MACHINE LEARNING (ML-7)

Dr. NEERAJ GUPTA, Department of CEA, GLA University, Mathura

AGENDA

- Logistic regression (Classification)

SRC : * Andrew NG

Classification

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

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

two

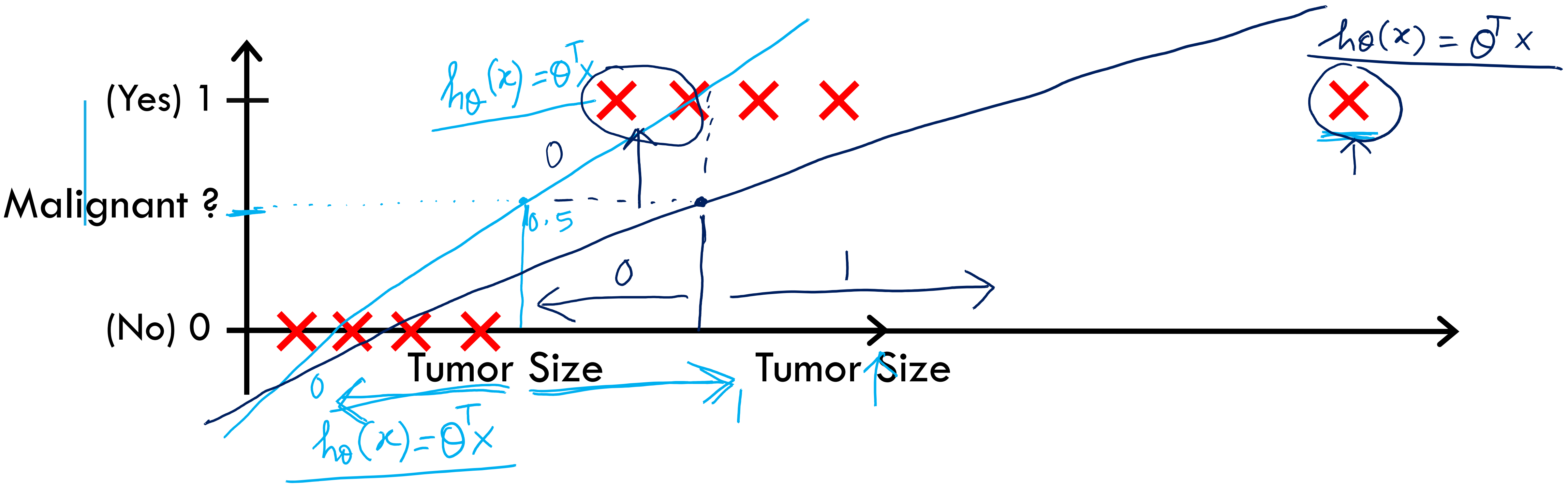
0: "Negative Class" (e.g., benign tumor)

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

→ Binary classification

$y \in \{0, 1, 2, 3\}$ → Multiclass Classification

four labels



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"

Classification: y = 0 or 1

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

Logistic Regression: $\underbrace{0 \leq h_\theta(x) \leq 1}$

Classification

Logistic Regression

HYPOTHESIS REPRESENTATION

Logistic Regression Model

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

$0 \le h_\theta(x) \le 1$ $\quad h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

