Machine learning algorithms

- Supervised Learning (Gözetimli)
- Unsupervised Learning (Gözetimsiz)
- Recomender Systems (Öneri sistemleri)
- Reinforcement Learning (Pekiştirmeli öğrenme)

Supervised Learning

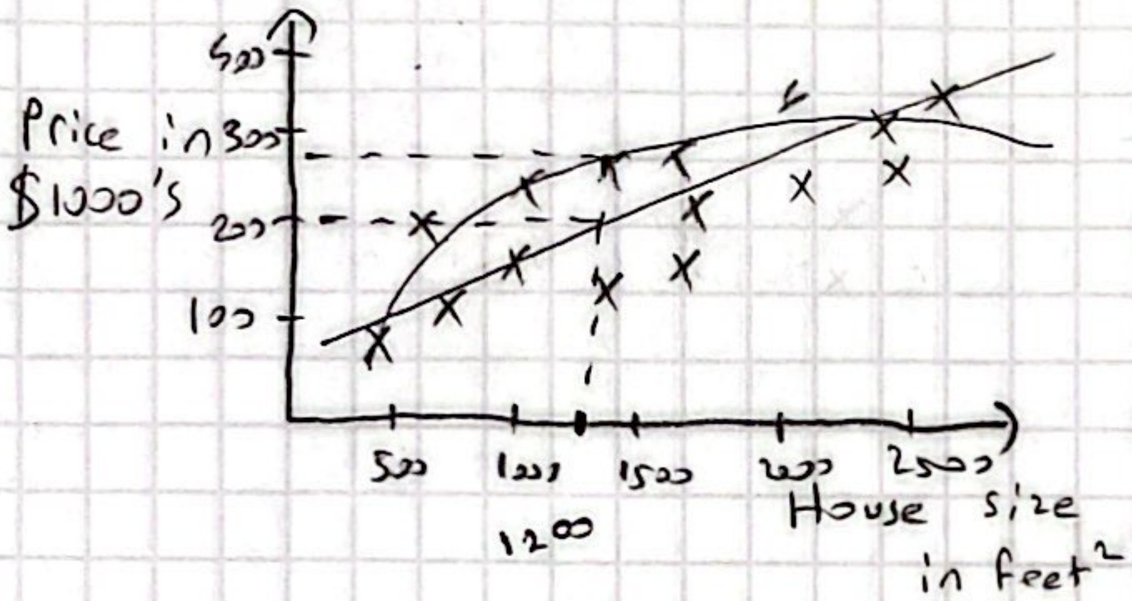
X
input

Y
output

X'leri ve Y'leri biz veriyoruz.
Sonrasında farklı bir X girdiğimizde en yakın Y değerini veriyor.

Input	Output	Application
email	spam?	spam filtering
audio	text transcription	speech recognition
English	Spanish	machine translation
ad/user info	click (0,1)	online advertising
image - radar info	position of other cars	self-driving car (otonom sürüş)
image of phone	defect (hata)	visual inspection

Regression's Housing price prediction



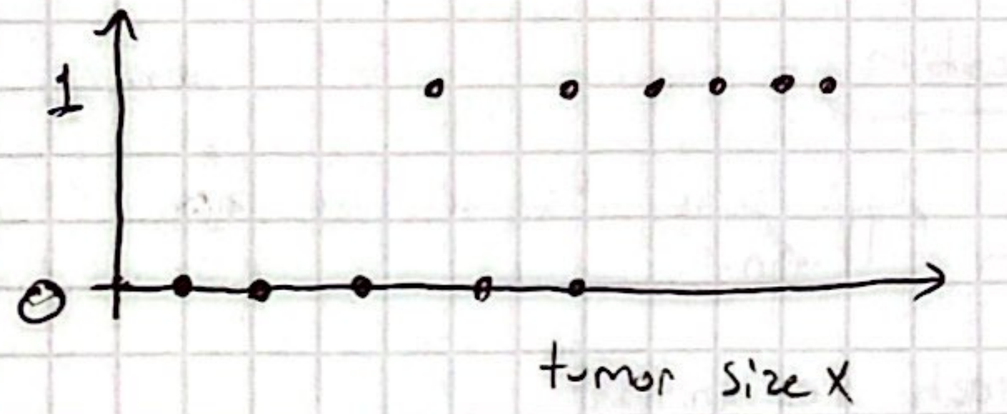
1200 m² evin fiyatı diğer farklı. 200 sıkıntı

" : : : eğri : 2700 sıkıntı

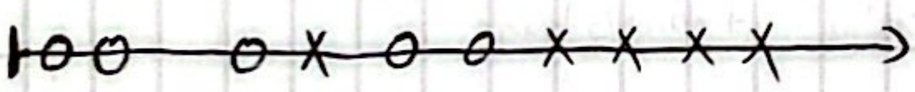
Regression : Predict number Sonsuz girdiye karşılık sonsuz çıktı

Classification : Breast cancer detetion

Sonsuz girdiye karşılık belli sayıda çıktı



O being (iyi huylu)
X malignant (kötü huylu)

