

Session 05

Logistic Regression

Machine Learning | Zahra Amini



Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

<https://zil.ink/zahraamini>

Gradient Descent

Stochastic Gradient Descent (SGD)

در این روش، گرادیان برای هر نمونه داده به صورت جداگانه محاسبه می‌شود

$$W_{\text{new}} = W_{\text{old}} - \alpha \frac{\delta L}{\delta W_{\text{old}}}$$

زهراءامینی

@zahraamini_ai

Batch Gradient Descent

در این روش، گرادیان کل مجموعه داده برای هر گام محاسبه می‌شود

Mini-batch Gradient Descent

این روش ترکیبی از دو روش قبلی است، به این صورت که گرادیان برای دسته‌های کوچکی از داده‌ها (مینی‌بچ‌ها) محاسبه می‌شود

یک دوره زمانی است که الگوریتم یادگیری ماشین یک بار کامل بر روی کل مجموعه داده‌ها آموزش می‌بیند

Epoch

یک بچ مجموعه‌ای از نمونه‌های داده است که در یک مرحله از الگوریتم گرادیان دیسنست استفاده می‌شود

Batch

Stochastic Gradient Descent (SGD)

```
# Set number of epochs and constant learning rate
n_epochs = 50
alpha = 0.1

# Initialize parameters randomly
w = np.random.randn(2,1)

# Training loop
for epoch in range(n_epochs):
    for i in range(m): # m should be defined as the number of samples in the dataset
        random_index = np.random.randint(m) # Select a random index
        xi = X_b[random_index:random_index+1] # Extract features for selected sample
        yi = y[random_index:random_index+1] # Extract target for selected sample

        # Compute gradient of loss function
        gradients = 2 * xi.T.dot(xi.dot(w) - yi)

        # Update parameters
        w = w - alpha * gradients # Apply gradient descent step
```

Batch Gradient Descent

```
n_epochs = 50
alpha = 0.1 # learning rate

w = np.random.randn(2,1) # random initialization

for epoch in range(n_epochs):
    gradients = 2/m * X_b.T.dot(X_b.dot(w) - y)
    w = w - alpha * gradients
```

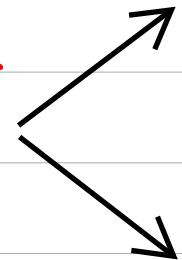
Mini-batch Gradient Descent

```
n_epochs = 50  
batch_size = 20 # size of mini-batch  
alpha = 0.1 # learning rate  
  
w = np.random.randn(2,1) # random initialization  
  
for epoch in range(n_epochs):  
    shuffled_indices = np.random.permutation(m)  
    X_b_shuffled = X_b[shuffled_indices]  
    y_shuffled = y[shuffled_indices]  
    for i in range(0, m, batch_size):  
        xi = X_b_shuffled[i:i+batch_size]  
        yi = y_shuffled[i:i+batch_size]  
        gradients = 2/batch_size * xi.T.dot(xi.dot(w) - yi)  
        w = w - alpha * gradients
```