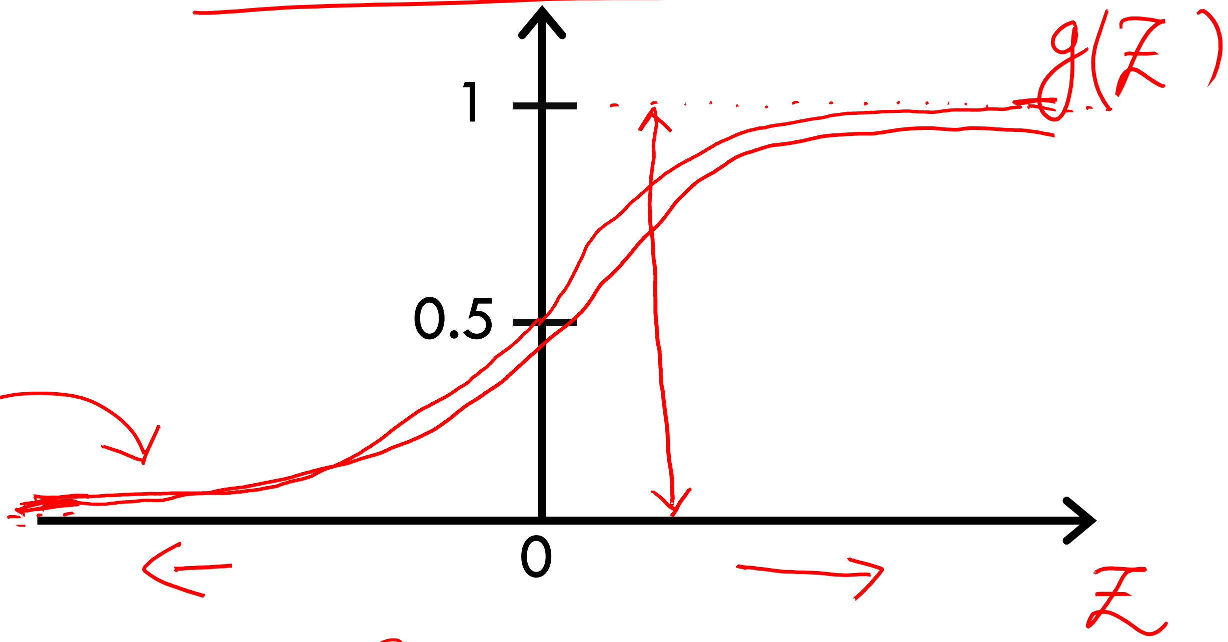
Want $0 \le h_\theta(x) \le 1$

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

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

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

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



→ Sigmoid function

=

→ Logistic function

Parameters → θ

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

Interpretation of Hypothesis Output

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

$h_\theta(x) =$ estimated probability that y = 1 on input x

Example: If $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$

$h_\theta(x) = 0.4$

$$\boxed{h_\theta(x) = 0.7}$$

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

Tell patient that 70% chance of tumor being malignant

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

"probability that y = 1, given x, parameterized by θ "

