Week 3

Classification

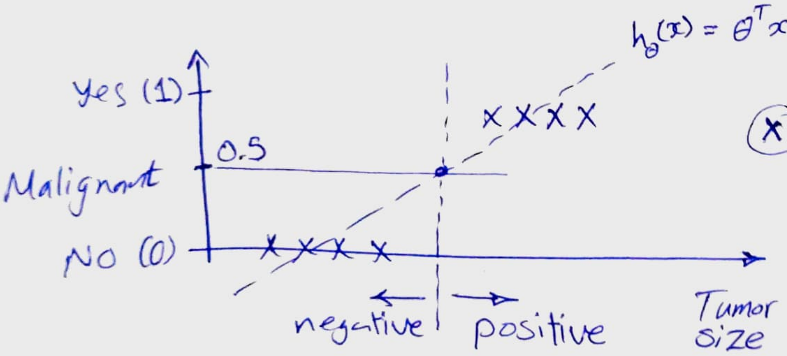
Examples :
- Email : Spam / Not spam
- Online Transaction : Fraudulent (Yes/No ?) قتال؟
- Tumor : Malignant / benign ?

Two classes → $y \in \{0, 1\}$ (Binary clf)

0 : Negative class (eg., benign tumor)
1 : Positive class (eg., malignant tumor)

The assignment of two classes to negative and positive doesn't really matter which one is positive and one is negative. but it often realized that negative class refers to the absence of something while positive class refers to the presence of that thing.

Multi_Classes $y \in \{0, 1, 2, 3\}$ → Later classes

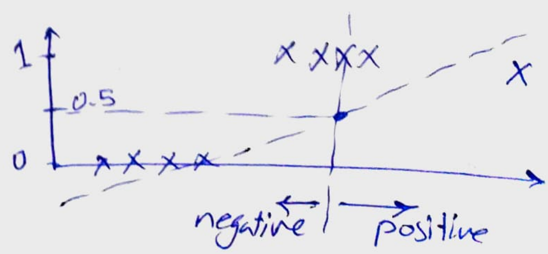


Applying linear regression on classification problem

(x) → Adding one training example doesn't actually change anything but it change $h_\theta(x)$

→ 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 "$y = 0$"

This seems a pretty bad thing for linear regression to have done. ←



So, Applying linear regression on a classification problem often is not a great idea.

- Another funny thing:

Classification : $Y = 0$ or 1

but, with Linear regression : $h_\theta(x)$ Can be > 1 or < 0

- Logistic Regression [Classification]:

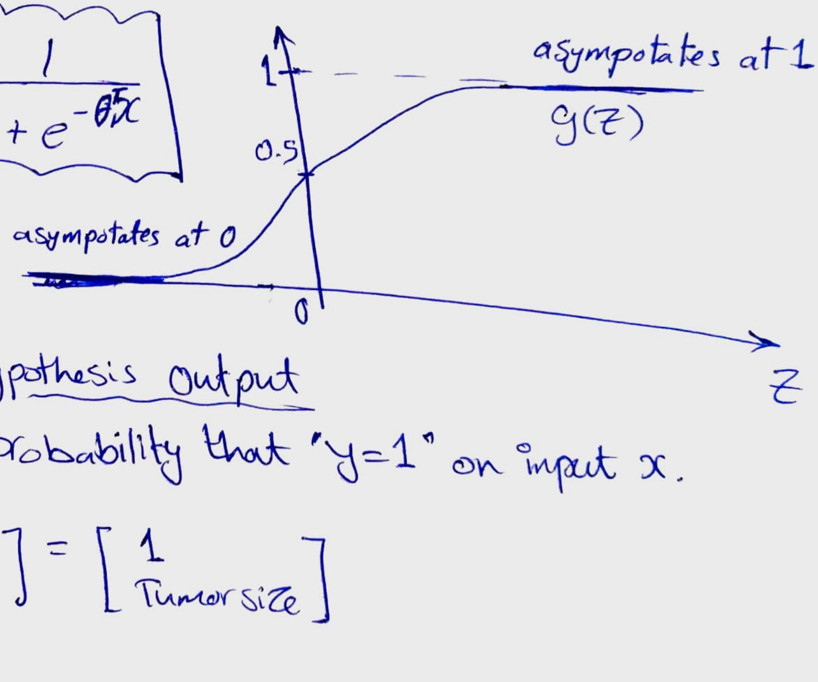
$$0 \leq h_\theta(x) \leq 1$$

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

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

Sigmoid Function
Logistic Function

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



- Interpretation of Hypothesis Output

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

Ex: if $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{Tumor size} \end{bmatrix}$

