Logistic regression: classification

Ngô Minh Nhựt

2023

Review

- Given a linear regression model with one variable

$$h_\theta(x) = \theta_0 + \theta_1 x$$

  - x is input and $h_\theta(x)$ is output
  - And a dataset of five samples

x	1	2	3	4	5
y	2	4	6	8	10

  - Run gradient descent algorithm on the above model and dataset and give results with $\alpha = 0.01$ and initial $\theta_0 = 2, \theta_1 = 1$.

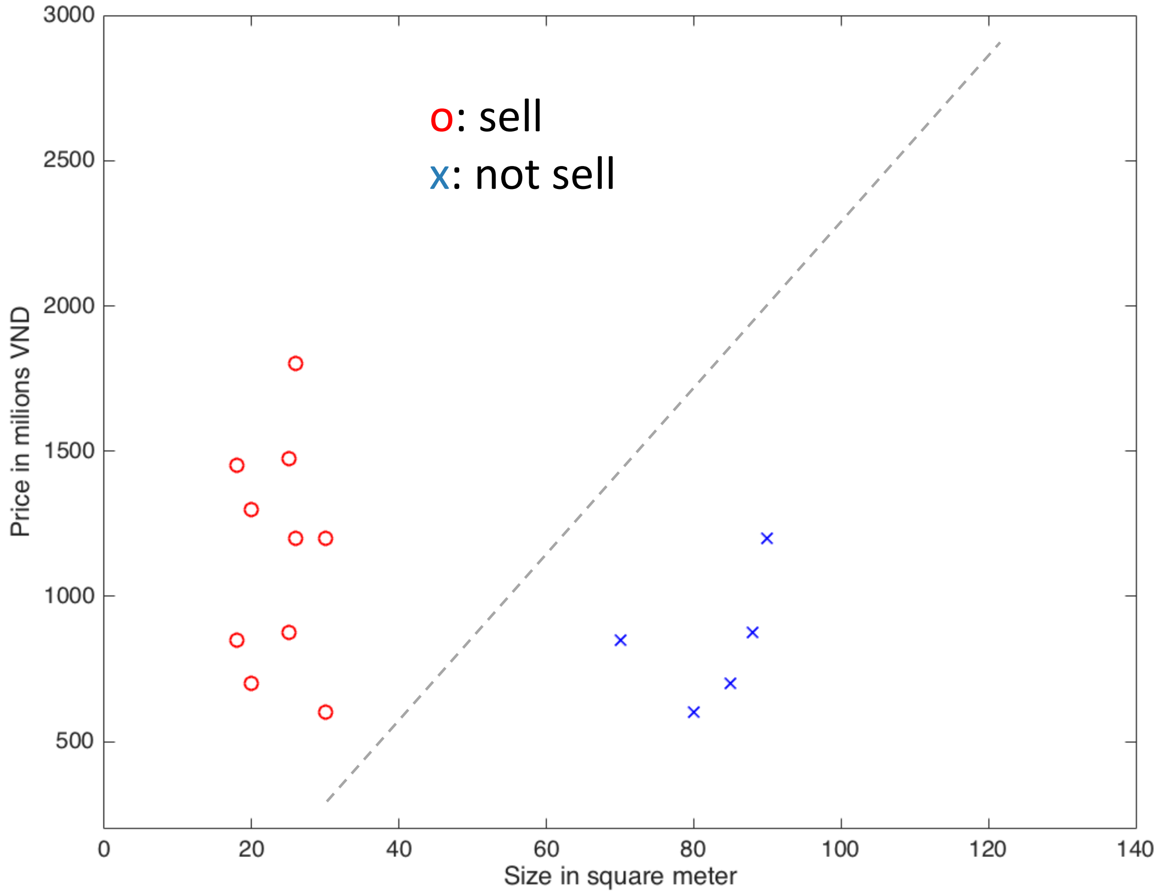
Iteration	θ_0	θ_1	Cost
0	2	1	1.5
1			
2			
3			

Classification

- Answer a question with **yes** or **no**
  - Check if an email is spam
  - Check if a transaction is anormal
  - Check if a person exposes to health risk
  - Check if an area of an image contains human face
  - Check if an area of an image contains character \`0\`
  - …

