2021 Spring: Machine Learning

Classification

Lecturer: Min Lu

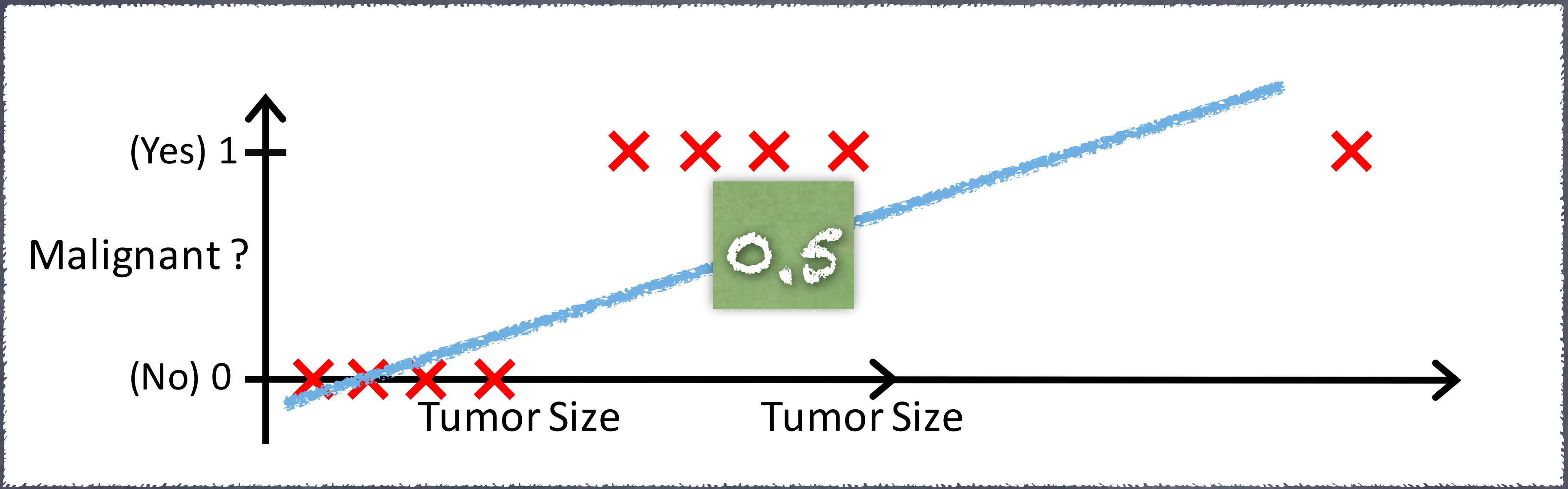
Email: lumin.vis@gmail.com

Classification

- Email: Spam / Not Spam?
- Online Transactions: Fraudulent (Yes / No)?
- Tumor: Malignant / Benign ?

