Session 4

Logistic Regression

AI2YOU

Virtual Learning

Supervised Learning → Regression

Supervised Learning → Classification → Email Spam

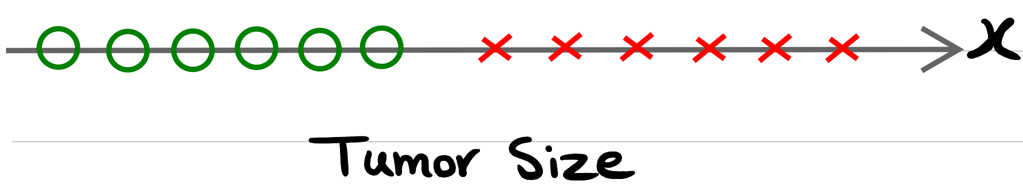
Transaction Fraudulent

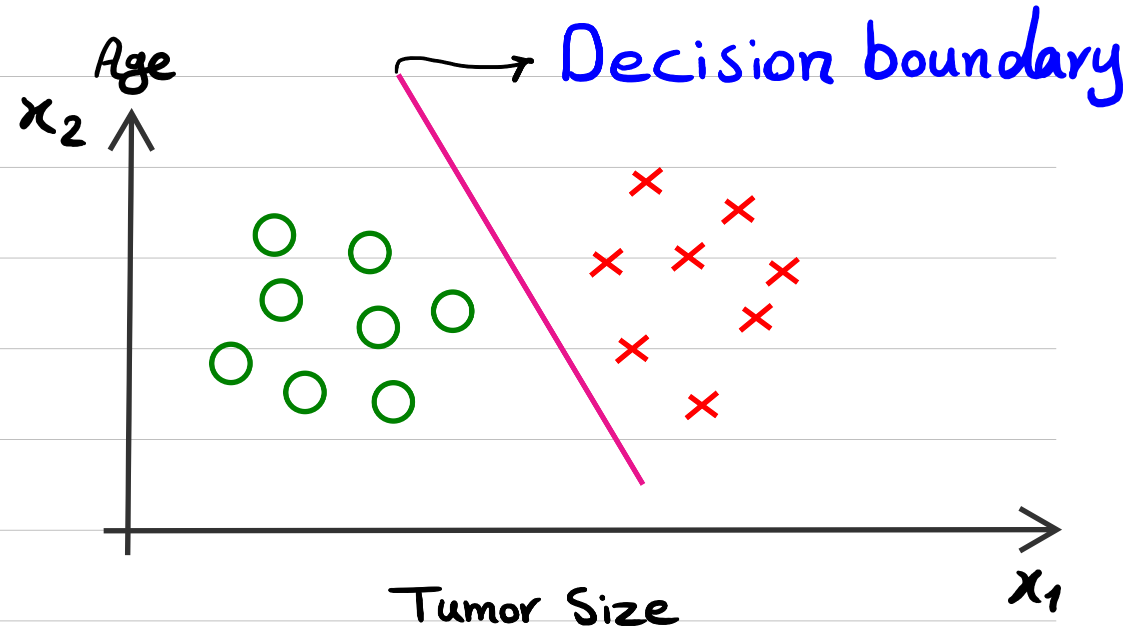
Tumor Maligant / cancer

Classification: Breast Cancer Detection

× Malignant بدخیم

○ Being خوش‌خیم





تفاوت بین classification و Regression چیست؟

در classification ما سعی داریم تعداد کمی از خروجی ها را پیش بینی کنیم. به عبارتی اعداد محدودی داریم.

$y \rightarrow 0$

$y \rightarrow 1$

مثل اینجا که فقط دو خروجی 1 (خوش خیم) و 0 (بدخیم) را داریم.

اما در Regression ما سعی داریم هر عددی را پیش بینی کنیم.

$$y \in [0, +\infty)$$

* در classification ما می توانیم بیش از دو عدد داشته باشیم مثلاً 4 ,3 ,2 ,1 ,0 اما نمی توانیم بگوییم تمام اعداد بین 0 تا 1 ($0, 0.0001, 0.2, \ldots, 1$).