$h_\theta(x) = 0.7$

↳ That means that 70% chance of tumor being malignant

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

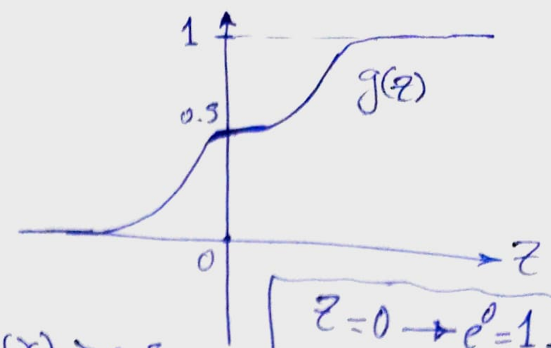
"probability that $y=1$, given x; parameterized by θ"

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

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

$\Rightarrow h_\theta(x) = g(\theta^T x) \geq 0.5$

when $\theta^T x \geq 0$



$z = 0 \rightarrow e^0 = 1 \rightarrow g = 0.5$

$z = \infty \rightarrow e^{-\infty} = 0 \rightarrow g = 1$

$z = -\infty \rightarrow e^{\infty} = \infty \rightarrow g = 0$

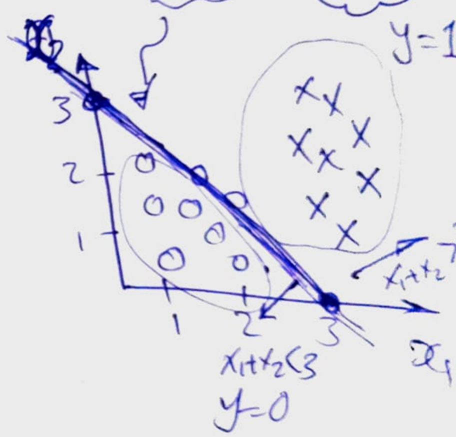
* prediction : predict "$y=1$" when $h_\theta(x) \geq 0.5$

$\boxed{\theta^T x \geq 0}$

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

$\boxed{\theta^T x < 0}$

* Decision Boundary



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

or $\boxed{\theta^T x \geq 0}$

Suppose $h_\theta(x) = -3 + x_1 + x_2$

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

so, "$y=1$" when $-3 + x_1 + x_2 \geq 0$

or $\boxed{x_1 + x_2 \geq 3}$

So, "$y=0$" if $\boxed{x_1 + x_2 < 3}$

The decision boundary $x_1 + x_2 = 3$ or $h_\theta(x) = 0.5$ separates the region where $h_\theta(x) \geq 0.5$ or "$y=1$" and the region where $h_\theta(x) < 0.5$ or "$y=0$"

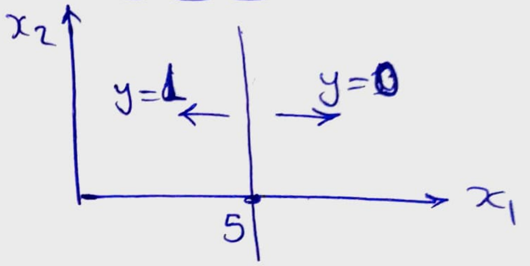
Note That

Decision Boundary is a property of the hypothesis and its parameters θ and it is not a property of the dataset

Example $\theta = \begin{bmatrix} 5 \\ -1 \\ 0 \end{bmatrix}$, $h_\theta(x) = g(5 - x_1)$

$y=1$ when $h_\theta(x) \geq 0.5$ or $\underline{\theta^T x \geq 0}$

$5 - x_1 \geq 0 \rightarrow \boxed{x_1 \leq 5}$

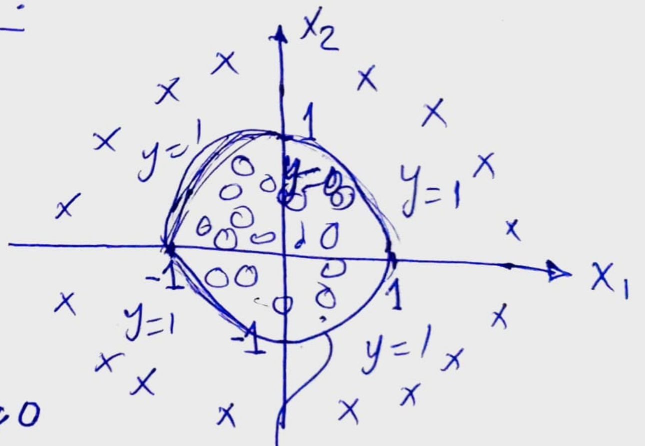


★ Non-linear Decision Boundary:

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

let $\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ → predict "$y=1$": if $-1 + x_1^2 + x_2^2 \geq 0$

→ $\boxed{x_1^2 + x_2^2 \geq 1}$



A circle equ
$x_1^2 + x_2^2 = 1$

→ More Complex