4. Logistic Regression

4.1 Classification

Krittameth Teachasrisaksakul

Classification (การจำแนกประเภท): ความเข้าใจพื้นฐาน

เราจะพิจารณา**ปัญหา binary classification** : ผล *y* เป็นไปได้ 2 ค่า คือ 0, 1 (False, True)

- Email: spam / ไม่ใช่ spam ?
- ธุรกรรมทางการเงินออนไลน์ (online transactions): fraudulent หลอกลวง (ใช่ / ไม่) ?
- เนื้องอก (tumor): malignant (ร้าย) / benign (ไม่ร้าย) ?

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

0: 'Negative Class' (*e.g.*, benign tumor)

1: 'Positive Class' (*e.g.*, malignant tumor)

Classification (การจำแนกประเภท): ความเข้าใจพื้นฐาน

เราจะพิจารณา**ปัญหา binary classification** : ผล y เป็นไปได้ 2 ค่า คือ 0, 1 (False, True)

- Email: spam / ไม่ใช่ spam ?
- ธุรกรรมทางการเงินออนไลน์ (online transactions): fraudulent หลอกลวง (ใช่ / ไม่) ?
- เนื้องอก (tumor): malignant (ร้าย) / benign (ไม่ร้าย) ?

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

0: ‘Negative Class’ (*e.g.*, benign tumor)

1: ‘Positive Class’ (*e.g.*, malignant tumor)

(ในบทเรียนหลังจากนี้) เราจะพิจารณากรณี multi-classes (มากกว่า 2 class) เช่น $y \in \{0,1,2,3\}$

ก็คือ **‘multi-classes classification problem’** (ปัญหาการจำแนกประเภทที่มีมากกว่า 2 class)

Classification : ความเข้าใจพื้นฐาน

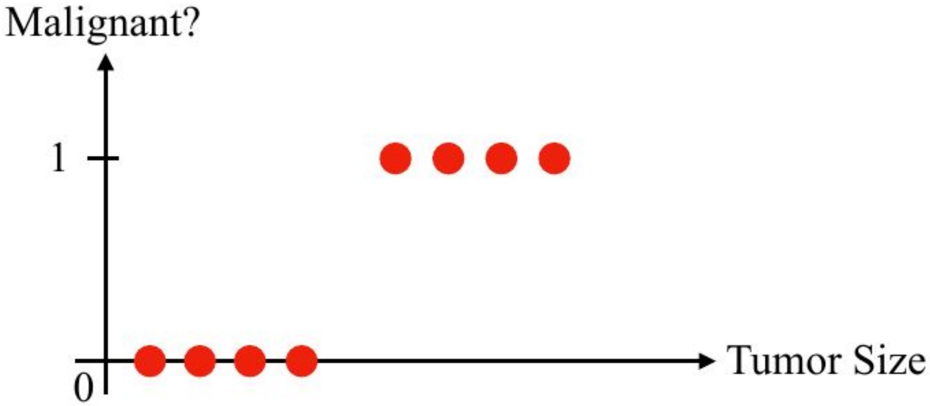
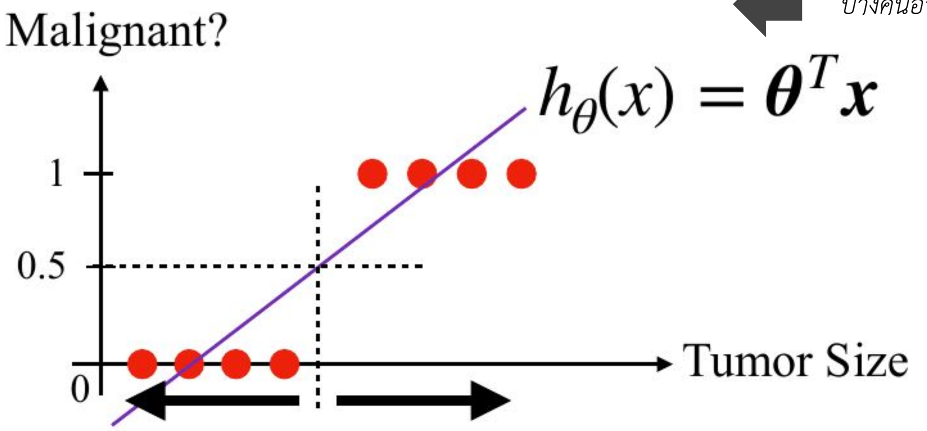


Figure Source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Classification : ความเข้าใจพื้นฐาน



บางคนอาจคิดว่า model นี้ น่าจะ work แต่อาจทำงานไม่ดี

วิธีหนึ่งที่จะทำ classification : **ใช้ linear regression เพื่อใช้ *x* ทำนาย *y*** โดย map (เชื่อมโยง) $h_\theta(x)$ ไปหา y

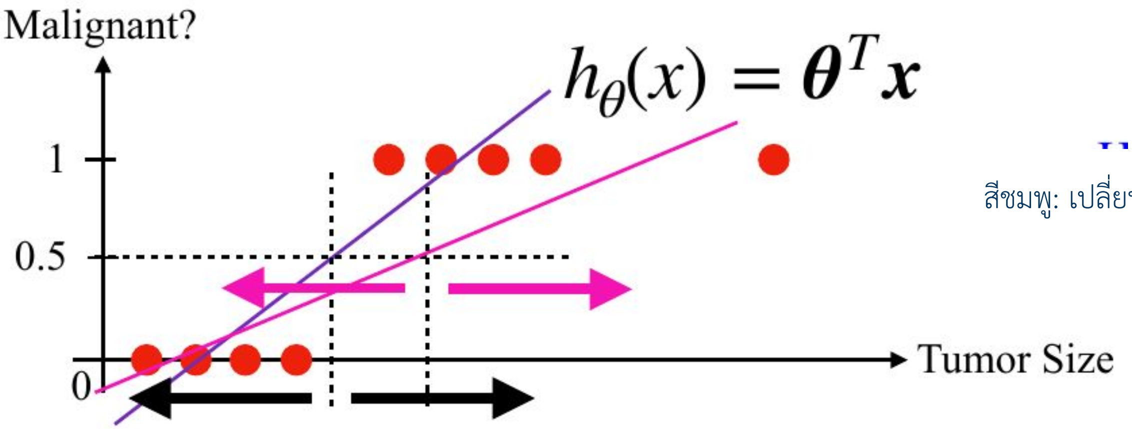
- $h_\theta(x) =$ ค่าที่ linear regresion ทำนาย
- $y =$ ผล (class)

ค่า threshold ของ $h_\theta(x)$ ที่ 0.5 เพื่อบอก classifier output (ผล/คำตอบจากตัวแยกประเภท):

- ถ้า $h_\theta(x) \geq 0.5 \quad \rightarrow$ ทำนาย $y = 1$
- ถ้า $h_\theta(x) \leq 0.5 \quad \rightarrow$ ทำนาย $y = 0$

Figure Source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Classification : ความเข้าใจพื้นฐาน



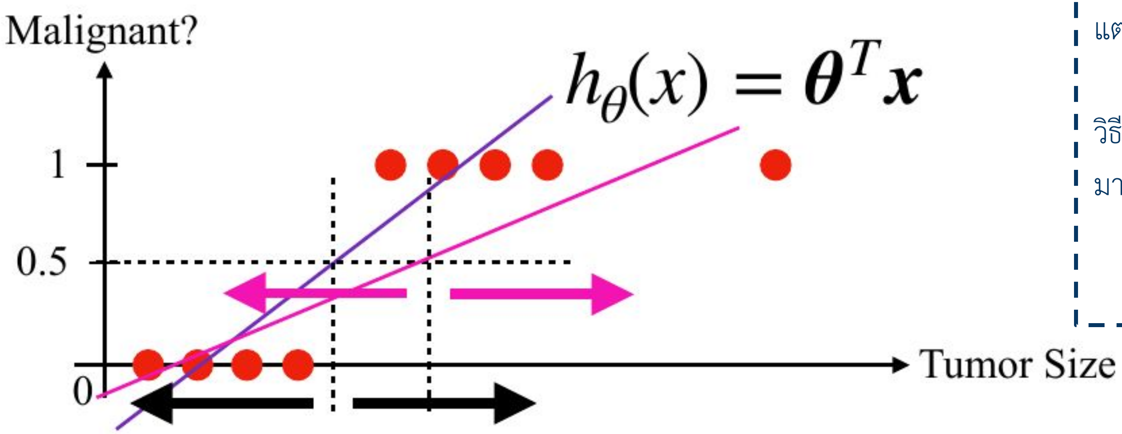
สีชมพู: เปลี่ยนค่า threshold

ค่า threshold ของ $h_\theta(x)$ ที่ 0.5 เพื่อบอก classifier output (ผล/คำตอบจากตัวแยกประเภท):

- ถ้า $h_\theta(x) \geq 0.5$ $\rightarrow$ ทำนาย $y = 1$
- ถ้า $h_\theta(x) \leq 0.5$ $\rightarrow$ ทำนาย $y = 0$

Figure Source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Classification : ความเข้าใจพื้นฐาน



$y \in \{0, 1\}$: ผล classification (classification output)

- y เป็นไปได้ 2 ค่า คือ 0 หรือ 1 เท่านั้น

แต่ $h_\theta(x)$ มีค่า > 1 หรือ < 0 ก็ได้

วิธีนี้ (ใช้ linear regression ทำ classification) จึงทำงานได้แย่มาก

ค่า threshold ของ $h_\theta(x)$ ที่ 0.5 เพื่อบอก classifier output (ผล/คำตอบจากตัวแยกประเภท):

- ถ้า $h_\theta(x) \geq 0.5$ $\rightarrow$ ทำนาย $y = 1$
- ถ้า $h_\theta(x) \leq 0.5$ $\rightarrow$ ทำนาย $y = 0$

Figure Source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

คำถาม

คำกล่าวข้อใดเป็นจริง ?