Supervised
learning

Regression



Classification → Email Spam



Classification: Breast Cancer Detection

✗ Malignant بُرخا

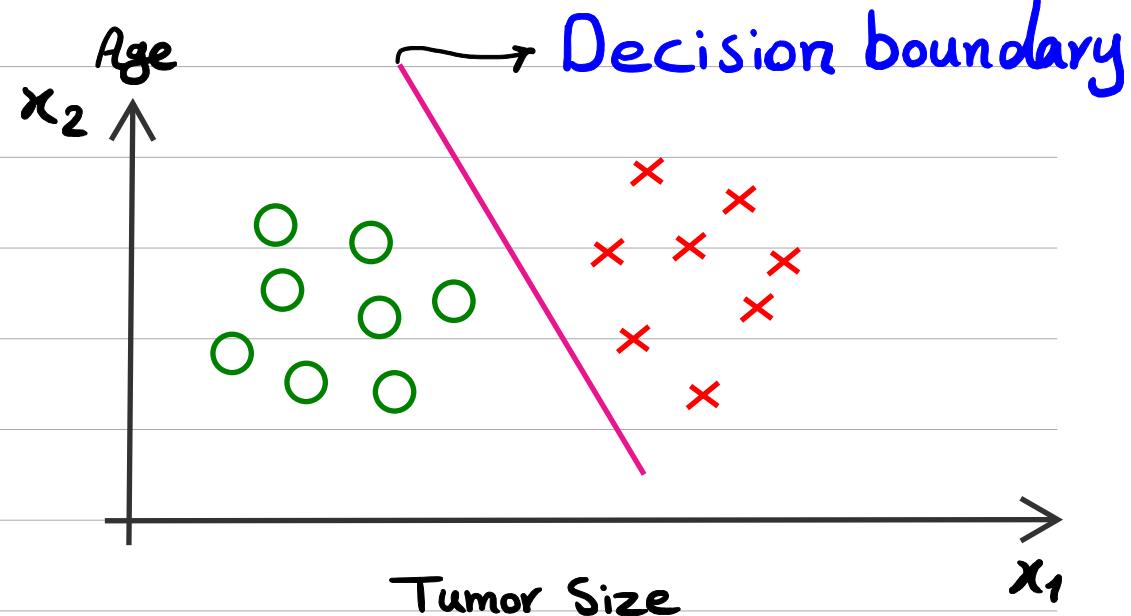
○ Benign خوبش



Tumor Size

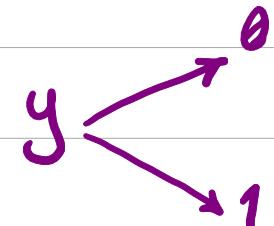
Transaction Fraudulent

Tumor Malignant / Cancer



جیسا کہ اسے جانتے ہیں Regression و classification

در classification ممکن است تعدادی از خروجی‌ها را پیش‌بینی کنیم. به عبارتی اعداد خروجی داریم.

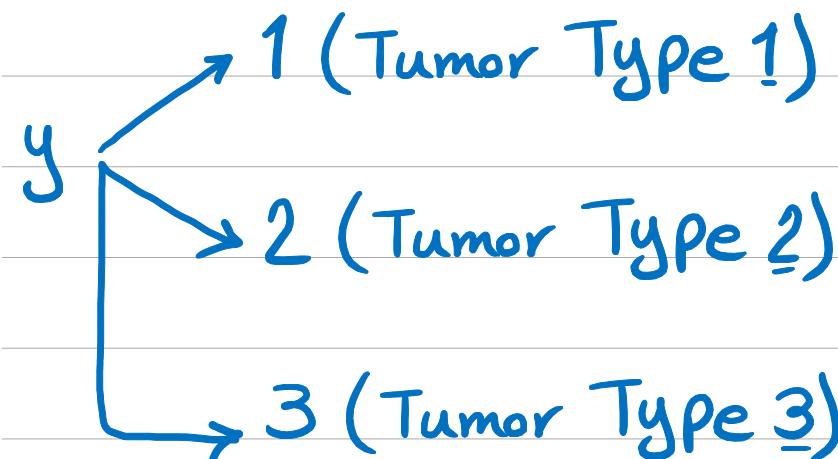


مثل اینجا که فقط دو خردمندی (خوش خشم) و ۰ (بدخشم) را داریم.

اما در Regression مابینی داریم هر عددی را پیش بینی کنیم.

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

* توانیم بیش از دو عدد را نه باشیم مثلاً $0, 1, 2, 3, 4$ اما نمی‌توانیم classification



بلوچیم تمام اعداد بین 0 تا 1 ($0, 0.0001, 0.2, \dots, 1$) .

Classification :