$y \rightarrow$ 1 (Tumor Type 1)

$y \rightarrow$ 2 (Tumor Type 2)

$y \rightarrow$ 3 (Tumor Type 3)

Clasification :

Question	Answer	
Is this Email Spam?	No	Yes
IS the transaction fraudulent?	No	Yes
Is the tumor malignant?	No	Yes

two classes → Binary Classification

category

False   True

0   1

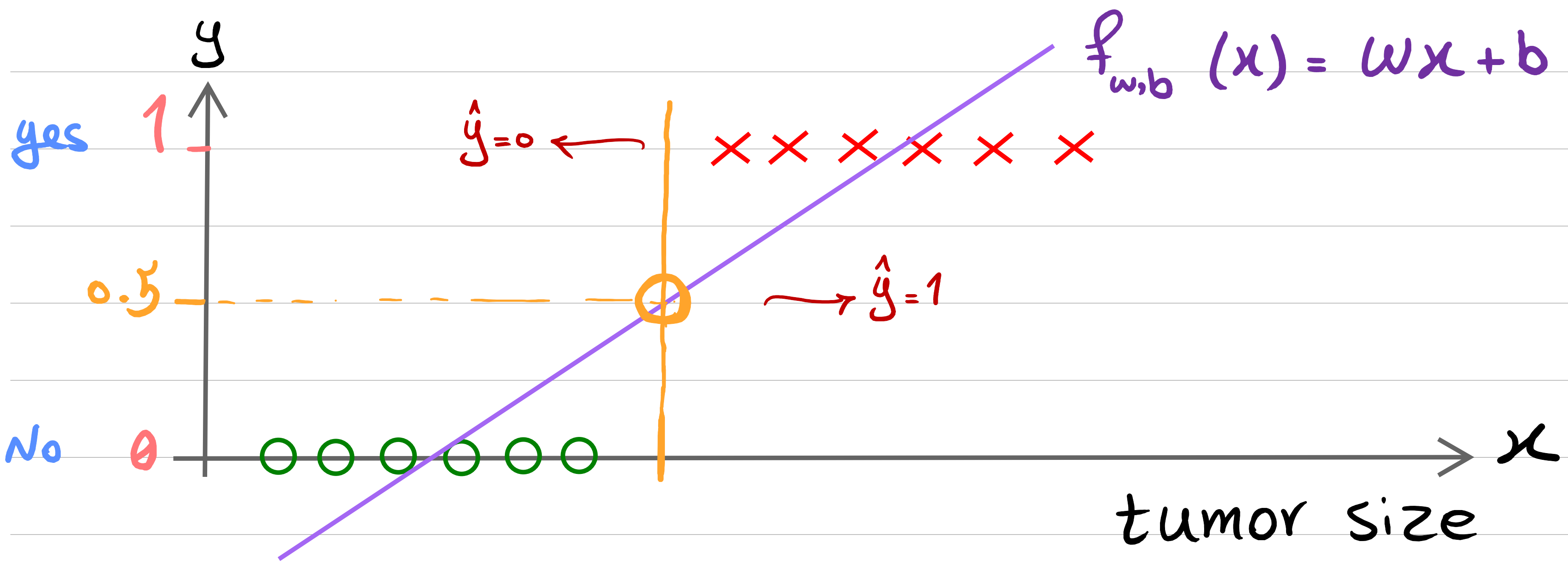
Negative class   Positive class

اگر بخواهیم به این سؤال پاسخ دهیم که "آیا تومور سرطانی ما بدخیم است؟" چه باید کرد؟

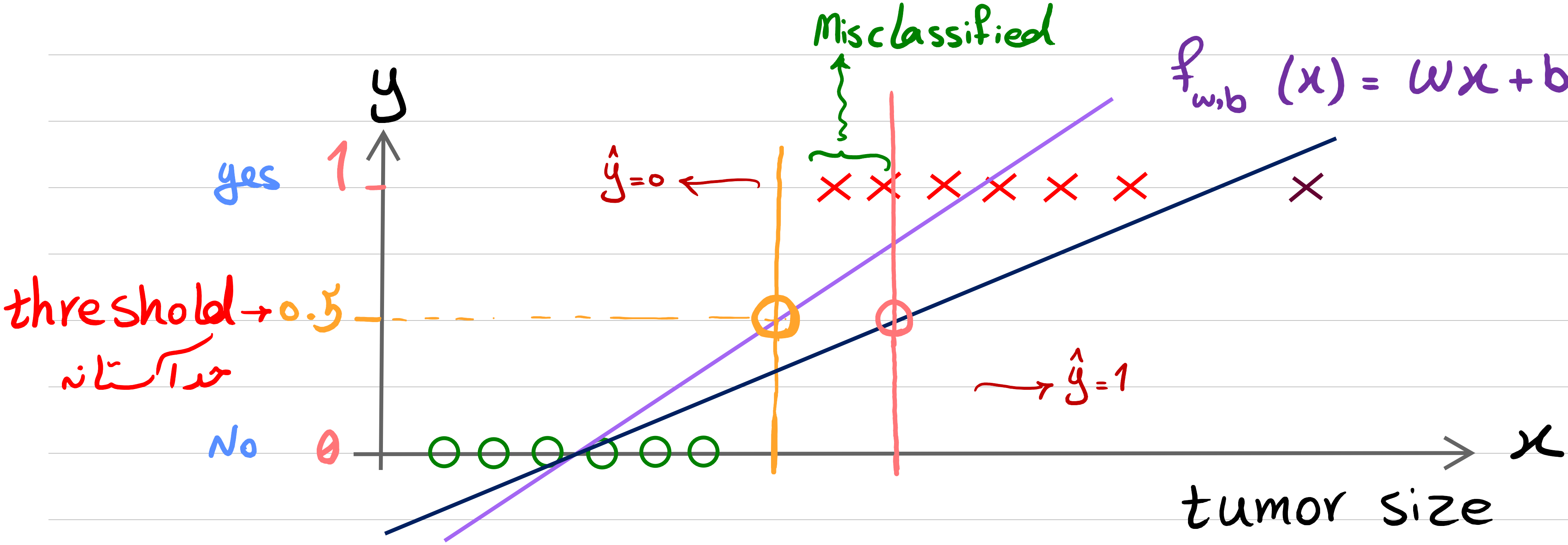
ج: چه طور داده هایمان را کلاسه بندی کنیم؟ ← classification