(i) ถ้า linear regression ใช้ทำ classification ไม่ได้ผล (ในตัวอย่างก่อนหน้า) การทำ feature scaling อาจช่วยได้

(ii) ถ้าชุดข้อมูล training set ทำให้ $0 \leq y^{(i)} \leq 1$ เป็นจริง สำหรับ training example ทุกอัน $(x^{(i)}, y^{(i)})$ แล้วผลทำนายของ linear regression จะทำให้ $0 \leq h_\theta(x) \leq 1$ เป็นจริง สำหรับทุกๆค่าของ x

(iii) ถ้ามี feature x ที่ทำนาย y ได้อย่างสมบูรณ์แบบ ก็คือ ถ้า y = 1 เมื่อ $x \geq c$ และ $y = 0$ เมื่อ $x < c$ (เมื่อ c เป็นค่าคงที่) แล้ว linear regression จะมี classification error เป็น 0 (ศูนย์)

(iv) ไม่มีคำกล่าวใดถูกต้อง

คำถาม

คำกล่าวข้อใดเป็นจริง ?

(i) ถ้า linear regression ใช้ทำ classification ไม่ได้ผล (ในตัวอย่างก่อนหน้า) การทำ feature scaling อาจช่วยได้

(ii) ถ้าชุดข้อมูล training set ทำให้ $0 \leq y^{(i)} \leq 1$ เป็นจริง สำหรับ training example ทุกอัน $(x^{(i)}, y^{(i)})$ แล้วผลทำนายของ linear regression จะทำให้ $0 \leq h_\theta(x) \leq 1$ เป็นจริง สำหรับทุกๆค่าของ x

(iii) ถ้ามี feature x ที่ทำนาย y ได้อย่างสมบูรณ์แบบ ก็คือ ถ้า $y = 1$ เมื่อ $x \geq c$ และ $y = 0$ เมื่อ $x < c$ (เมื่อ c เป็นค่าคงที่) แล้ว linear regression จะมี classification error เป็น 0 (ศูนย์)

(iv) ไม่มีคำกล่าวใดถูกต้อง

4. Logistic Regression

4.2 Hypothesis Representation

(การเขียนอธิบาย Hypothesis
ด้วยสัญลักษณ์คณิตศาสตร์)

Krittameth Teachasrisaksakul

Logistic Regression Model

$0 \leq h_\theta(x) \leq 1$

การปรับปรุง: เปลี่ยนรูปแบบของ hypothesis function $h_\theta(x)$ เพื่อ**จำกัดค่า**ของมันอยู่**ในช่วง** **[0, 1]**
โดยแทนค่า $\theta^T x$ ใน logistic function

$$h_\theta(\mathbf{x}) = g(\boldsymbol{\theta}^T \mathbf{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \mathbf{x}}}$$

นิยามของ function g :

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

Function นี้ เรียกว่า logistic sigmoid function หรือ sigmoid function

Logistic Regression Model

อยากให้ $0 \le \boldsymbol{h_\theta(x)} \le 1$

เพราะอยากทำ binary classification ที่ผลเป็นไปได้ 2 ค่า คือ 0, 1 (true, false)

การปรับปรุง: เปลี่ยนรูปแบบของ hypothesis function h เพื่อ**จำกัดค่า**ของมันอยู่**ในช่วง** **[0, 1]**:

$$h_\theta(\boldsymbol{x}) = g(\boldsymbol{\theta}^T \boldsymbol{x}) = \frac{1}{1 + e^{\boxed{-\boldsymbol{\theta}^T \boldsymbol{x}}}}$$

<ins>เป้าหมาย</ins>: หาค่า parameter θ ที่เหมาะกับข้อมูล

(fit parameter θ to the data)

นิยามของ function g :

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

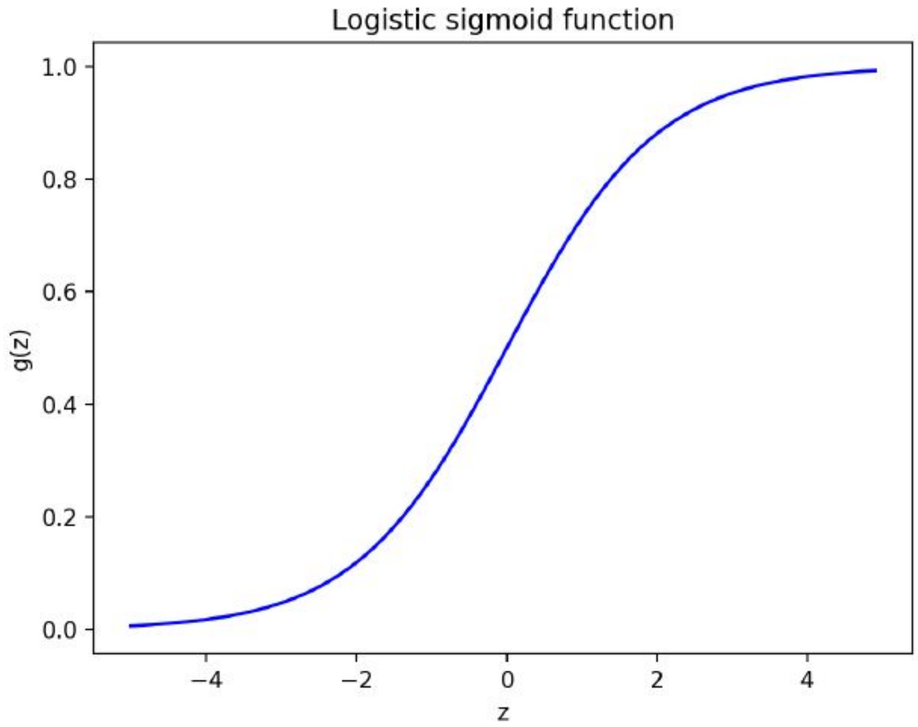
Function นี้ เรียกว่า **logistic sigmoid function** หรือ **sigmoid function**

Logistic Sigmoid Function

```python
import numpy
import matplotlib.pyplot as plt

def f(z):
    return 1/(1 + numpy.exp(-z))

z = numpy.arange(-5, 5, 0.1)
plt.plot(z, f(z), 'b')
plt.xlabel('z')
plt.ylabel('g(z)')
plt.title('Logistic sigmoid function')
```



Slide from: *Teeradaj Racharak; AI Practical Development Bootcamp [L 3.2]*

การอ่านผล / output ของ Hypothesis

$$h_\theta(\boldsymbol{x}) = g(\boldsymbol{\theta}^T\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T\boldsymbol{x}}}$$

การประมาณค่า ความน่าจะเป็น (probability) ที่ $y = 1$ เมื่อมี input x

การอ่านผล / output ของ Hypothesis

$$h_\theta(\boldsymbol{x}) = g(\boldsymbol{\theta}^T\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T\boldsymbol{x}}}$$

การประมาณค่า ความน่าจะเป็น (probability) ที่ $y = 1$ เมื่อมี input x

ตัวอย่าง: สมมติ

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \textbf{tumorSize} \end{bmatrix} \text{ and } h_\theta(x) = 0.7$$

นี่หมายความว่า มีโอกาส 70% ที่เนื้องอกจะเป็นเนื้อร้าย (malignant)

การอ่านผล / output ของ Hypothesis

$$h_\theta(\boldsymbol{x}) = g(\boldsymbol{\theta}^T\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T\boldsymbol{x}}}$$

การประมาณค่า ความน่าจะเป็น (probability) ที่ $y = 1$ เมื่อมี input x

ตัวอย่าง: สมมติ

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \textbf{tumorSize} \end{bmatrix} \text{ and } h_\theta(x) = 0.7$$

นี่หมายความว่า มีโอกาส 70% ที่เนื้องอกจะเป็นเนื้อร้าย (malignant)

เขียนในทางคณิตศาสตร์ว่า

$$h_\theta(x) = P(y = 1 \mid x; \theta)$$

'ความน่าจะเป็นที่ $\boldsymbol{y}$ = 1 เมื่อรู้ค่า $\boldsymbol{x}$ และ มี parameter เป็น $\boldsymbol{\theta}$ '

คำถาม

สมมติ เราอยากทำนายจากข้อมูล x เกี่ยวกับเนื้องอกว่ามันเป็นเนื้อร้าย ($y = 1$) หรือไม่ร้าย ($y = 0$) ตัวแบ่งประเภท (classifier) แบบ logistic regression บอกผลของเนื้องอกแต่ละอัน $h_\theta(x) = P(y = 1 \mid x; \theta) = 0.7$ ดังนั้น เราประมาณค่าว่ามีโอกาส 70% ว่าเนื้องอกจะเป็นเนื้อร้าย ค่าประมาณของ $P(y = 0 \mid x; \theta)$ ความน่าจะเป็นที่เนื้องอกไม่ร้าย