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

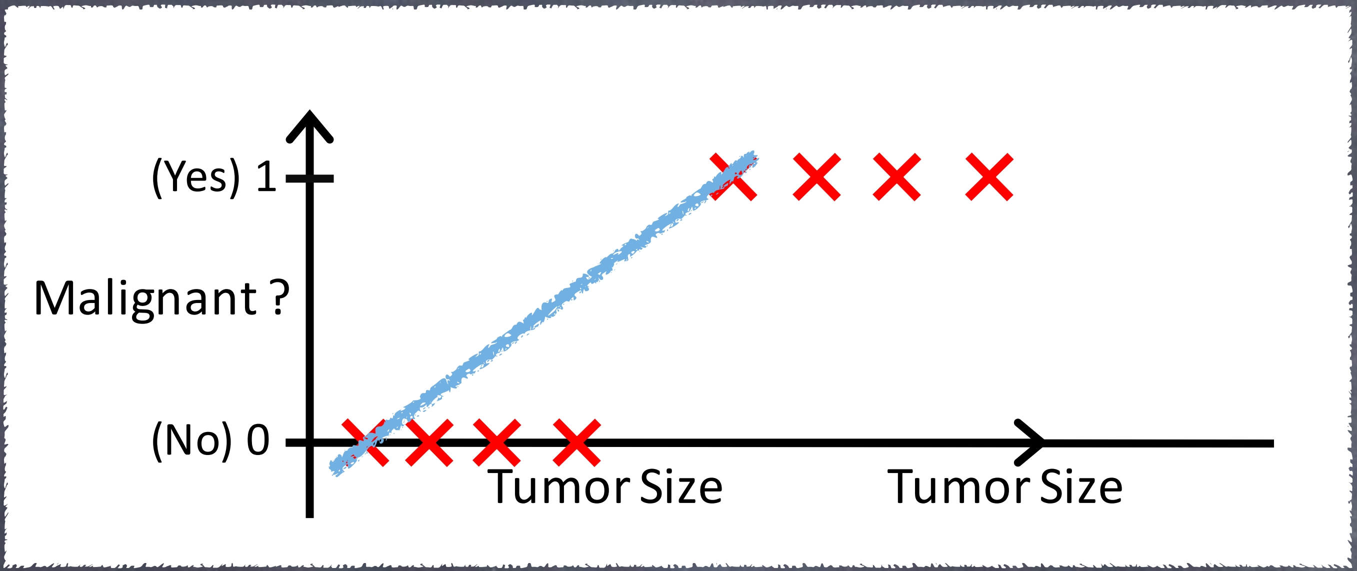
Multi-Classification

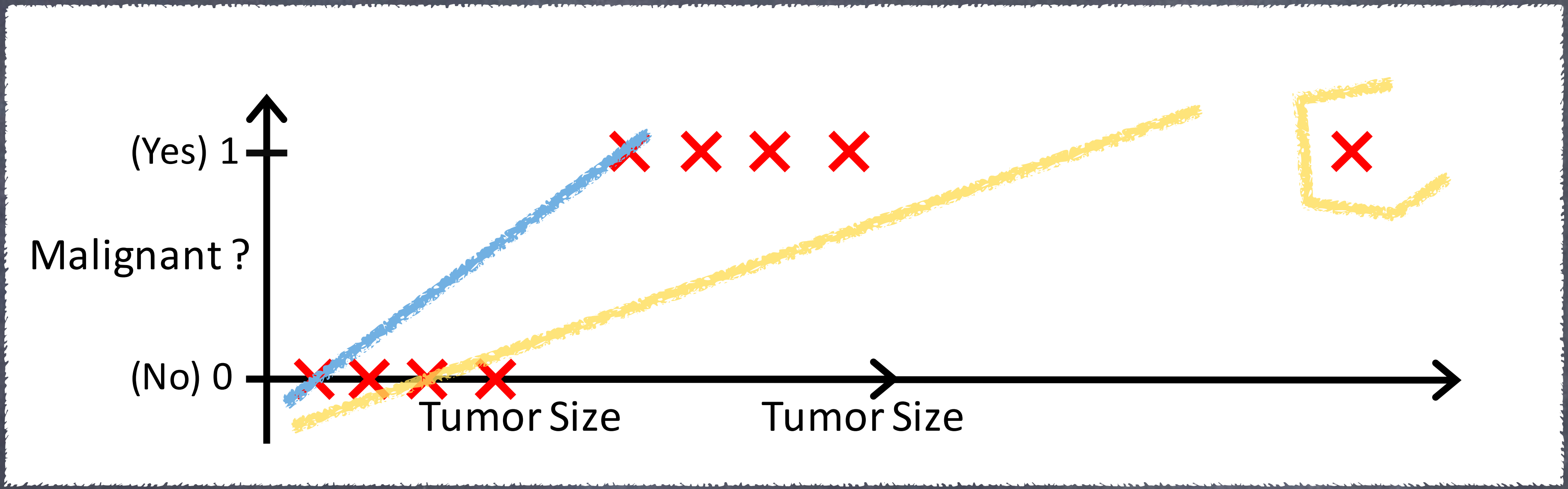


Threshold classifier output $h_\theta(x)$ at 0.5:

