Classification

Classification and Representation

Logistic Regression

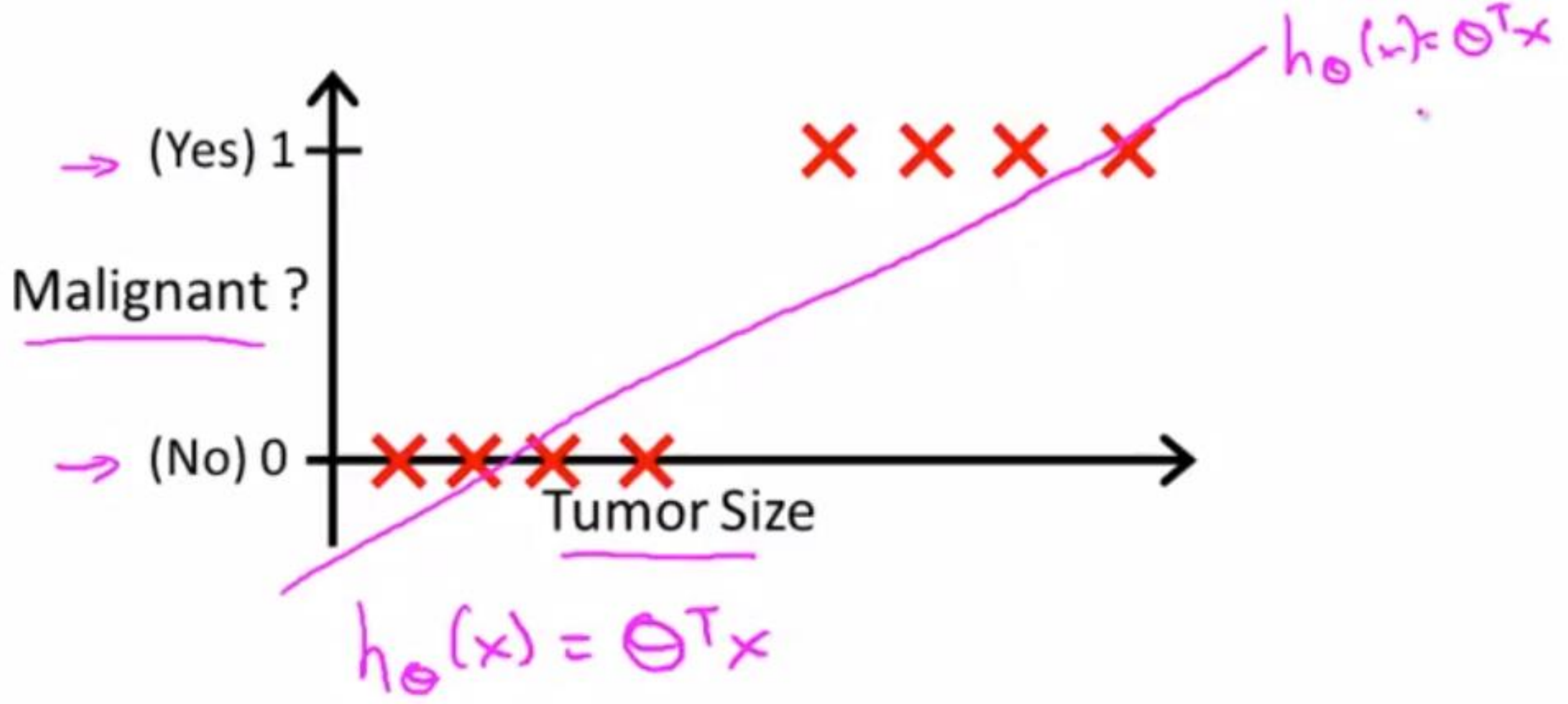
Classification

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

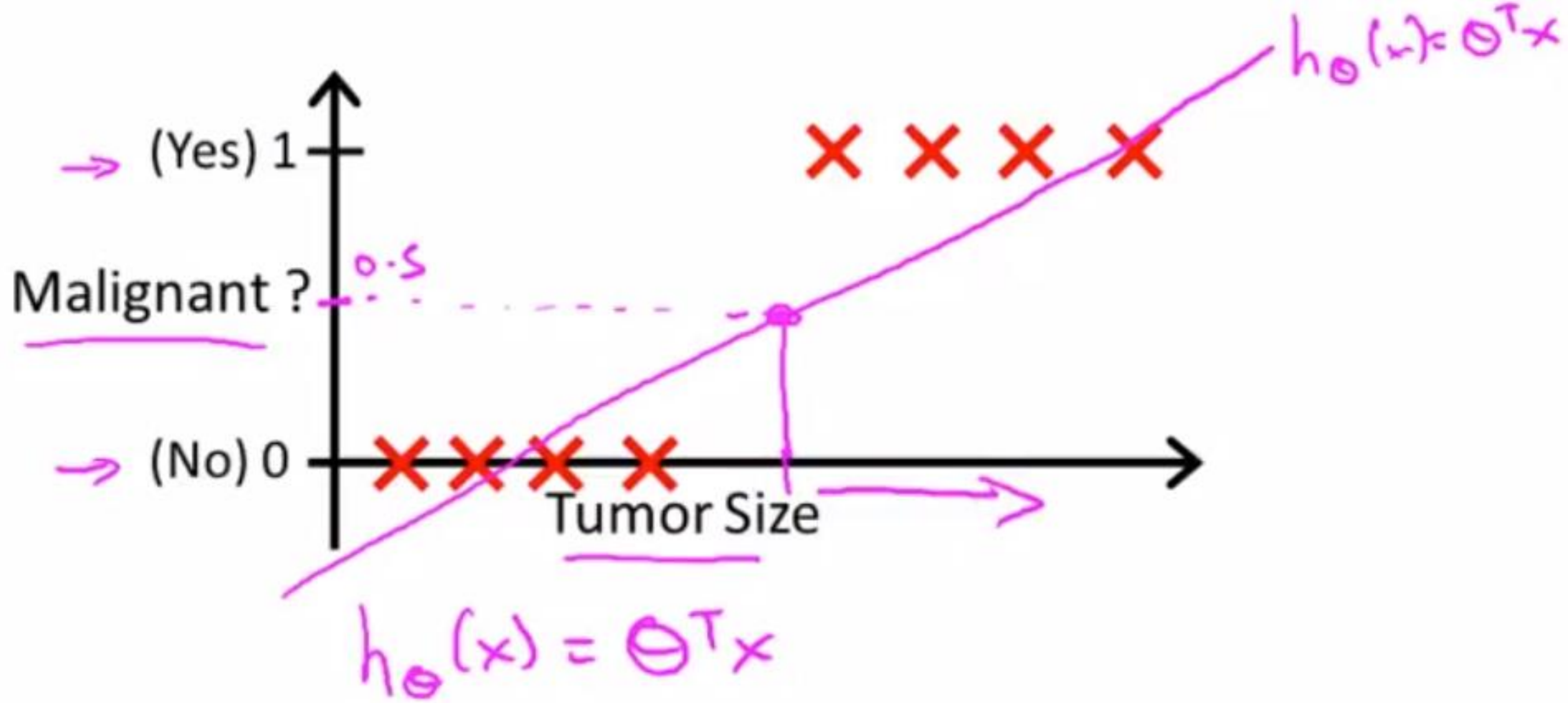
$y \in \{0, 1\}$

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

$y \in \{0, 1, 2, 3\}$



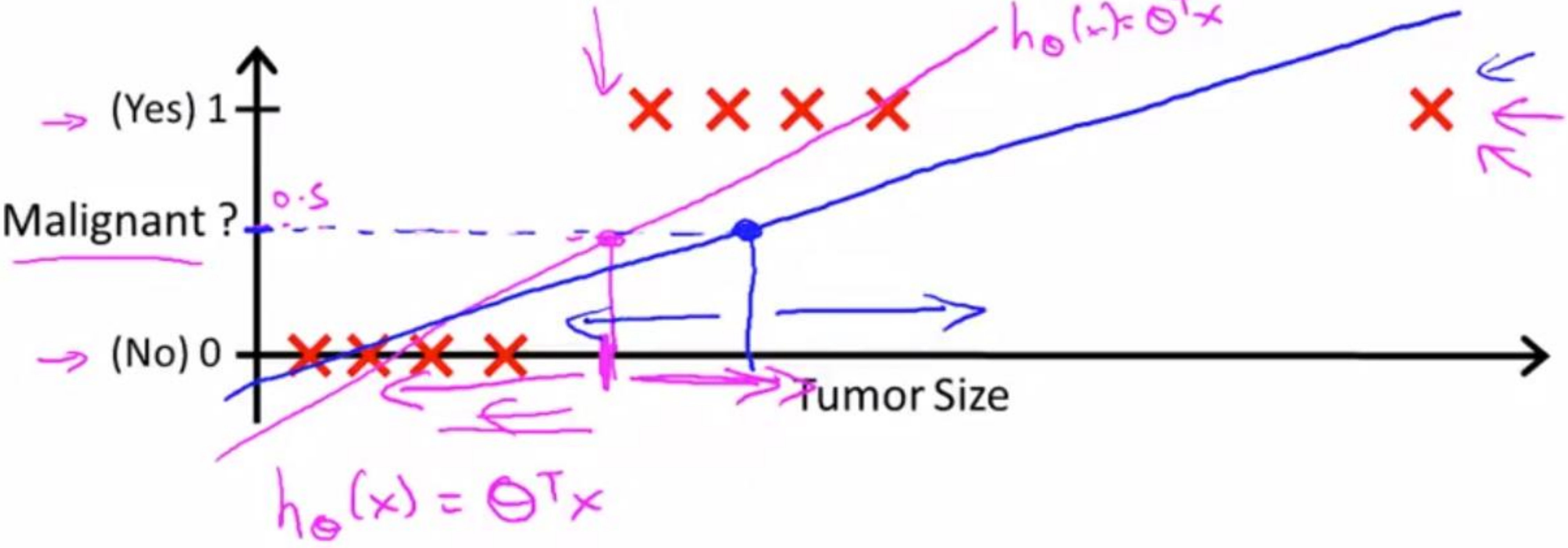
(Yes) 1
Malignant ?
(No) 0
Tumor Size
$h_\theta(x) = \theta^T x$
$h_\theta(x) = \theta^T x$



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"



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"

Exercise

Which of the following statements is true?

- If linear regression doesn't work on a classification task as in the previous example, applying feature scaling may help.
- If the training set satisfies $0 \leq y(i) \leq 1$ for every training example $(x(i),y(i))$, then linear regression's prediction will also satisfy $0 \leq h(x) \leq 1$ for all values of x.
- If there is a feature x that perfectly predicts y, i.e if $y=1$ when $x \geq c$ and $y=0$ whenever $x<c$ (for some constant c), then linear regression will obtain zero classification error.
- None of the above statements are true.