(i) $P(y = 0|x;\, \theta) = 0.3$

(ii) $P(y = 0|x;\, \theta) = 0.7$

(iii) $P(y = 0|x;\, \theta) = 0.7^2$

(iv) $P(y = 0|x;\, \theta) = 0.3 \times 0.7$

คำใบ้:

$P(y = 0|x;\, \theta) + P(y = 1|x;\, \theta) = 1$

คำถาม

สมมติ เราอยากทำนายจากข้อมูล x เกี่ยวกับเนื้องอกว่ามันเป็นเนื้อร้าย ($y = 1$) หรือไม่ร้าย ($y = 0$) ตัวแบ่งประเภท (classifier) แบบ logistic regression บอกผลของเนื้องอกแต่ละอัน $h_\theta(x) = P(y = 1 \mid x; \theta) = 0.7$ ดังนั้น เราประมาณค่าว่ามีโอกาส 70% ว่าเนื้องอกจะเป็นเนื้อร้าย ค่าประมาณของ $P(y = 0 \mid x; \theta)$ ความน่าจะเป็นที่เนื้องอกไม่ร้าย

(i) $P(y = 0|x; \theta) = 0.3$

(ii) $P(y = 0|x; \theta) = 0.7$

(iii) $P(y = 0|x; \theta) = 0.7^2$

(iv) $P(y = 0|x; \theta) = 0.3 \times 0.7$

คำใบ้:

$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$

4. Logistic Regression

4.3 Decision Boundary

(เส้นขอบเขตตัดสินใจ)

Krittameth Teachasrisaksakul

Decision boundary : ความเข้าใจพื้นฐาน

พฤติกรรมของฟังก์ชั่น logistic regression หรือ g คือ
เมื่อ $\boldsymbol{input \geq 0}$ จะได้ $\boldsymbol{output \geq 0.5}$

$\because$ เมื่อ $z \geq 0 \rightarrow \quad g(z) \geq 0.5$

$\therefore$ เมื่อ $\theta^T x \geq 0 \quad \rightarrow \quad h_\theta(x) = g(\theta^T x) \geq 0.5$

Hypothesis:

$$P(y = 1 \mid x; \theta)$$

$$h_\theta(\boldsymbol{x}) = g(\boldsymbol{\theta}^T \boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}}}$$

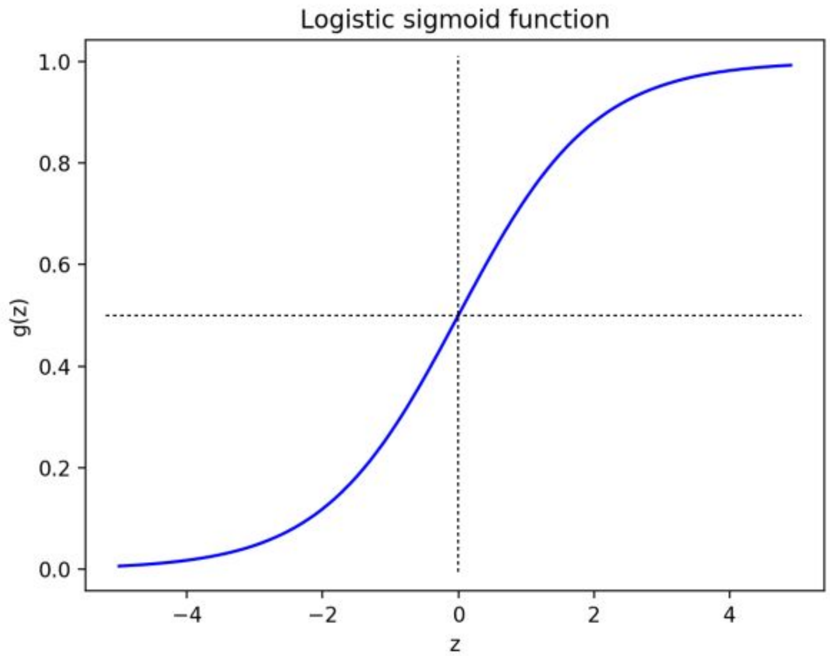


Figure source: *Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]*

Decision boundary

เพื่อทำ classification ให้ได้ค่า 0 หรือ 1 เราจะแปลง output ของ hypothesis function

Hypothesis:

$P(y=1|x;\theta)$

$$h_\theta(\mathbf{x}) = g(\boldsymbol{\theta}^T\mathbf{x}) = \frac{1}{1+e^{-\boldsymbol{\theta}^T\mathbf{x}}}$$

$\because$ เมื่อ $z \geq 0 \rightarrow \quad g(z) \geq 0.5$

$\therefore$ เมื่อ $\theta^T x \geq 0 \quad \rightarrow \quad h_\theta(x)=g(\theta^T x) \geq 0.5$

- ถ้า $h_\theta(x) \geq 0.5 \rightarrow$ ทำนาย '$y = 1$'
  - $\theta^T x \geq 0$
- ถ้า $h_\theta(x) < 0.5 \rightarrow$ ทำนาย '$y = 0$'
  - $\theta^T x < 0$

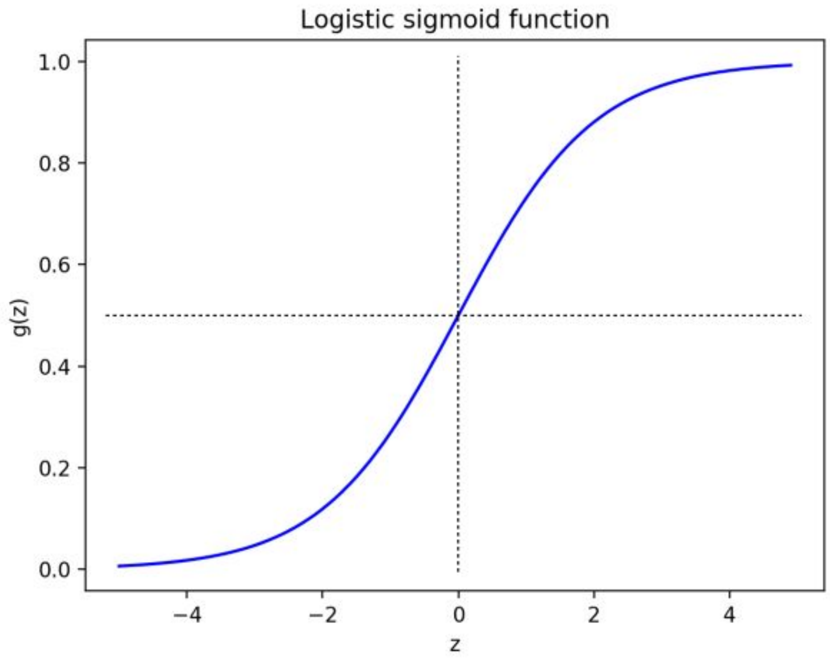
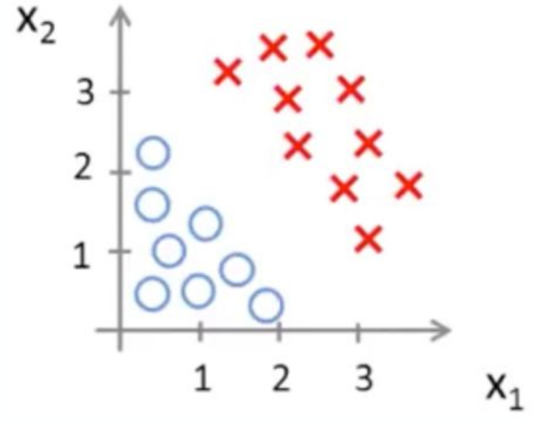


Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary : ความเข้าใจพื้นฐาน



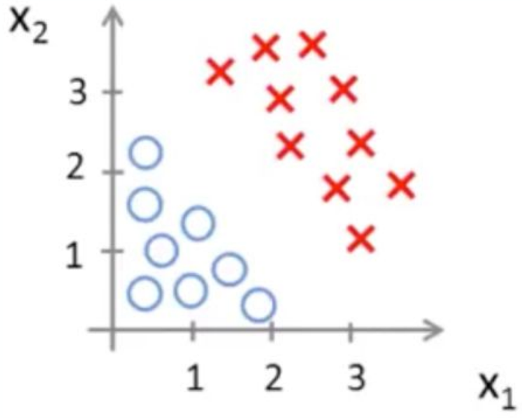
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

-3 1 1

$$i.e.\ \ \theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary : ความเข้าใจพื้นฐาน



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

-3 1 1

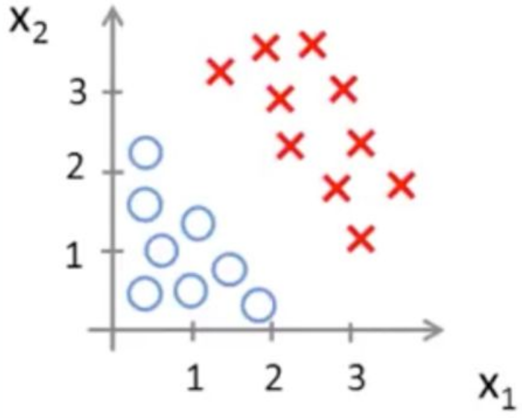
$$i.e.\ \ \theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$\because \theta^T x = [-3 \quad 1 \quad 1] \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = -3 + x_1 + x_2$$

