Classification

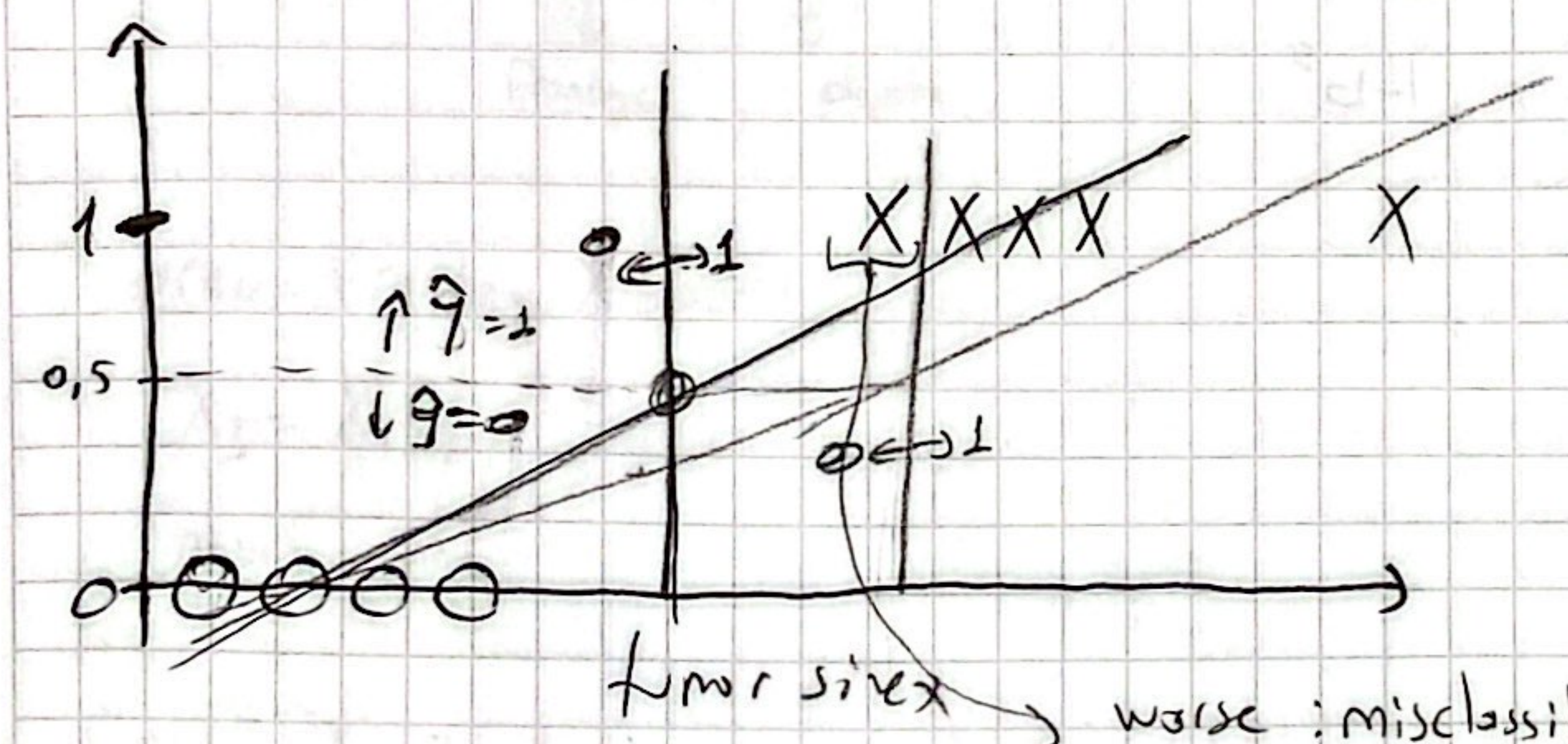
Motivation

Question

	y	
email spam	yes	no
transaction fraudulent	yes	no
tumor malignant	yes	no

no	yes
false	true
0	1

y can only be one of two values
"binary classification"
(class = category)



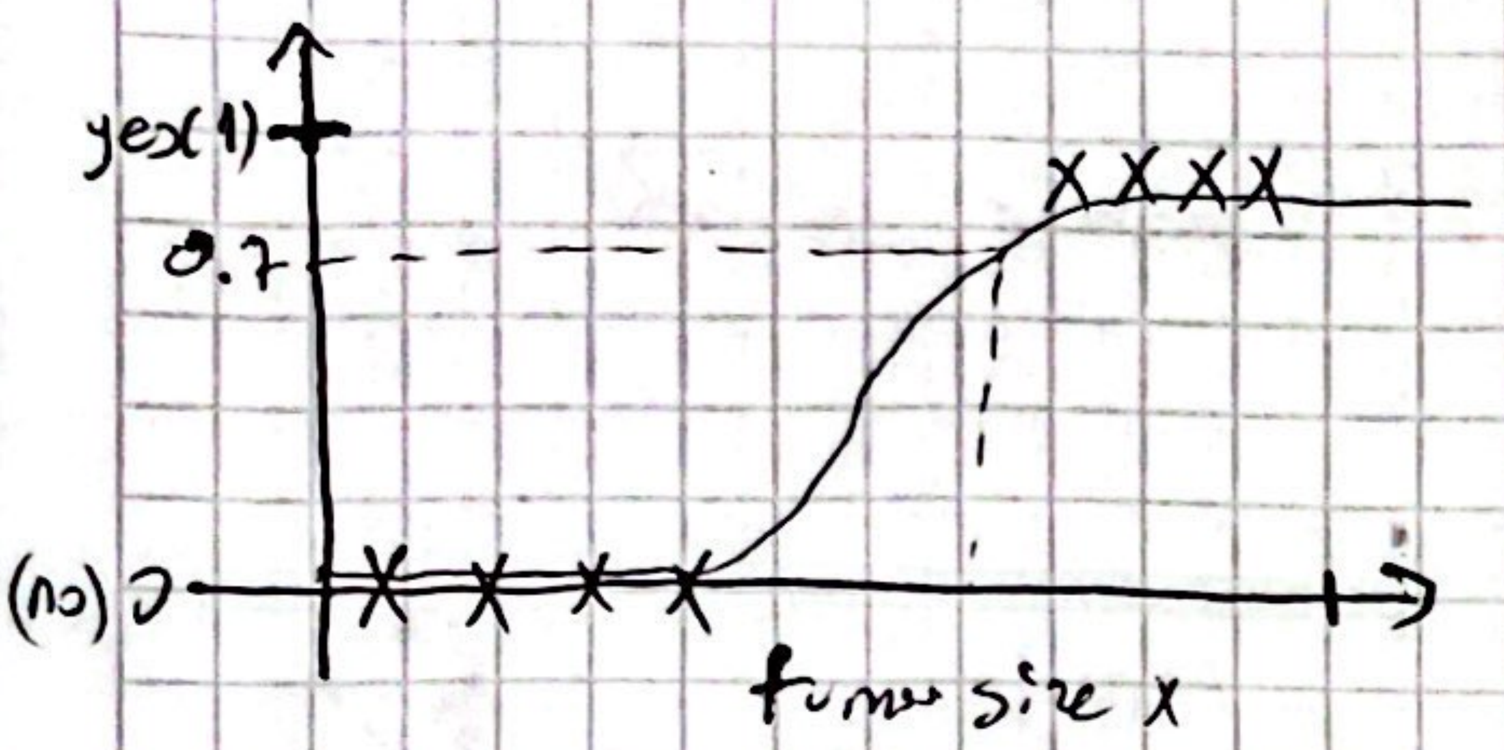
worse: misclassified

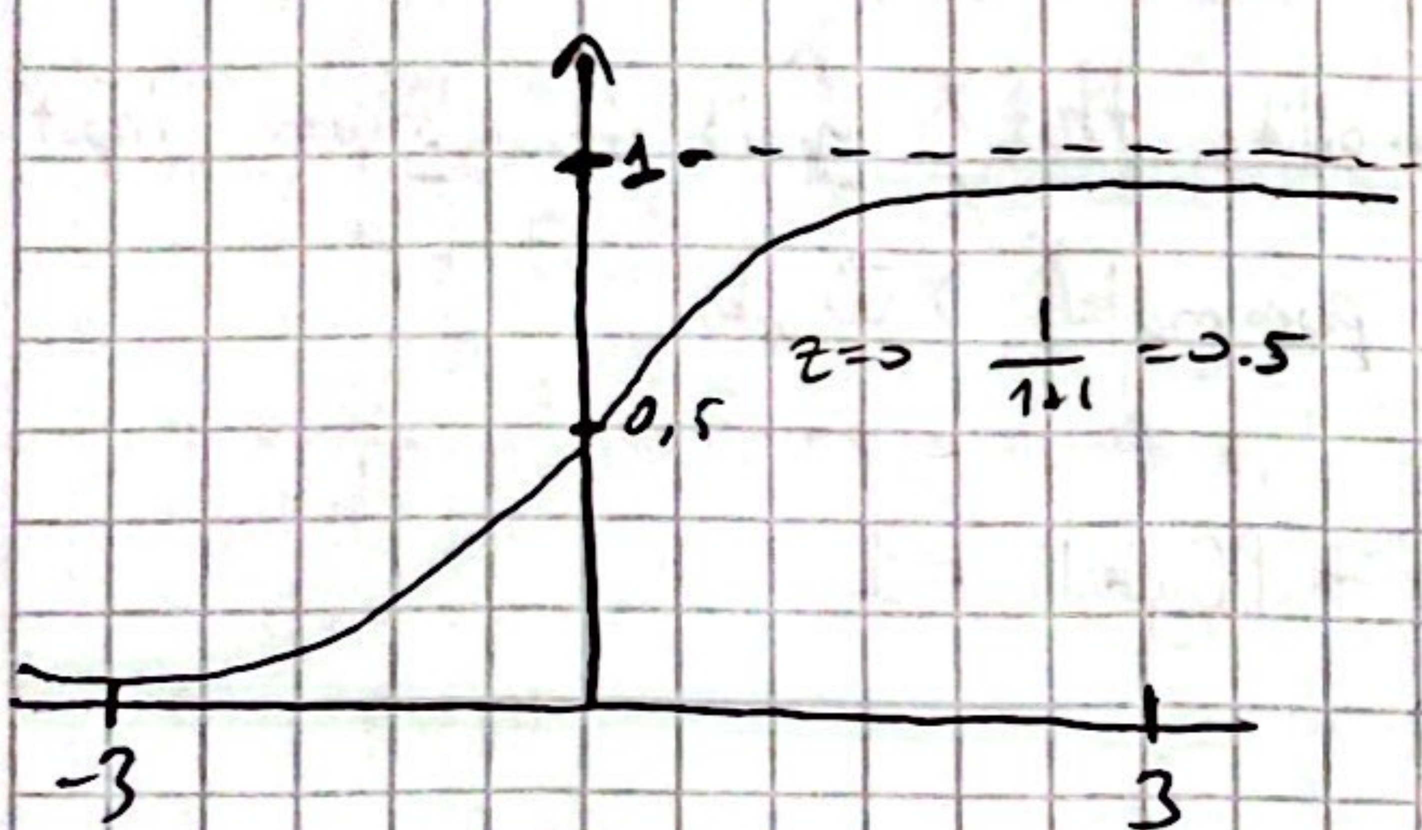
if $f_{w,b}(x) < 0{,}5 \rightarrow \hat{y} = 0$