Classification: $y = 0$ or 1

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

Logistic Regression: $0 \le h_\theta(x) \le 1$

Logistic Regression Model

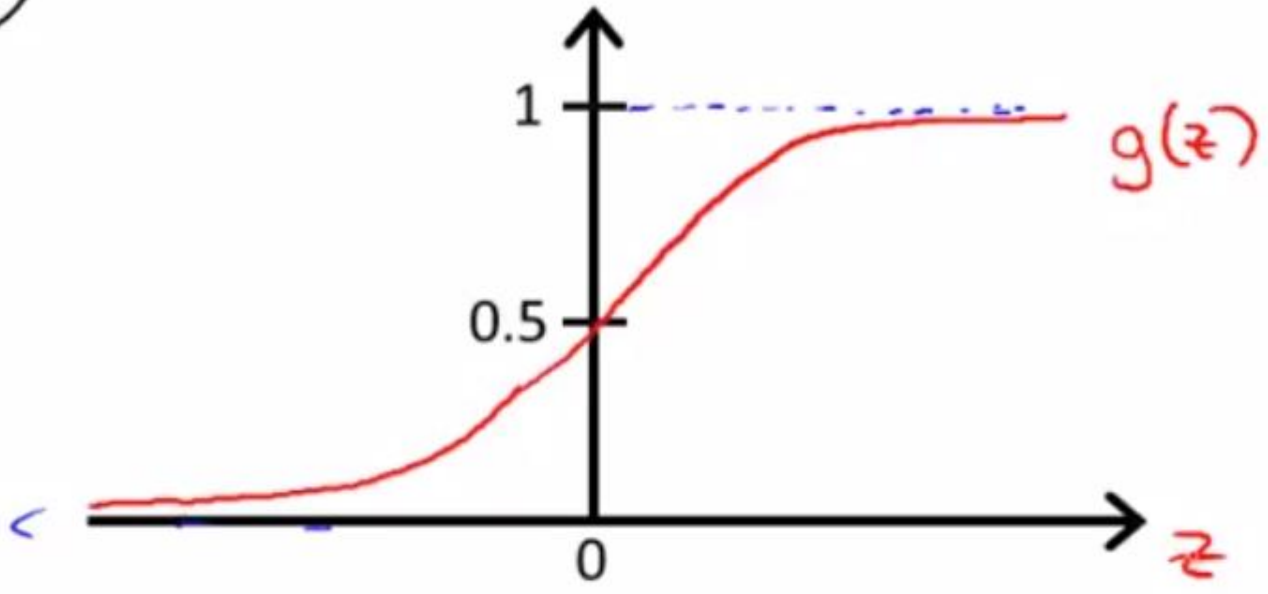
Want $0 \le h_\theta(x) \le 1$

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

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

$\theta^T x$

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



Sigmoid function

Logistic function

Interpretation of Hypothesis Output

$h_\theta(x)$

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

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

$$h_\theta(x) = 0.7$$

$y=1$

Tell patient that 70% chance of tumor being malignant

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

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

Exercise

- Suppose we want to predict, from data ***x*** about a tumor, whether it is malignant (*y*=1) or benign (*y*=0).
- Our logistic regression classifier outputs, for a specific tumor, $h_\theta(x) = P(y = 1|x; \theta) = 0.7$, so we estimate that there is a 70% chance of this tumor being malignant.
- What should be our estimate for the probability the tumor is benign?
  - 0.7^2
  - 0.7-0.3
  - 0.7-0.5
  - 0.3

Interpretation of Hypothesis Output

$h_\theta(x)$

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

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

$$h_\theta(x) = 0.7 \qquad y=1$$

Tell patient that 70% chance of tumor being malignant

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

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

$y = 0$ or 1

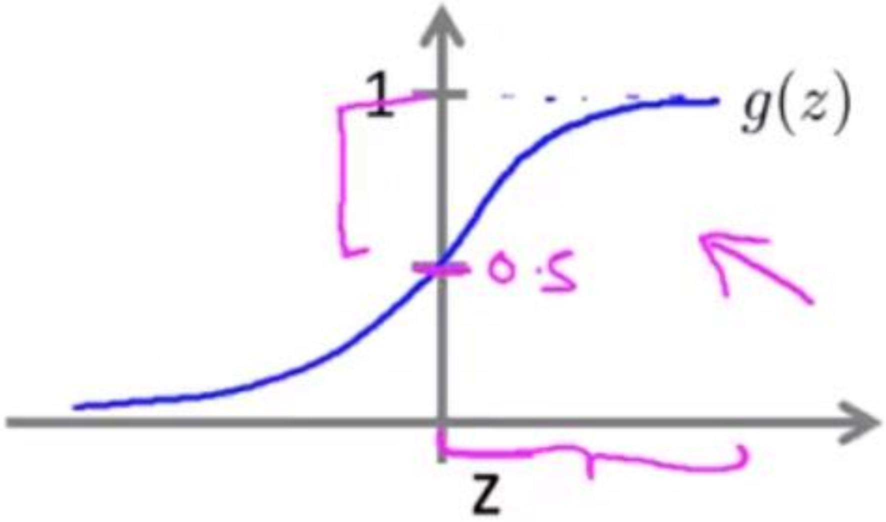
$$P(y = 0|x;\theta) + P(y = 1|x;\theta) = 1$$
$$P(y = 0|x;\theta) = 1 - P(y = 1|x;\theta)$$