$\therefore$ Predict '$y = 1$' if $-3 + x_1 + x_2 \geq 0$

Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary : ความเข้าใจพื้นฐาน



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

-3 1 1

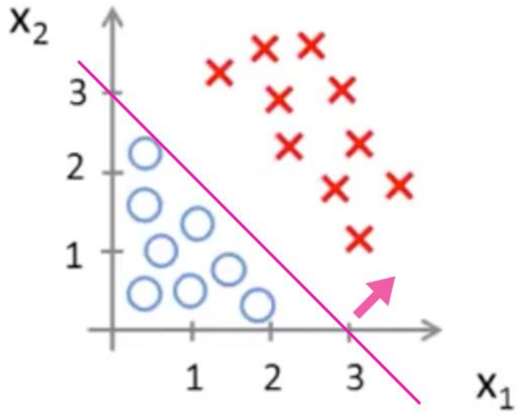
$$i.e.\ \ \theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$\because \theta^T x = [-3 \quad 1 \quad 1] \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = -3 + x_1 + x_2$$

$$\therefore \text{ Predict } `y = 1\text{' if } -3 + x_1 + x_2 \geq 0 \Longleftrightarrow x_1 + x_2 \geq 3$$

Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary : ความเข้าใจพื้นฐาน



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

-3 1 1

$$i.e.\ \theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

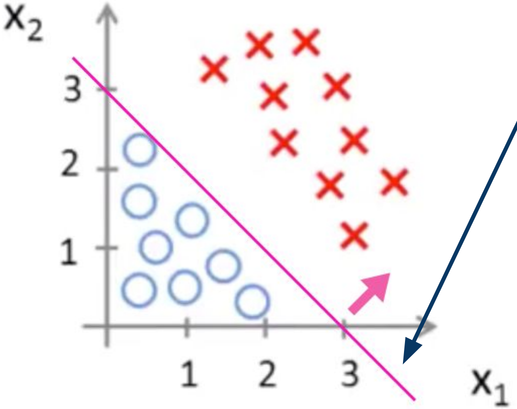
$$\because \theta^T x = [-3 \quad 1 \quad 1] \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = -3 + x_1 + x_2$$

$$\therefore \text{ Predict `}y = 1\text{' if } -3 + x_1 + x_2 \geq 0 \Longleftrightarrow \boxed{x_1 + x_2 \geq 3}$$

Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary

Decision Boundary : เส้นที่แบ่งพื้นที่ที่ $y = 0$ และพื้นที่ที่ $y = 1$ เส้นนี้ถูกสร้างด้วย hypothesis function



$$h_\theta(x) = g(\underbrace{\theta_0}_{-3} + \underbrace{\theta_1}_{1} x_1 + \underbrace{\theta_2}_{1} x_2)$$

$$i.e.\ \ \theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$\because \theta^T x = [-3 \quad 1 \quad 1] \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = -3 + x_1 + x_2$$

$$\therefore \text{ Predict '}y = 1\text{' if } -3 + x_1 + x_2 \geq 0 \Longleftrightarrow \boxed{x_1 + x_2 \geq 3}$$

Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

คำถาม

พิจารณา logistic regression ที่มี feature 2 ตัว: x_1 และ x_2
สมมติ $\theta_0 = 5, \theta_1 = -1, \theta_2 = 0$ ถ้าเราต้องการสร้างโมเดล
ที่มี hypothesis function: $h_\theta(x) = g\,(5 - x_1)$
ภาพไหนแสดง decision boundary ของ $h_\theta(x)$

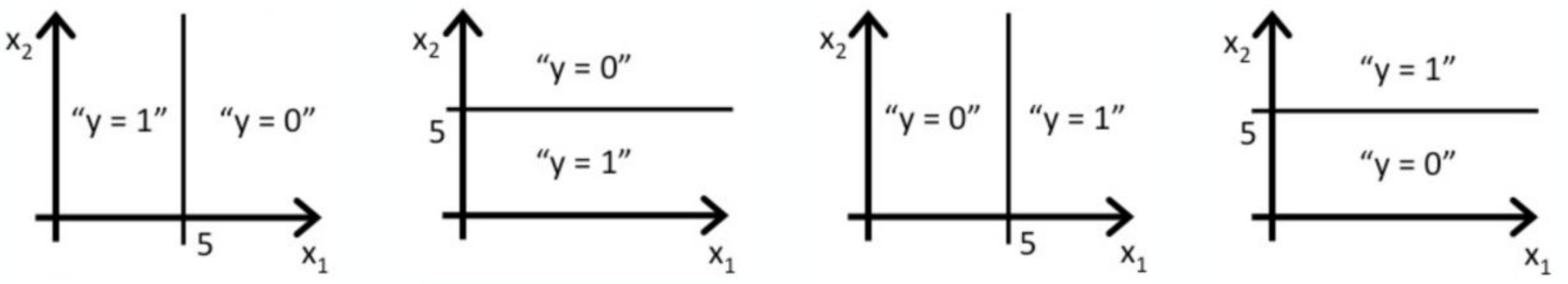


Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

คำถาม

พิจารณา logistic regression ที่มี feature 2 ตัว: x_1 และ x_2
สมมติ $\theta_0 = 5$, $\theta_1 = -1$, $\theta_2 = 0$ ถ้าเราต้องการสร้างโมเดล
ที่มี hypothesis function: $h_\theta(x) = g\,(5 - x_1)$
ภาพไหนแสดง decision boundary ของ $h_\theta(x)$

วิธีทำ

$$\theta = \begin{bmatrix} 5 \\ -1 \\ 0 \end{bmatrix} \qquad \theta^T x \geq 0$$

$$y = 1 \; if \; 5 + (-1)x_1 + 0x_2 \geq 0$$
$$5 - x_1 \geq 0$$
$$-x_1 \geq -5$$
$$x_1 \leq 5$$

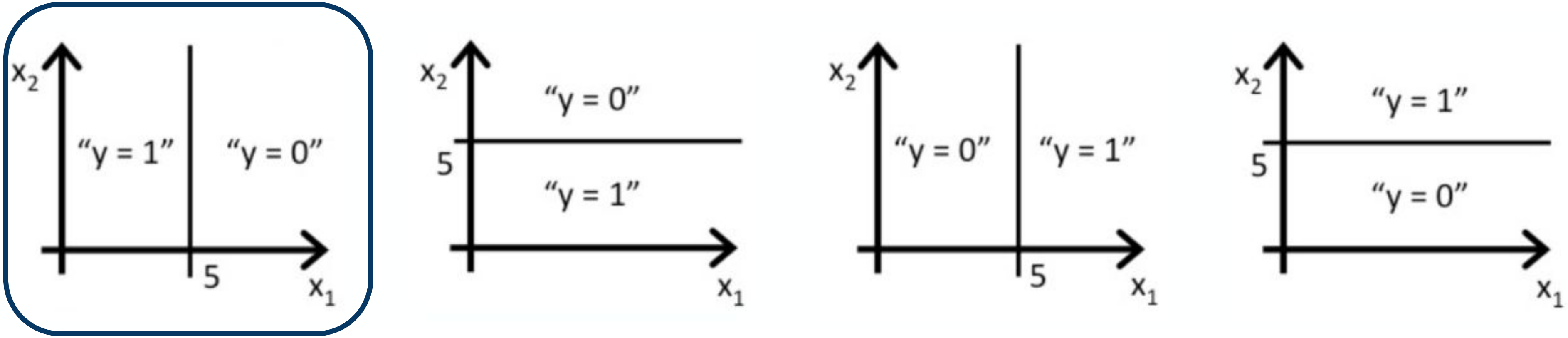
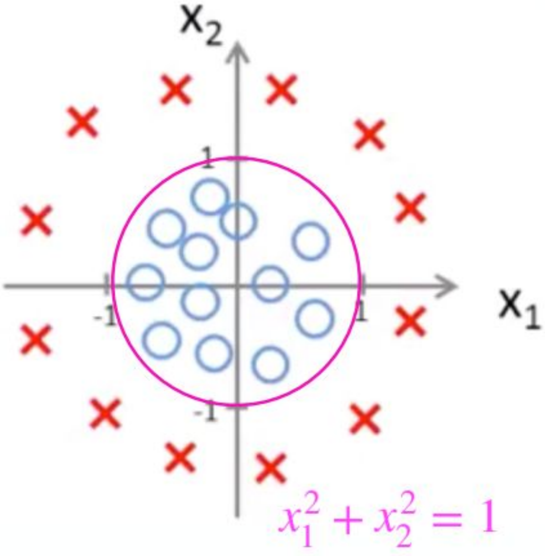


Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary ที่ไม่ใช่เส้นตรง (Non-linear)



x : positive example

o : negative example

$$h_\theta(x) = g(\underset{-1}{\theta_0} + \underset{0}{\theta_1} x_1 + \underset{0}{\theta_2} x_2 + \underset{1}{\theta_3} x_1^2 + \underset{1}{\theta_4} x_2^2)$$

$$i.e.\ \ \theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\therefore \text{ Predict '}y = 1\text{' if} -1 + x_1^2 + x_2^2 \geq 0$$

$$\iff x_1^2 + x_2^2 \geq 1$$

Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

Decision boundary ที่ไม่ใช่เส้นตรง (Non-linear)

Decision boundary ไม่จำเป็นต้องเป็นฟังก์ชั่นเส้นตรง (linear function)