× Malignant بدخیم

o Being خوش خیم



$$
\text{if} \begin{cases} f_{w,b}(x) < 0.5 & \hat{y}=0 \\ f_{w,b}(x) \geq 0.5 & \hat{y}=1 \end{cases}
$$



؟: آیا می‌توان از Regression برای Classification استفاده کرد؟

مشکلات استفاده از Regression:

1. Misclassified

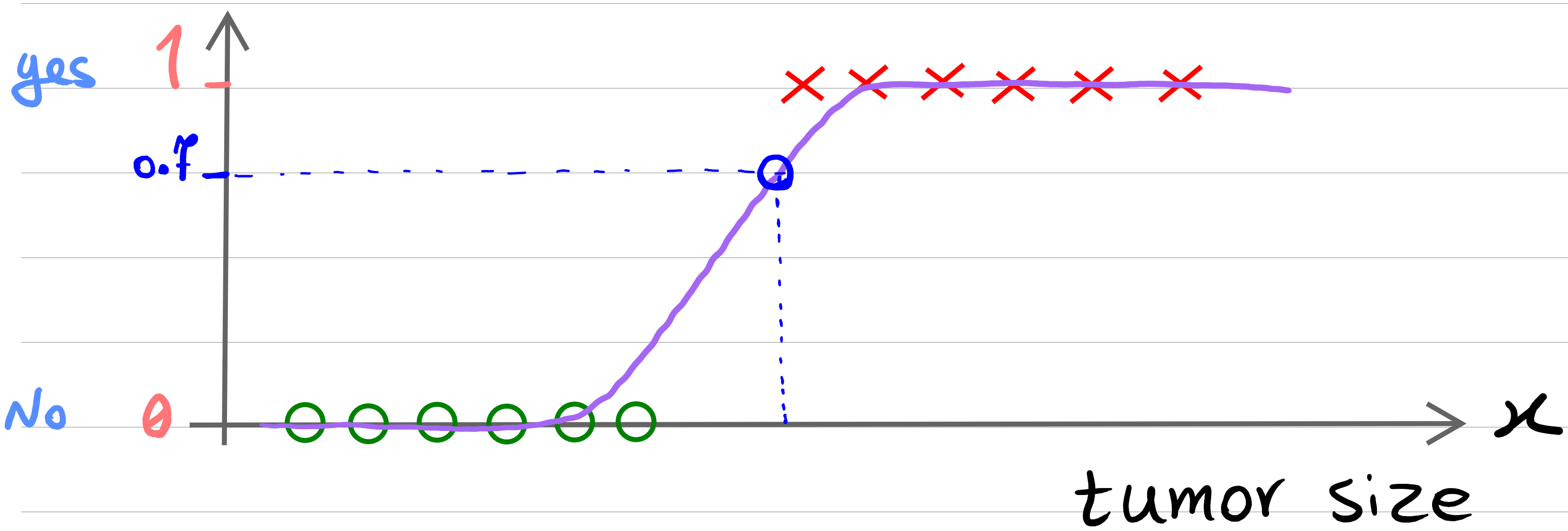
2. $F_{w,b}(x) \in (-\infty, +\infty)$

Logistic Regression Model Equation: $\log\left(\frac{P}{1-P}\right) = W_1x_1 + W_2x_2 + \dots + W_nx_n + b$