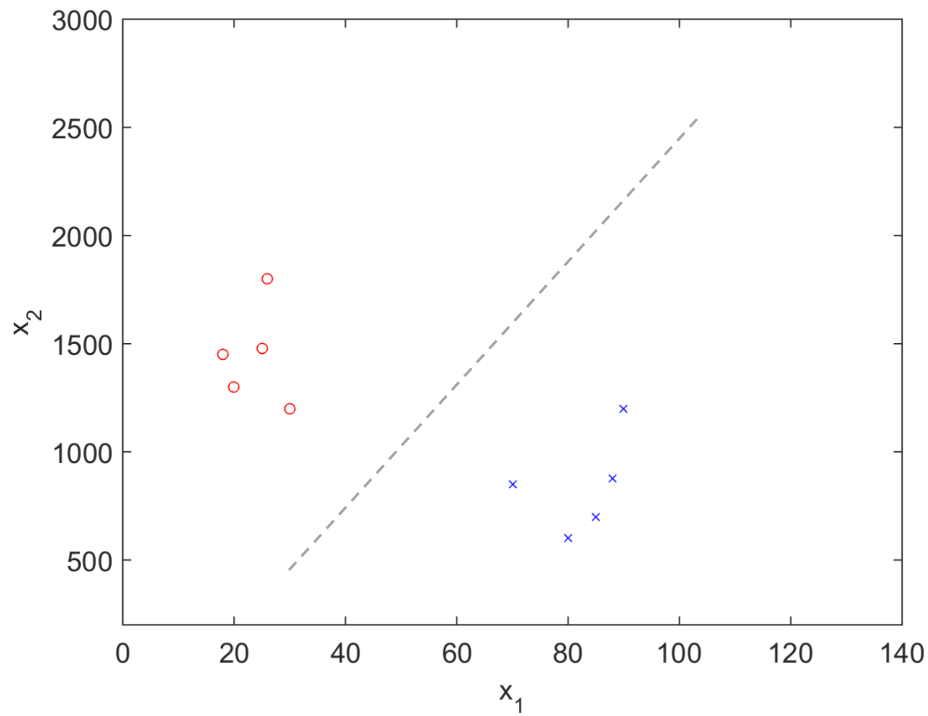
Classification

Size	Price	Sell?
80	600	No
30	1200	Yes
70	850	No
26	1200	No
…	…	…



Output

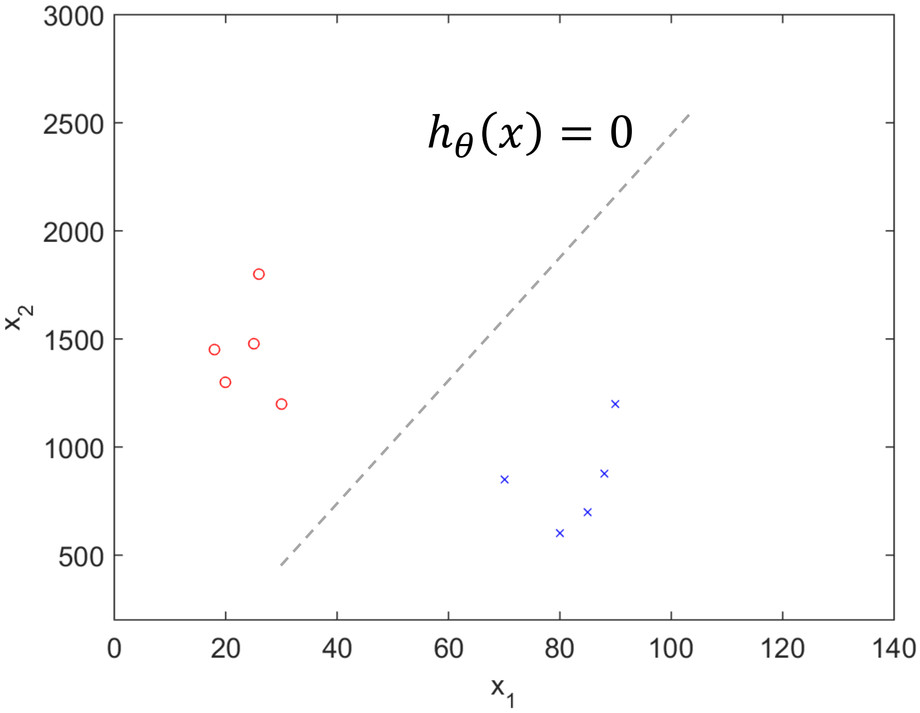
- Output = {yes, no}
- Y = {1, 0}
  - 1: positive
  - 0: negative