Decision boundary สามารถเป็น function ที่แทนวงกลม หรือ รูปร่างอื่นๆ ที่เหมาะกับข้อมูลของเรา

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1^2 x_2 + \theta_5 x_1^2 x_2^2 + \theta_6 x_1^3 x_2 + \ldots)$$

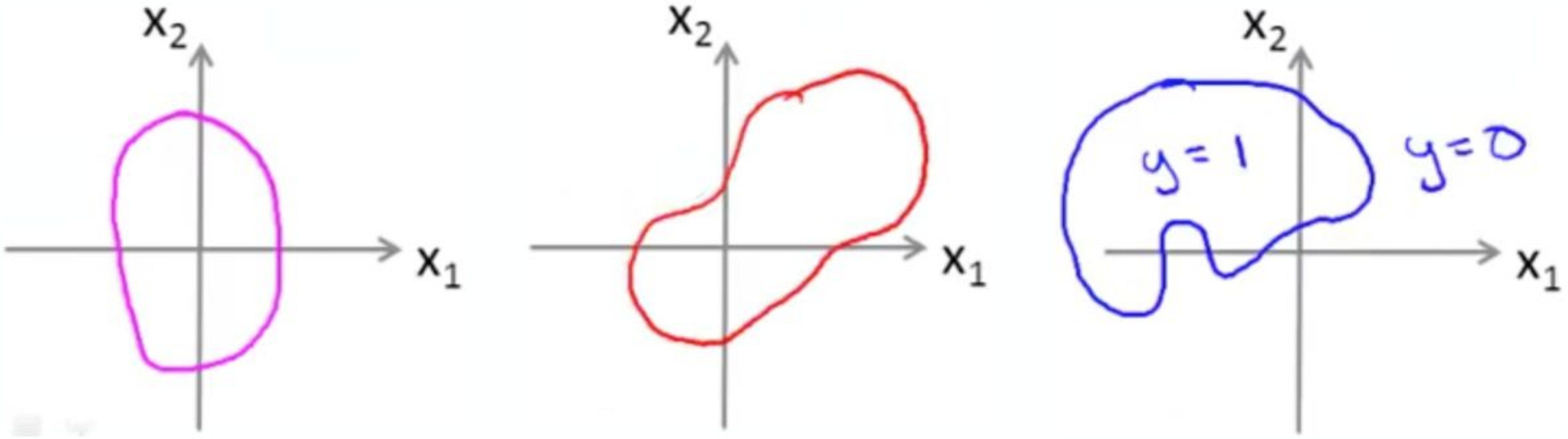


Figure source: Teeradaj Racharak; AI Practical Develop- ment Bootcamp [L 3.2]

4. Logistic Regression

4.4 Cost Function

Krittameth Teachasrisaksakul

เรามี

ชุดข้อมูล Training set: $$\{(\boldsymbol{x}^{(1)}, y^{(1)}), (\boldsymbol{x}^{(2)}, y^{(2)}), \ldots, (\boldsymbol{x}^{(m)}, y^{(m)})\}$$

เมื่อ $$\boldsymbol{x}^{(i)} \in \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathbb{R}^{n+1}}, \quad x_0 = 1,\ y^{(i)} \in \{0,1\}$$

Hypothesis: $$\boxed{h_\theta(\boldsymbol{x}) = \frac{1}{1 + e^{-\theta^T \boldsymbol{x}}}}$$

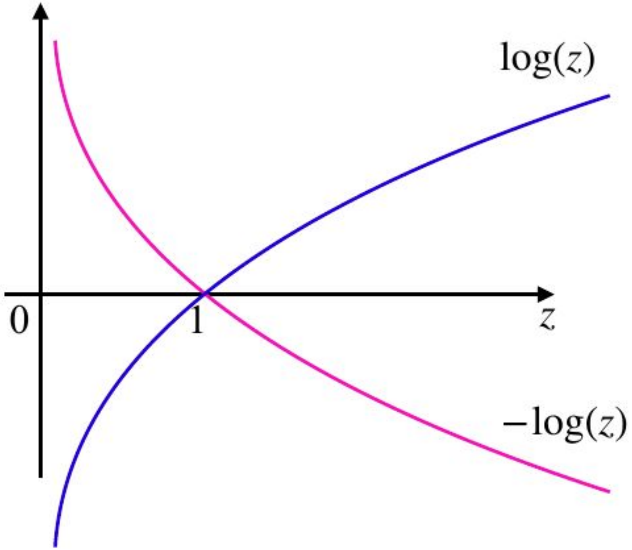
เป้าหมาย: เลือก parameter ยังไง ให้ minimize cost error (ทำ cost error ให้น้อยที่สุด)

Cost error function หน้าตาเป็นยังไง ?(เป็น $\mathbf{cost}(h_\theta(x^{(i)}), y^{(i)})$ หรือเปล่า ?)

Cost Function ของ Logistic Regression

สิ่งที่ต้องการ:

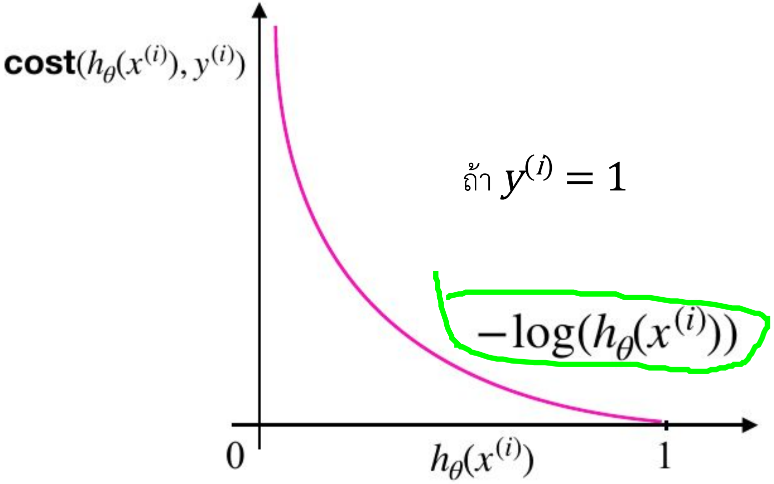
- ถ้า $y = 1$ และ $h_\theta(x) = 1 \longrightarrow$ แล้ว cost ต้องเป็น 0
- ถ้า $y = 1$ และ $h_\theta(x) = 0 \longrightarrow$ แล้ว cost ควรจะมีค่า ใหญ่มาก
- ถ้า $y = 0$ และ $h_\theta(x) = 1 \longrightarrow$ แล้ว cost ควรจะมีค่า ใหญ่มาก
- ถ้า $y = 0$ และ $h_\theta(x) = 0 \longrightarrow$ แล้ว cost ต้องเป็น 0
- เมื่อ y เป็น output จริง และ
  $h_\theta(x)$ เป็น output ที่ประมาณค่า (ได้จาก classifier)



Cost Function ของ Logistic Regression

สิ่งที่ต้องการ: (2 กรณีแรก)

- ถ้า $y = 1$ และ $h_\theta(x) = 1 \longrightarrow$ แล้ว cost ต้องเป็น 0
- ถ้า $y = 1$ และ $h_\theta(x) = 0 \longrightarrow$ แล้ว cost ควรจะมีค่า ใหญ่มาก

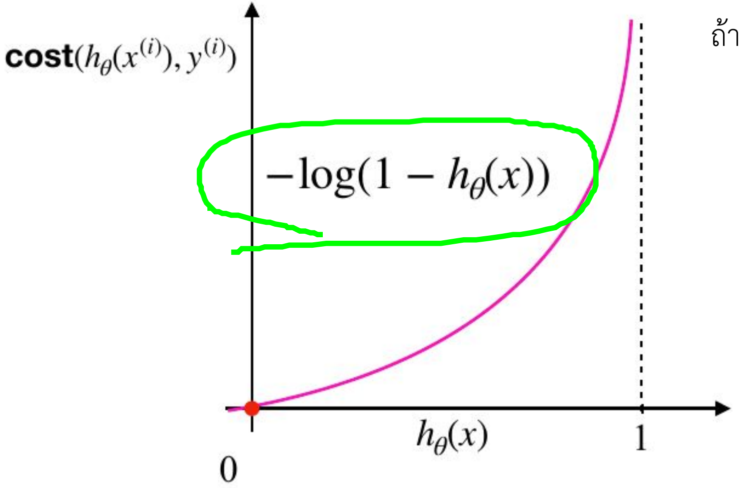


Source: Teeradaj Racharak; AI Practical Development Bootcamp [L 3.2]

Cost Function ของ Logistic Regression

สิ่งที่ต้องการ: (2 กรณีหลัง)

- ถ้า $y = 0$ และ $h_\theta(x) = 1 \longrightarrow$ แล้ว cost ควรจะมีค่า ใหญ่มาก
- ถ้า $y = 0$ และ $h_\theta(x) = 0 \longrightarrow$ แล้ว cost ต้องเป็น 0



ถ้า $y^{(i)} = 0$

Source: Teeradaj Racharak; AI Practical Development Bootcamp [L 3.2]

Cost Function ของ Logistic Regression

$$\textbf{cost}(h_\theta(\boldsymbol{x}^{(i)}), y^{(i)}) = \begin{cases} -\log(h_\theta(\boldsymbol{x}^{(i)})) & \textbf{if } y^{(i)} = 1 \\ -\log(1 - h_\theta(\boldsymbol{x}^{(i)})) & \textbf{if } y^{(i)} = 0 \end{cases}$$