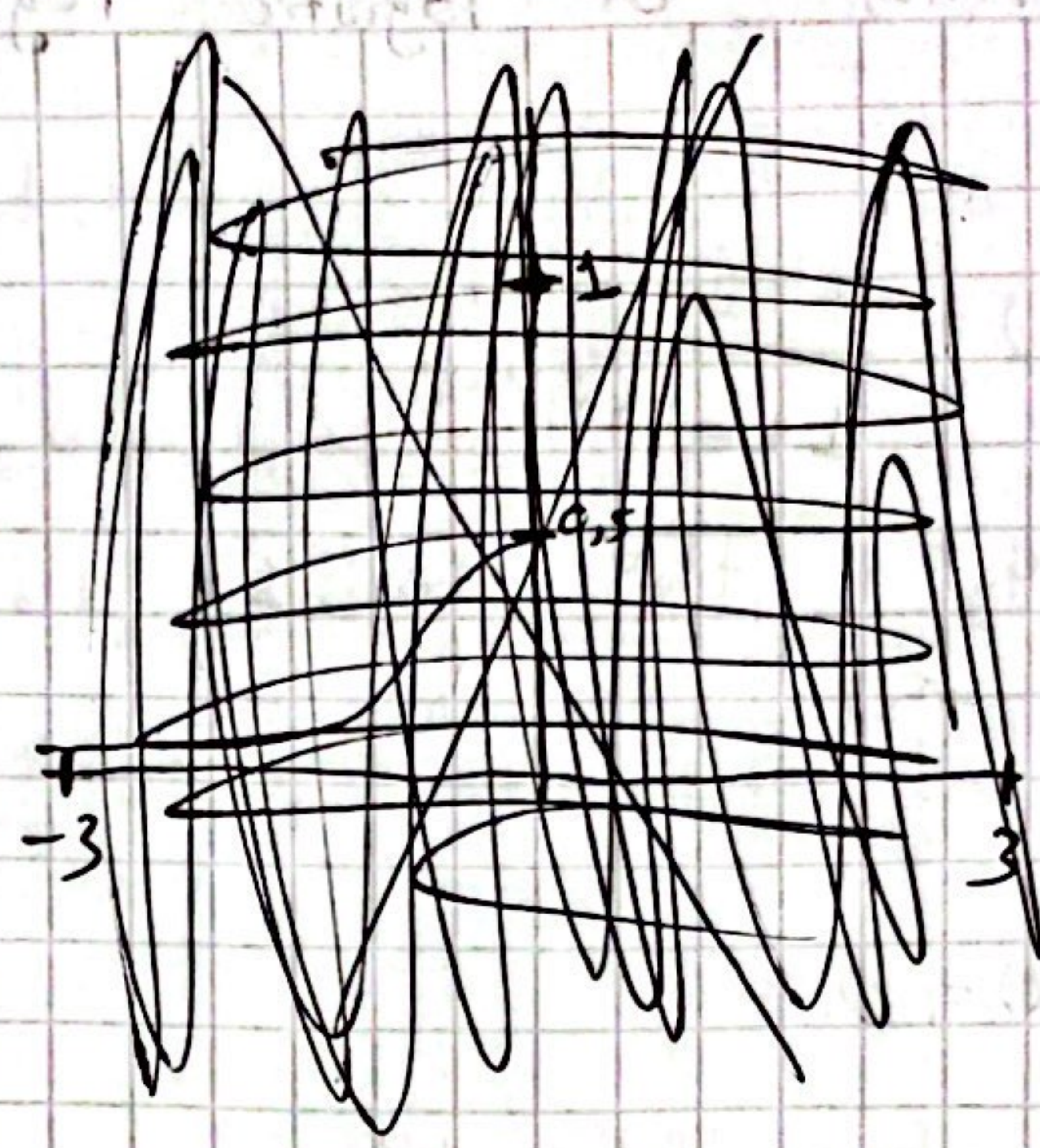
if $f_{w,b}(x) > 0{,}5 \rightarrow \hat{y} = 1$

next: $\underbrace{\text{logistic regression}}_{\text{classification}}$

Logistic Regression







sigmoid function

logistic "

outputs between 0 and 1

$g(z) = \frac{1}{1+e^{-z}}$ $\quad 0 < g(z) < 1$

$f_{\vec{w},b}(x)$ =

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

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

$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w}\cdot\vec{x}+b}_{z}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$

