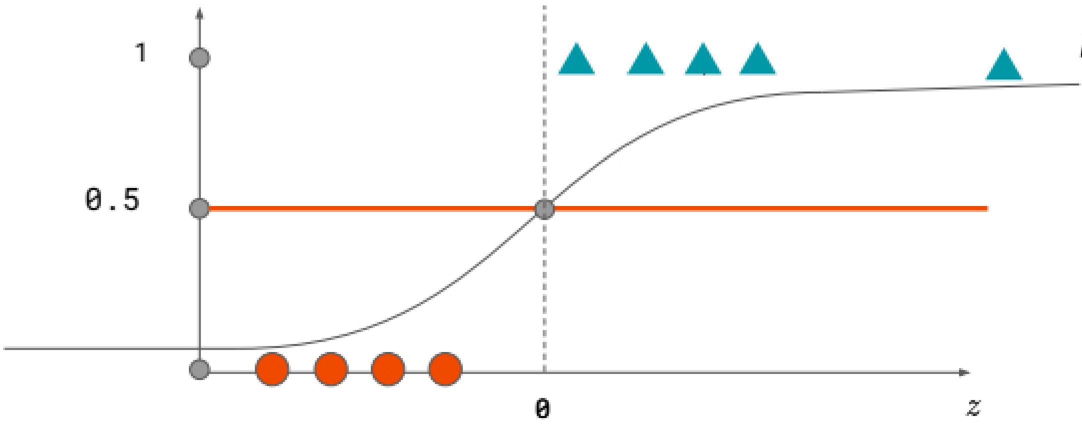
Áp dụng Linear Regression thường sẽ **không hiệu quả** với bài toán phân loại


1
0
10
20
30
40
50
300
400
800
x
$z = \theta^T \mathbf{x}$
$z = \theta^T \mathbf{x}$
1
0.5
z

Hàm Sigmoid

Sigmoid Function

Hàm giả thiết $h_\theta(\mathbf{x}) = \sigma(\theta^T\mathbf{x}) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-\theta^T\mathbf{x}}}$



$h_\theta(\mathbf{x})$

$z = \theta^T\mathbf{x} \in (-\infty, +\infty)$

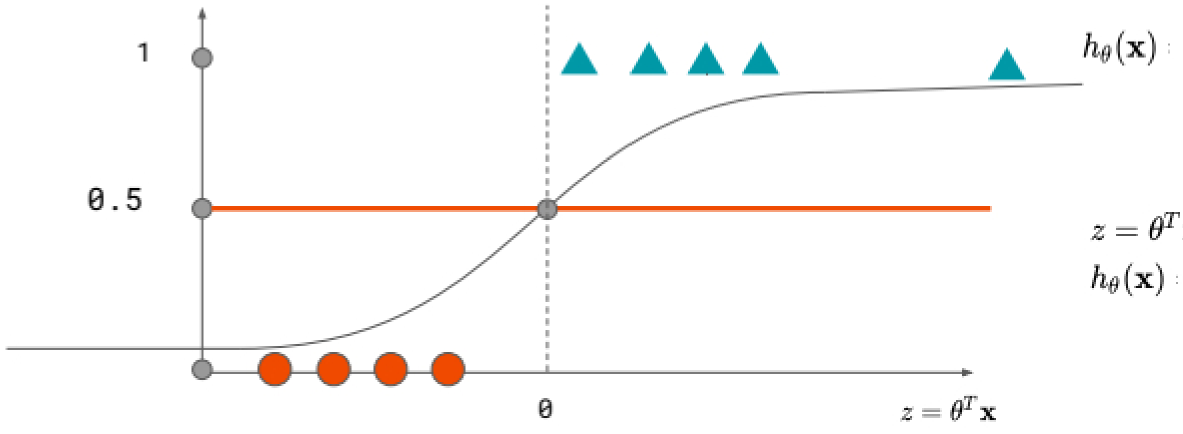
$h_\theta(\mathbf{x}) = \sigma(z) \in [0, 1]$