$y = 0$ or 1

$P(y=0 \mid x; \theta)$ 0.3 30%

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

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

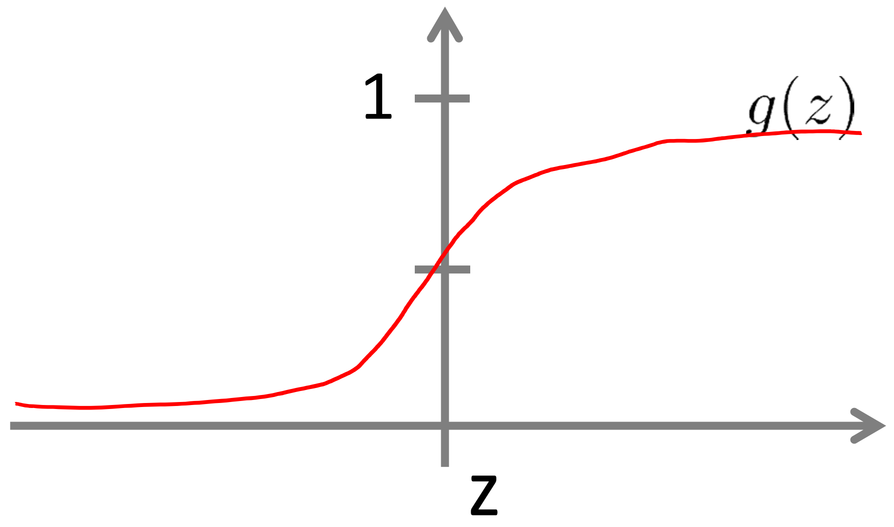
Logistic Regression

DECISION BOUNDARY

Logistic regression

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

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



Suppose predict "$y = 1$" if $\boxed{h_\theta(x) \geq 0.5}$

predict "$y = 0$" if $\boxed{h_\theta(x) < 0.5}$

$h_\theta(x) = 0.7$

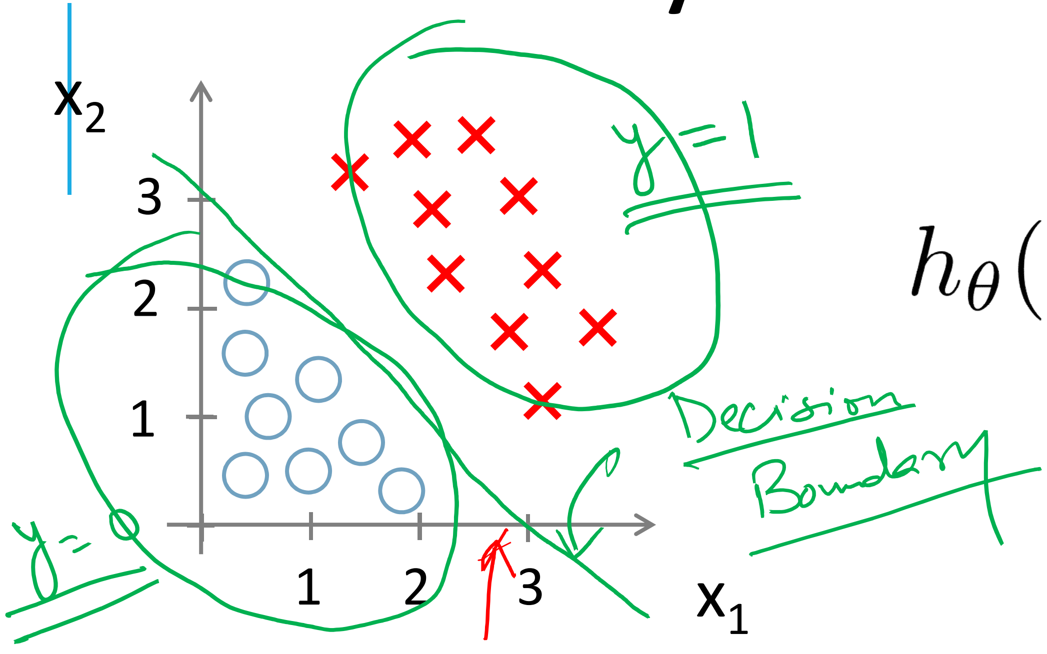
$y \in \{0, 1\}$

$y = 1$

$h_\theta(x) = 0.4$

$y = 0$

Decision Boundary



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

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

$$-3 + x_1 + x_2$$

Predict "$y = 1$" if $\underbrace{-3 + x_1 + x_2}_{\theta^T x} \geq 0$

$x_1 + x_2 \geq 3$

$x_1 + x_2 < 3$

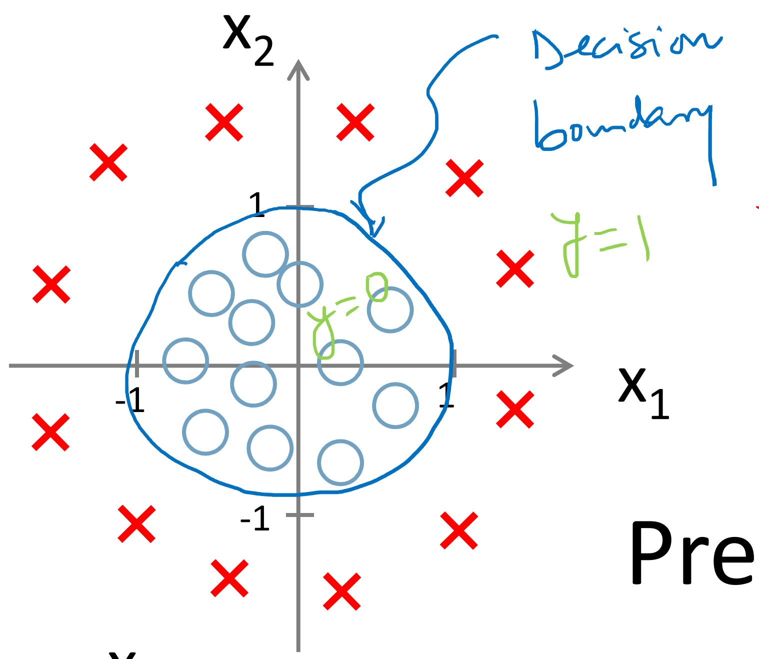
$y = 0$

$x_1 + x_2 \geq -3$

$x_1 + x_2 = -3$

$x_1 = 0$

Non-linear decision boundaries



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

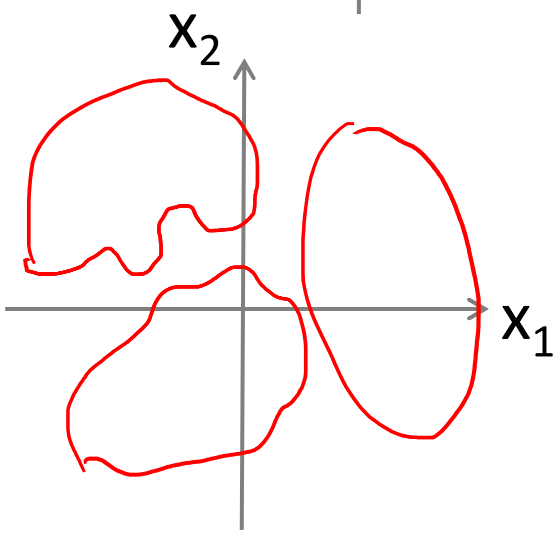
$$\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Predict "$y = 1$" if $-1 + x_1^2 + x_2^2 \geq 0$

$$\boxed{x_1^2 + x_2^2 = 1}$$

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

$$x_1^2 + x_2^2 < 1 \quad \boxed{y = 0}$$



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

Logistic Regression

COST FUNCTION

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

m examples $\quad x \in \begin{bmatrix} x_0 \\ x_1 \\ \cdots \\ x_n \end{bmatrix} \quad \boxed{x_0 = 1}, y \in \{0, 1\}$

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

How to choose parameters θ?