logistic regression

Interpretation of logistic regression output

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

"probability" that class is 1

$f_{\vec{w},b}(x) = 0.7$

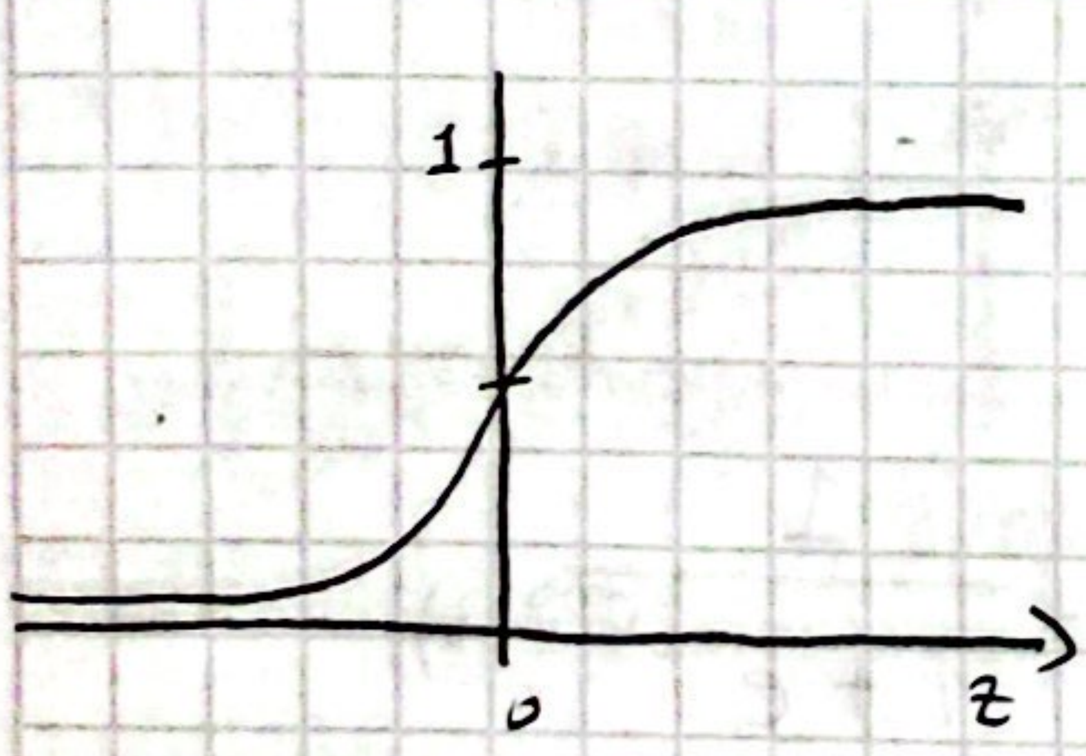
%70 chance that y is 1

$f_{\vec{w},b}(\vec{x}) = P(y=1 \mid \vec{x}; \vec{w}, b)$

probability that y is 1, given input $\vec{x}$, parameters $\vec{w}, b$

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

Decision Boundary



$f_{\vec{w},b}(\vec{x})$

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

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

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

if $f_{\vec{w},b}(\vec{x}) \geq 0.5$

Yes: $\hat{y}=1$ No: $\hat{y}=0$

When is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

$g(z) \geq 0,5$

$z \geq 0$

$\vec{w}\cdot\vec{x}+b \geq 0$ $\vec{w}\cdot x+b < 0$

$\hat{y}=1$ $\hat{y}=0$

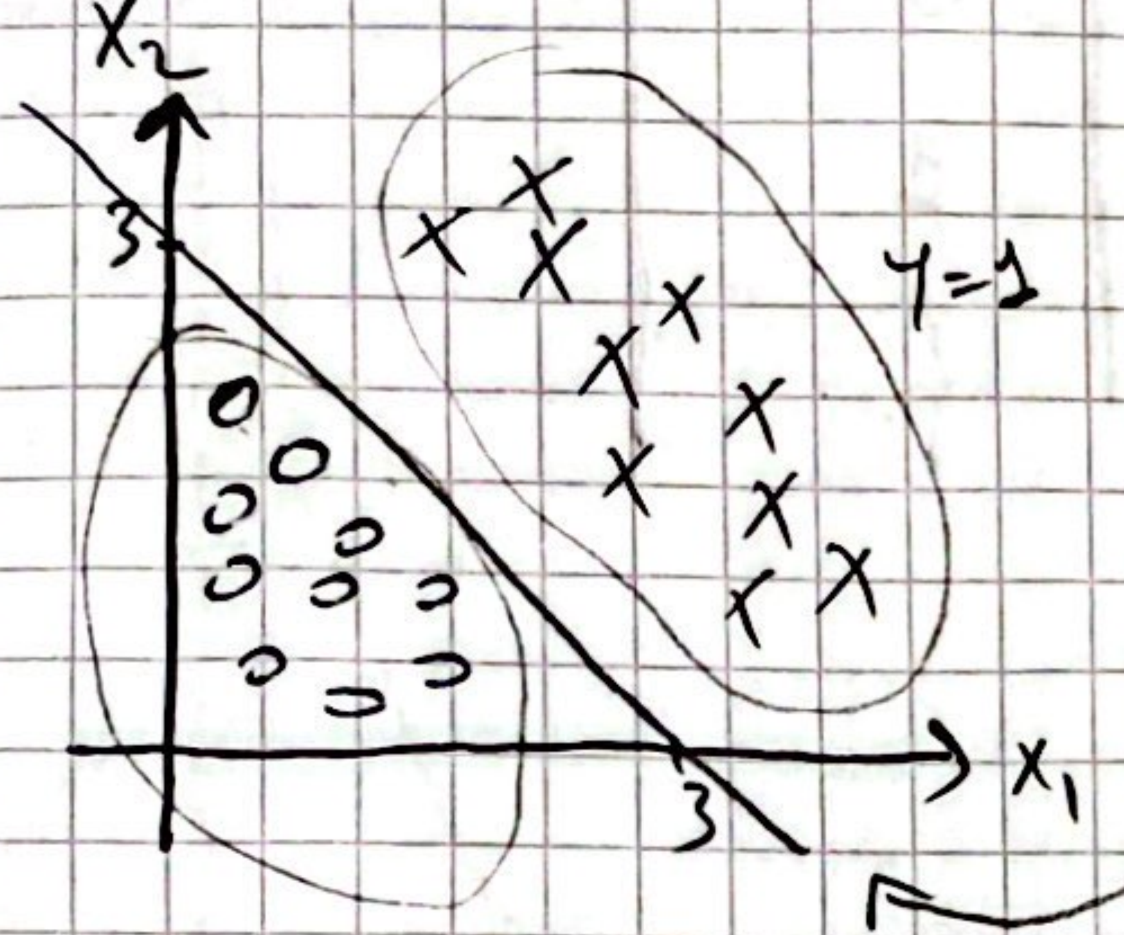
$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\underset{1}{w_1} x_1 + \underset{1}{w_2} x_2 + \underset{-3}{b})$$

Decision boundry

$$z = w_1 x_1 + w_2 x_2 + b = 0$$

$$z = x_1 + x_2 - 3 = 0$$

$$z = x_1 + x_2 = 3$$

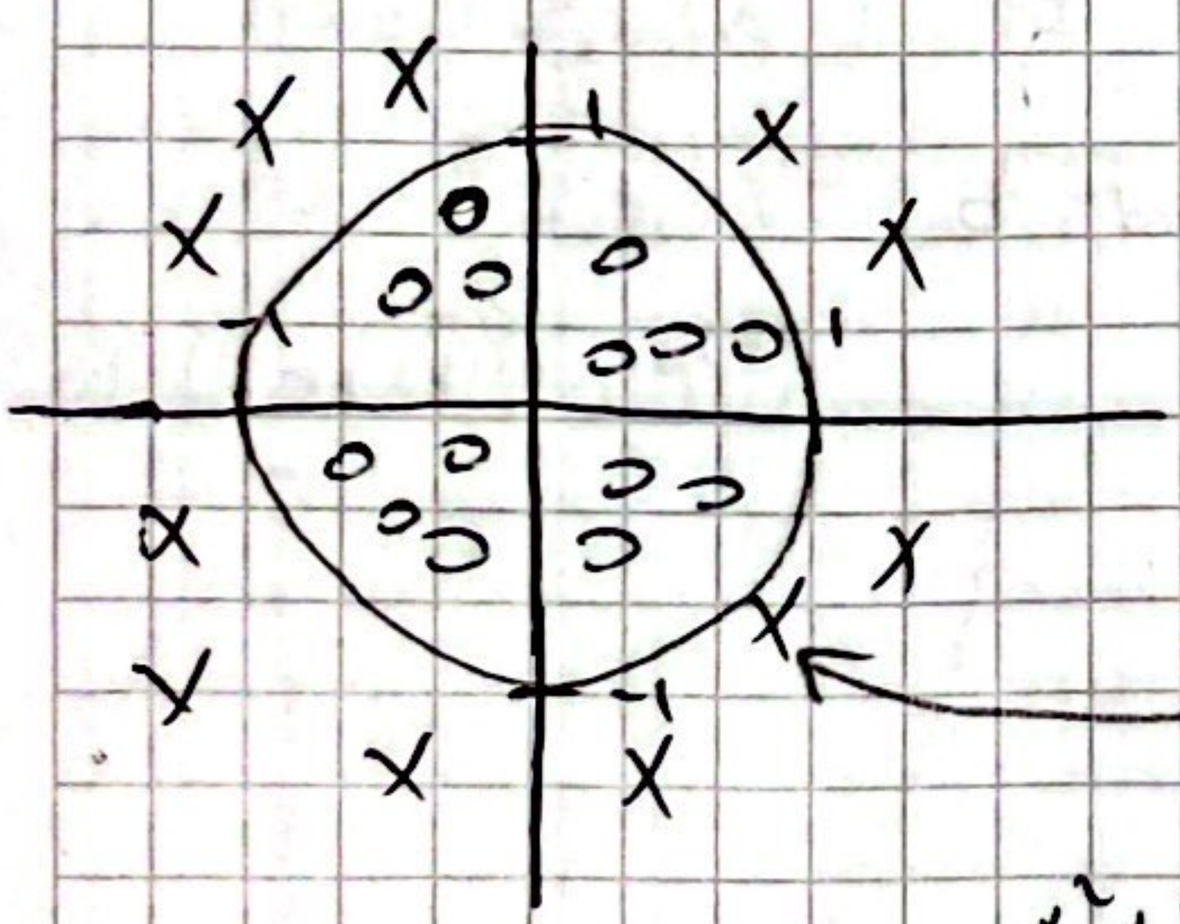


$y=0$

$\boxed{\text{Non-linear decision boundaries}}$

$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\underset{1}{w_1} x_1^2 + \underset{1}{w_2} x_2^2 + \underset{-1}{b})$$

decision boundry $z = x_1^2 + x_2^2 = 1$



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

$\hat{y} = 1$

$x_1^2 + x_2^2 < 1$

$y = 0$

☆ $f_{\vec{w},b}(\vec{x}) = g(w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_1 x_2 + w_5 x_2^2)$

olabilir

Cost Function for Logistic Regression

	x_1 tumor size (cm)	...	x_n patient's age	malignant
$i=1$	10		52	1
	2		73	0
	5		55	0
	12		49	1
$i=m$	⋮		⋮	⋮

$i = 1 ... m$ ← training examples

$j = 1 ... n$ ← features

target y is 0 or 1

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

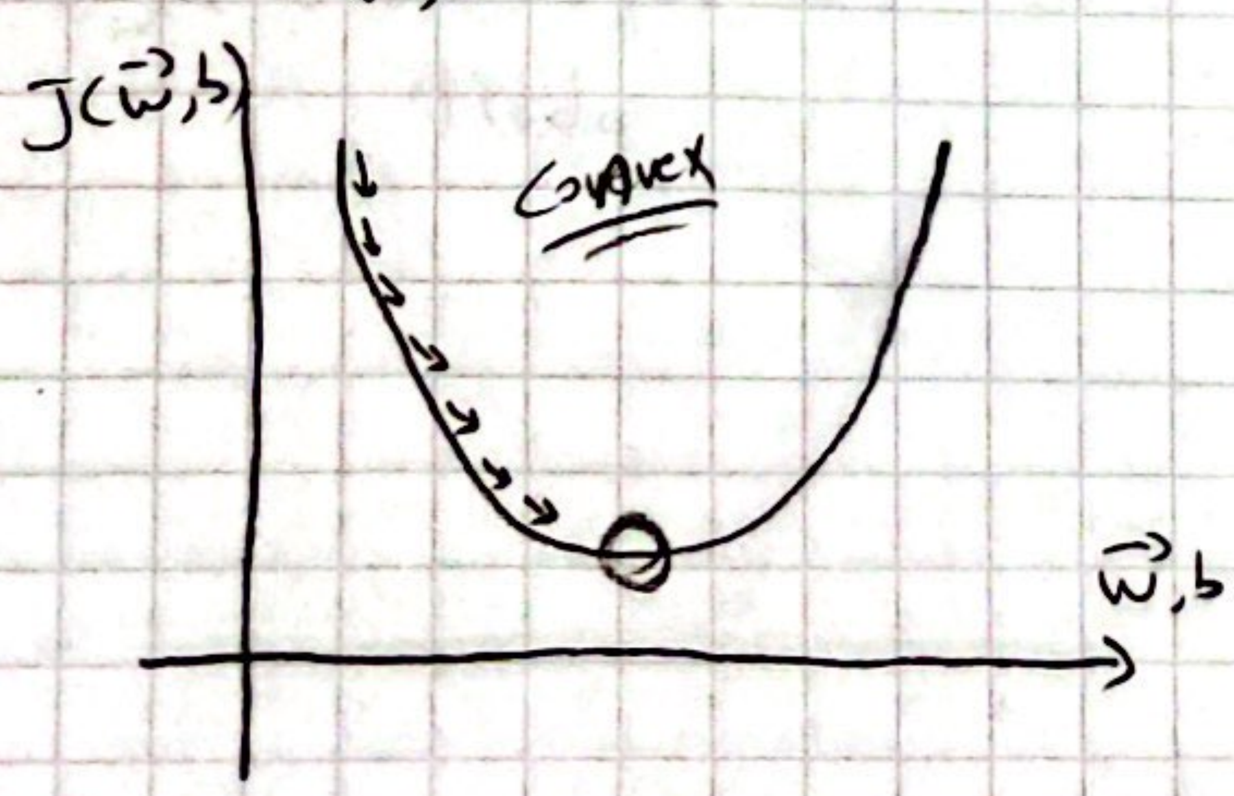
How to chose $\vec{w} = [w_1 \; w_2 \; \cdots \; w_n]$ and b