If $h_\theta(x) \geq 0.5$, predict "y = 1"

If $h_\theta(x) < 0.5$, predict "y = 0"


(Yes) 1
Malignant ?
(No) 0
Tumor Size
Tumor Size


(Yes) 1
Malignant ?
(No) 0
Tumor Size
Tumor Size

Linear Regression

$h_\theta(x)$ can be > 1 or < 0

Logistic Regression

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

Classification!

Logistic Regression

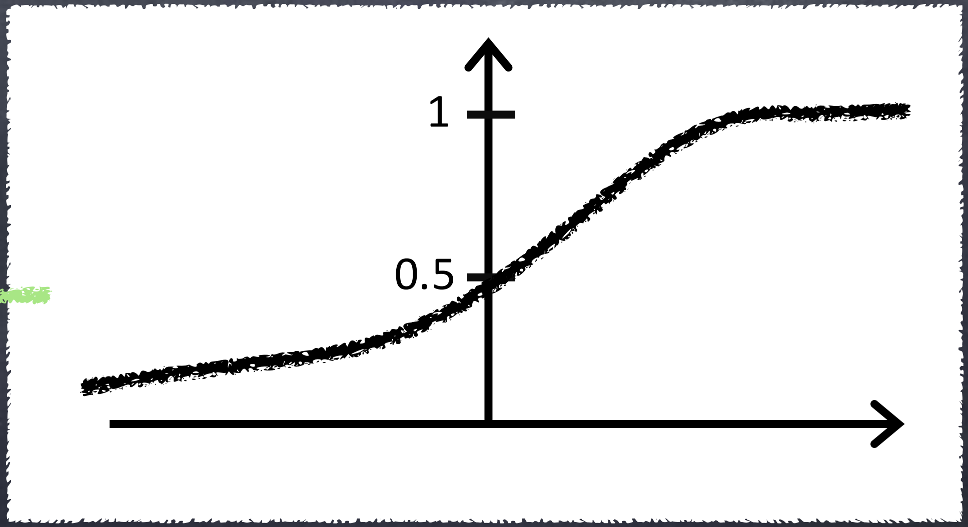
$$0 \leq h_\theta(x) \leq 1$$

$$h_\theta(x) = g(\ \theta^T x\ )$$

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