Hàm sigmoid có khả năng **chuyển phân phối** có những giá trị trong khoảng **[âm vô cùng, dương vô cùng]** sang phân phối có những giá trị nằm trong

Hàm Sigmoid

Sigmoid Function

Hàm giả thiết $h_\theta(\mathbf{x}) = \sigma(\theta^T \mathbf{x}) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-\theta^T \mathbf{x}}}$



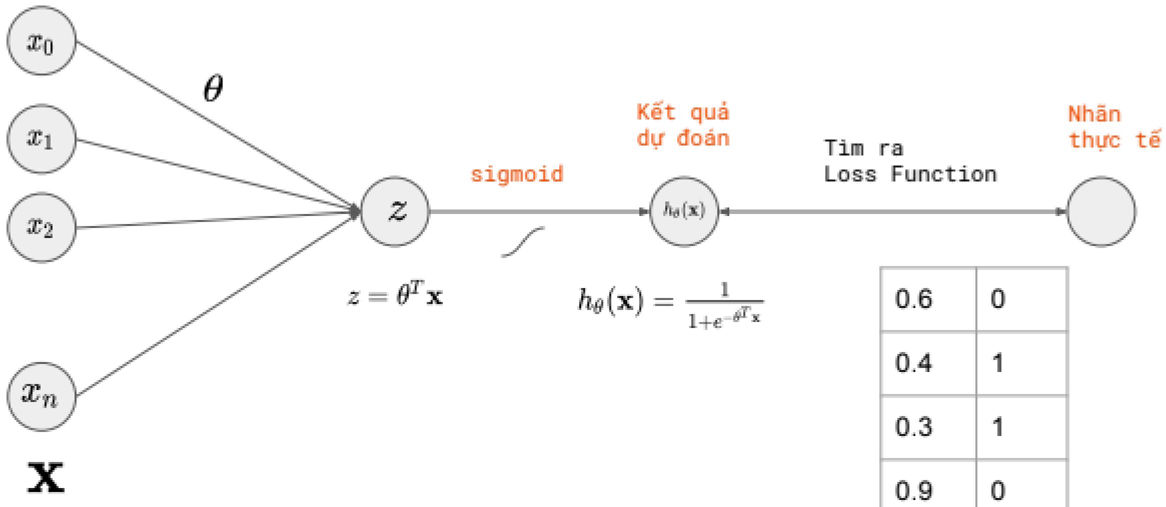
$z = \theta^T \mathbf{x} \in (-\infty, +\infty)$

$h_\theta(\mathbf{x}) = \sigma(z) \in [0, 1]$

Khi z **càng lớn** thì $h_\theta(\mathbf{x})$ **càng gần tới 1**
Khi z **càng nhỏ** thì $h_\theta(\mathbf{x})$ **càng gần tới 0**

Quy trình biến đổi trên 1 điểm dữ liệu

Transformation Progress on a data point



0.6	0
0.4	1
0.3	1
0.9	0
0.2	1

Điểm dữ liệu này giờ là một vector có n features

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

$z = \theta^T \mathbf{x}$

1

Giả thiết

0.6

$$P(y=1|\mathbf{x}) = h_\theta(\mathbf{x}) = \sigma(\theta^T \mathbf{x}) = \sigma(z)$$

Xác suất **dữ liệu là nhãn 1** khi cho X

$$P(y=0|\mathbf{x}) = 1 - h_\theta(\mathbf{x}) = 1 - \sigma(z)$$

Xác suất **dữ liệu là nhãn 0** khi cho X

0.4

Giống **phân phối** nào?