P: احتمال وقوع $y=1$

$1-P$: احتمال عدم وقوع $y=1$ یا احتمال وقوع $y=0$

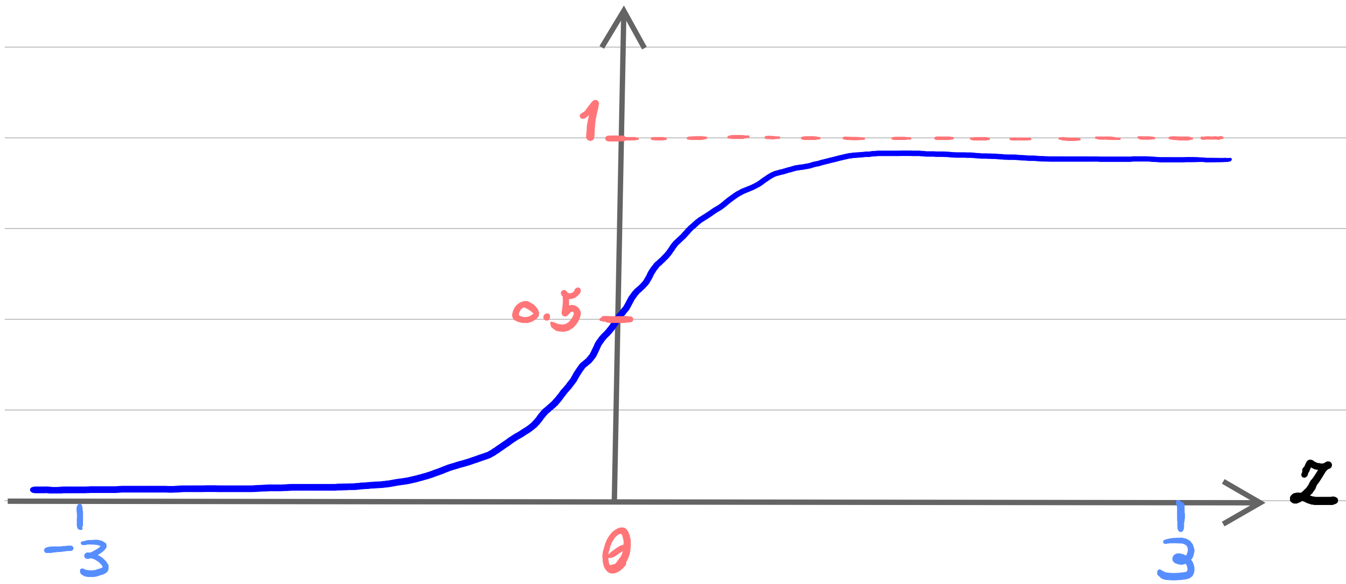
ج: برای رفع مشکلات Regression چه باید کرد؟



از تابع Sigmoid استفاده می‌کنیم. Sigmoid تابعی است که هر عددی را به عنوان ورودی بگیرد در خروجی به ما احتمالش (یعنی عددی بین ۰ و 1) را می‌دهد.

Sigmoid Function:

$0 < \text{outputs} < 1$



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

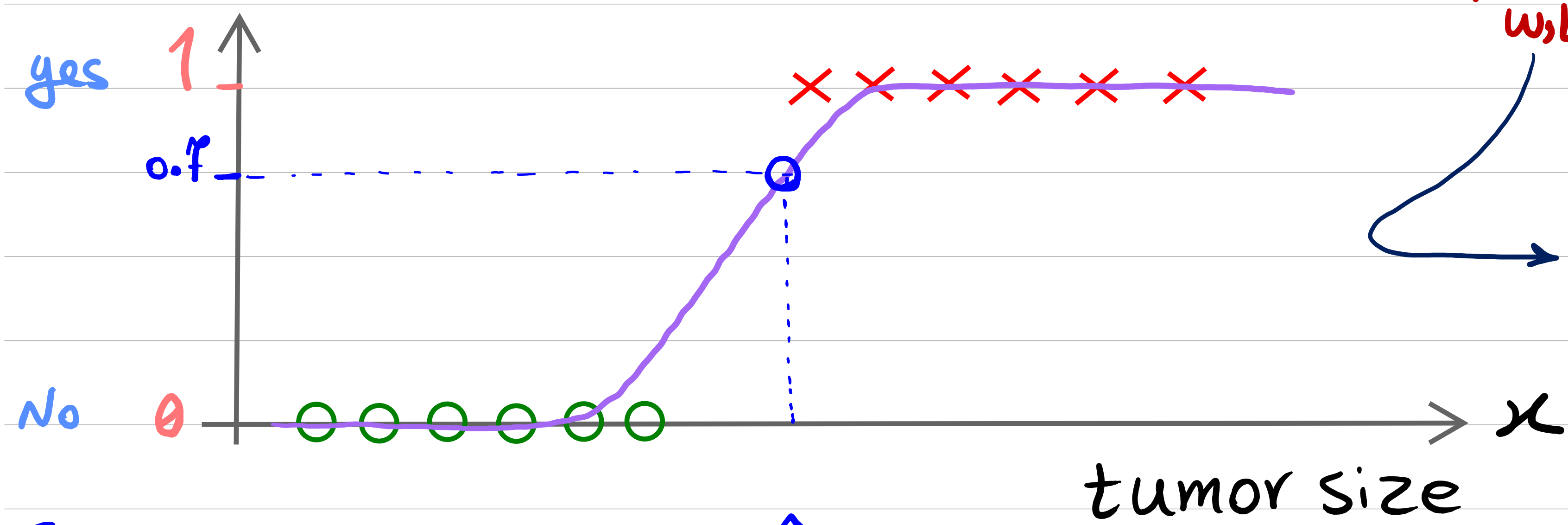
$$0 < g(z) < 1$$

$$f_{w,b}(x) = wx + b \rightsquigarrow z = wx + b \xrightarrow{\text{Sigmoid}} g(z)$$

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

$\boxed{\text{Logistic Regression}}$

$$f_{w,b}(x) = g(wx+b) = \frac{1}{1+e^{-(wx+b)}}$$



$$F_{w,b}(x) = \frac{1}{1 + e^{-(wx+b)}}$$

Probability that class is 1

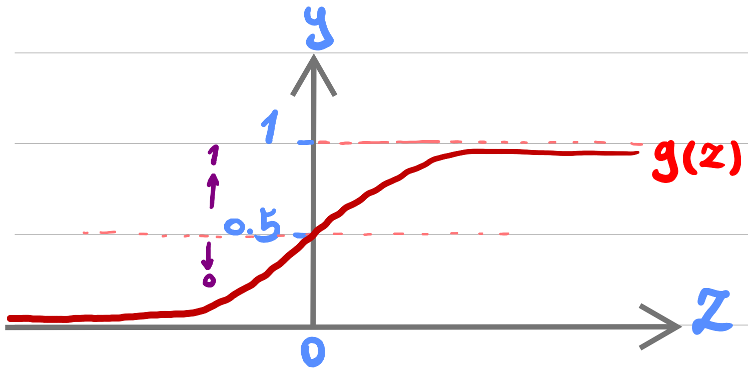
$F_{w,b}(x) = 0.7 \rightsquigarrow \hat{y} = ?$

$F_{w,b}(x) = P(y=1 \mid x ; w,b) \longrightarrow$ احتمال اینکه $\hat{y}=1$ بشود به‌ازای x و با
پارامترهای w,b

$P(y=0) + P(y=1) = 1$

$$\begin{cases} F_{w,b}(x) \geq 0.5 & \hat{y} = 1 \\ F_{w,b}(x) < 0.5 & \hat{y} = 0 \end{cases}$$

$$\underbrace{P(y=0)}_{0.3} + \underbrace{P(y=1)}_{0.7} = 1$$



$$z = w.x + b$$

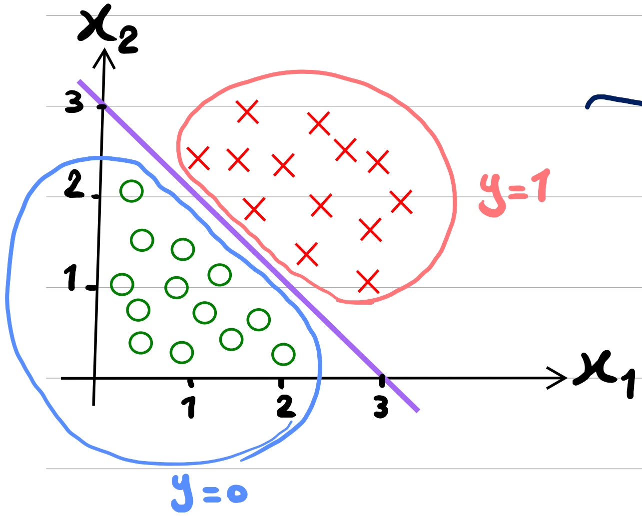
$$z$$

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

$$F_{\vec{w},b}(\vec{x}) = g(\vec{w}.\vec{x} + b) = \frac{1}{1+e^{-(\vec{w}.\vec{x}+b)}} = P(y=1 \mid x; \vec{w}, b)$$