Cost function

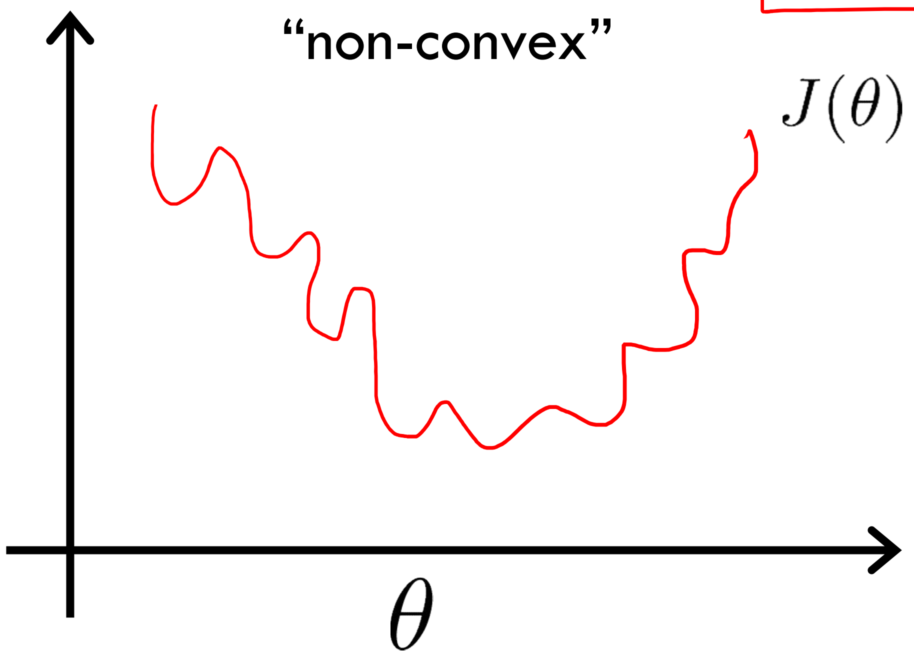
~~Linear~~ regression: $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

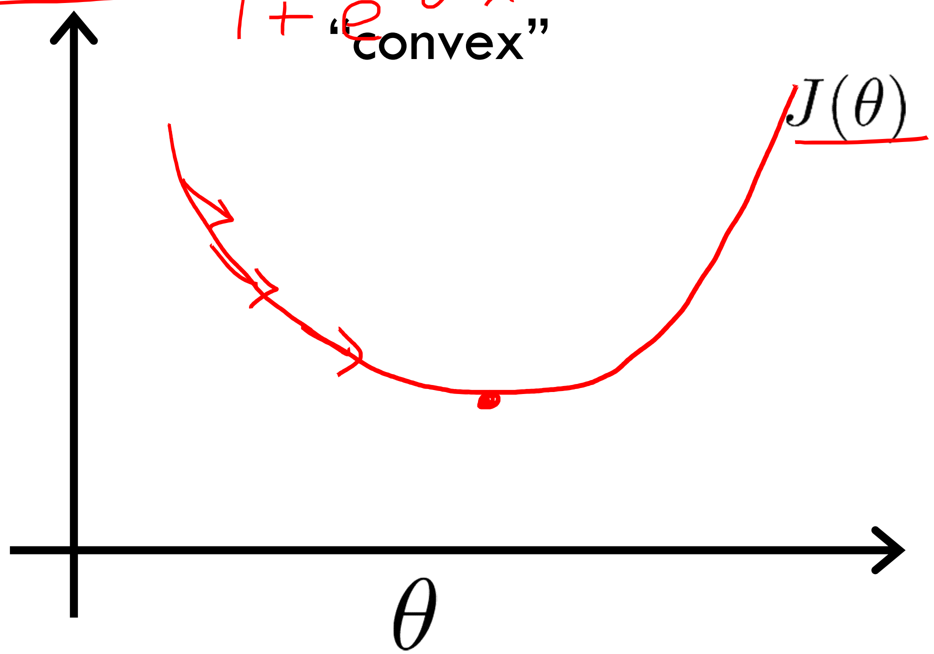
Logistic

$\mathrm{Cost}\left(h_\theta(x^{(i)}), y\right)$

$$\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)}) = \frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