Trả lời nhanh tay
nhận ngay ly nước :D

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

$z = \theta^T \mathbf{x}$

1

Giả thiết: tại 1 điểm dữ liệu

$$P(y=1|\mathbf{x}) = h_\theta(\mathbf{x}) = \sigma(\theta^T \mathbf{x}) = \sigma(z)$$

Xác suất **dữ liệu là nhãn 1** khi cho X

$$P(y=0|\mathbf{x}) = 1 - h_\theta(\mathbf{x}) = 1 - \sigma(z)$$

Xác suất **dữ liệu là nhãn 0** khi cho X

Viết gọn:

$$p(y|\mathbf{x}) = (\sigma(z))^y (1 - \sigma(z))^{(1-y)}$$

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

$z = \theta^T \mathbf{x}$

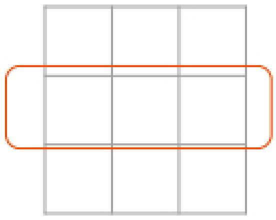
1. Giả thiết

$$p(y|\mathbf{x}) = (\sigma(z))^y (1 - \sigma(z))^{(1-y)}$$

2. Hàm Likelihood trên toàn tập dữ liệu X

$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})$

$\mathbf{x}^{(i)}$



$$
\begin{aligned}
L(\theta) &= p(\mathbf{y}|X;\theta) \\
&= \prod_{i=1}^{m} p(y^{(i)}|\mathbf{x}^{(i)};\theta) \\
&= \prod_{i=1}^{m} (\sigma(z^{(i)}))^{y^{(i)}} (1 - \sigma(z^{(i)}))^{(1-y^{(i)})} \\
&= \prod_{i=1}^{m} (h_\theta(\mathbf{x}^{(i)}))^{y^{(i)}} (1 - h_\theta(\mathbf{x}^{(i)}))^{(1-y^{(i)})}
\end{aligned}
$$

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

$z = \theta^T \mathbf{x}$

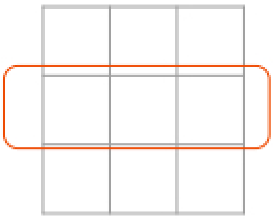
① Giả thiết

$$p(y|\mathbf{x}) = (\sigma(z))^y(1 - \sigma(z))^{(1-y)}$$

② Hàm Likelihood trên toàn tập dữ liệu X

$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})$

$\mathbf{x}^{(i)}$



$$\begin{aligned} L(\theta) &= p(\mathbf{y}|X;\theta) \\ &= \prod_{i=1}^{m} p(y^{(i)}|\mathbf{x}^{(i)};\theta) \\ &= \prod_{i=1}^{m} (\sigma(z^{(i)}))^{y^{(i)}} (1 - \sigma(z^{(i)}))^{(1-y^{(i)})} \\ &= \prod_{i=1}^{m} (h_\theta(\mathbf{x}^{(i)}))^{y^{(i)}} (1 - h_\theta(\mathbf{x}^{(i)}))^{(1-y^{(i)})} \end{aligned}$$

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

$z = \theta^T \mathbf{x}$

1. Giả thiết

$$p(y|\mathbf{x}) = (\sigma(z))^y (1 - \sigma(z))^{(1-y)}$$

2. Hàm Likelihood

$$L(\theta) = \prod_{i=1}^{m} (h_\theta(\mathbf{x}^{(i)}))^{y^{(i)}} (1 - h_\theta(\mathbf{x}^{(i)}))^{(1-y^{(i)})}$$

3. Hàm Log Likelihood

$$LL(\theta) = \sum_{i=1}^{m} y^{(i)} \log h_\theta(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(\mathbf{x}^{(i)}))$$

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

2. Hàm Likelihood

$$L(\theta) = \prod_{i=1}^{m} (h_\theta(\mathbf{x}^{(i)}))^{y^{(i)}} (1 - h_\theta(\mathbf{x}^{(i)}))^{(1-y^{(i)})}$$

3. Hàm Log Likelihood

$$LL(\theta) = \sum_{i=1}^{m} y^{(i)} \log h_\theta(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(\mathbf{x}^{(i)}))$$