Decision boundary

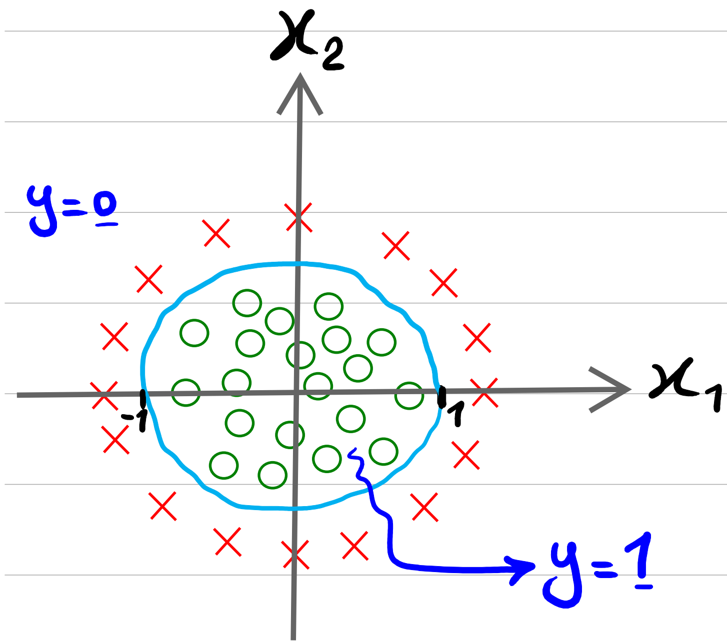
حالا می‌خواهیم بریم سراغ حالتی که مسأله‌ی classification ما دو ویژگی (x_1, x_2) دارد.



$$F_{\vec{w},b}(\vec{x}) = g(z) = g(w_1x_1 + w_2x_2 + b)$$

$$z = \vec{w}\cdot\vec{x} + b = 0 \qquad \text{if } \; w_1=1,\; w_2=1,\; b=-3$$

$$w_1x_1 + w_2x_2 + b = 0 \rightsquigarrow x_1 + x_2 - 3 = 0 \rightsquigarrow \begin{cases} x_1=0, & x_2=3 \\ x_1=3, & x_2=0 \end{cases}$$



$$z = x_1^2 + x_2^2 - 1 = 0$$

$$\rightarrow x_1^2 + x_2^2 = 1$$

$$\boxed{\begin{aligned} &\text{if } x_1 = 0 \\ &\cancel{w_1x_1} + w_2x_2 + b = 0 \\ &\Rightarrow w_2x_2 = -b \Rightarrow \\ &x_2 = \frac{-b}{w_2} = \frac{-(-3)}{1} = 3 \end{aligned}}$$

tumor size x_1	...	Patient's age x_n	malignant? y
10		52	1
2		73	0
5		55	0
12		49	1
⋮		⋮	⋮

$i = 1, \dots, m$

$j = 1, \dots, n$

target y is $\underline{0}$ or $\underline{1}$

$$F_{w,b}(x) = \frac{1}{1+e^{-(wx+b)}}$$

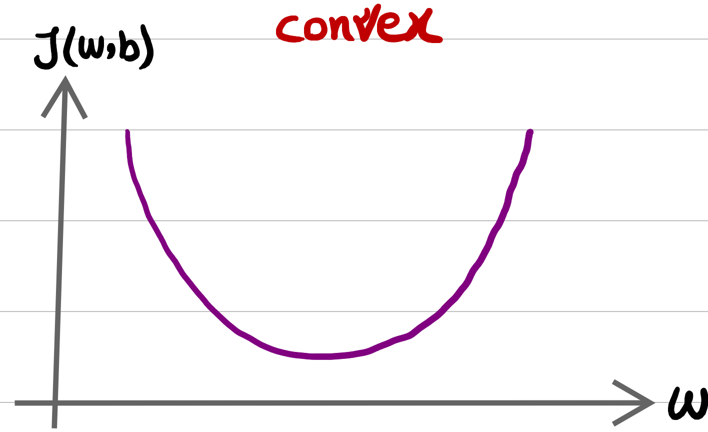
$$\vec{w} = [w_1 \ w_2 \ \cdots \ w_n], \quad b$$

Linear Regression

Cost Function ← چه طور بهترین w و b را انتخاب کنیم؟

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{F_{w,b}(x^{(i)})}_{\hat{y}^{(i)}} - y^{(i)} \Big)^2$$