Decision Boundary

Logistic regression

$$\rightarrow h_\theta(x) = g(\theta^T x) = P(y=1|x;\theta)$$

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

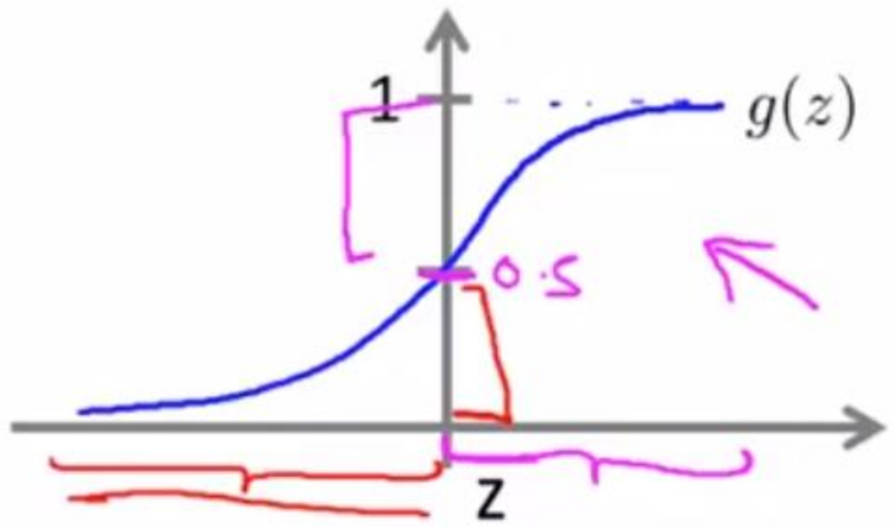


Suppose predict "$y = 1$" if $h_\theta(x) \geq 0.5$

$g(z) \geq 0.5$
when $z \geq 0$

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

Logistic regression



$$\rightarrow h_\theta(x) = g(\theta^T x) = P(y=1|x;\theta)$$

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

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

$\theta^T x \geq 0$

$g(z) \geq 0.5$
when $z \geq 0$

$h_\theta(x) = \underline{g(\theta^T x)} \geq 0.5$
whenever $\theta^T x \geq 0$

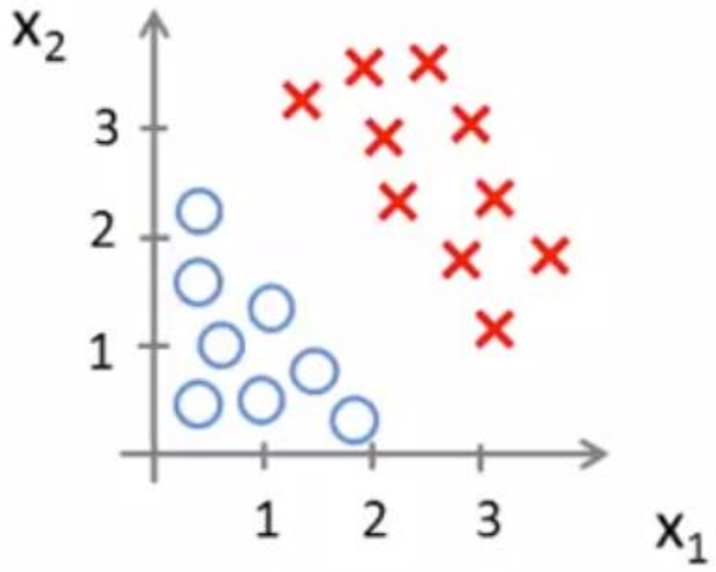
predict "$\underline{y = 0}$" if $\underline{h_\theta(x) < 0.5}$