Boundary

- Classes: Y = {1, 0}
- Boundary: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0$
- Classification rule

$$\begin{cases} \text{If } h_\theta(x) \geq 0, y = 1 \\ \text{If } h_\theta(x) < 0, y = 0 \end{cases}$$

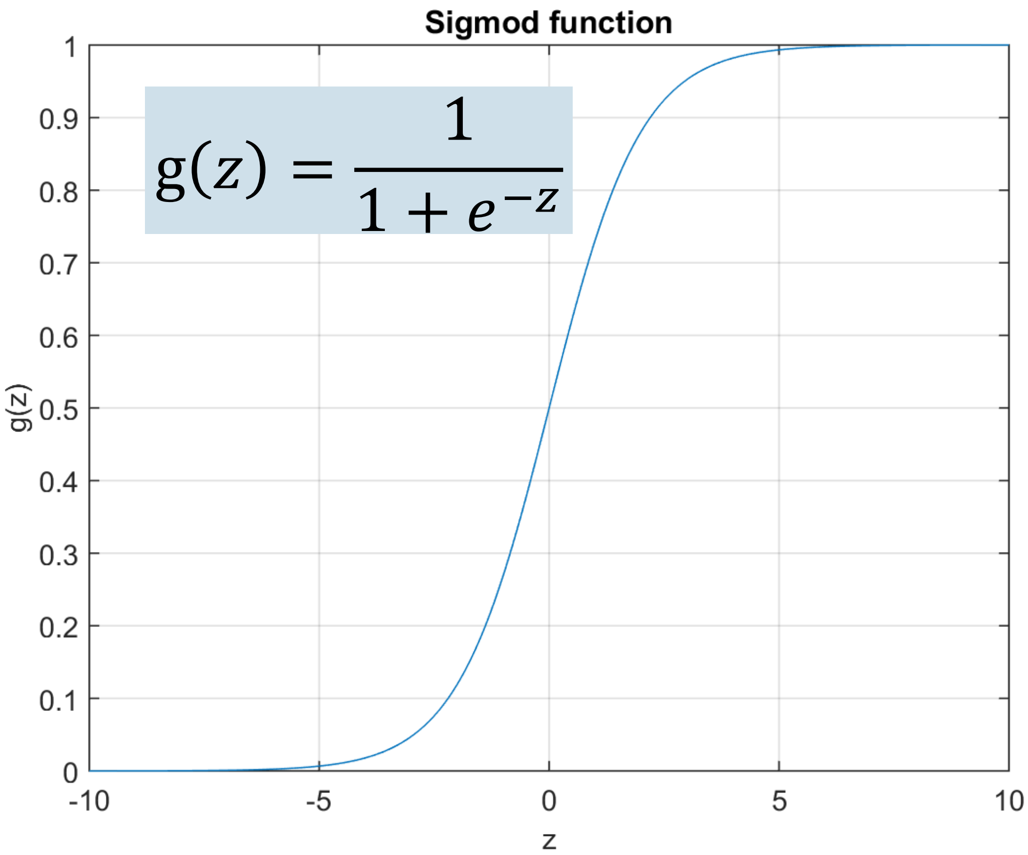


Hypothesis

- Boundary: $\theta^T x = 0$
  - $\begin{cases} \text{If } \theta^T x \geq 0, y = 1 \\ \text{If } \theta^T x < 0, y = 0 \end{cases}$
- We need: $\begin{cases} y = 1, h_\theta(x) \to 1 \\ y = 0, h_\theta(x) \to 0 \end{cases}$
- A new model: $h_\theta(x) = g(\theta^T x)$
  - $\theta^T x$ is much bigger than 0 then $g(\theta^T x)$ approaches to 1
  - $\theta^T x$ is much smaller than 0 then $g(\theta^T x)$ approaches to 0