Squared error cost

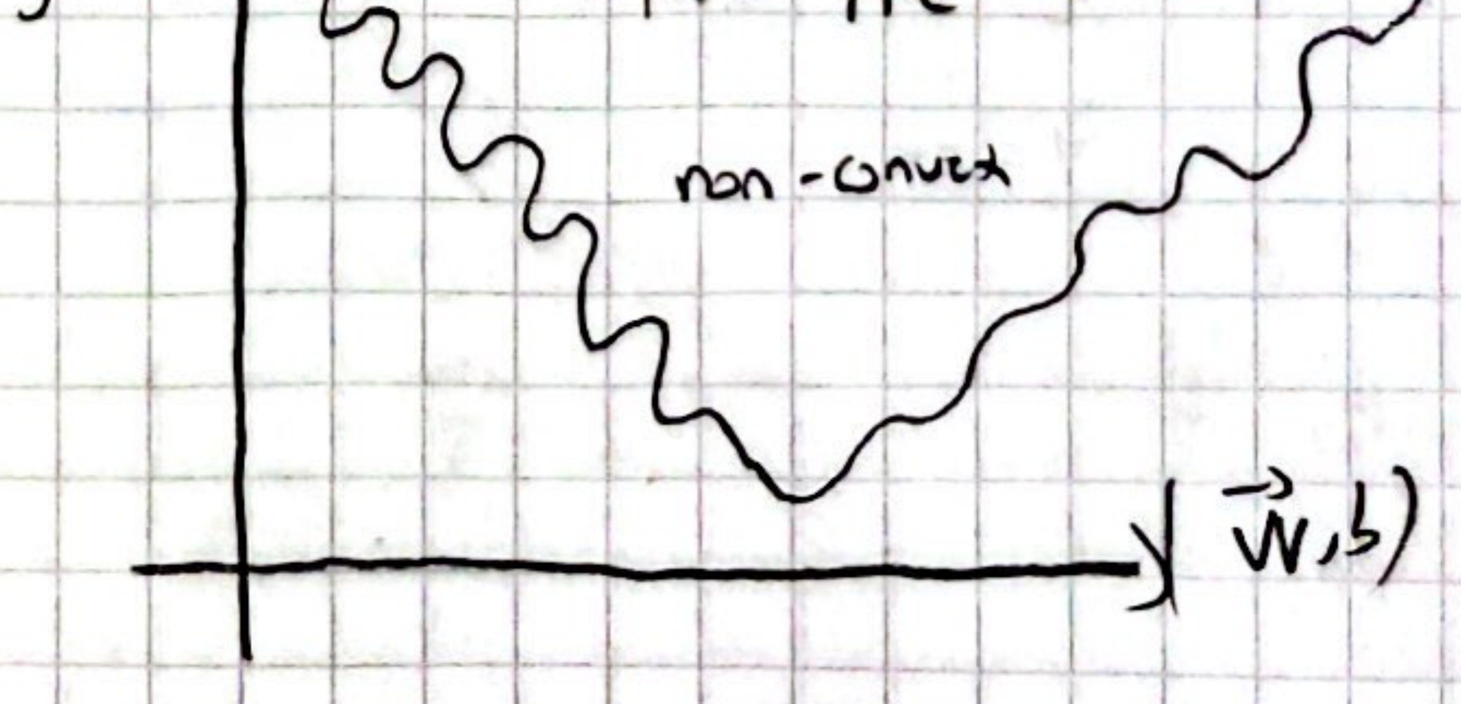
$$J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2$$