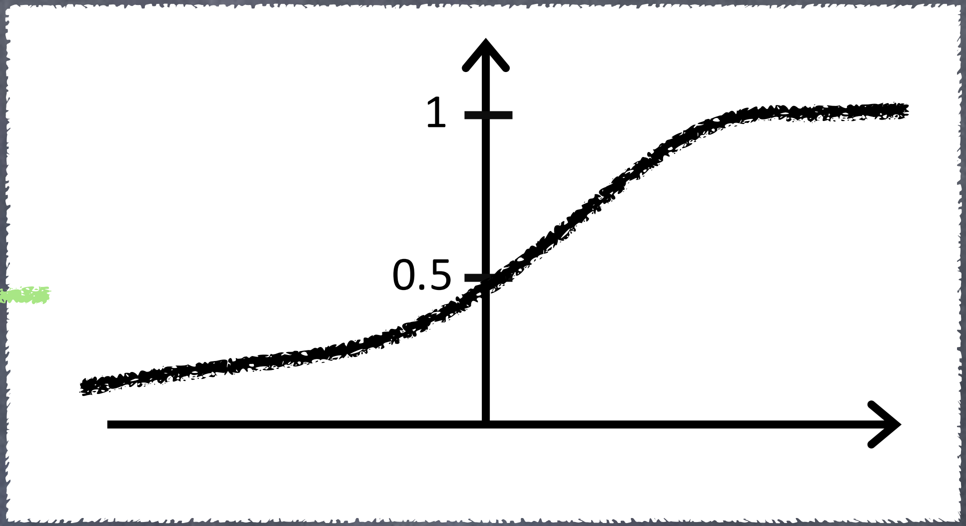
Sigmod function
Logistic function

estimated probability
that y = 1 on input x

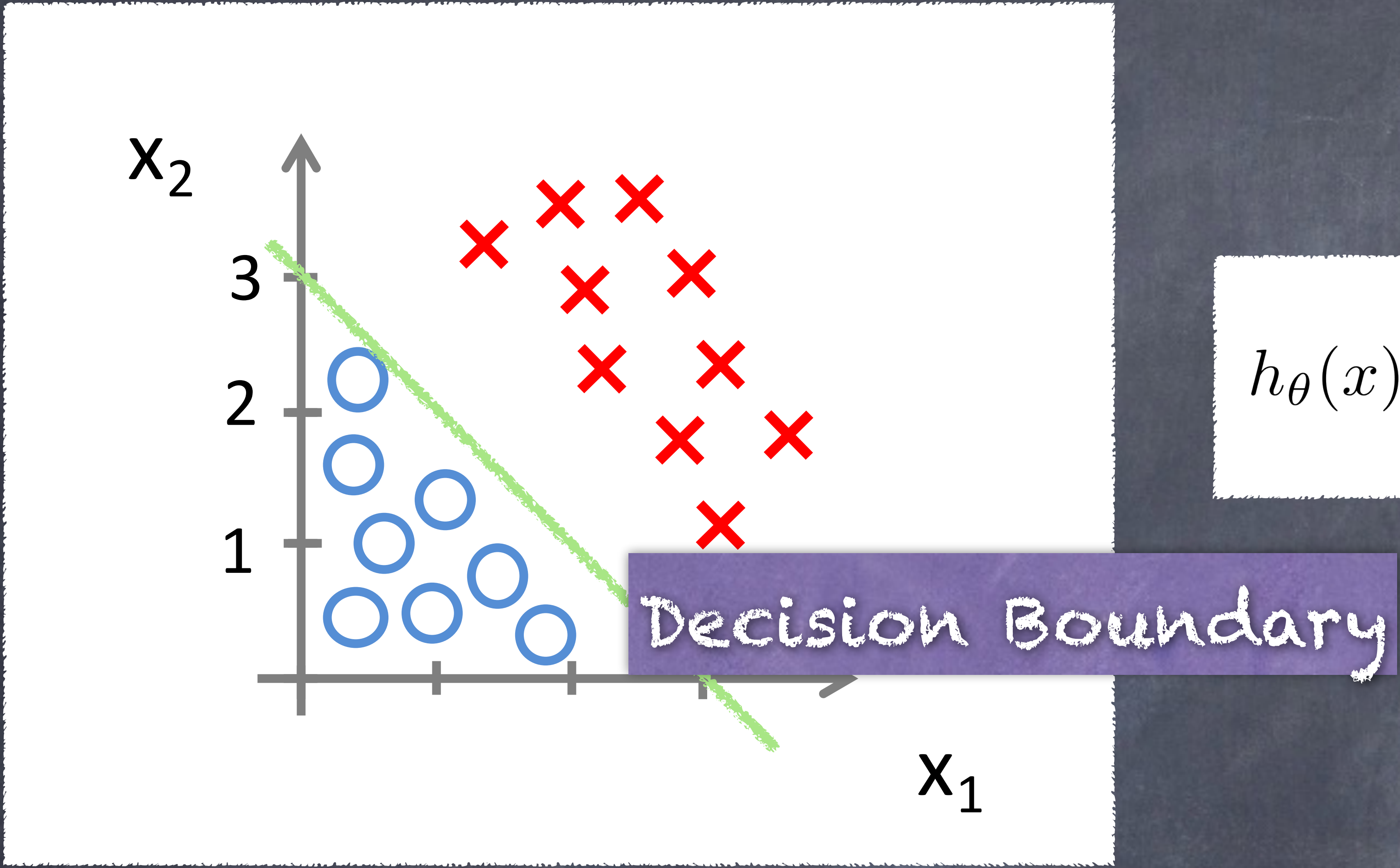


Decision Boundary

estimated probability
that y = 1 on input x



If h >= 0.5, predict y = 1
If h < 0.5, predict y = 0

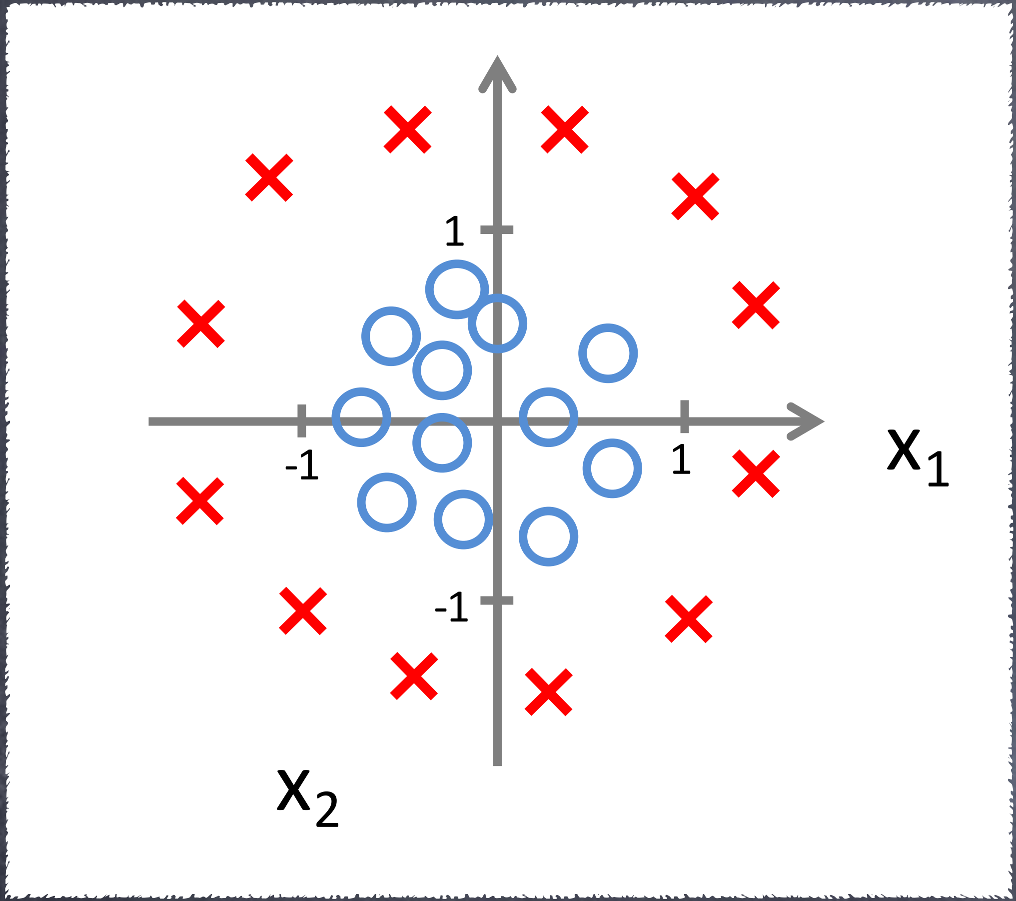


$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

$-3, \quad 1, \quad 1$

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