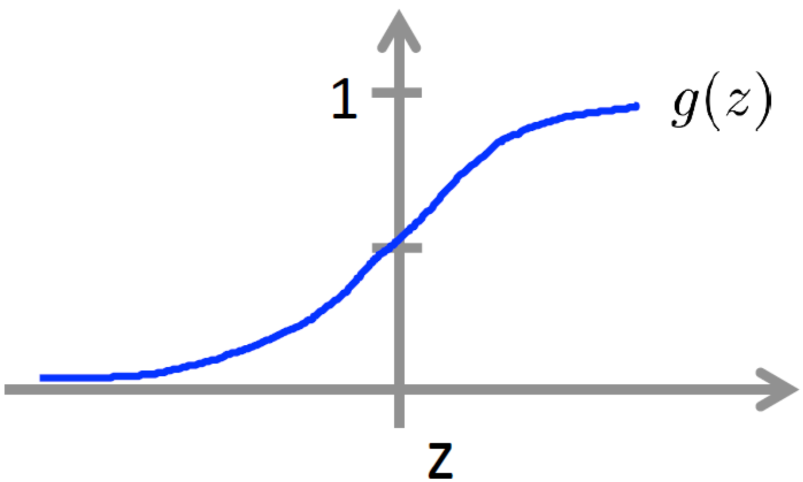
Hypothesis: sigmoid function

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



Boundary

- $h_\theta(x) = g(\theta^T x)$
- $g(z) = \frac{1}{1+e^{-z}}$
- Boundary:
  - y = 1 if $h_\theta(x) \geq 0.5$
    $\rightarrow \theta^T x \geq 0$
  - y = 0 if $h_\theta(x) < 0.5$
    $\rightarrow \theta^T x < 0$

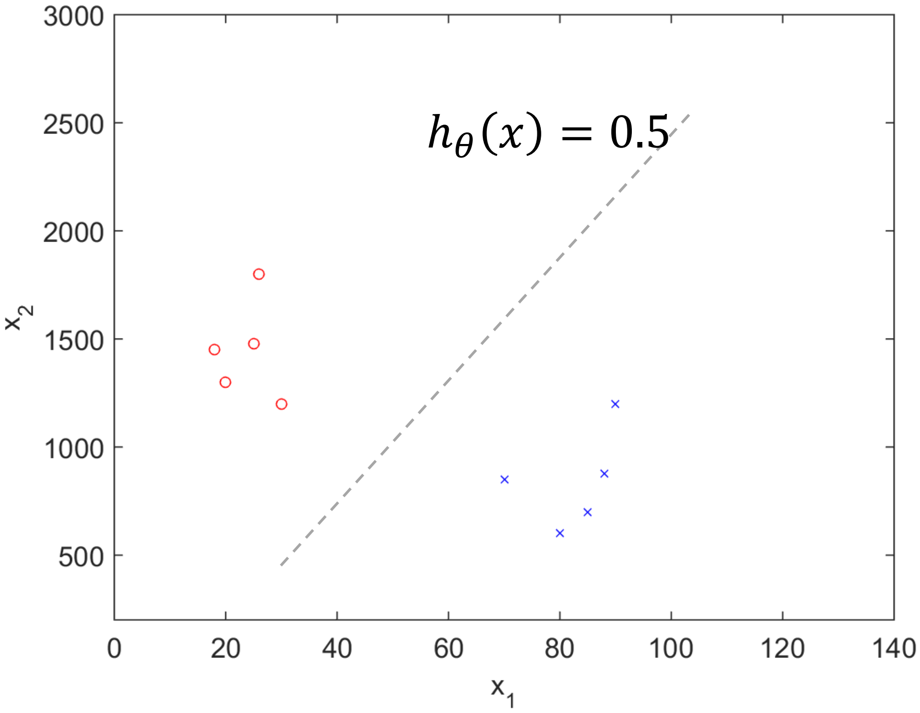


- $h_\theta(x)$ is actually probability y=1
  $h_\theta(x) = P(y = 1|x; \theta)$

Source: Andrew Ng

Boundary

- Boundary:
  - y = 1 if $h_\theta(x) \geq 0.5$
    $\rightarrow \theta^T x \geq 0$
  - y = 0 if $h_\theta(x) < 0.5$
    $\rightarrow \theta^T x < 0$