linear regression

$f_{\vec{w},b} = \vec{w}\vec{x}+b$

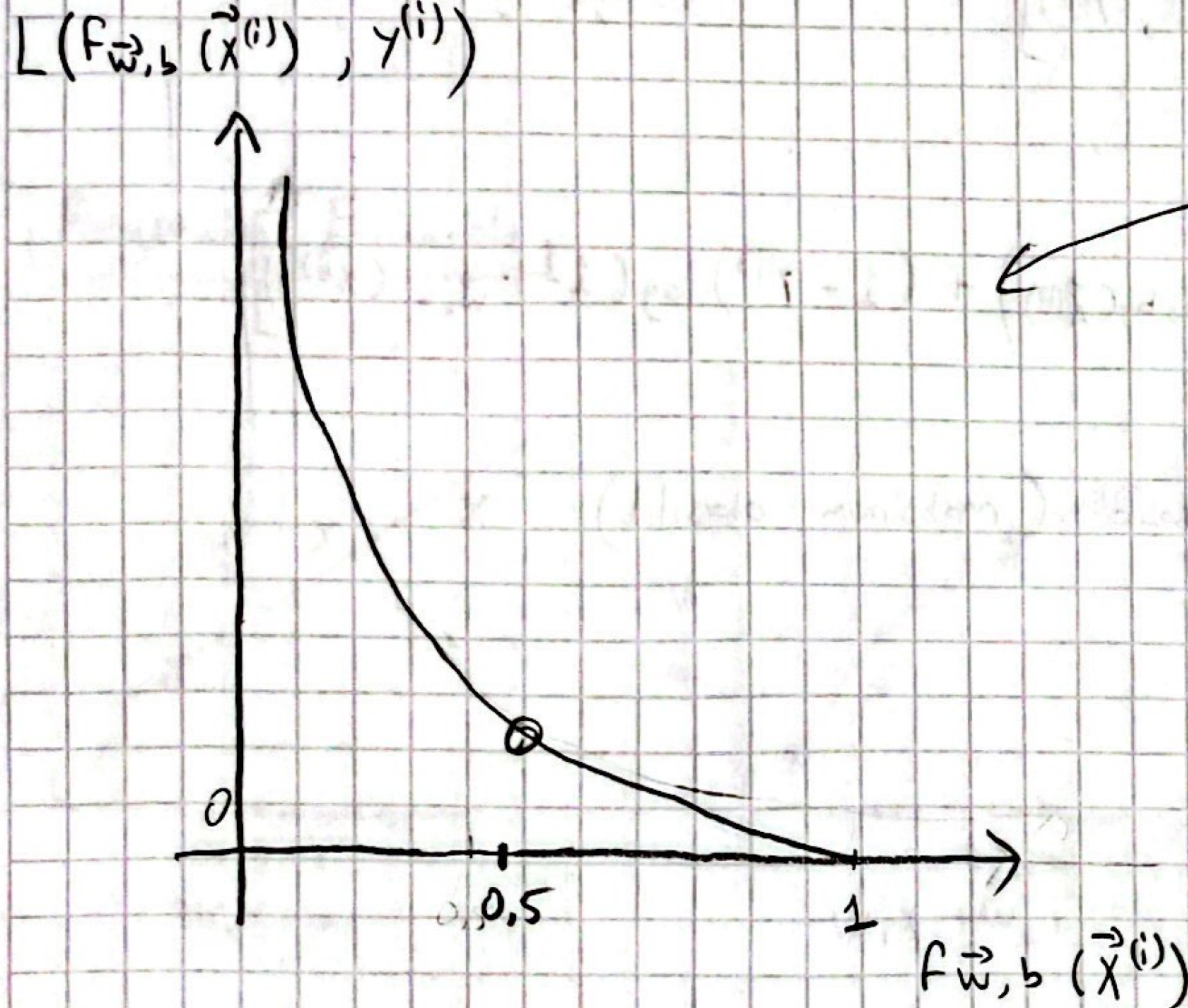


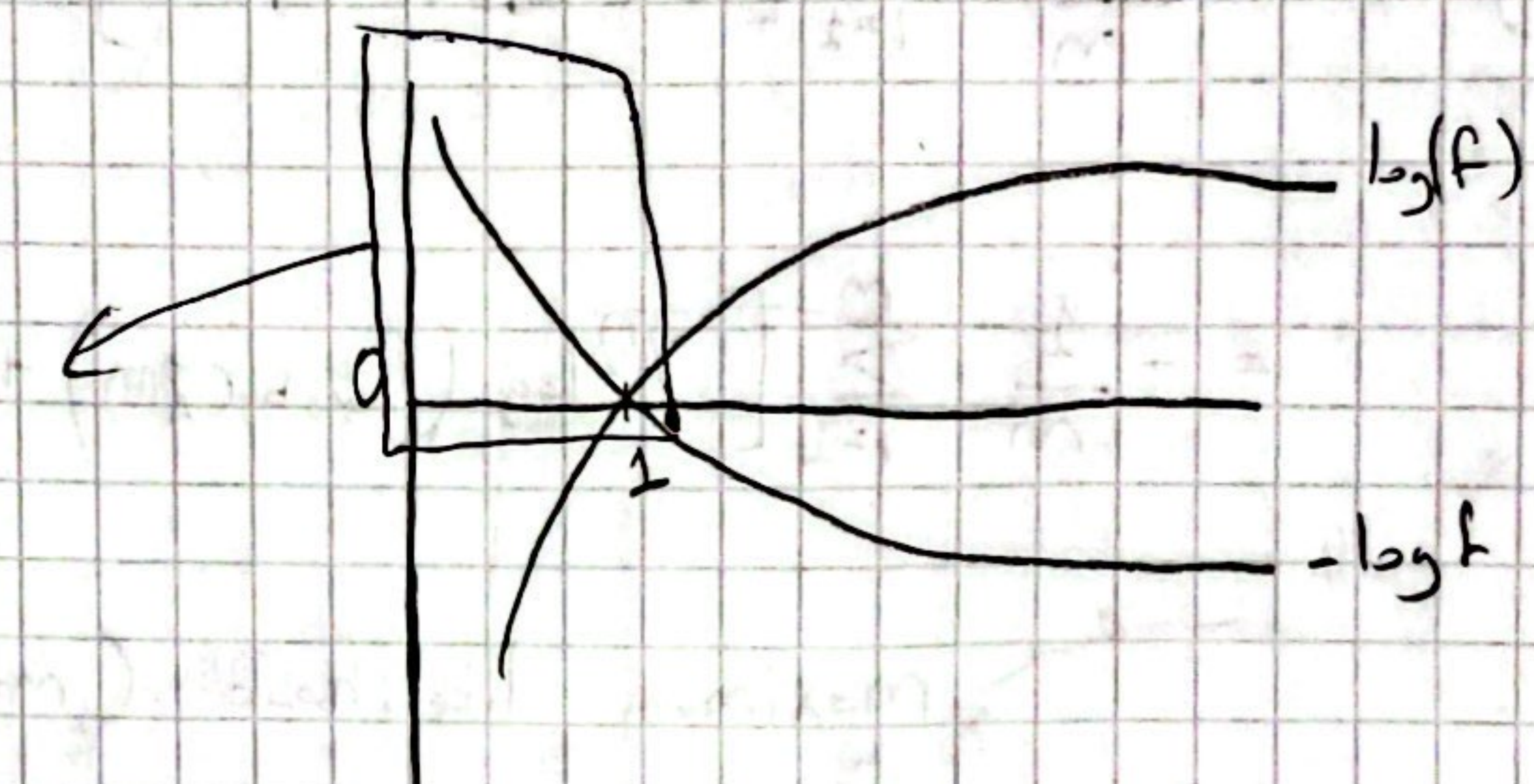
logistic regression



Logistic Loss function

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



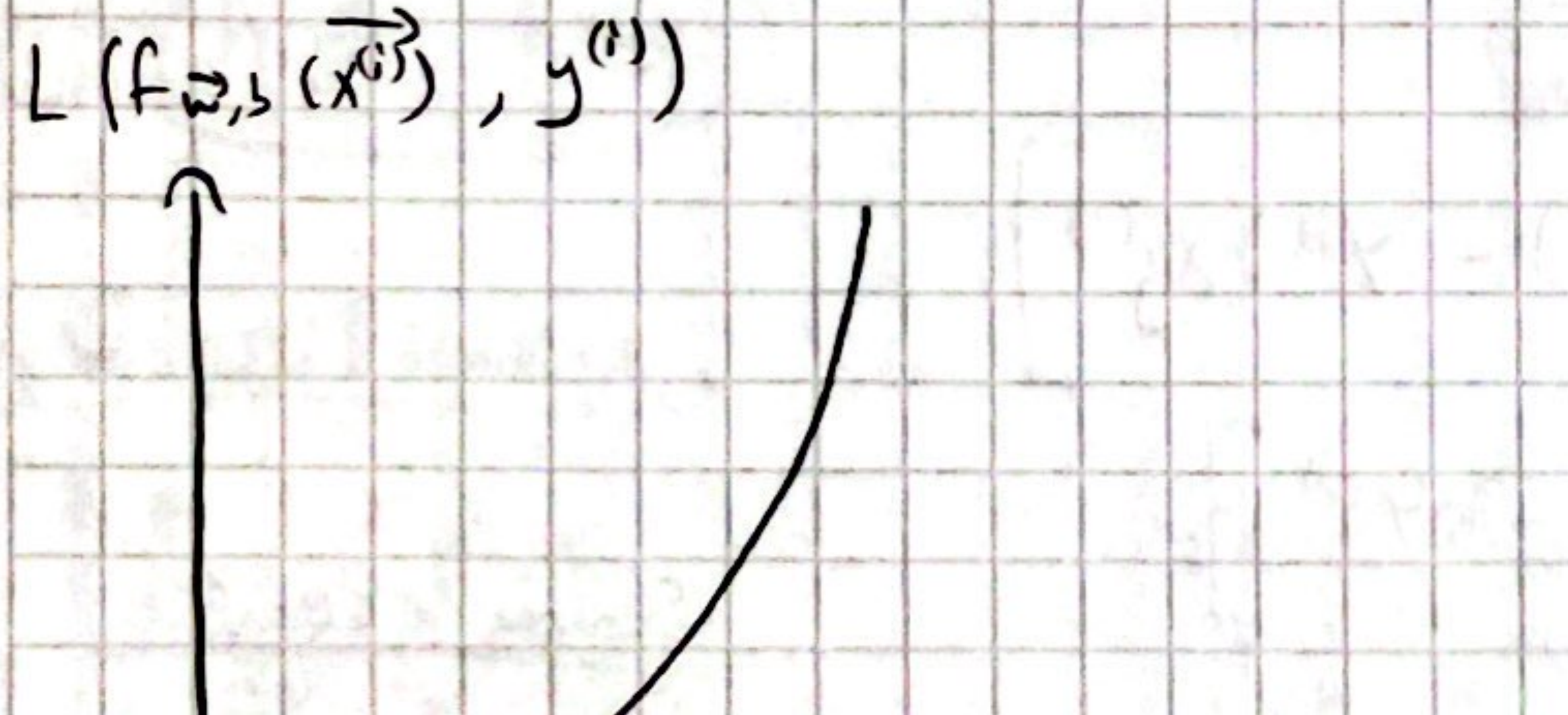


if $y^{(i)} = 1$

As $f_{\vec{w},b}(\vec{x}^{(i)}) \to 1$ then loss $\to 0$

As $f_{\vec{w},b}(\vec{x}^{(i)}) \to 0$ then loss $\to \infty$

$f_{\vec{w},b}(\vec{x}^{(i)})$ y'ye yakın bir tahminde bulunduğunda loss endüşüktür.



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

As $f_{\vec{w},b}(\vec{x}^{(i)}) \to 1$ then loss $\to \infty$

Simplified Cost Function

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

convex
(single gobal minimum)

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

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

maximum likelihood (maksimum olasılık)

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$$

Logistik Regresyon için Dereceli azalma

repeat {

$$w_j = w_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}\right]$$

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

}

Liner regresyon $f_{\vec{w},b}(\vec{x}) = \vec{w}\cdot\vec{x} + b$

Logistic " $f_{\vec{w},b}(\vec{x}) = \dfrac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$

Same Concept

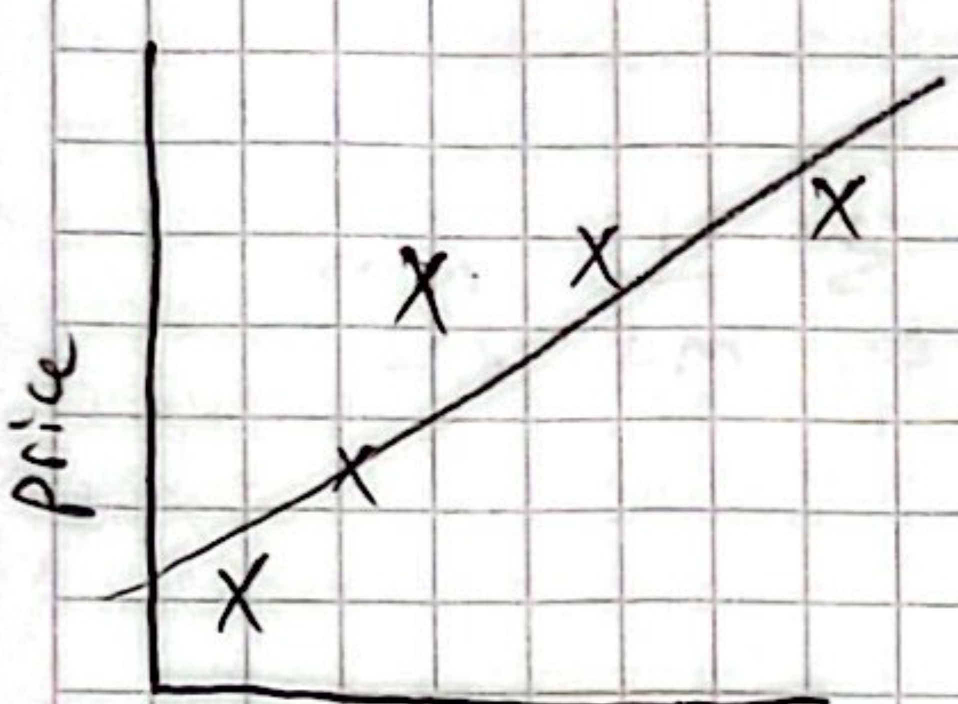
- Monitor gradient descent
- Vectorized implementation
- Feature scaling