Non-linear decision boundaries



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

-1, 0, 0, 1, 1

More complex decision boundary

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

Cost Function

Training set

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

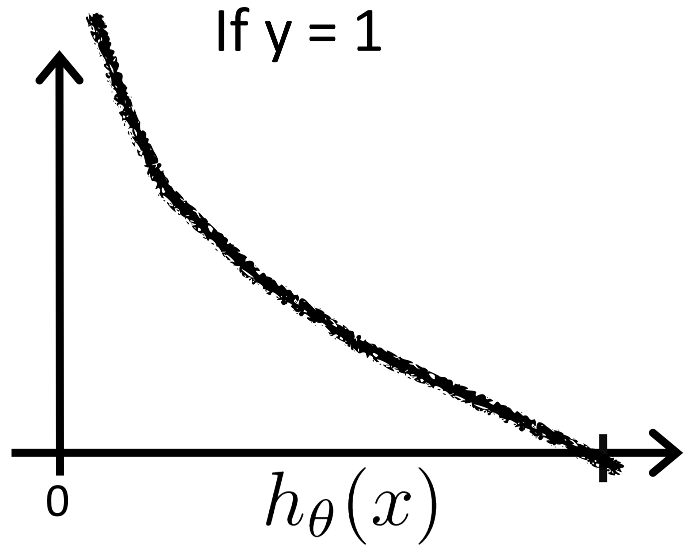
Cost Function
in Linear Regression

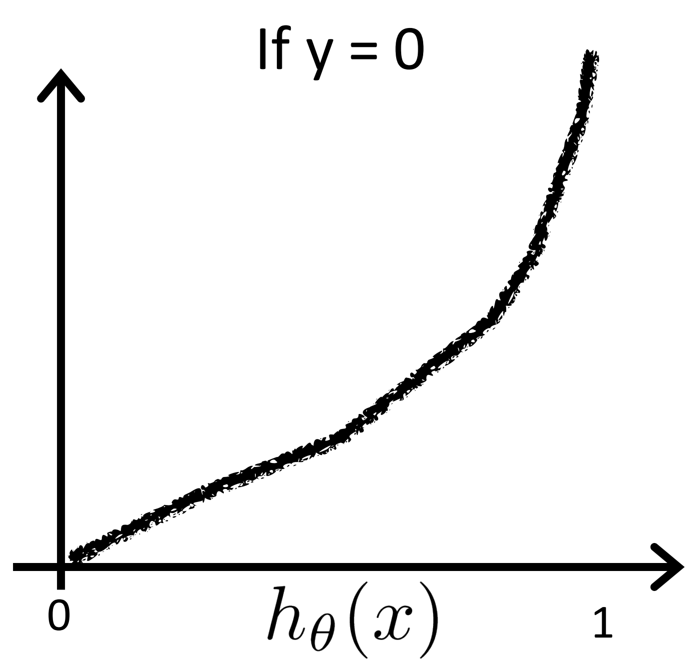
$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Cost Function
in Logistic Regression

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

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





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

$$\text{Cost}(h_\theta(x), y) = -\log(h_\theta(x)) * y \ -\log(1 - h_\theta(x)) * (1-y)$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}(h_\theta(x^{(i)}), y^{(i)})$$
$$= -\frac{1}{m} [\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))]$$

Gradient Descent

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

Want $\min_\theta J(\theta)$:

Repeat $\{$

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

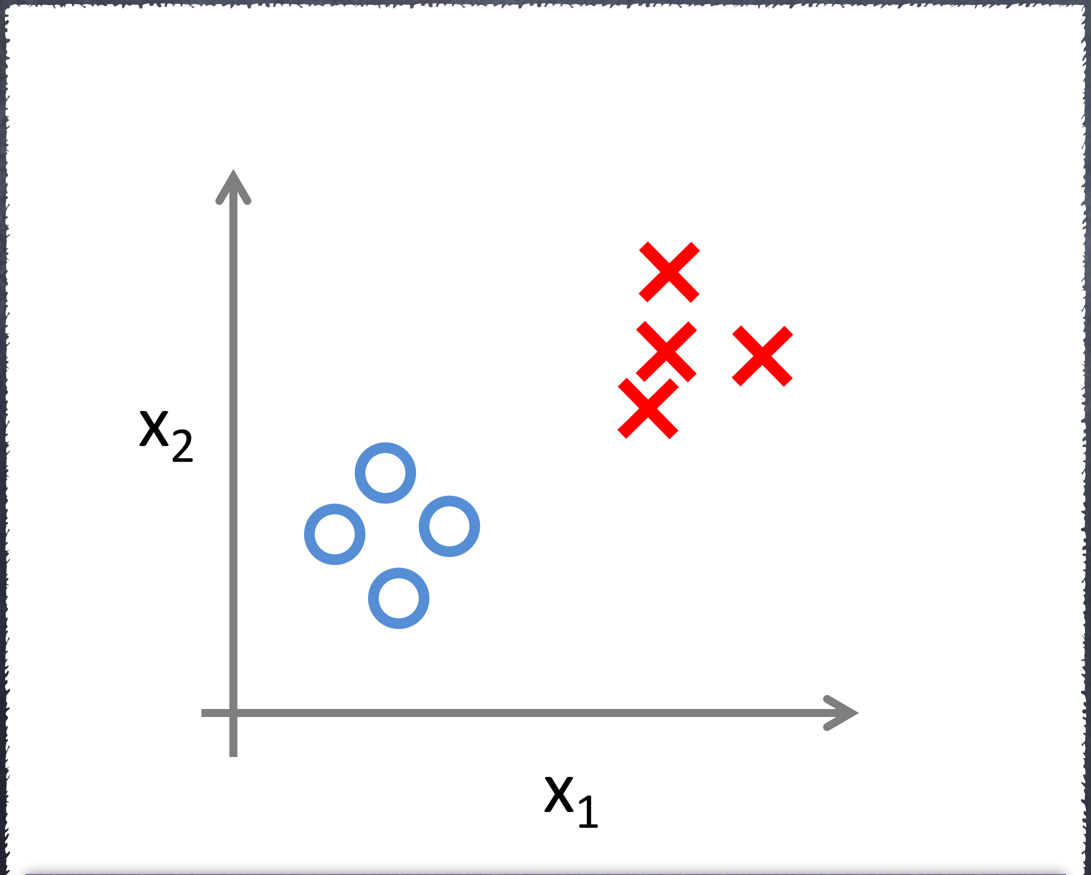
(simultaneously update all θ_j)