ยุบรวม cost function ที่มี 2 กรณี ให้เหลือสมการเดียว

เขียนใหม่ ในรูปแบบที่กระชับ: cost function ของ 1 training example

$$\textbf{cost}(h_\theta(\boldsymbol{x}^{(i)}), y^{(i)}) = -\boxed{y^{(i)}\log(h_\theta(\boldsymbol{x}^{(i)}))} - \boxed{(1 - y^{(i)})\log(1 - h_\theta(\boldsymbol{x}^{(i)}))}$$

ถ้า $y = 0 \rightarrow$
พจน์ที่ 1 จะเป็น 0 และ ไม่มีผลกับผลลัพธ์

ถ้า $y = 1 \rightarrow$
พจน์ที่ 2 จะเป็น 0 และ ไม่มีผลกับผลลัพธ์

Cost Function ของ Logistic Regression

เพราะ $\quad \textbf{cost}(h_\theta(\boldsymbol{x}^{(i)}), y^{(i)}) = -y^{(i)}\log(h_\theta(\boldsymbol{x}^{(i)})) - (1-y^{(i)})\log(1-h_\theta(\boldsymbol{x}^{(i)}))$

ดังนั้น เขียน cost function แบบเต็มๆ ได้เป็น

$$J(\theta) = \frac{1}{m}\Sigma_{i=1}^{m}\textbf{cost}(h_\theta(\boldsymbol{x}^{(i)}), y^{(i)})$$

$$= -\frac{1}{m}\left[\Sigma_{i=1}^{m} y^{(i)}\log(h_\theta(\boldsymbol{x}^{(i)})) + (1-y^{(i)})\log(1-h_\theta(\boldsymbol{x}^{(i)}))\right]$$

cost function ของ training example ทั้งหมด : training example ที่ $i=1$ ถึง $i=m$

เรามี

Cost function:

$$J(\theta) = -\frac{1}{m}\left[\Sigma_{i=1}^{m} y^{(i)}\log(h_\theta(\boldsymbol{x}^{(i)})) + (1-y^{(i)})\log(1-h_\theta(\boldsymbol{x}^{(i)}))\right]$$

เป้าหมาย:

$$\min_{\theta} J(\theta)$$

เพื่อทำนายผล โดยใช้ค่า $\boldsymbol{x}$ ใหม่:

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}}$$

ก็คือ ความน่าจะเป็นที่ $y = 1$ เมื่อรู้ค่า x และมี parameter เป็น θ

เรามี

เป้าหมาย: $$\min_{\theta} J(\theta)$$

Minimize cost function ด้วยวิธี gradient descent อย่างไร ?

ใช้ algorithm เดียวกับที่ใช้ใน linear regression

ทำซ้ำ จน converge {

$$\theta_j := \theta_j - \alpha \underbrace{\left[\frac{1}{m}\Sigma_{i=1}^{m}(h_\theta(\boldsymbol{x}^{(i)}) - y^{(i)})\boldsymbol{x}_j^{(i)}\right]}_{\nabla_J(\boldsymbol{\theta})}$$

}

(ปรับค่า θ_j ทุกตัวพร้อมกัน)

คำถาม

สมมติเรา run gradient descent เพื่อหา parameter $\theta \in \mathbb{R}^{n+1}$ ของโมเดล logistic regression ข้อใดที่เป็นวิธีที่เหมาะสมที่ทำให้แน่ใจว่าค่า learning rate α เป็นค่าที่เหมาะสม และ gradient descent ทำงานอย่างถูกต้อง

(i) Plot

$$J(\theta) = (1/m)\Sigma_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

เป็นฟังก์ชั่นของจำนวน iterations (การวนซ้ำ) (ก็คือ แก ... ลังทุก iteration

(ii) Plot

$$J(\theta) = (-1/m)\Sigma_{i=1}^{m}[y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))]$$

เป็นฟังก์ชั่

(iii) Plot $J(\theta)$ เป็นฟังก์ชั่นของ θ และทำให้แน่ใจว่า $J(\theta)$ ลดลง หลังทุก iteration

(iv) Plot $J(\theta)$ เป็นฟังก์ชั่นของ θ และทำให้แน่ใจว่า $J(\theta)$ เป็นกราฟ convex

คำถาม

สมมติเรา run gradient descent เพื่อหา parameter $\theta \in \mathbb{R}^{n+1}$ ของโมเดล logistic regression ข้อใดที่เป็นวิธีที่เหมาะสมที่ทำให้แน่ใจว่าค่า learning rate α เป็นค่าที่เหมาะสม และ gradient descent ทำงานอย่างถูกต้อง

(i) Plot $J(\theta) = (1/m)\Sigma_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$ เป็นฟังก์ชั่นของจำนวน iterations (การวนซ้ำ)
(ก็คือ แก ... ลังทุก iteration

(ii) Plot
เป็นฟังก์ชั่ $J(\theta) = (-1/m)\Sigma_{i=1}^{m}[y^{(i)}\log h_\theta(x^{(i)}) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))]$

(iii) Plot $J(\theta)$ เป็นฟังก์ชั่นของ θ และทำให้แน่ใจว่า $J(\theta)$ ลดลง หลังทุก iteration

(iv) Plot $J(\theta)$ เป็นฟังก์ชั่นของ θ และทำให้แน่ใจว่า $J(\theta)$ เป็นกราฟ convex

4. Logistic Regression

4.5 Multi-class Classification

(การจำแนกประเภทที่มีมากกว่า 2 class)

Krittameth Teachasrisaksakul

Multi-class classification : ความเข้าใจพื้นฐาน

- **การจัดกลุ่ม email ใส่ folder:** งาน, เพื่อน, ครอบครัว, งานอดิเรก

$$y = 1 \quad y = 2 \quad y = 3 \quad y = 4$$

- **การวินิจฉัยโรค:** ไม่ป่วย, เป็นหวัด, เป็นไข้

$$y = 1 \quad y = 2 \quad y = 3$$

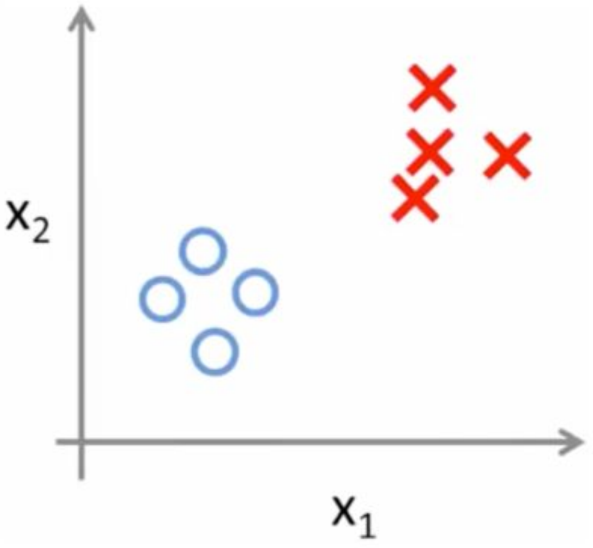
- **สภาพอากาศ:** มีแดดมาก, มีเมฆมาก, ฝนตก

$$y = 1 \quad y = 2 \quad y = 3$$

Source: Teeradaj Racharak; AI Practical Development Bootcamp [L 3.2]

Binary vs. Multi-class

Binary classification



Multi-class classification

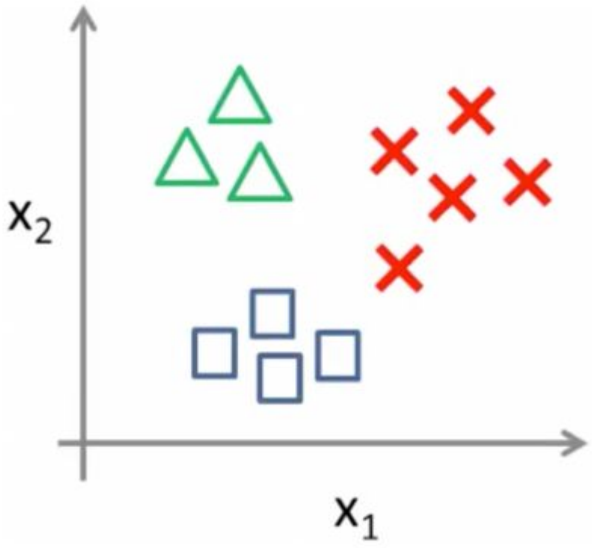
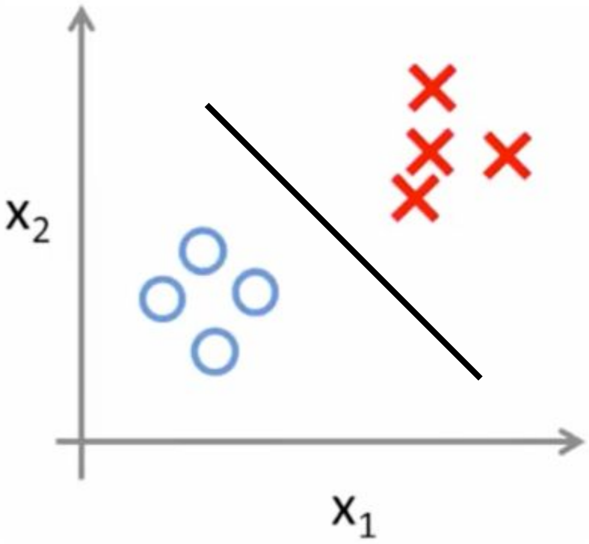