The Problem of Overfitting

overfitting = aşırı uyum

underfitting = yetersiz uyum

Regression Example

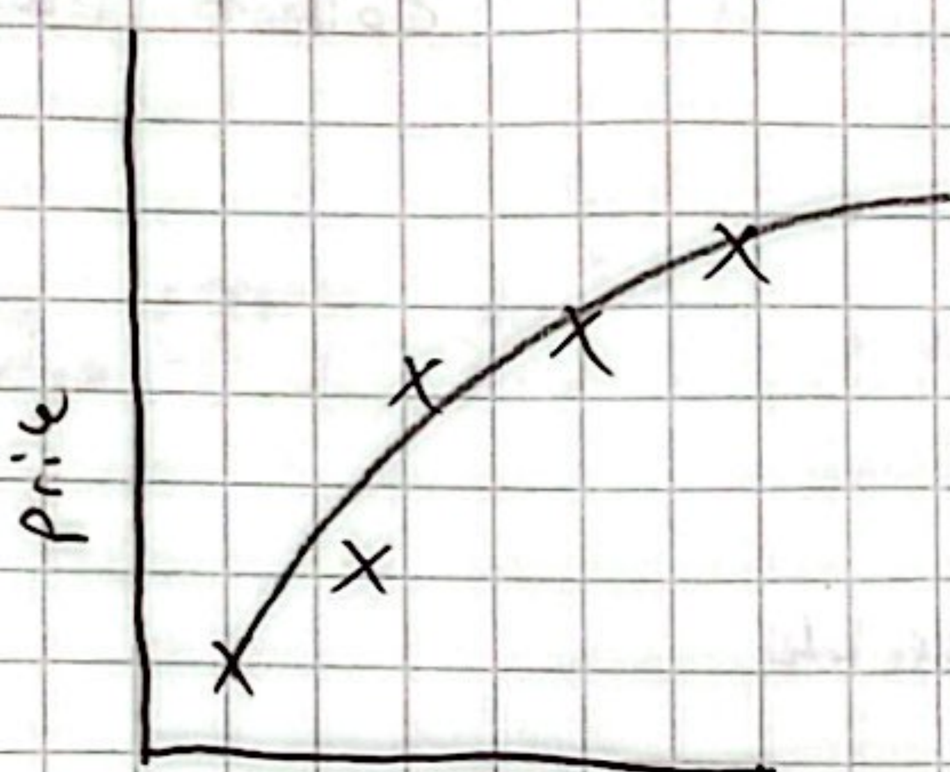


w_1x+b

underfit

- Does not fit the training set well

high bias

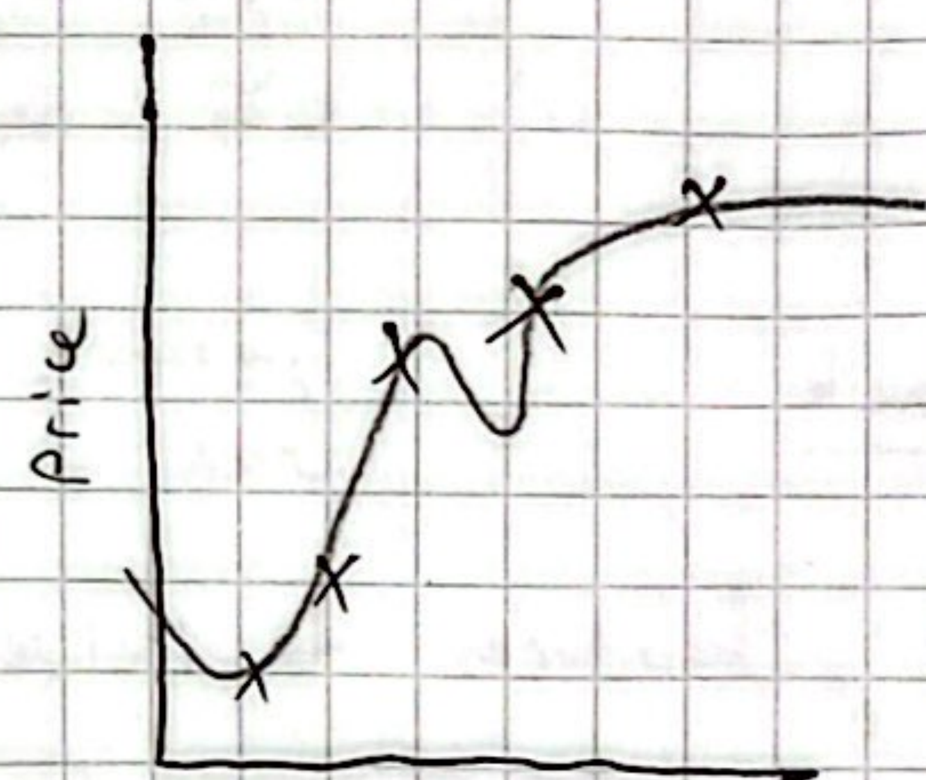


$w_1x+w_2x^2+b$

just right ✓

- Fits training set pretty well

generalization (genelleştirme)