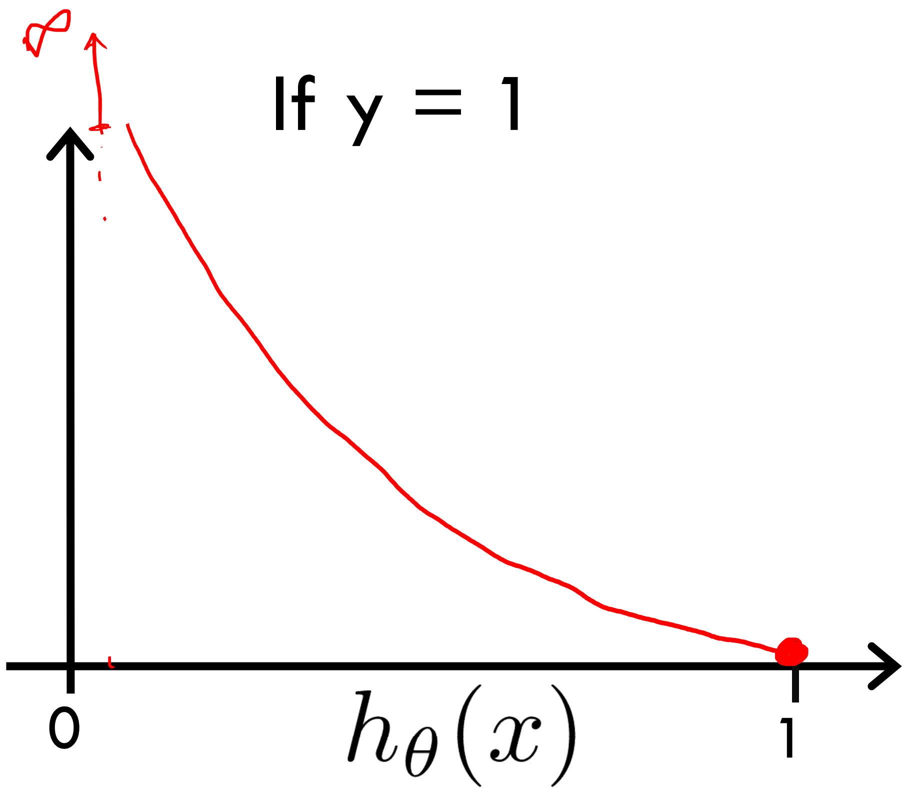
$\frac{1}{1+e^{-\theta^T x}}$





Logistic regression cost function

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



Cost $= 0$ if $y = 1, h_\theta(x) = 1$

But as $h_\theta(x) \rightarrow 0$

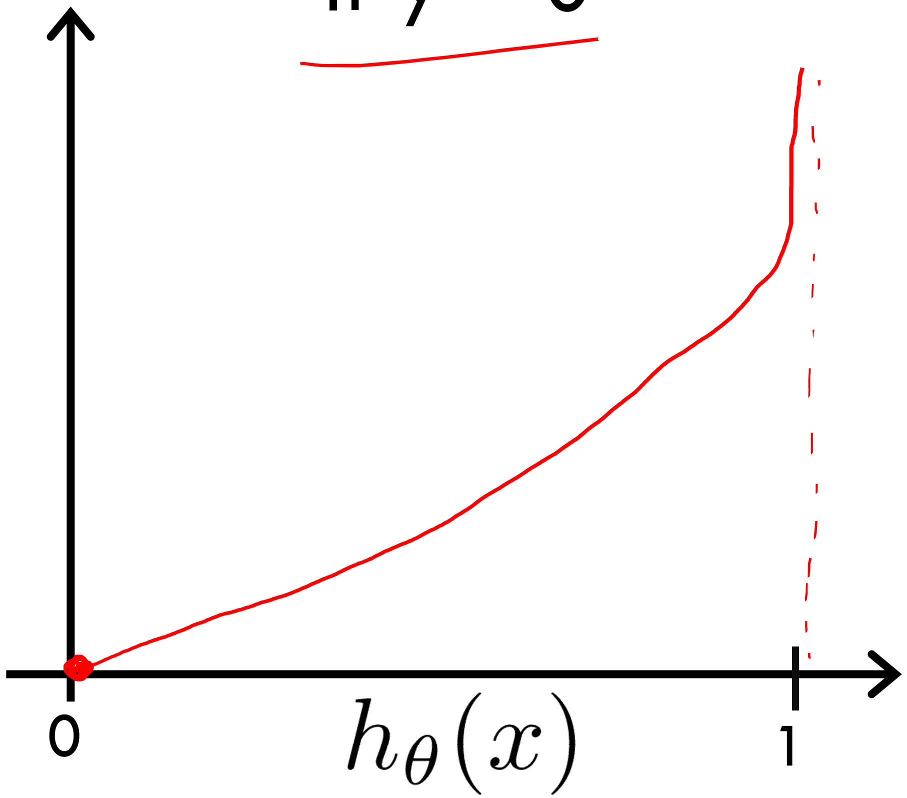
$$\boxed{Cost \rightarrow \infty}$$

Captures intuition that if $h_\theta(x) = 0$, (predict $P(y = 1|x;\theta) = 0$), but $y = 1$, we'll penalize learning algorithm by a very large cost.