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

$\frac{0}{1}$

Negative class

$\frac{1}{1}$

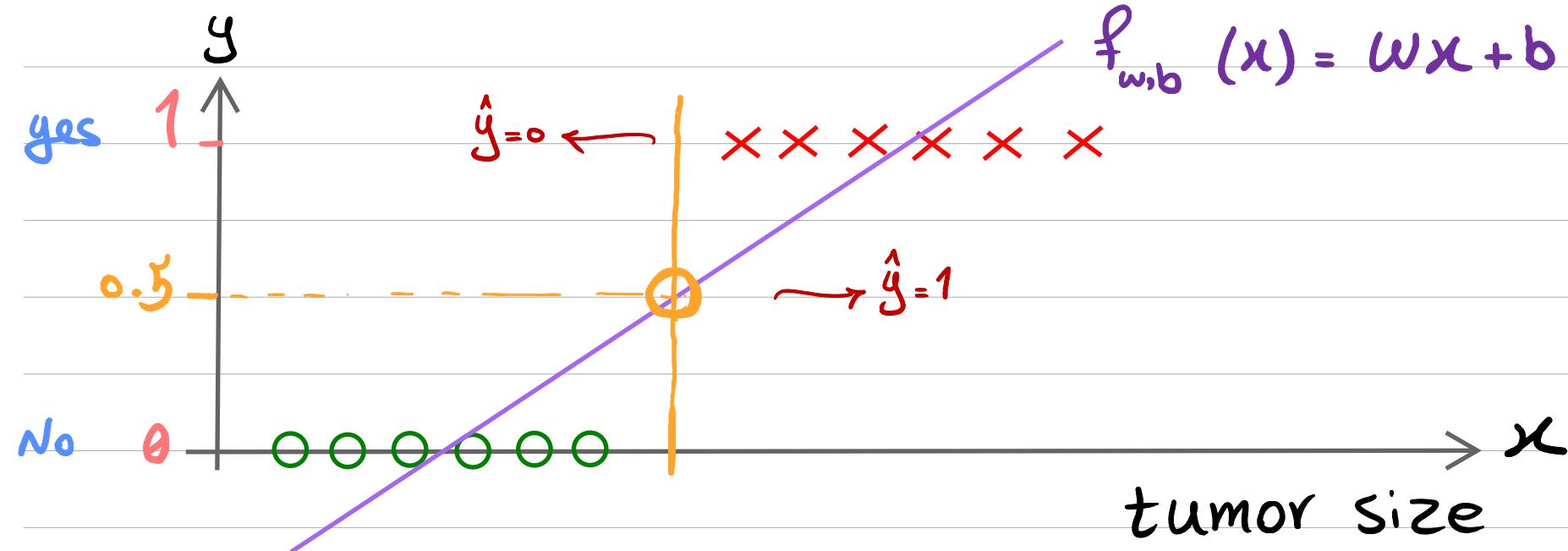
Positive class

اگر بخواهیم بین سؤال پاسخ دهیم که "آیا تومور سرطانی مبتداست یا" چه باید کرد؟

Malignant بدشیم

Being خوششیم

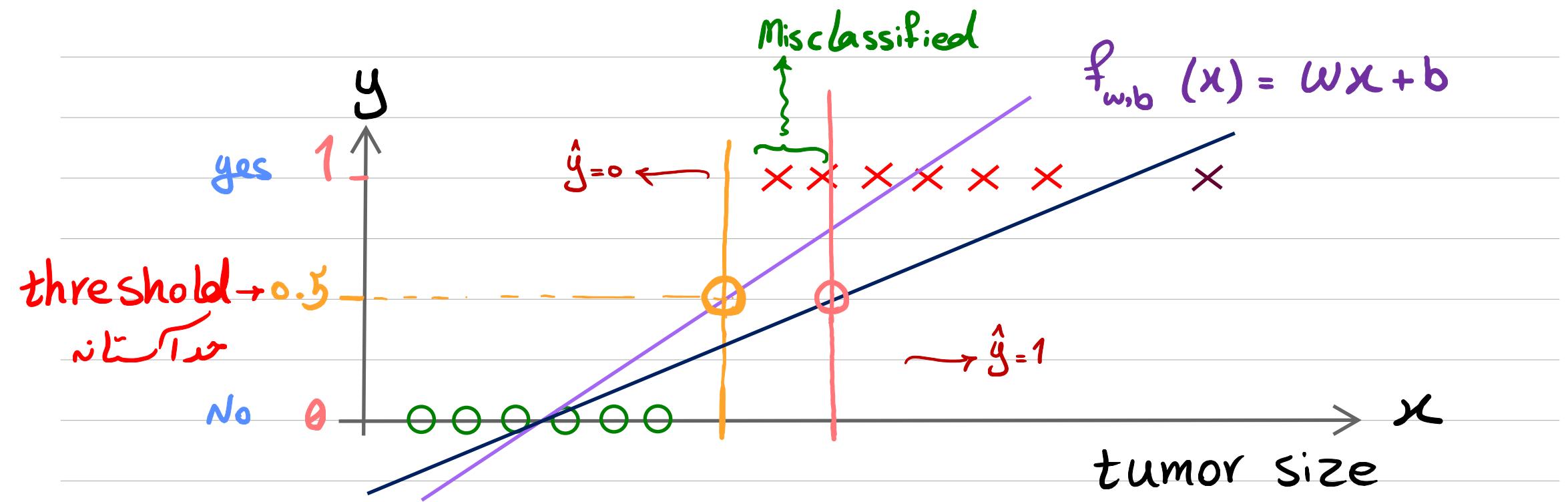
ج: چه طور داده هایمان را کلاس بندی کنیم؟ classification