$h_\theta(x) = g(\underline{\theta^T x})$

$\underline{g(z) < 0.5}$

$\theta^T x < 0$

Decision Boundary



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

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

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

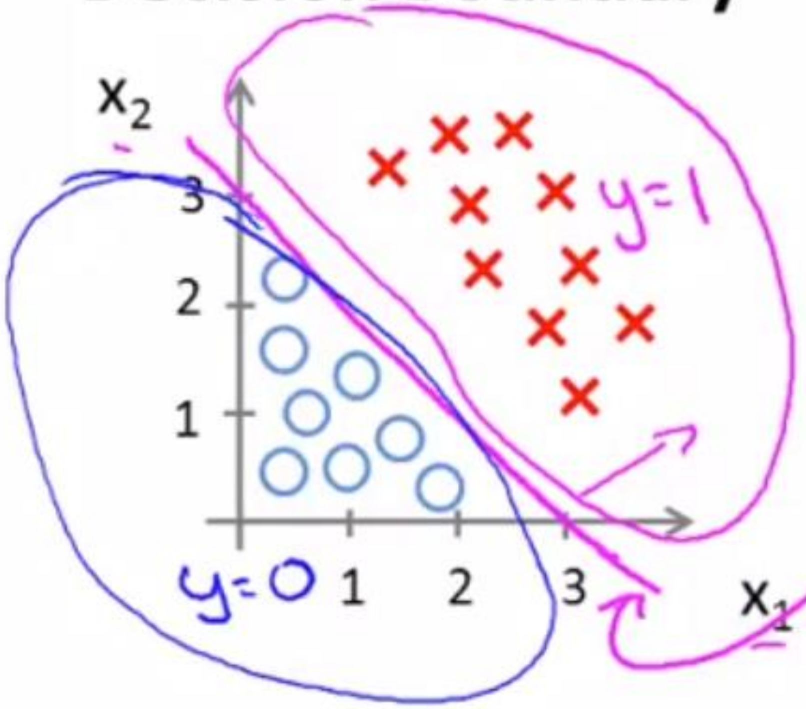
x_1, x_2

$x_1 + x_2 = 3$

$\rightarrow x_1 + x_2 \geq 3$

Decision Boundary

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



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

Decision boundary

x_1, x_2

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

$x_1 + x_2 = 3$

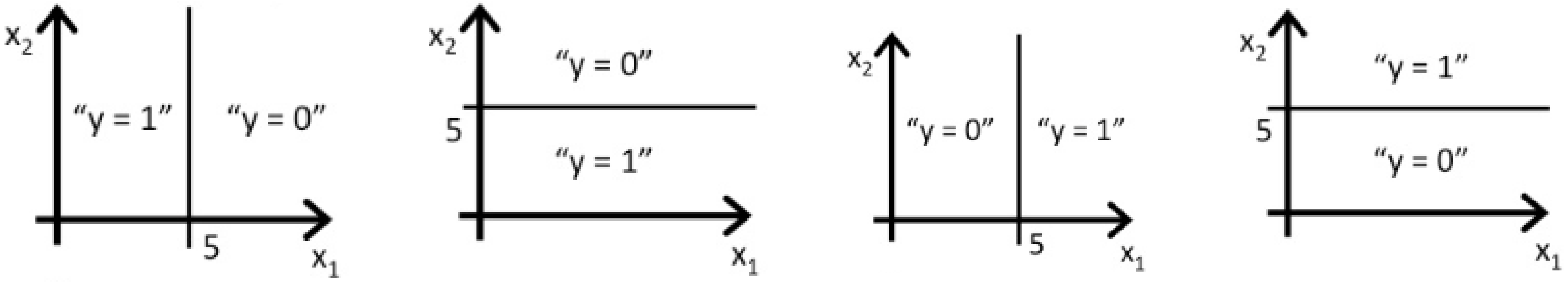
$\rightarrow x_1 + x_2 \geq 3$

$x_1 + x_2 < 3$

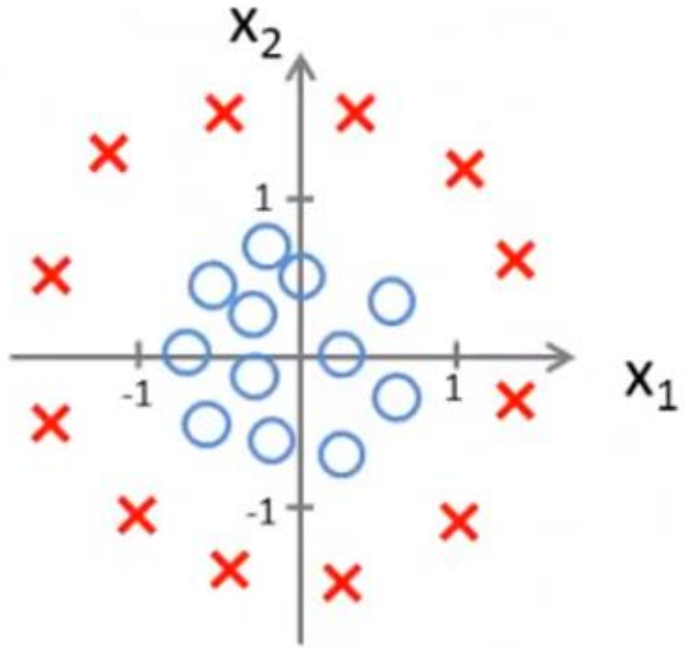
$y = 0$

Exercise

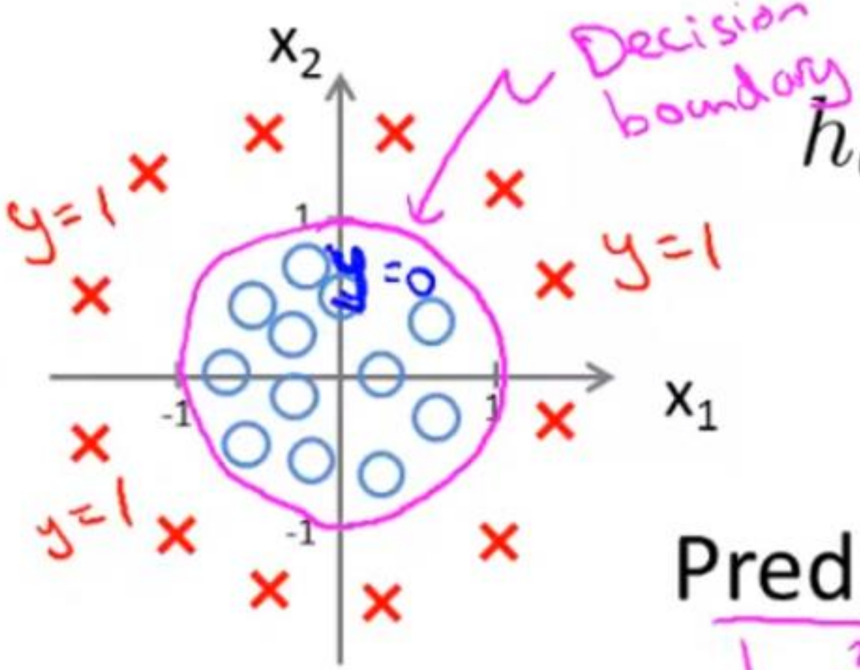
- Consider logistic regression with two features x1 and $x2$. Suppose $\theta_0 = 5$ and $\theta_1 = -1$, θ_2=0, so that $h_\theta(x) = g(5 - x_1)$. Which of these shows the decision boundary?