4. Cực đại hóa Log Likelihood. 3 phương pháp

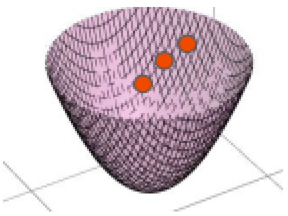
- Cực tiểu hóa $-LL(\theta)$ sử dụng Gradient Descent
- Cực đại hóa $LL(\theta)$ sử dụng Gradient Ascent
- Tính Gradient $\nabla LL(\theta)$ và cho bằng 0 rồi tìm theta

Hàm mất mát trên 1 điểm dữ liệu

Loss Function

4

Cực đại hóa Log Likelihood

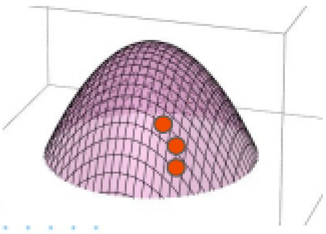


Cách 1: Cực tiểu hóa $-LL(\theta)$ sử dụng Gradient Descent

Cross Entropy

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(\mathbf{x}^{(i)}))$$

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(\mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$



Cách 2: Cực đại hóa $LL(\theta)$ sử dụng Gradient Ascent

$$LL(\theta) = \sum_{i=1}^{m} y^{(i)} \log h_\theta(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(\mathbf{x}^{(i)}))$$

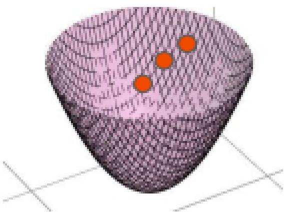
$$\theta_j := \theta_j + \alpha \frac{\partial LL(\theta)}{\partial \theta_j} := \theta_j + \alpha \sum_{i=1}^{m} (y^{(i)} - h_\theta(\mathbf{x}^{(i)})) x_j^{(i)}$$

Gradient Descent

Gradient Descent

4

Cực đại hóa Log Likelihood



Cực tiểu hóa $-LL(\theta)$ sử dụng Gradient Descent

Cross Entropy

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta(\mathbf{x}^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(\mathbf{x}^{(i)}))$$

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(\mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

Công thức tổng quát trên toàn tập dữ liệu X

$$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})$$

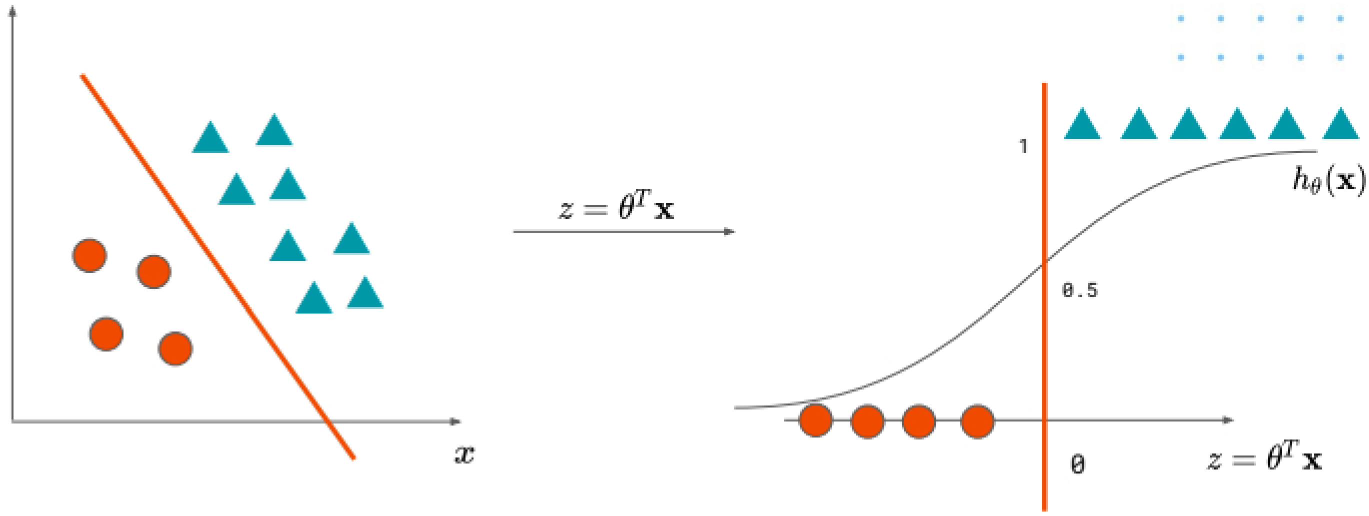
$$J(\theta) = -\frac{1}{m} \sum (\mathbf{y} \odot \log(h_\theta(X)) + (1 - \mathbf{y}) \odot (1 - \log(h_\theta(X))))$$