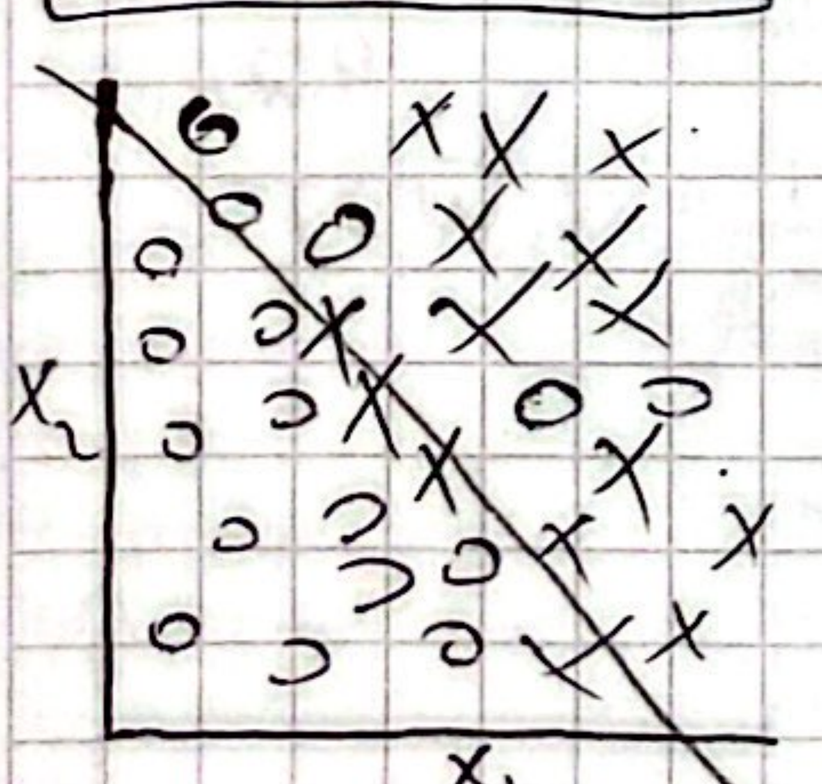
$w_1x+w_2x^2+w_3x^3+w_4x^4+b$

overfit

- Fits the training set extremely well

high variance

Classification

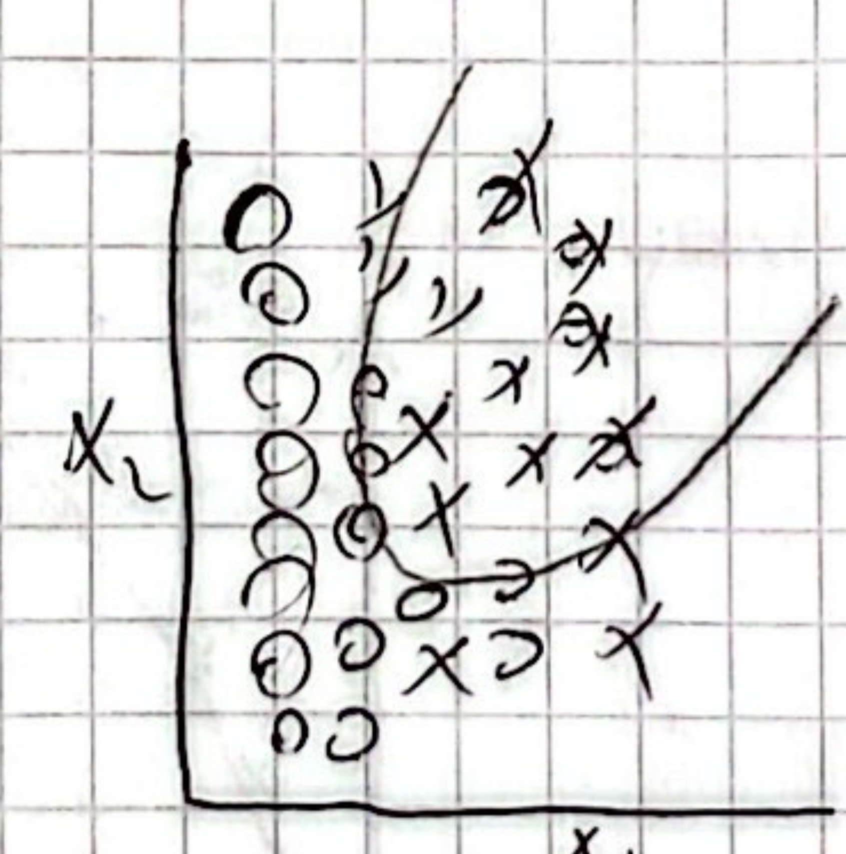


$z=w_1x_1+w_2x_2+b$

$f_{\vec{w},b}(\vec{x})=g(z)$

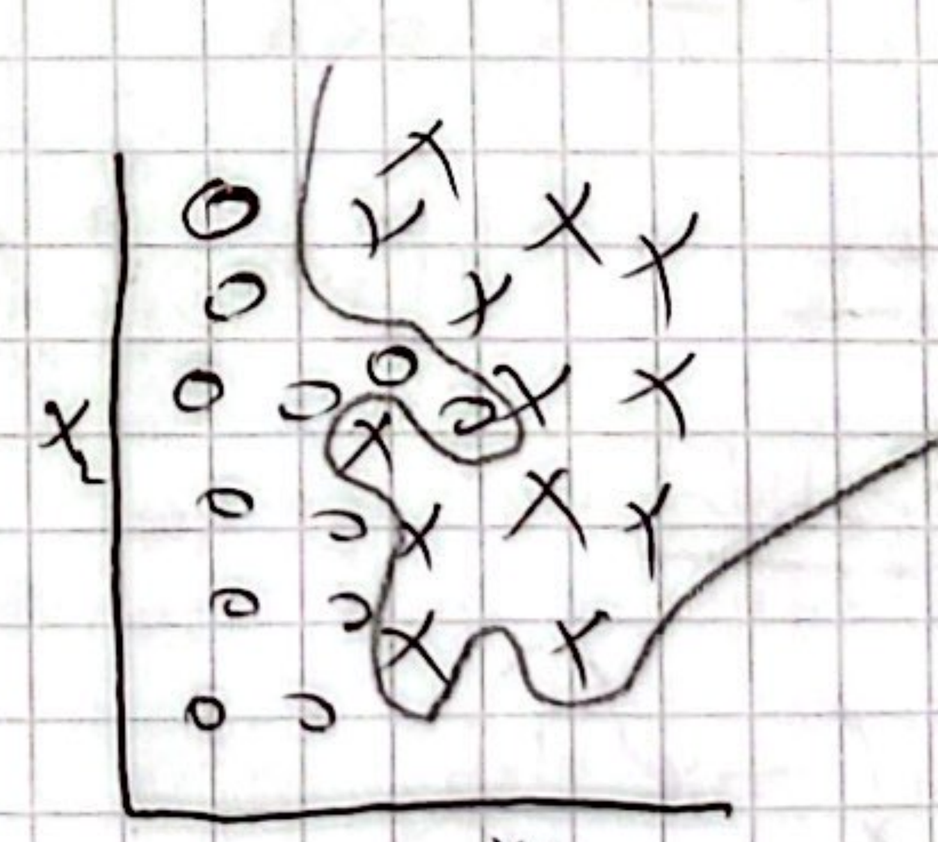
underfit

high bias



$z=w_1x_1+w_2x_2+w_3x_1^2+w_4x_2^2+w_5x_1x_2+b$

just right ✓

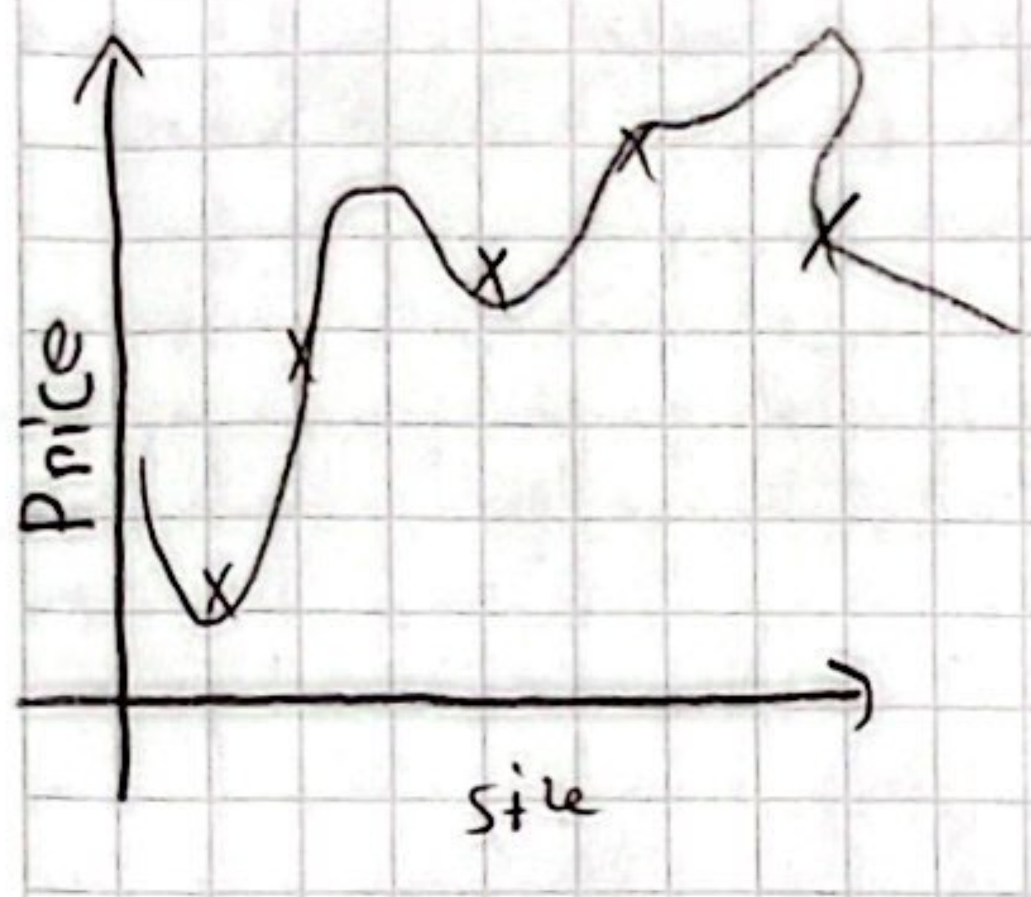


$z=w_1x_1+w_2x_2+w_3x_1^2x_2+w_4x_1^2x_2^2+w_5x_1^2x_2^3+w_6x_1^3x_2+\cdots+b$

overfit

high variance

Adressing Overfitting



1. seçenek

Collect more training examples

2. seçenek

Select features to include/exclude

Size	bedrooms	floors	age	avg income	....	distance to 6th shop	price
x_1	x_2	x_3	x_4	x_5	--	x_{100}	

all features
+
insuffucient data
⇓
overfit

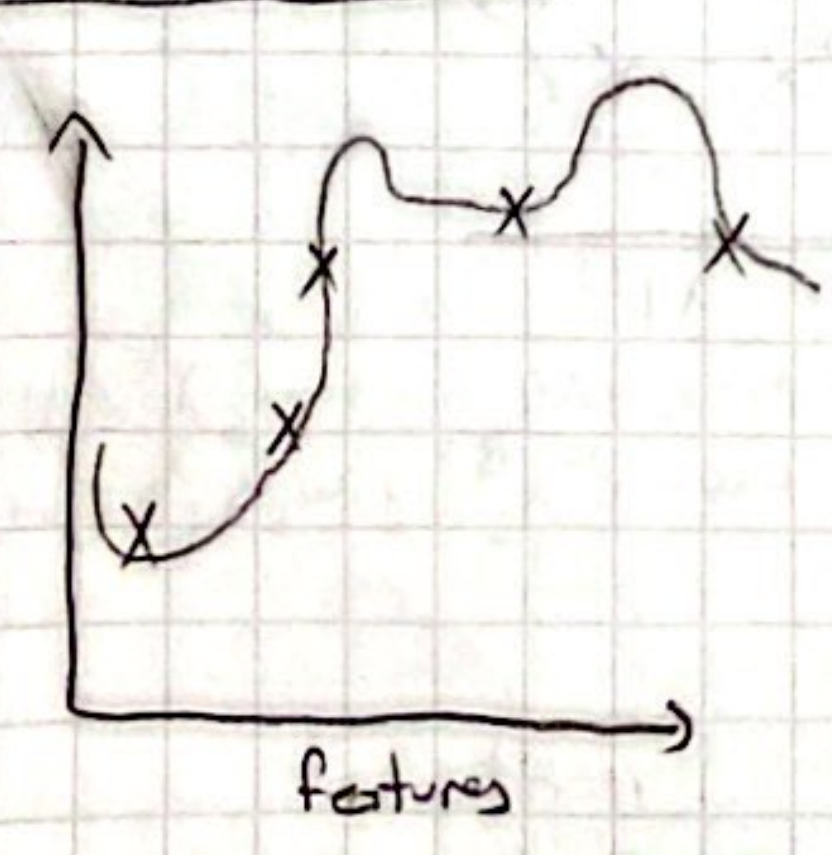
selected features
- size
- bedrooms
- age

just right ✓

disadvantage
↓
useful features could be lost

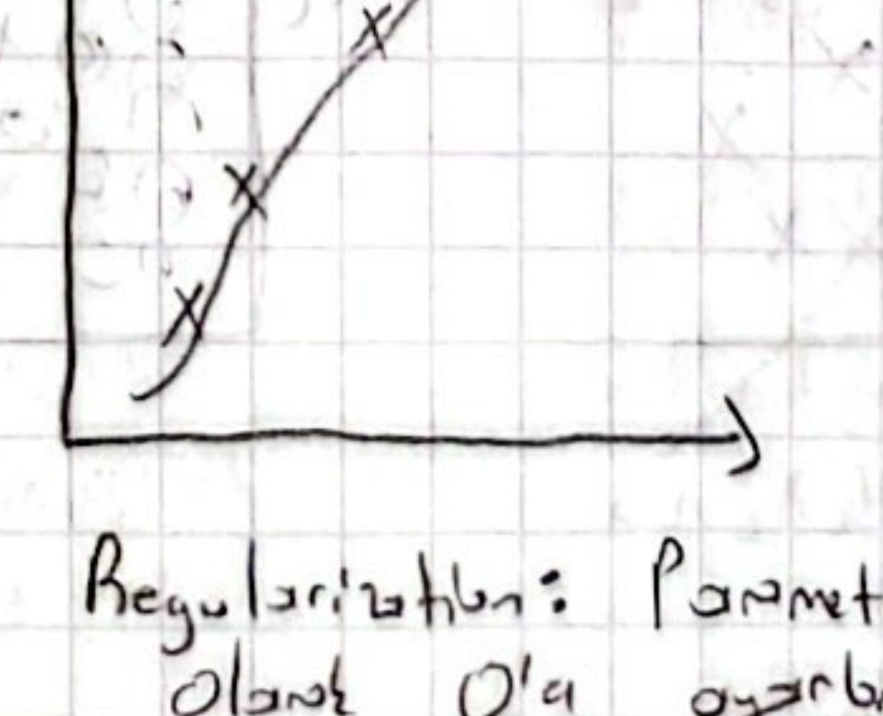
3. seçenek

Regularization



$f(x) = 28x - 385x^2 + 39x^3 - \underbrace{\cancel{174x^4}}_{0} + 100$

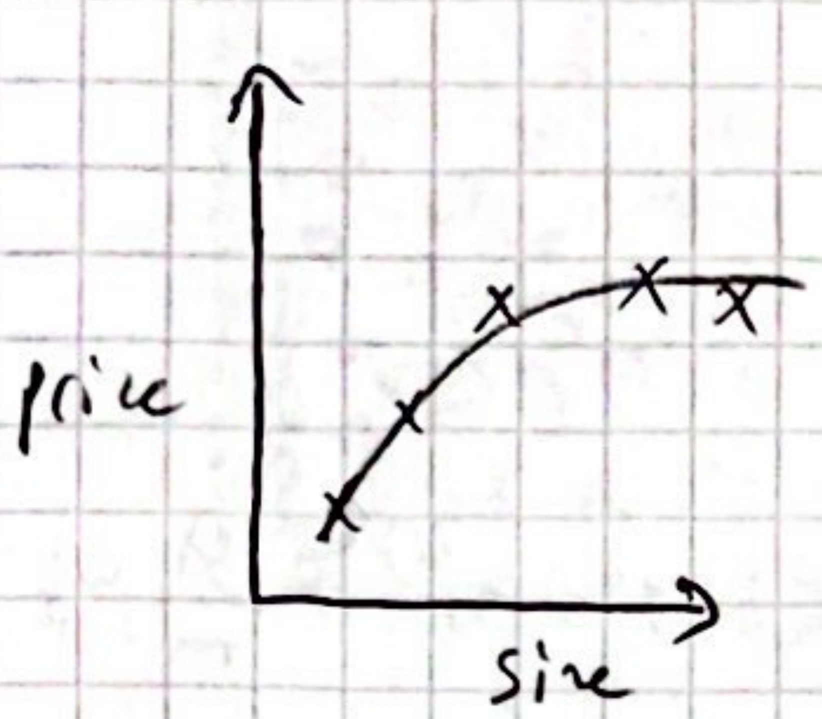
eliminate feature) 2. seçeneklerin eşdeğer.