Non-linear decision boundaries



Non-linear decision boundaries



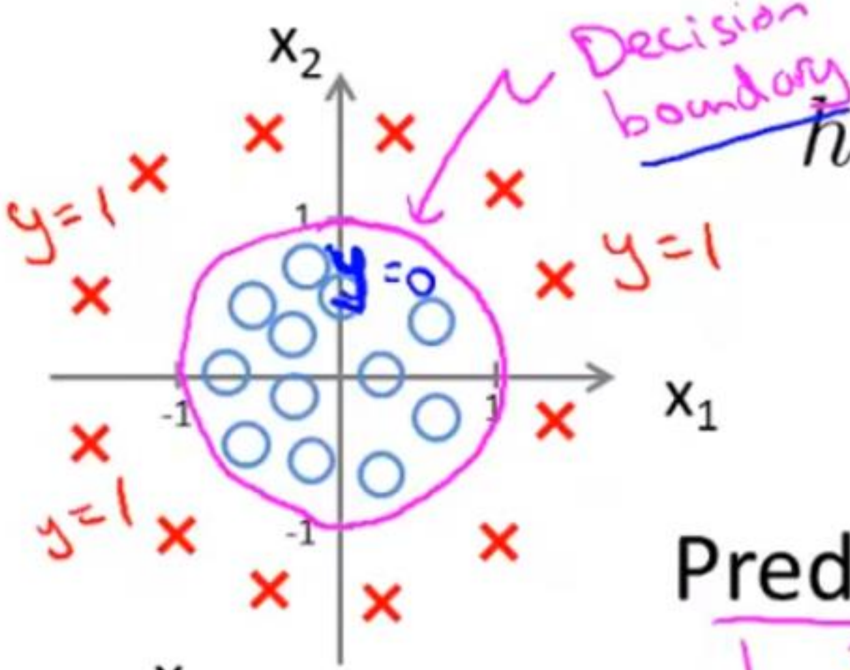
$$h_\theta(x) = g(\overset{=-1}{\theta_0} + \overset{=0}{\theta_1} x_1 + \overset{=0}{\theta_2} x_2 + \underset{=1}{\theta_3} x_1^2 + \underset{=1}{\theta_4} x_2^2)$$

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

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

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

Non-linear decision boundaries



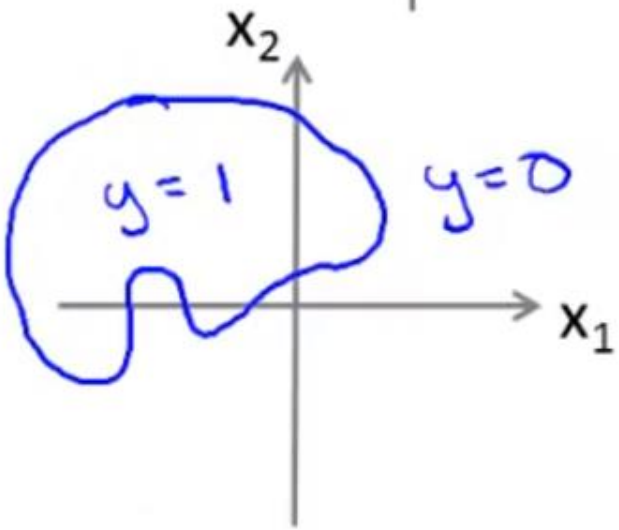
$$h_\theta(x) = g(\overset{=-1}{\theta_0} + \overset{=0}{\theta_1} x_1 + \overset{=0}{\theta_2} x_2 + \underset{=1}{\theta_3} x_1^2 + \underset{=1}{\theta_4} x_2^2)$$