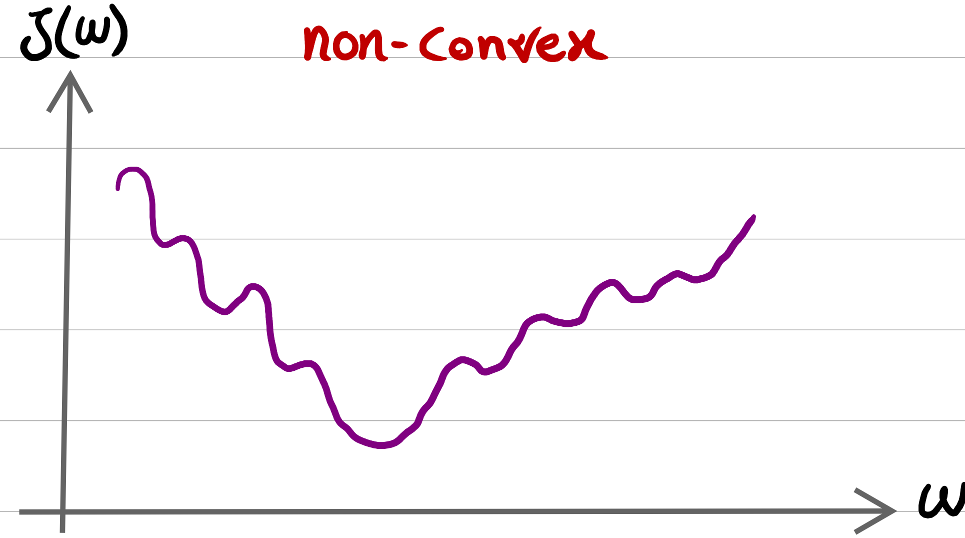
convex



$$F_{w,b}(x) = w.x + b$$

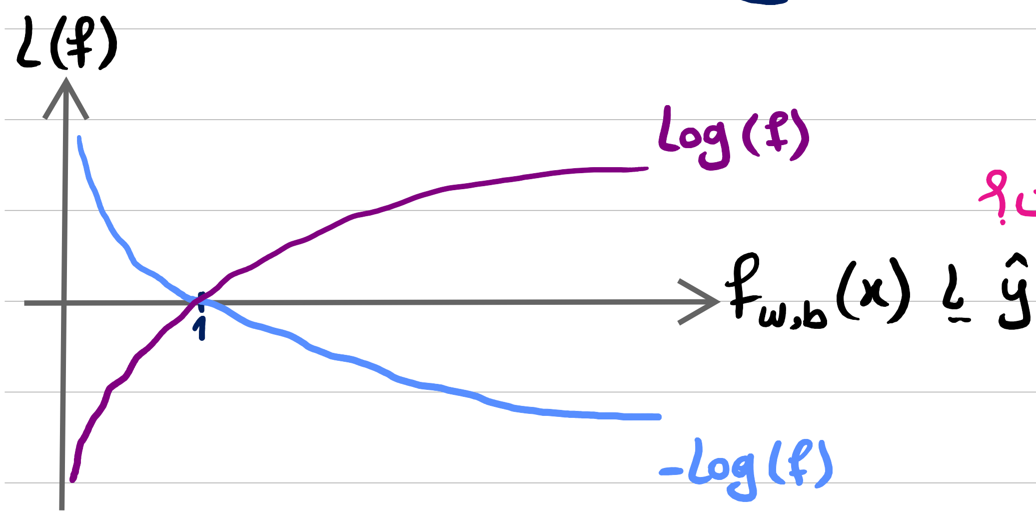
Logistic Regression

$$f_{w,b}(x) = \frac{1}{1+e^{-(wx+b)}}$$



Logistic Loss Function

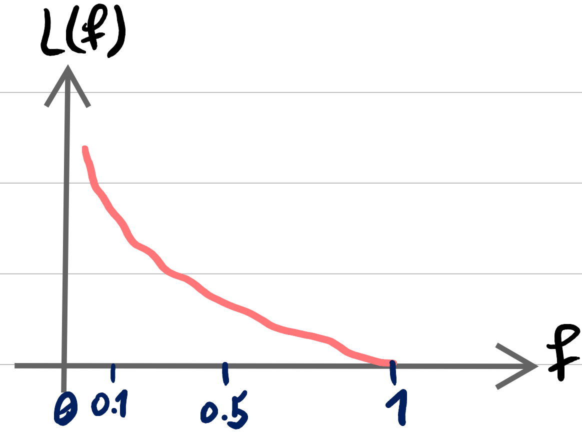
$$L\left(f_{w,b}(x^{(i)}),\ y^{(i)}\right) = \begin{cases} -\log\left(f_{w,b}(x^{(i)})\right) & \text{if } y^{(i)}=1 \\ -\log\left(1-f_{w,b}(x^{(i)})\right) & \text{if } y^{(i)}=0 \end{cases}$$