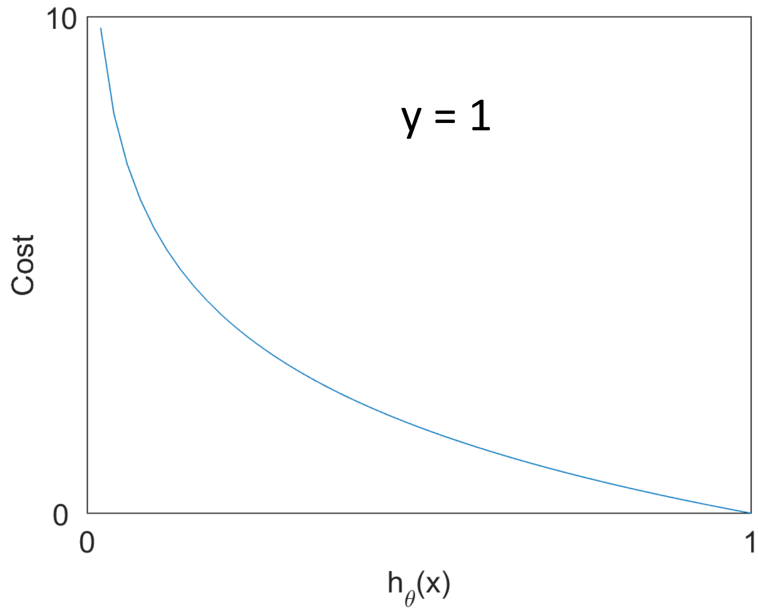


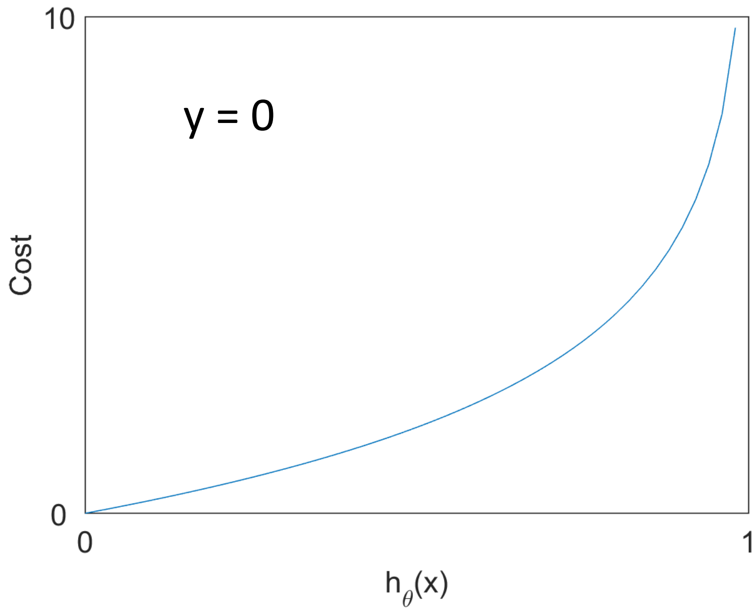
Cost function

- Hypothesis: $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
  - $0 \leq h_\theta(x) \leq 1$
- $\mathrm{Cost}(h(x), y) = \begin{cases} -\log h_\theta(x) & \text{if } y = 1 \\ -\log\big(1 - h_\theta(x)\big) & \text{if } y = 0 \end{cases}$
  - y =1: if $h_\theta(x)$ → 1, cost → 0, if $h_\theta(x)$ → 0, cost → infinity
  - y =0: if $h_\theta(x)$ → 0, cost → 0, if $h_\theta(x)$ → 1, cost → infinity

Cost function

$$\mathrm{Cost}(h(x), y) = \begin{cases} -\log h_\theta(x) & \text{if } y = 1 \\ -\log\big(1 - h_\theta(x)\big) & \text{if } y = 0 \end{cases}$$





Cost function

- Hypothesis: $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
- $\text{Cost}(h(x), y) = \begin{cases} -\log h_\theta(x) & \text{if } y = 1 \\ -\log\left(1 - h_\theta(x)\right) & \text{if } y = 0 \end{cases}$
- In a new form

$$\text{Cost}(h(x), y) = -y^{(i)} \log h\left(x^{(i)}\right) - \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)$$

  - y = 1, 1 - y = 0 ➔ Cost(h(x), y) = ?
  - y = 0, 1 - y = 1 ➔ Cost(h(x), y) = ?