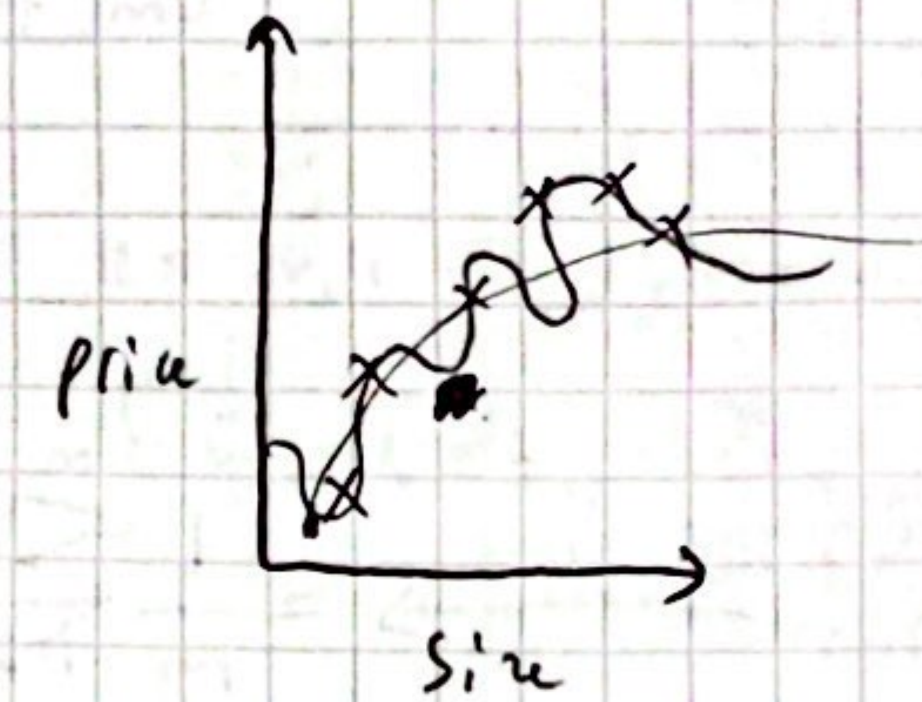


Regularization: Parametrenin tam olarak 0'a ayarlanmasını talep etmeden parametrelerin değerini küçültmeye teşvik etmektir.

$f(x) = 13x - 0.23x^2 + 0.000014x^3 - 0.0001x^4 + 10$

Cost Function with Regularization





$w_1x + w_2x^2 + b$

$w_1x + w_2x^2 + \underbrace{\cancel{w_3x^3}}_{\approx 0} + \underbrace{\cancel{w_4x^4}}_{\approx 0} + b$

$$\min_{\vec{w},b} \frac{1}{2m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})^2 + \underbrace{1000 w_3^2}_{0,001} + \underbrace{1000 w_4^2}_{0,002}$$

Size x_1	bedrooms x_2	floors x_3	......	distance to coffeeshop x_{100}	price y
1	2	3		$n=100$	

$w_1, w_2, w_3, \ldots, w_{100}, b$

$$J(\vec{w},b) = \frac{1}{2m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})^2 + \overbrace{\frac{\lambda}{2m} \sum_{j=1}^{n} w_j^2}^{\text{regularization term}} + \underbrace{\frac{\lambda}{2m} b^2}_{\text{include or exclude}}$$

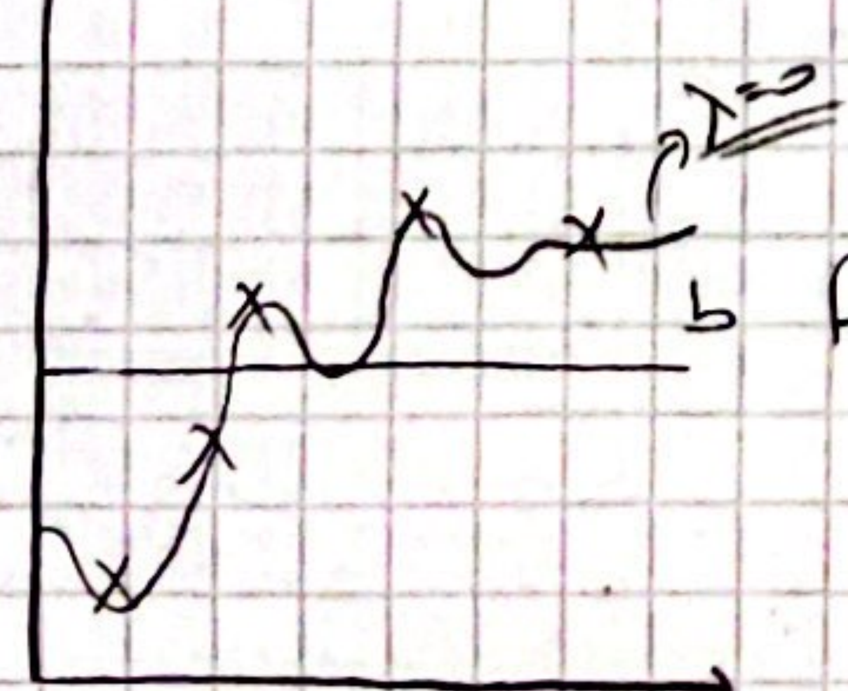
"lambda"
regularization parameter
$\lambda > 0$

$$\min_{\vec{w},b} J(\vec{w},b) = \min_{\vec{w},b} \left[\frac{1}{2m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} w_j^2 \right]$$

fit data

Keep w_j small

λ balances both goals



choose $\lambda: 10^{10}$

$f_{\vec{w},b}(\vec{x}) = \underbrace{w_1x}_{\approx 0} + \underbrace{w_2x^2}_{\approx 0} + \underbrace{w_3x^3}_{\approx 0} + \underbrace{w_4x^4}_{\approx 0} + b$

$f(x) = b$

Regularized Linear Regression

$$j(\vec{w},b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

Gradient descent

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} j(\vec{w},b) \longrightarrow = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m} w_j$$

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

Imlementing gradient descent

repeat {

$$w_j = w_j - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m} w_j\right]$$

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

} simultaneous update

opsiyonel

$$w_j = \underbrace{1 \cdot w_j - \alpha\frac{\lambda}{m} w_j}_{} - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

$$= w_j\left(1 - \alpha\frac{\lambda}{m}\right)$$

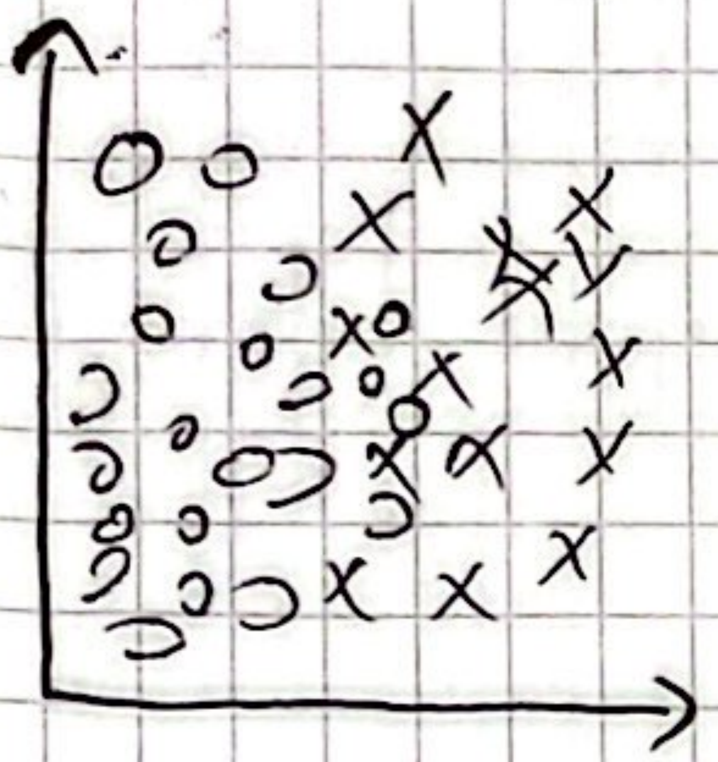
$\alpha = 0{,}01 \quad \lambda = 1 \quad m = 50$

$\alpha\frac{\lambda}{m} = \frac{0{,}01 \cdot 1}{50} = 0{,}0002$

$w_j(1 - 0{,}0002)$

$\underline{0{,}9998\, w_j}$

Regularized Logistic Regression



$$z = w_1 x_1 + w_2 x_2 + w_3 x_1^2 x_2 + w_4 x_1^2 x_2^2 + w_5 x_1^2 x_2^3 + \cdots + b$$

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

Cost function

$$J(\vec{w},b) = -\frac{1}{m}\left[y^{(i)} \log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1-y^{(i)}) \log\left(1-f_{\vec{w},b}(\vec{x}^{(i)})\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j$$

Gradient descent

repeat {

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w},b) \longrightarrow = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}\Big] + \frac{\lambda}{m} w_j$$

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