Logistic regression cost function

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

If y = 0



Logistic Regression

SIMPLIFIED COST FUNCTION AND GRADIENT DESCENT

Logistic regression cost function

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

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

Note: $y = 0$ or 1 always

$$\text{Cost}(h_\theta(x), y) = -\underbrace{y}_{=0,\ =1}\log(h_\theta(x)) - \underbrace{(1-y)}_{=0,\ =1}\log(1 - h_\theta(x))$$

$1-1=0$

If $y=1$: $\text{cost}(h_\theta(x), y) = -\log(h_\theta(x)) - 0$

If $y=0$: $\text{cost}(h_\theta(x), y) = 0 - (1)\log(1 - h_\theta(x))$

$= -\log(1 - h_\theta(x))$

Logistic regression cost function

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

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

To fit parameters θ:

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

Get θ

To make a prediction given new x:

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

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

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 $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} \boxed{J(\theta)}$$

(simultaneously update all θ_j)

$\}$

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 $\{$

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

(simultaneously update all θ_j)

$\}$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

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

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

logistic

Algorithm looks identical to linear regression!

Logistic Regression

MULTI-CLASS CLASSIFICATION: ONE-VS-ALL

Binary

Two Classes

0 1

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

$y \in \{1, 2\}$

Multiclass classification

App

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

$y \in 1$ $y \in 2$ $y \in 3$ $y \in 4$

$y \in \{1, 2, 3, 4\}$

Medical diagrams: Not ill, Cold, Flu

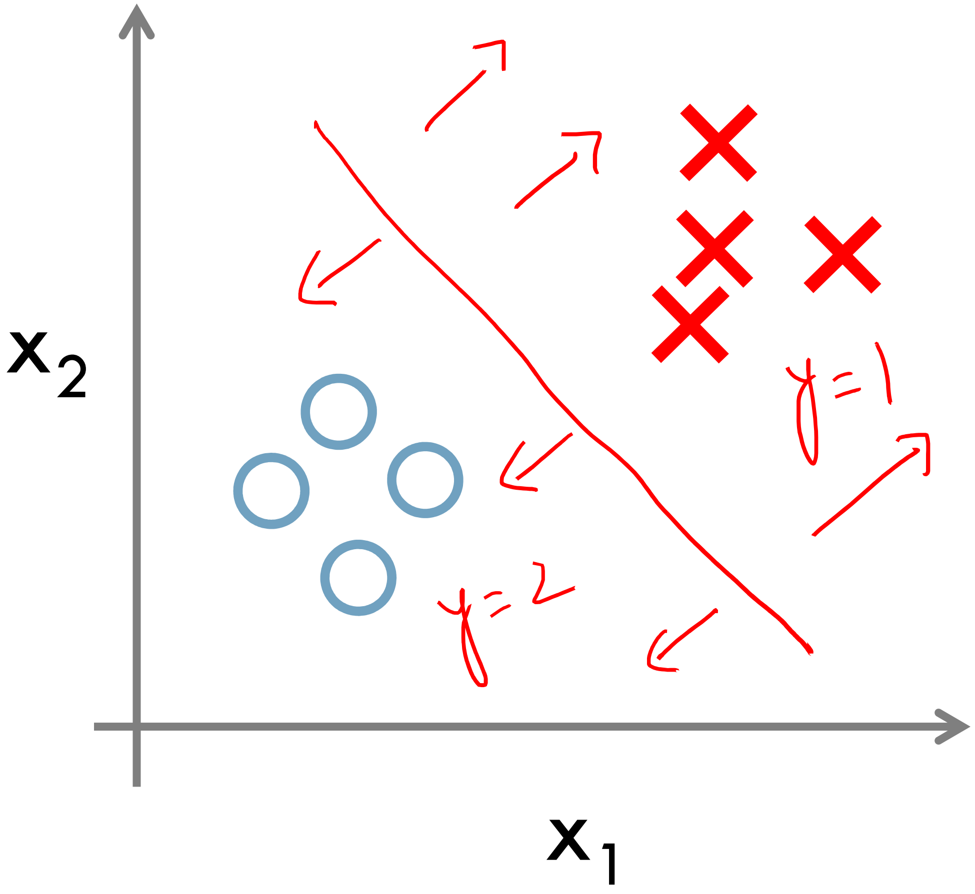
$y=1$ $y=2$ $y=3$ 3-class

$y=1$ $y=2$ $y=3$ $y=4$

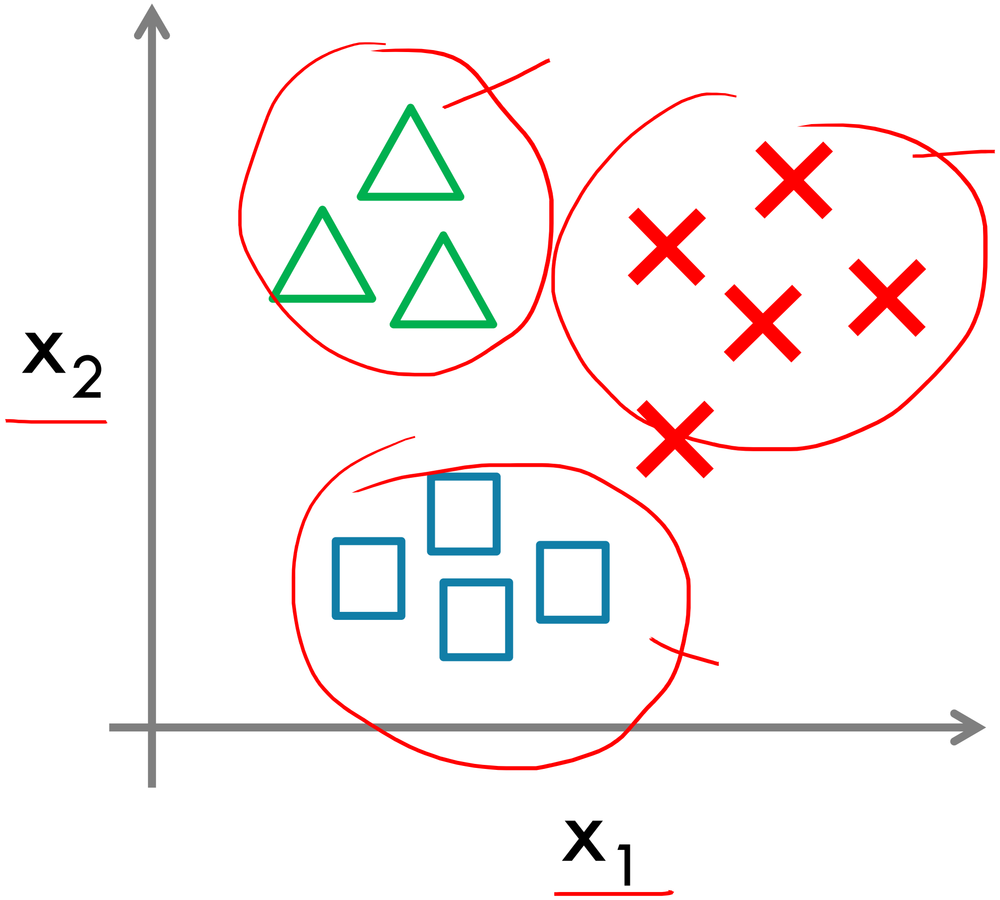
Weather: Sunny, Cloudy, Rain, Snow

$y \in \{1, 2, 3, 4\}$

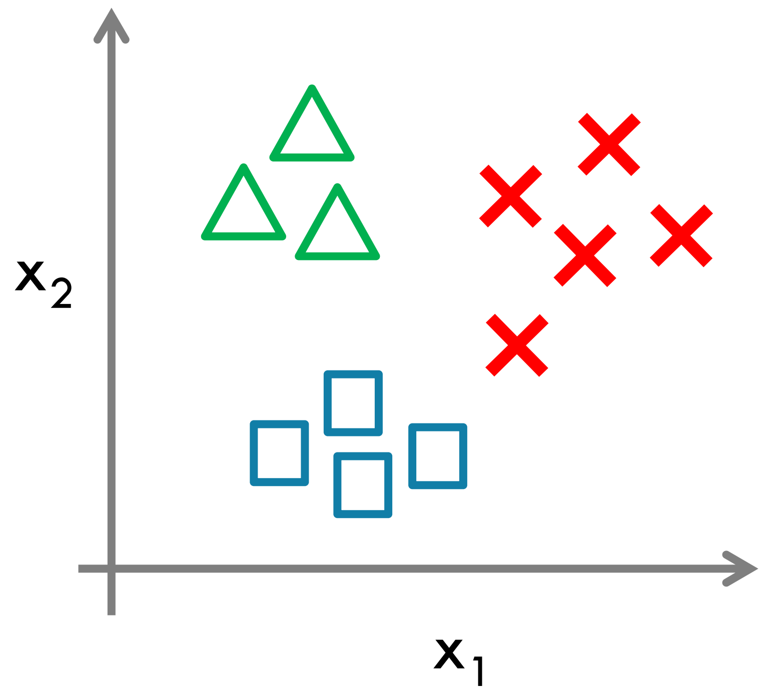
Binary classification:

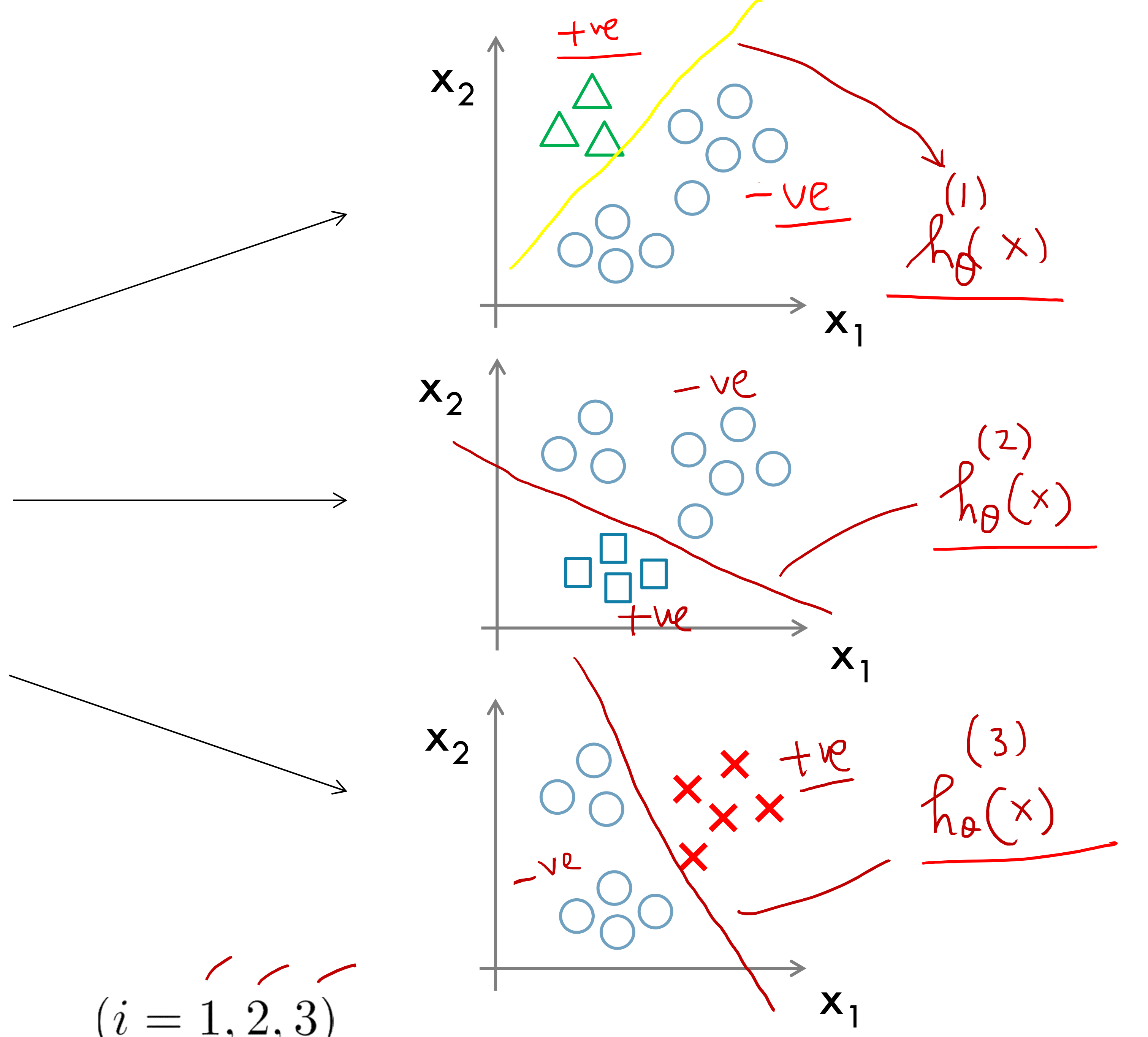


Multi-class classification:



One-vs-all (one-vs-rest):





Class 1: △

Class 2: □

Class 3: ✕

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

One-vs-all

$i = 1, 2, 3$

Train a logistic regression classifier $\boxed{h_\theta^{(i)}(x)}$ for each class i to predict the probability that $y = i$.

On a new input x, to make a prediction, pick the class i that maximizes

$$\boxed{\max_i h_\theta^{(i)}(x)}$$

$x \Leftarrow$ data given for prediction

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

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

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

$\boxed{y = \max(h_\theta^{(i)}(x))}$

THANKS


Keep Learning
Keep Growing


Dr. Neeraj Gupta
Assistant Professor, Dept. of CEA
neeraj.gupta@gla.ac.in