Cost function

- Hypothesis: $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
- Const function on one sample

$$\text{Cost}(h(x), y) = -y^{(i)} \log h\left(x^{(i)}\right) - \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)$$

- Cost function

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)$$

Partial derivative

- $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
- $J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$
- $\frac{dJ}{d\theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$

Gradient descent

- Cost function:

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

- Find θ so that $J(\theta)$ reaches minimal
- Predict for a new input
  - Output: $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

Gradient descent

- Vector gradient:
  - $\frac{dJ}{d\theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$
  - $j = 0, 1, 2, \dots, n$
- Repeat until convergence

$$\{$$

$$\theta_j = \theta_j - \alpha \frac{dJ}{d\theta_j}$$

$$\}$$

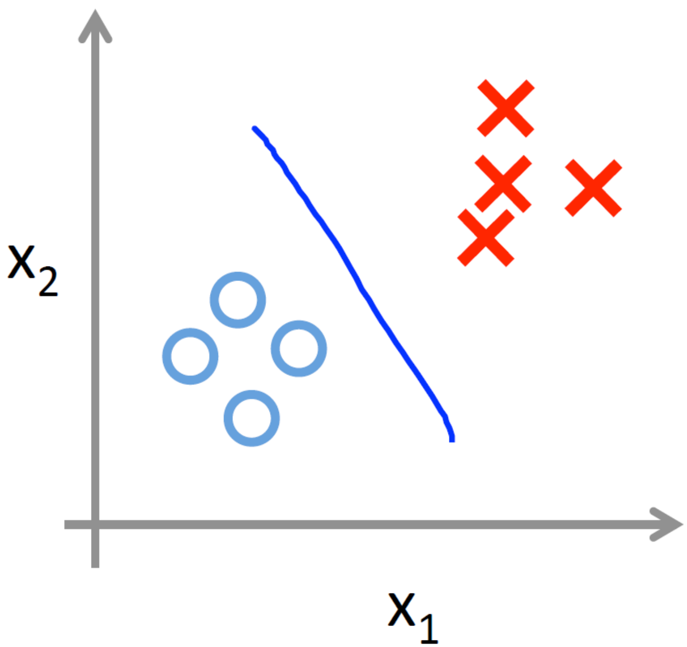
Logistic regression with multi-classes

- Weather: sunny, cloudy, rain, heavy rain
- Digit: 0, 1, ..., 9
- Object: human, cat, house, landscape

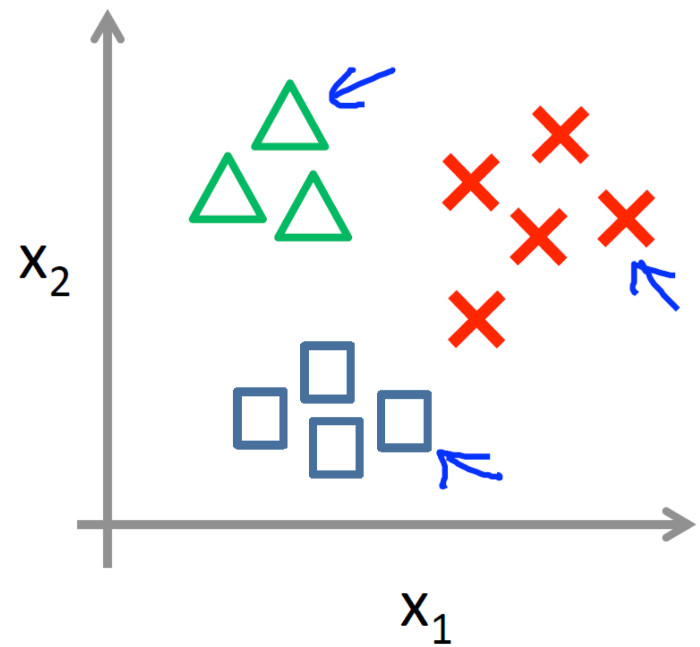
➔ $y = \{1, 2, 3, ...\}$

Logistic regression with multi-classes

Binary classification:



Multi-class classification:

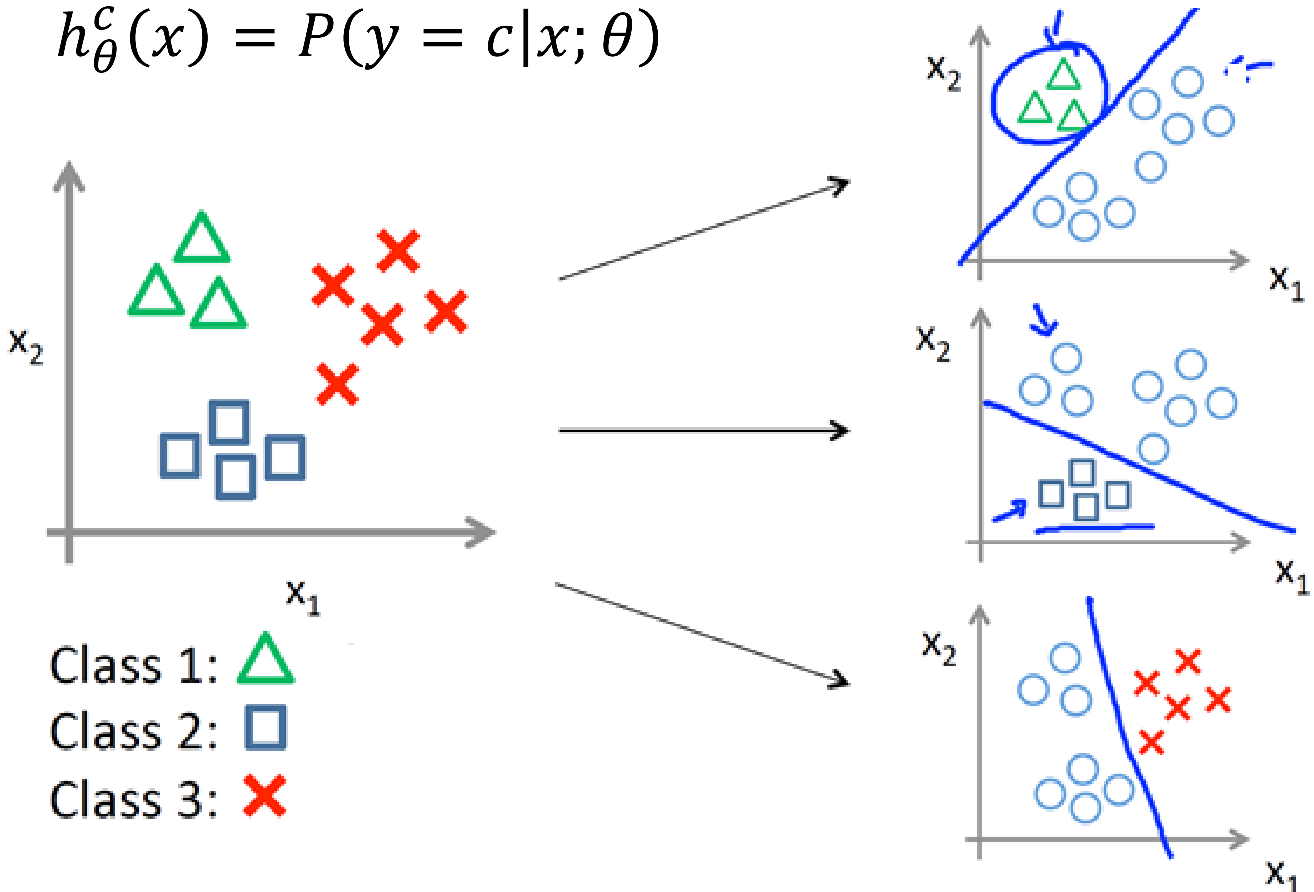


Source: Andrew Ng

Logistic regression with multi-classes

- Train classifier for each class $h_\theta^c(x)$

$$h_\theta^c(x) = P(y = c|x;\theta)$$



Source: Andrew Ng

Logistic regression with multi-classes

- Train classifier for each class $h_\theta^c(x)$

$$h_\theta^c(x) = P(y = c|x;\theta)$$

- Predict for a new input
  - $y = \max_c h_\theta^c(x)$