Figure Source: Andrew Ng's Machine Learning course

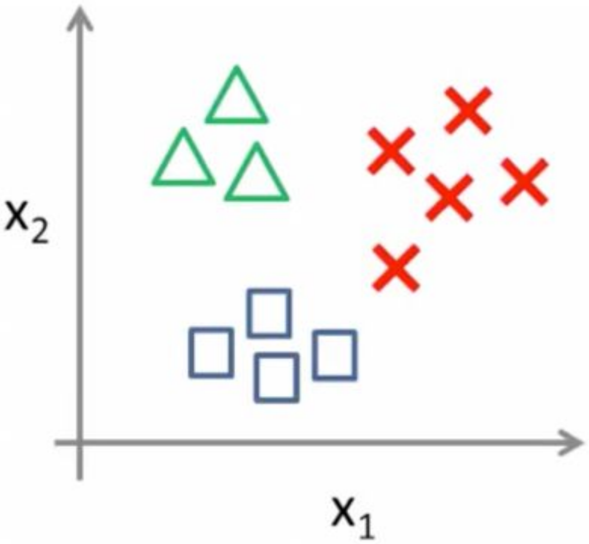
Binary vs. Multi-class

Binary classification



Regression algorithm

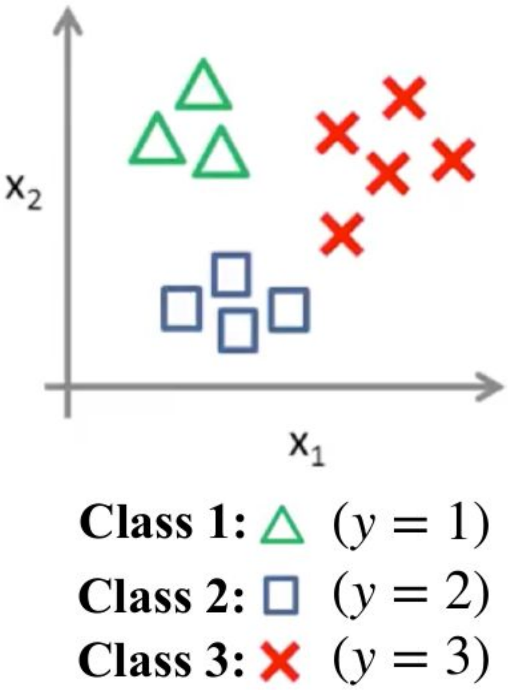
Multi-class classification



One-vs-all algorithm

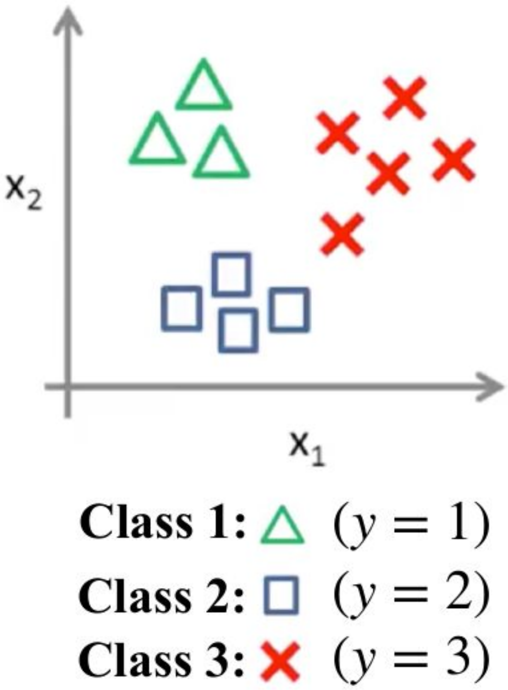
Figure Source: Andrew Ng's Machine Learning course

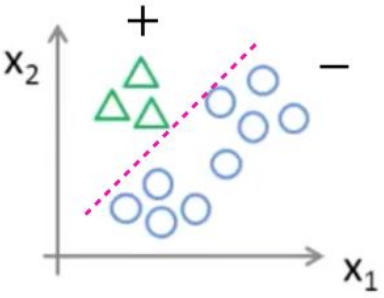
One-vs-all (One-vs-rest)



One-vs-all หรือ One-vs-rest คือ การแยกประเภท class 1 class ออกจาก class ที่เหลือทั้งหมด

One-vs-all (One-vs-rest)

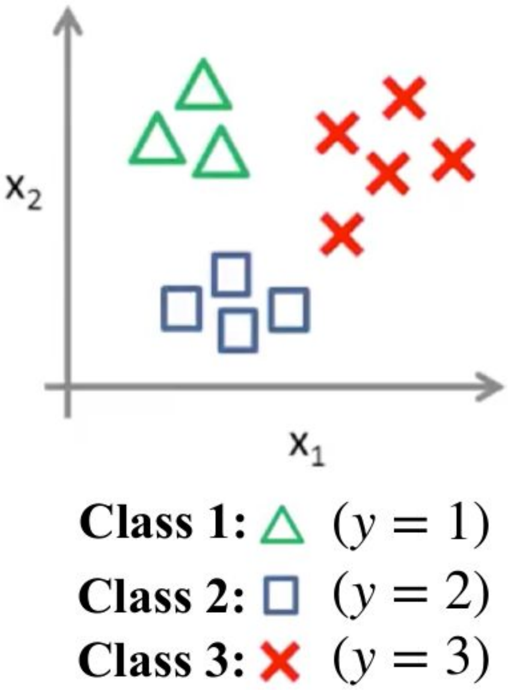


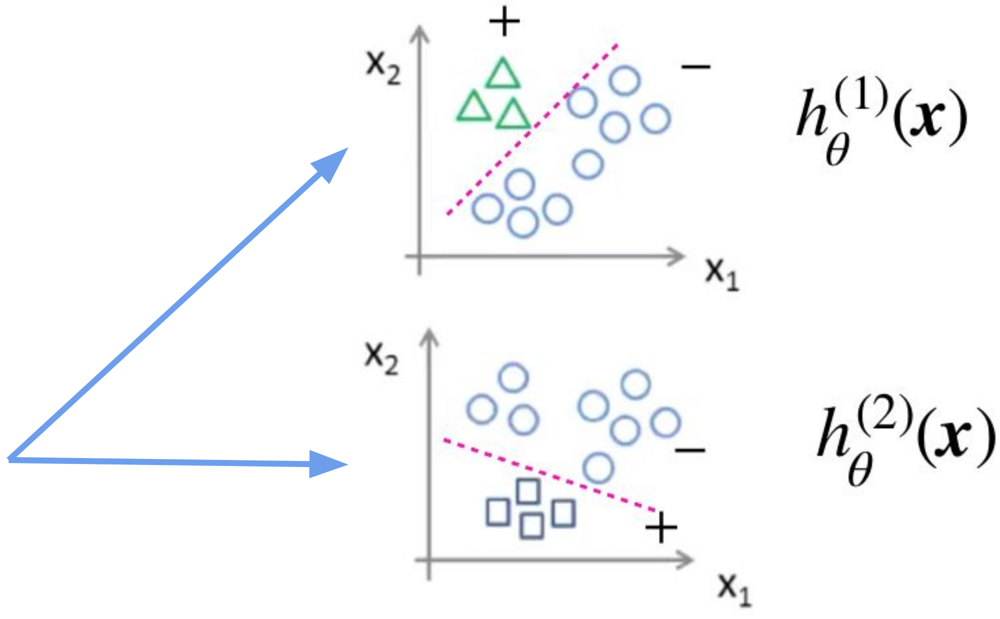


$$h_\theta^{(1)}(\boldsymbol{x})$$

แยก class 1 จาก class อื่นๆ ทั้งหมด (class 2, 3)