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

زهرا امینی

@zahraamini_ai



1. Misclassified

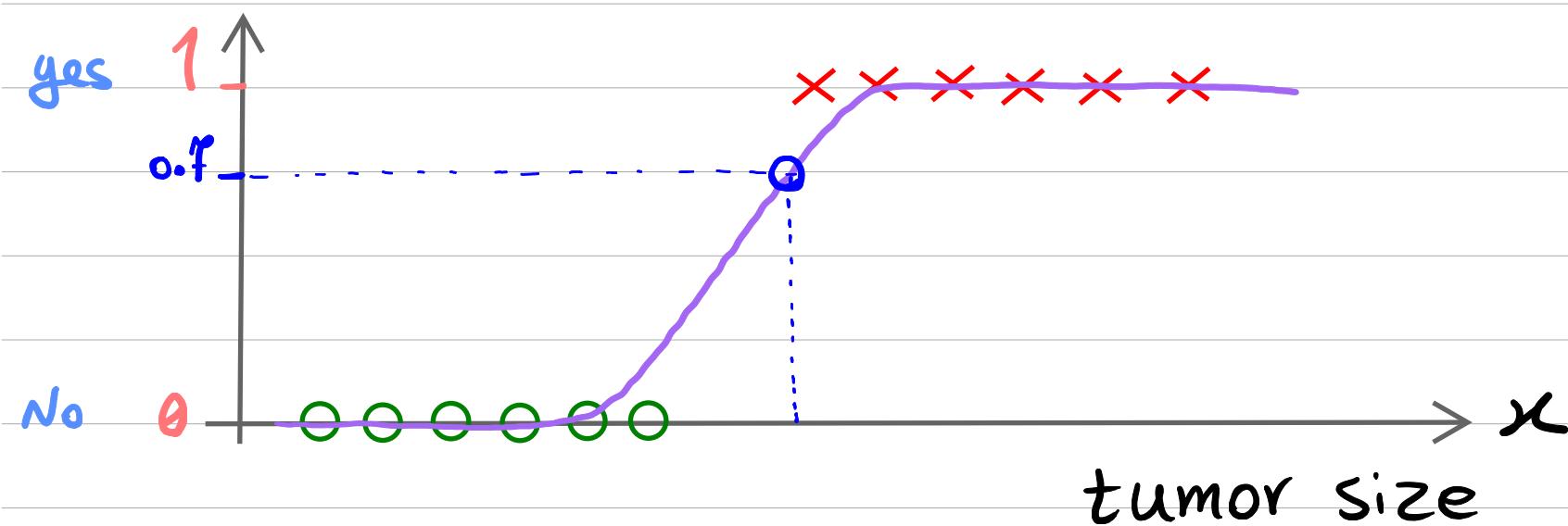
2. $F_{w,b}(x) \in (-\infty, +\infty)$

زهرا امینی
@zahraamini.ai

نکات استفاده از Regression

۸: آیا می توان از Classification برای Regression استفاده نمود؟

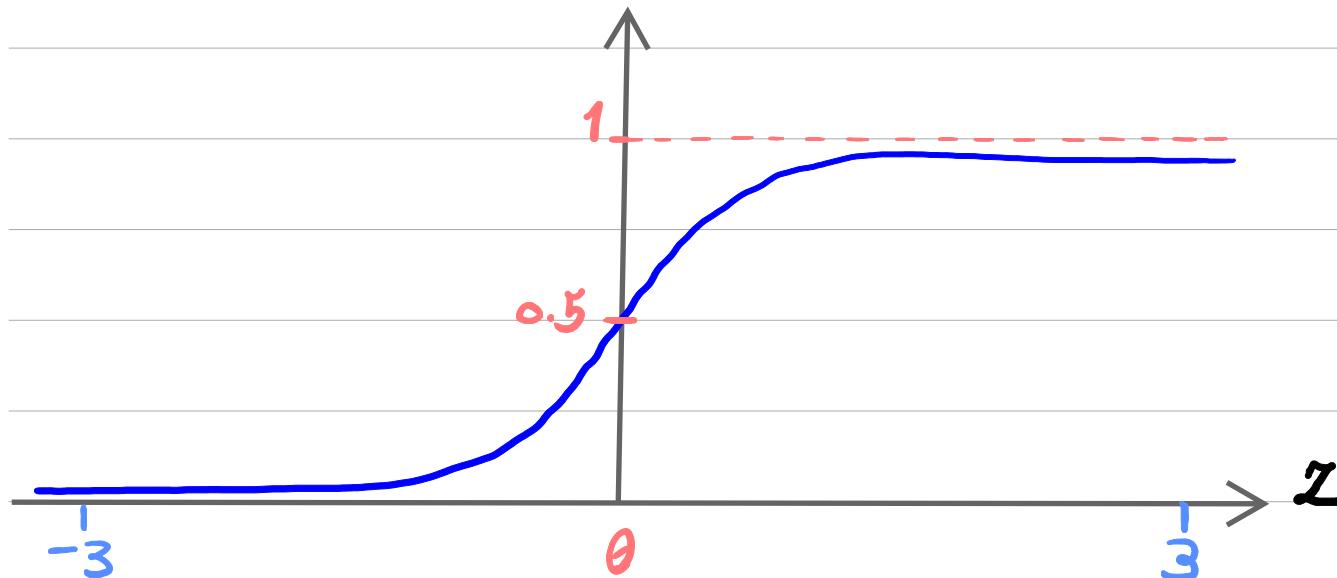
۲. برای رفع مشکلات چه باید نمود؟ Regression



از تابع Sigmoid استفاده می‌کنیم. Sigmoid تابعی است که هر عددی را به مسیر در درجی به ماقامش (یعنی عددی بین ۰ و ۱) راهی می‌دارد.

Sigmoid Function:

0 < outputs < 1



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

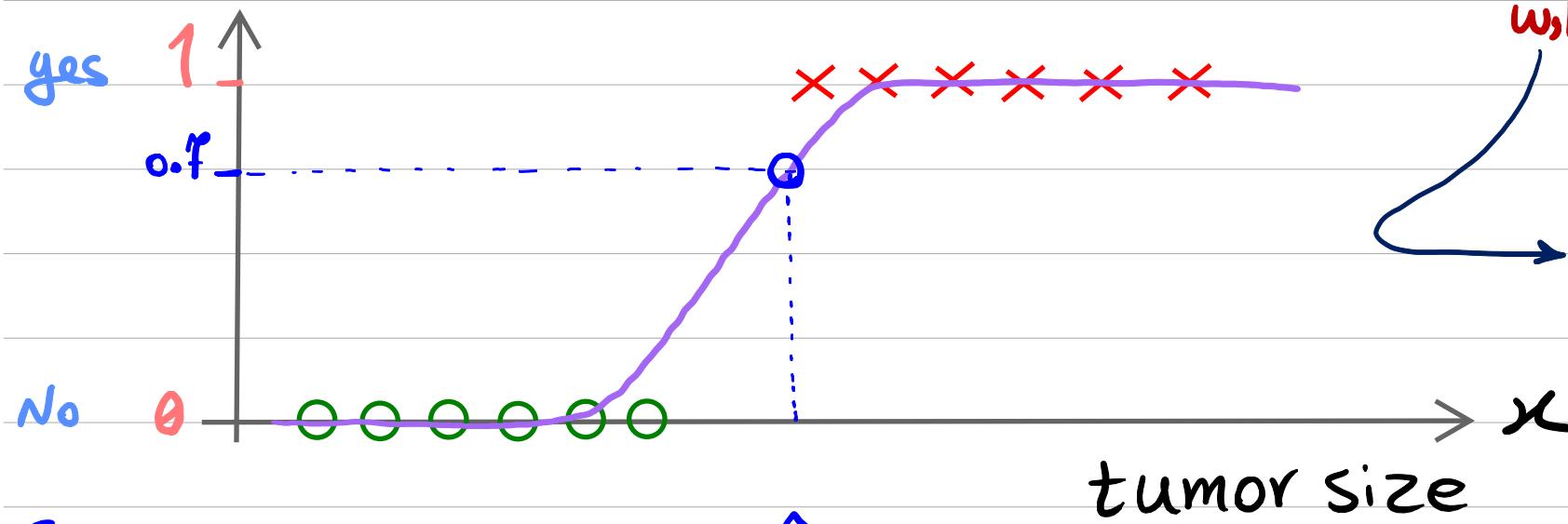
0 < g(z) < 1

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

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

Logistic Regression

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



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

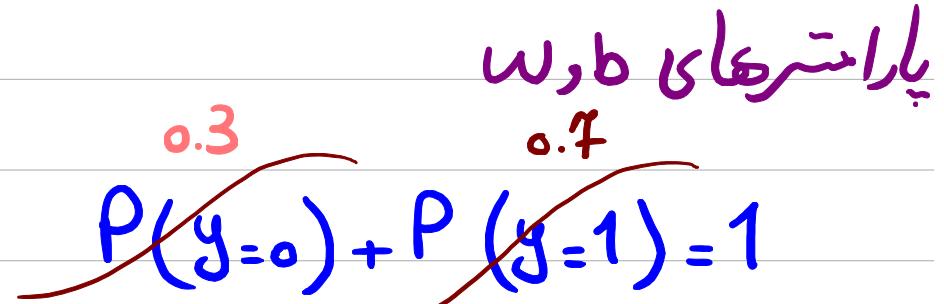
Probability that
class is 1

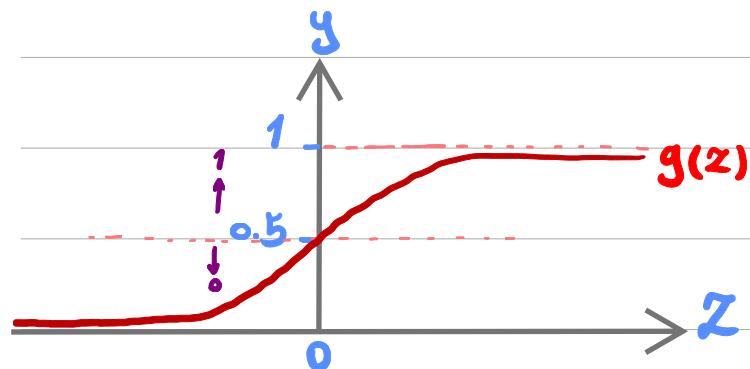
$$F_{w,b}(x) = 0.7 \rightsquigarrow \hat{y} = ?$$