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

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

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

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



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

Multiclass Classification: One vs All

Multiclass Classification

Logistic Regression

Multiclass classification

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

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

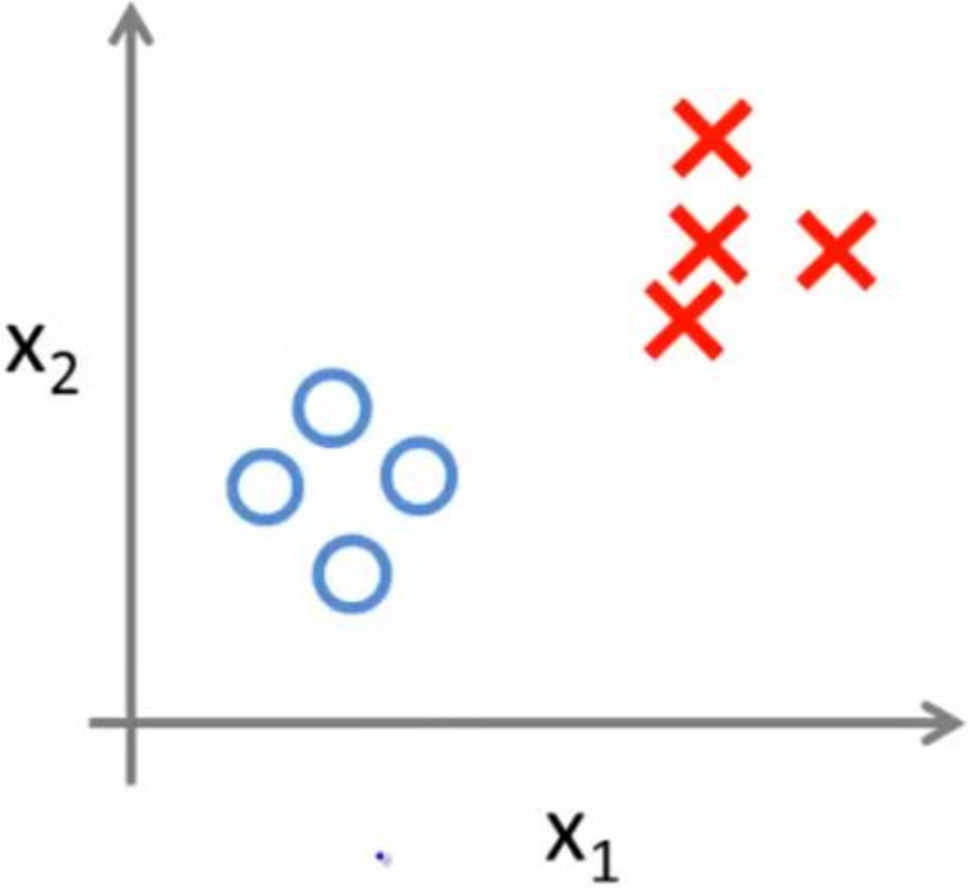
Medical diagrams: Not ill, Cold, Flu

$y=1$ 2 3

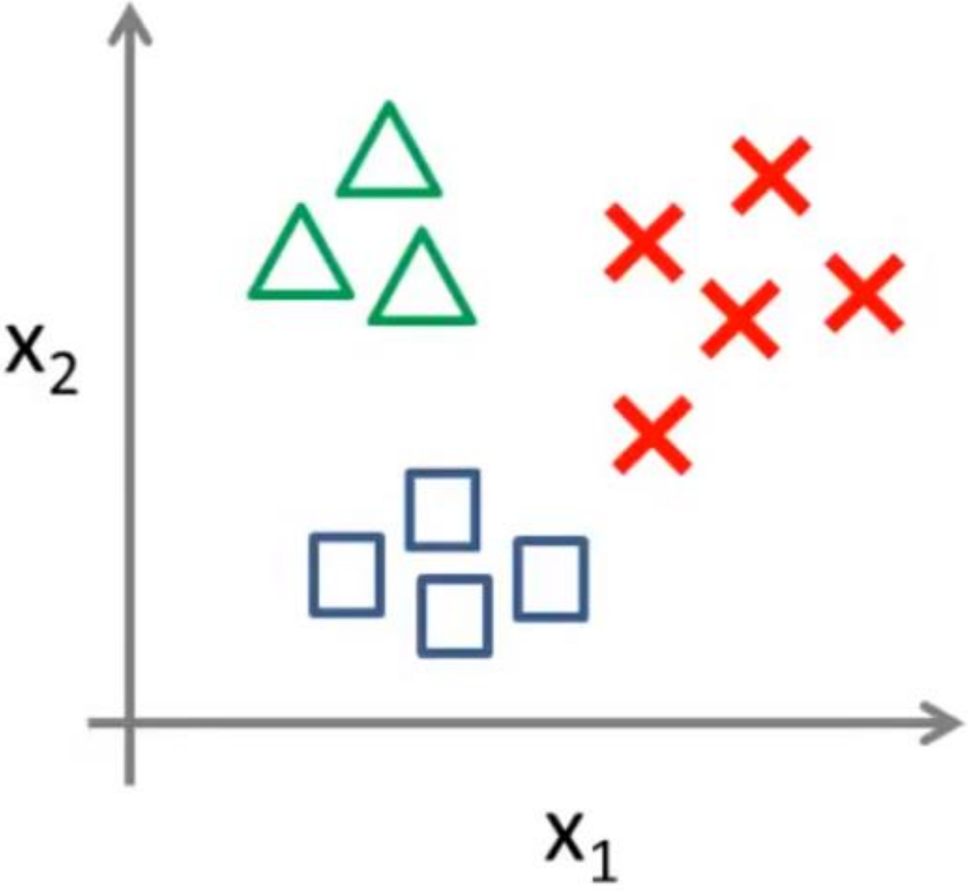
Weather: Sunny, Cloudy, Rain, Snow

$y=1$ 2 3 4

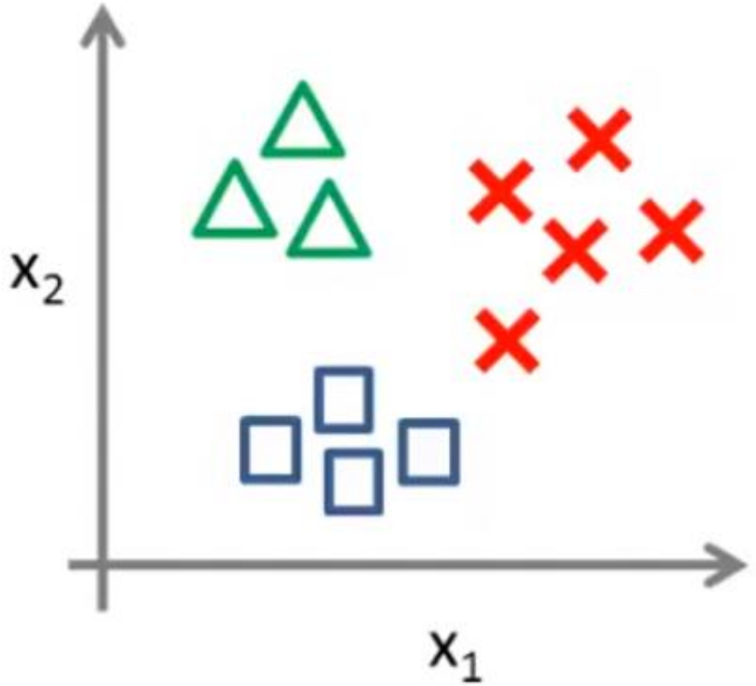
Binary classification:

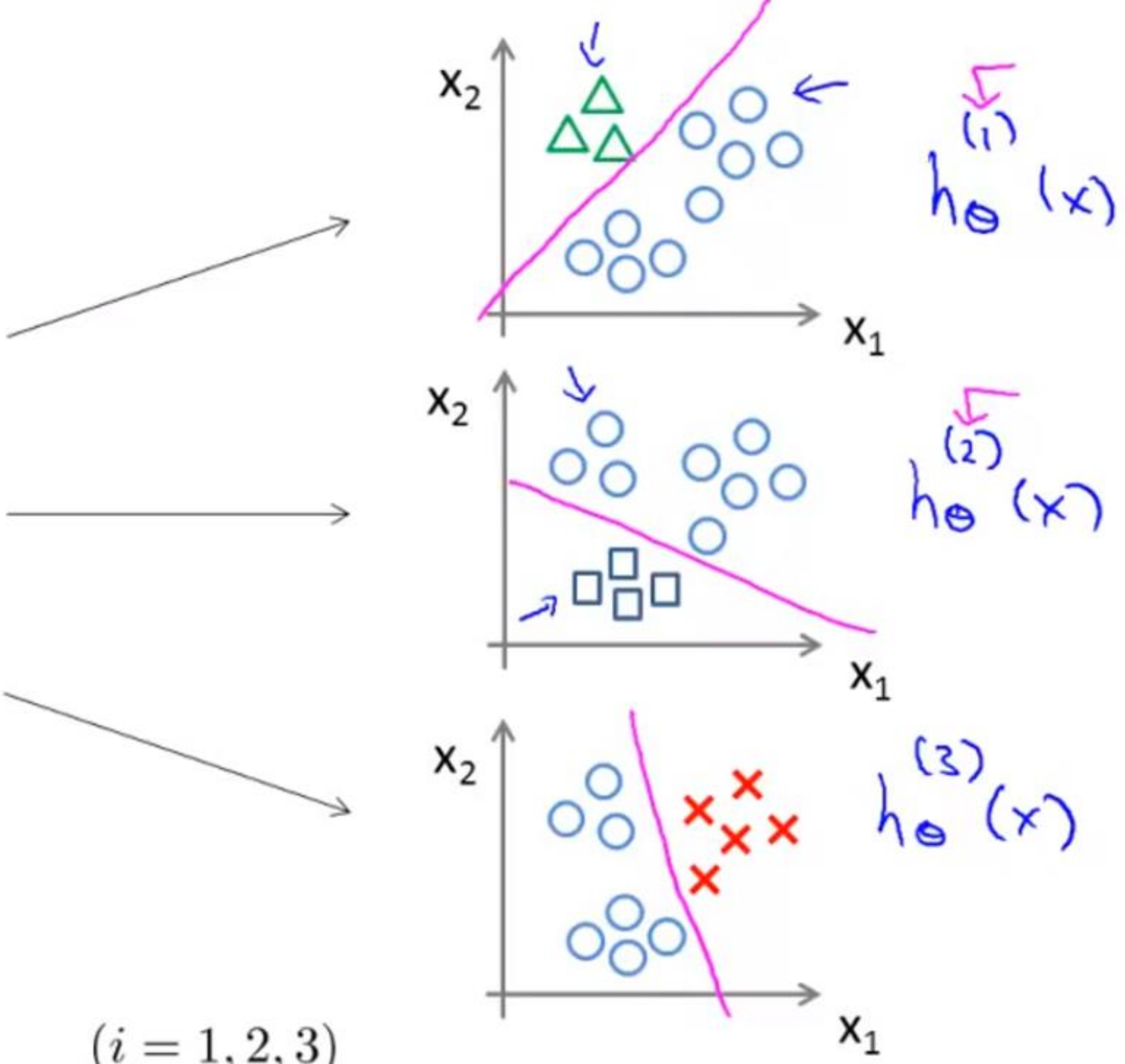


Multi-class classification:



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

Train a logistic regression classifier $h_\theta^{(i)}(x)$ 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

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

Exercise

- Suppose you have a multi-class classification problem with *k* classes $y \in \{1,2,\ldots,k\}$). Using the 1-vs.-all method, how many different logistic regression classifiers will you end up training?
  - K-1
  - K
  - K+1
  - Approximately $\log_2(k)$

Logistic regression cost function

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

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

Note: $y = 0$ or 1 always

How can we write this function in a single line?

Logistic regression cost function

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

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