$\}$

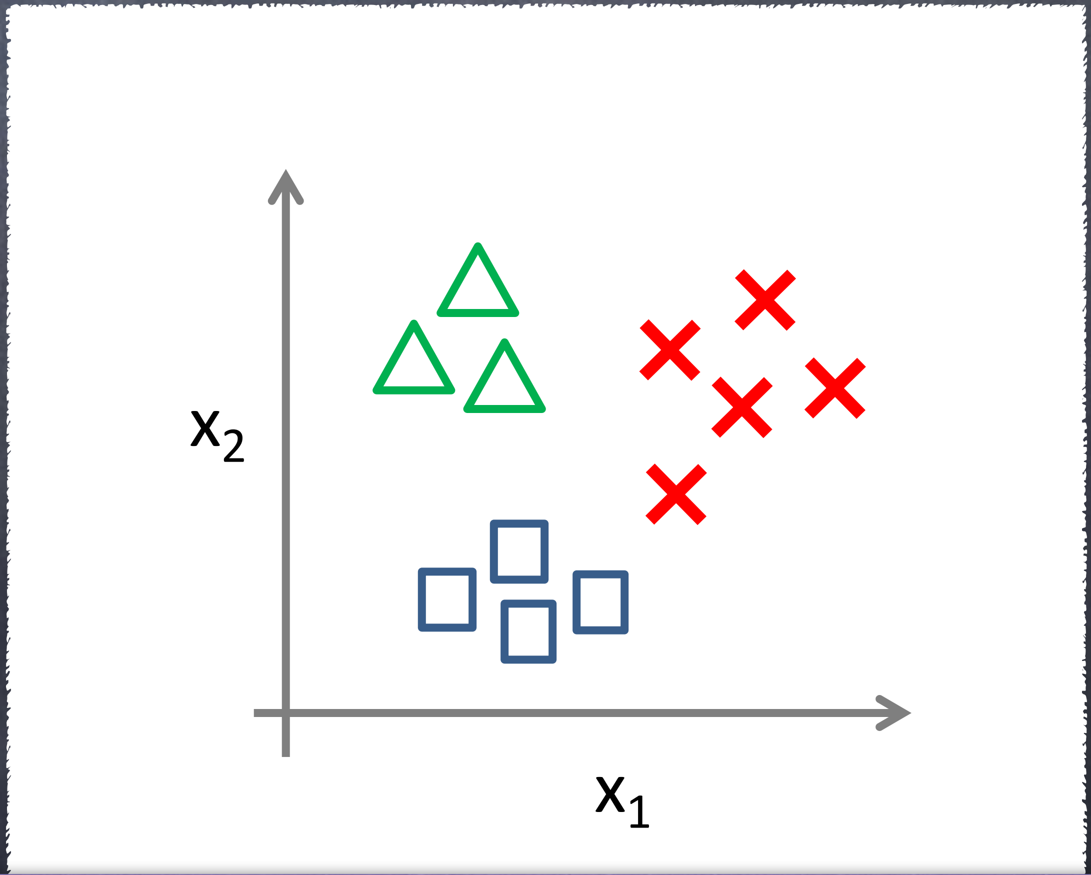
Multi-classification

An example:

Email foldering/tagging: Work, Friends, Family, Hobby

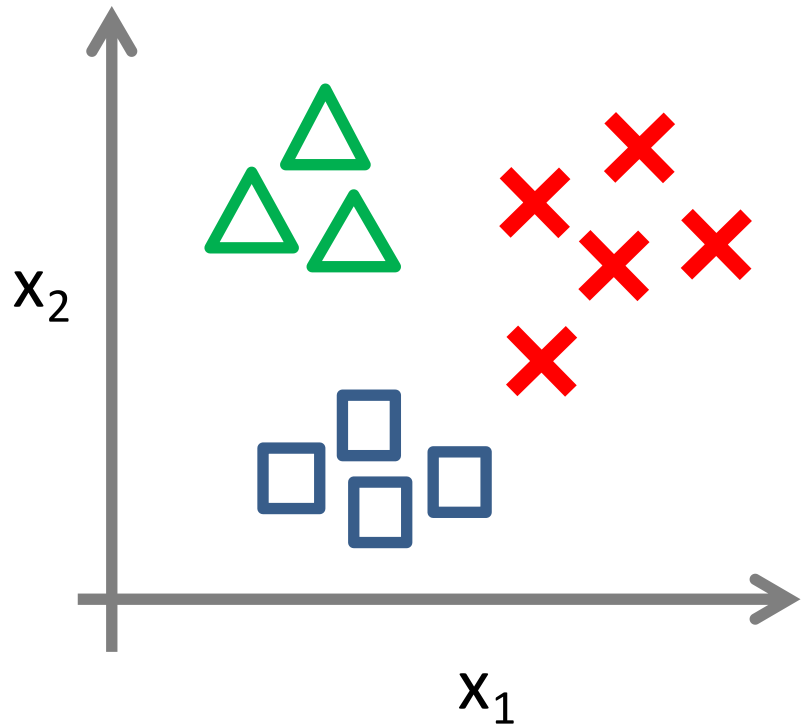


Binary Classification



Multi-class Classification

One-Vs-ALL



Class 1: △

Class 2: □

Class 3: ×

$$h_\theta^{(i)}(x) = P(y = i | x; \theta)$$