$$f_{w,b}(x) = P(y=1 | x; w, b) \rightarrow \text{احتمال اینکه } \hat{y}=1 \text{ بشود بازای } x \text{ داده شود}$$

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

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





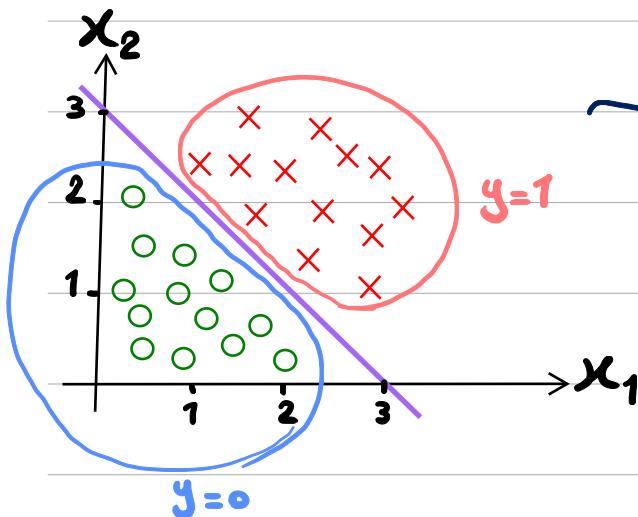
$$z = w \cdot x + b$$

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

$$f_{\vec{w}, b}(\vec{x}) = g(\vec{w} \cdot \vec{x} + b) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}} = P(y=1 | x; \vec{w}, b)$$

Decision boundary

حالات خواهیم بود که سایر حالتی را که دو دسته (classification) دارد.