Quá trình training. Ở mỗi vòng lặp (epoch)

???

Ranh giới quyết định

Decision Boundary



Giả sử mô hình dự đoán là nhãn 1

$$\longrightarrow \quad h_\theta(\mathbf{x}) \geq \frac{1}{2} \Leftrightarrow e^{-z} \leq 1 \Leftrightarrow z \geq 0 \Leftrightarrow \theta^T \mathbf{x} \geq 0$$

Giả sử mô hình dự đoán là nhãn 0

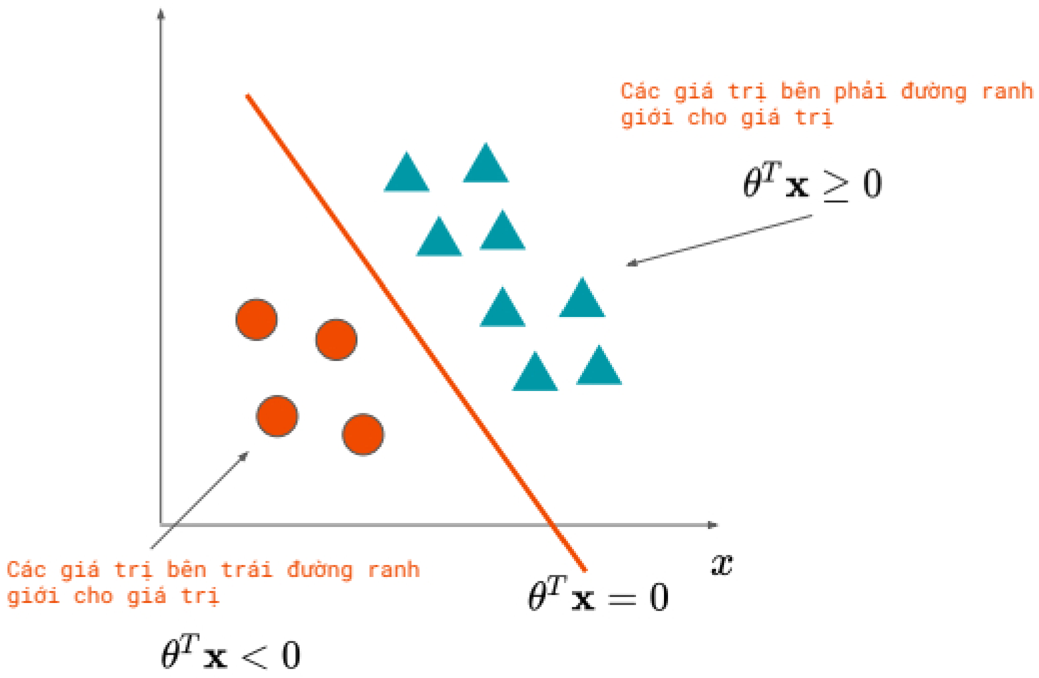
$$\longrightarrow \quad h_\theta(\mathbf{x}) < \frac{1}{2} \Leftrightarrow \theta^T \mathbf{x} < 0$$

$$\theta^T \mathbf{x} = 0$$

Chính là đường ranh giới quyết định phân chia nhãn

Ranh giới quyết định

Decision Boundary



Practice

Thực hành

PROTONX