ج: آیا استفاده از این تابع برای Logistic درست است؟

$L(f_{w,b}(x^{(i)}), y^{(i)})$

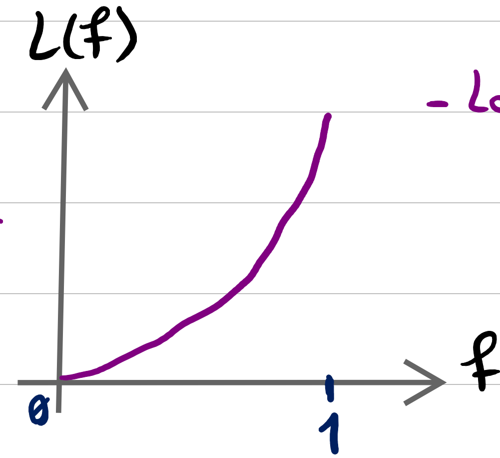
if $y=1$



$\hookrightarrow$ if $f_{w,b}(x^{(i)}) \to 1$ then $\text{Loss} \to 0$

but if $f_{w,b}(x^{(i)}) \to 0$ then $\text{Loss} \to \infty$

$-\log(1 - f_{w,b}(x^{(i)}))$ $\boxed{\text{if } y^{(i)} = 0}$



$-\log(1-f) \rightsquigarrow [0,1]$

if $f_{w,b}(x^{(i)}) \to 0$ then $\text{Loss} \to 0$

$f_{w,b}(x^{(i)}) \to 1$ then $\text{Loss} \to \infty$

Logistic Regression Cost function:

$$J(w,b) = \frac{1}{m} \sum_{i=1}^{m} L\left(f_{w,b}(x^{(i)}), y^{(i)}\right)$$

$$\begin{cases} -\log\left(f_{w,b}(x^{(i)})\right) & \text{if } y^{(i)}=1 \\ -\log\left(1-f_{w,b}(x^{(i)})\right) & \text{if } y^{(i)}=0 \end{cases}$$

$$L\left(f_{w,b}(x^{(i)}),\ y^{(i)}\right) = \begin{cases} -\log\left(f_{w,b}(x^{(i)})\right) & \text{if } y^{(i)}=1 \\ -\log\left(1-f_{w,b}(x^{(i)})\right) & \text{if } y^{(i)}=0 \end{cases}$$

چ: حالا اگر یک تابع یکپارچه بخواهیم باید چه کنیم؟

$$L\left(f_{w,b}(x^{(i)}),\ y^{(i)}\right) = \overbrace{-y^{(i)}\log\left(f_{w,b}(x^{(i)})\right)-\left(1-y^{(i)}\right)\log\left(1-f_{w,b}(x^{(i)})\right)}^{\text{Convex}}$$

if $y^{(i)}=1$:

$$-(1)\log\left(f_{w,b}(x^{(i)})\right)-\cancelto{0}{(1-(1))\log\left(1-f_{w,b}(x^{(i)})\right)}$$

if $y^{(i)}=0$:

$$\cancelto{0}{-(0)\log\left(f_{w,b}(x^{(i)})\right)}-(1-(0))\log\left(1-f_{w,b}(x^{(i)})\right)$$

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left[L\left(f_{w,b}(x^{(i)}),\ y^{(i)}\right)\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\underbrace{\left[y^{(i)}\log\left(f(x^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{w,b}(x^{(i)})\right)\right]}_{\text{Log likelihood}}$$

حالا ما باید به دنبال w و b باشیم که J cost را minimum کنند.

ج: چه طور مقدار w و b را پیدا کنیم؟

دو روش رایج داریم:

1. Newton's Method

2. Gradient Descent

Gradient Descent :

$$J(w,b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\,\mathrm{Log}\left(f_{w,b}(x^{(i)})\right) + (1-y^{(i)})\,\mathrm{Log}\left(1-f_{w,b}(x^{(i)})\right)\right]$$

repeat

{

$$w_j = w_j - \alpha \boxed{\frac{\partial}{\partial w_j} J(w,b)} \rightsquigarrow \frac{\partial}{\partial w} J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

$$b = b - \alpha \boxed{\frac{\partial}{\partial b} J(w,b)} \rightsquigarrow \frac{\partial}{\partial b} J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)$$

}

$$f_{w,b}(x) = \frac{1}{1+e^{-(wx+b)}}$$