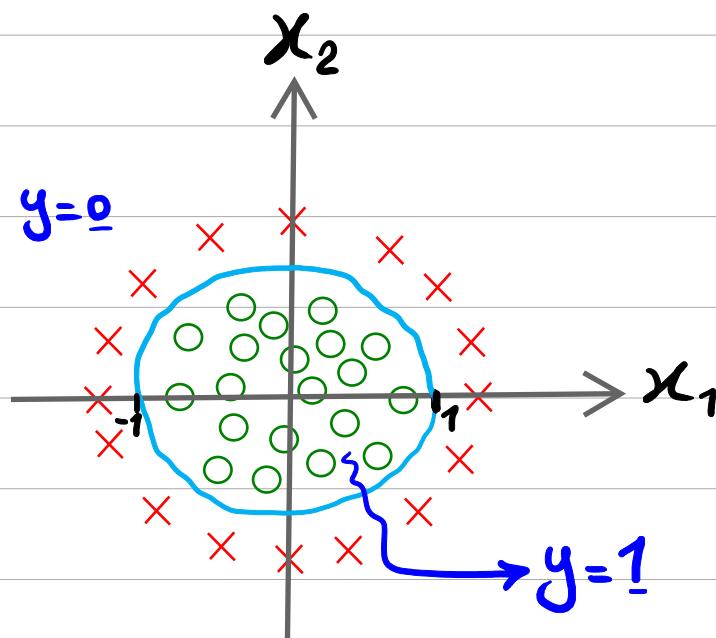


$$F_{\vec{w}, b}(\vec{x}) = g(z) = g(w_1 x_1 + w_2 x_2 + b)$$

$$z = \vec{w} \cdot \vec{x} + b = 0$$

$$\text{if } w_1=1, w_2=1, b=-3$$

$$w_1 x_1 + w_2 x_2 + b = 0 \rightsquigarrow x_1 + x_2 - 3 = 0 \rightsquigarrow$$



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

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

$$\begin{cases} x_1=0, x_2=3 \\ x_1=3, x_2=0 \end{cases}$$

if $x_1=0$
 ~~$w_1 x_1 + w_2 x_2 + b = 0$~~
 $\Rightarrow w_2 x_2 = -b \Rightarrow$
 $x_2 = \frac{-b}{w_2} = \frac{-(-3)}{1} = 3$

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

$i=1, \dots, m$

$j=1, \dots, n$

target y is 0 or 1

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

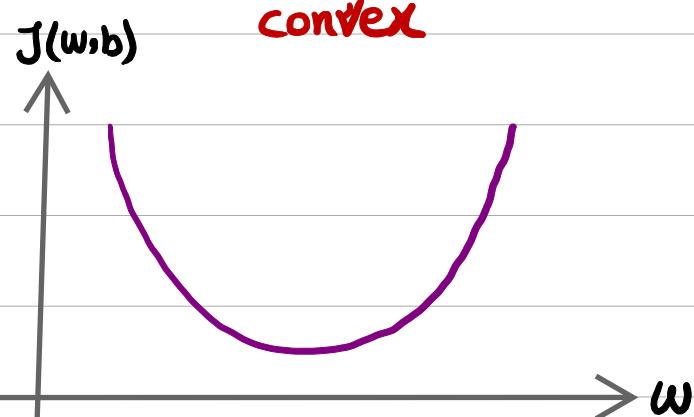
$$\vec{W} = [w_1 \ w_2 \ \dots \ w_n], \ b$$