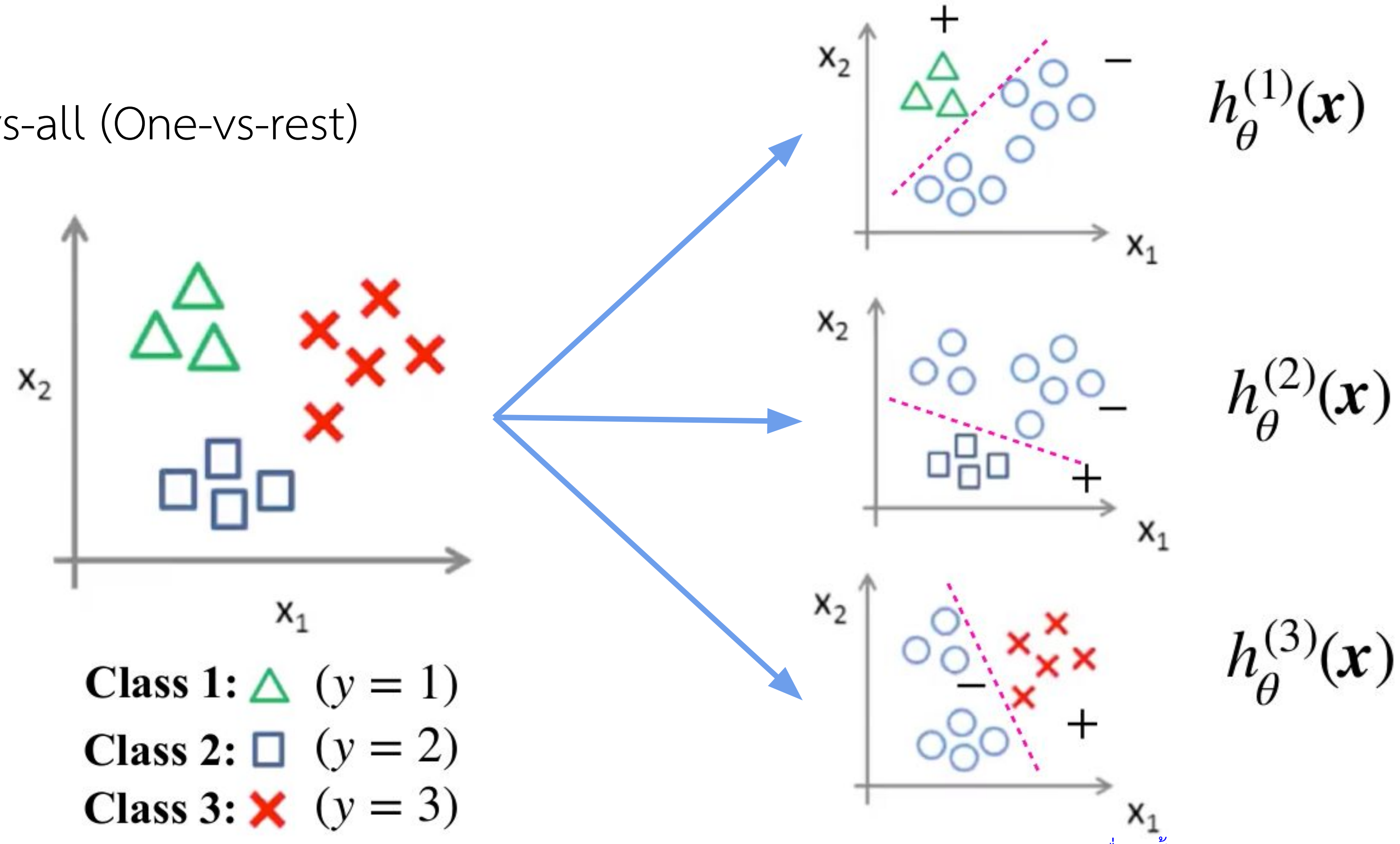
One-vs-all (One-vs-rest)





แยก class 2 จาก class อื่นๆ ทั้งหมด (class 1, 3)

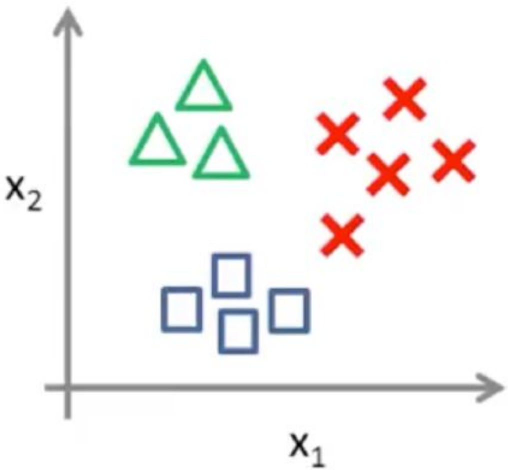
One-vs-all (One-vs-rest)



Class 1: △ $(y = 1)$
Class 2: □ $(y = 2)$
Class 3: ✕ $(y = 3)$

แยก class 3 จาก class อื่นๆ ทั้งหมด (class 1, 2)

One-vs-all (One-vs-rest)

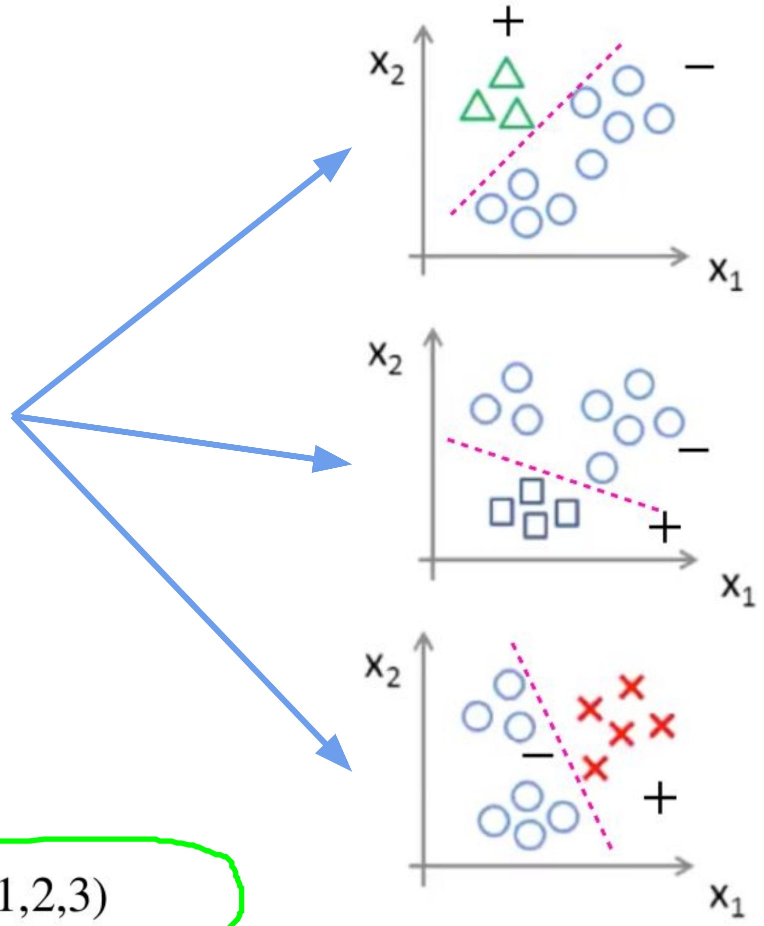


Class 1: △ $(y = 1)$

Class 2: □ $(y = 2)$

Class 3: ✕ $(y = 3)$

$$h_\theta^{(i)}(x) = P(y = i \,|\, x; \theta) \qquad (i = 1,2,3)$$



$h_\theta^{(1)}(\boldsymbol{x})$ 0.7

$h_\theta^{(2)}(\boldsymbol{x})$ 0.2

$h_\theta^{(3)}(\boldsymbol{x})$ 0.1

One-vs-all : ขั้นตอน multiclass classification โดยใช้ binary logistic regression

- ฝึก (Train) / สร้าง logistic regression classifier (ตัวจำแนก) $h_\theta^{(i)}(x)$ ของแต่ละ class (ประเภท) i เพื่อทำนายความน่าจะเป็นที่ $y = i$ (หมายความว่า ตัวอย่างข้อมูลนี้ จัดเป็นประเภท i)
- เพื่อทำนาย class ของ input ใหม่ $x \longrightarrow$ **เลือก class i ที่ทำให้ $h_\theta^{(i)}(x)$ มีค่าสูงสุด** ก็คือ

$$y = \arg\max_i h_\theta^{(i)}(\boldsymbol{x})$$

y = 1

คำถาม

สมมติ เรามีปัญหา multi-class classification (การแบ่งประเภทมากกว่า 2 ประเภท) ที่มี k classes (ประเภท) (ดังนั้น $y \in \{1,2,\ldots, k\}$) ถ้าใช้วิธี one-vs-all จะต้อง train logistic regression classifier ทั้งหมดกี่ตัว

(i) $k - 1$

(ii) k

(iii) $k + 1$

(iv) ประมาณ $\log_2(k)$

คำถาม

สมมติ เรามีปัญหา multi-class classification (การแบ่งประเภทมากกว่า 2 ประเภท) ที่มี k classes (ประเภท) (ดังนั้น $y \in \{1,2,..., k\}$) ถ้าใช้วิธี one-vs-all จะต้อง train logistic regression classifier ทั้งหมดกี่ตัว

(i) $k - 1$

(ii) k

(iii) $k + 1$

(iv) ประมาณ $\log_2(k)$

สรุป

1. ถ้ามี $\mathcal{X}$ ที่มีค่าต่อเนื่อง (continuous) และอยากทำนาย $\mathcal{Y}$ ที่มีค่า**ต่อเนื่อง**
   ให้ลองใช้ **linear regression** model เป็นอันแรก
2. ถ้ามี $\mathcal{X}$ ที่มีค่าต่อเนื่อง (continuous) และอยากทำนาย $\mathcal{Y}$ ที่มีค่า**ไม่ต่อเนื่อง (discrete)**
   ให้ลองใช้ **logistic regression** model เป็นอันแรก

วิธีที่เราเรียน คือ การเรียนรู้ $p(y \mid x)$ โดยตรง ก็คือ เมื่อให้ input x เราจะแปลง (map) ไปเป็น target (เป้าหมาย / ผล) y โดยตรง

วิธีพวกนี้ เรียกว่า **discriminative** learning algorithms

Source: Teeradaj Racharak; AI Practical Development Bootcamp [L 3.2]

References

1. Andrew Ng, Machine Learning, Coursera.
2. Teeradaj Racharak, AI Practical Development Bootcamp.
3. What is Machine Learning?, https://www.digitalskill.org/contents/5