Linear Regression

Cost Function

چه طور کترین w, b را انتخاب کنم؟

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

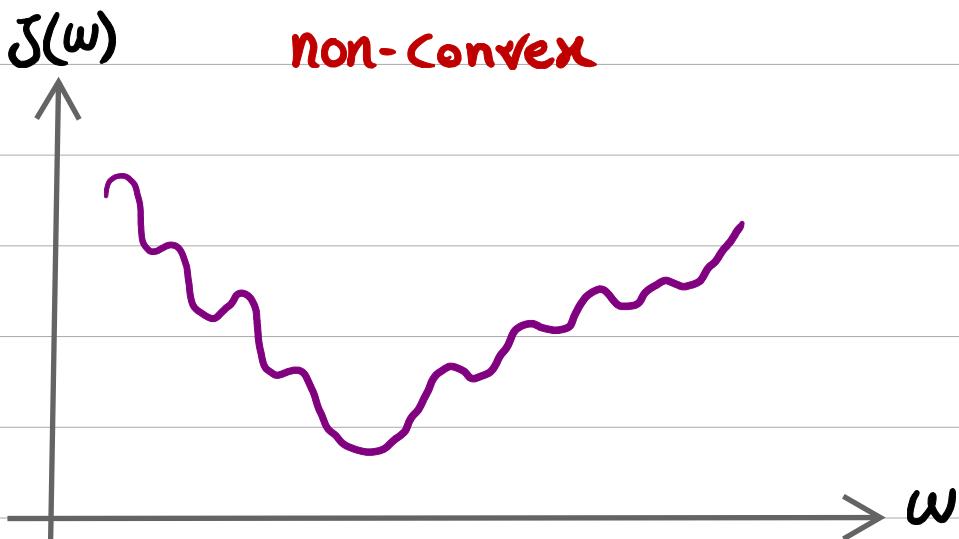


$$F_{w,b}(x) = w \cdot x + b$$

زنگنه
@zahraamini-ai

Logistic Regression

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



Logistic Loss Function

زهراء امینی
@zahraamini_ai

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

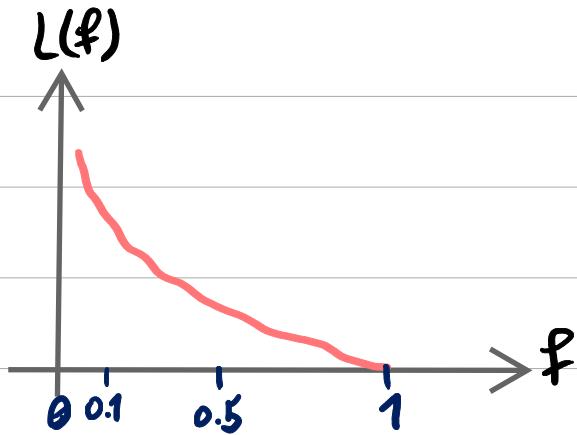


ج: آیا استفاده از این تابع برای Logistic درست است؟

$$f_{w,b}(x) \leq \hat{y}$$

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

if $y=1$



زهراء امینی
@zahraamini_ai

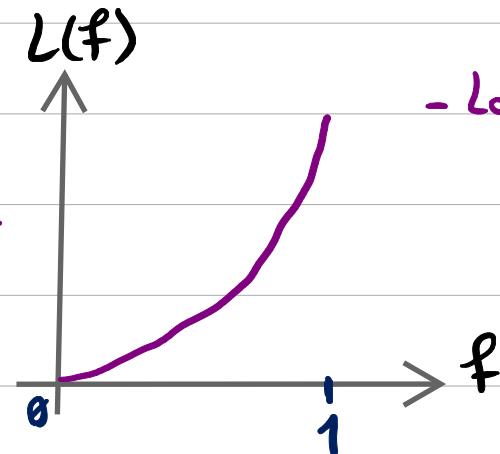
↳ if $f_{w,b}(x^{(i)}) \rightarrow 1$ then Loss $\rightarrow 0$

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

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

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

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



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

$f_{w,b}(x^{(i)}) \rightarrow 1$ then Loss $\rightarrow 0$

Logistic Regression Cost function:

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

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

زهراء أميني

@zahraamini_ai

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

ج: حالاً الريک تابع کیا رچہ بخواهیم باید چہ کنم؟

$$L(f_{w,b}(x^{(i)}), y^{(i)}) = -y^{(i)} \log(f_{w,b}(x^{(i)})) - (1-y^{(i)}) \log(1-f_{w,b}(x^{(i)}))$$

Convex

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

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

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

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

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

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

حالا ما باید به ذیل ج می باشیم که minimum J, L cost را با شرایط w, b را پیدا کنیم.

چه طور متدار w, b را پیدا کنیم؟

1. Newton's Method

دروش رایج داریم:

2. Gradient Descent

Gradient Descent :

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

repeat

{

$$w_j = w_j - \alpha \boxed{\frac{\partial}{\partial w_j} J(w, b)}$$

$$\frac{\partial}{\partial w} J(w, b) =$$

$$\frac{1}{m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$b = b - \alpha \boxed{\frac{\partial}{\partial b} J(w, b)}$$

$$\frac{\partial}{\partial b} J(w, b) =$$

$$\frac{1}{m} \sum_{i=1}^m (f_{w,b}(x^{(i)}) - y^{(i)})$$

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

Zahra Amini

@zahraamini_ai