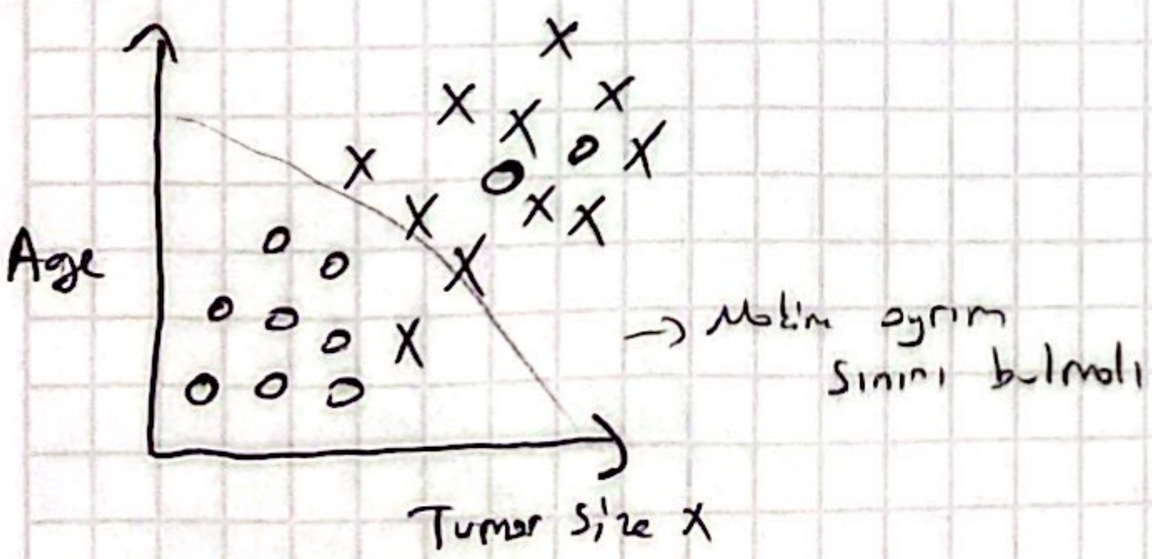


output = class = category

input = image

output = cat - dog -

cat, dog, benign, malign, 0-1-2-3)

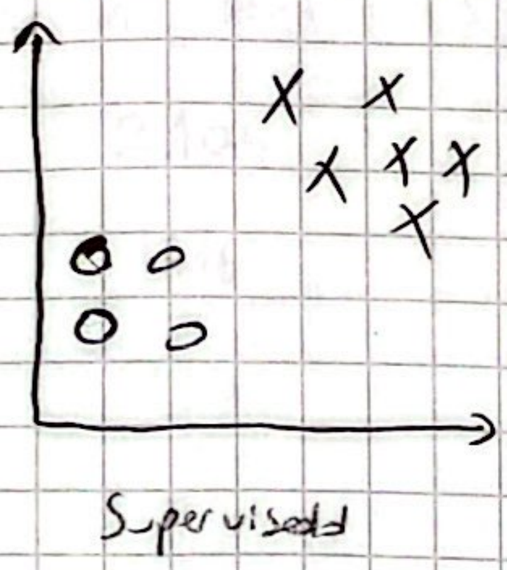


Unsupervised Learning

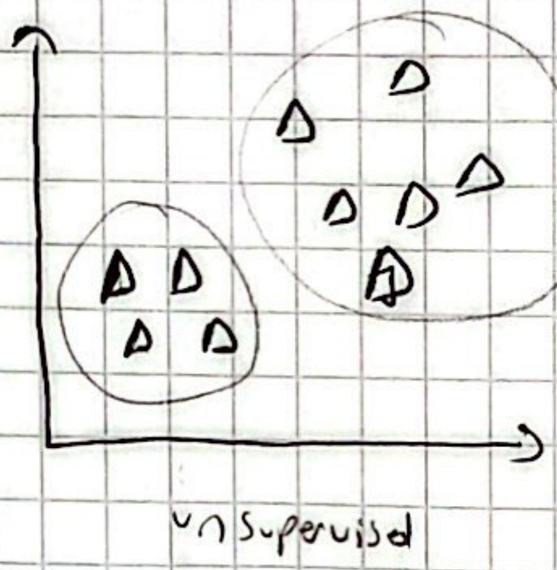
Now

X'leri biz veriyoruz

Yleri onun bulmasını istiyoruz



Supervised



Clustering

unsupervised

. Clustering : Google News

↳ Grup similar data points together

Anomaly Detections

Find unusual data points.

Dimensionality Reduction. (bayt indirgeme

Compress data using fewer numbers.

Regresyon Modeli

Training Set: Data used to train the model.

Notation:

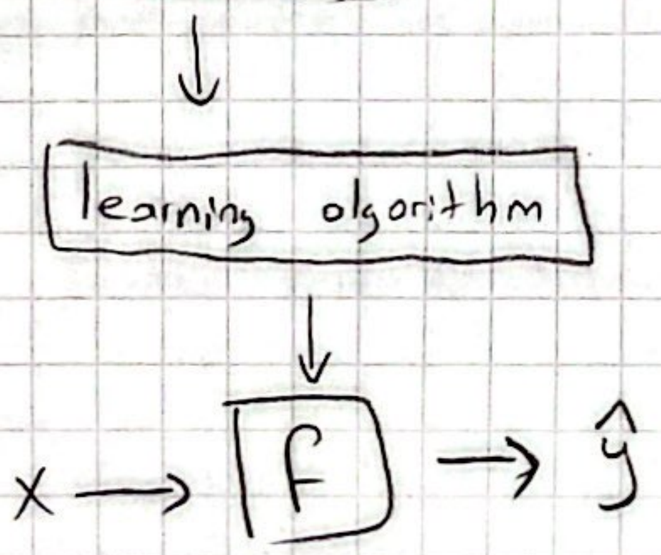
X = "input" variable Feature
    Feature

y = "output" variable
    "target" variable

m = number of training examples

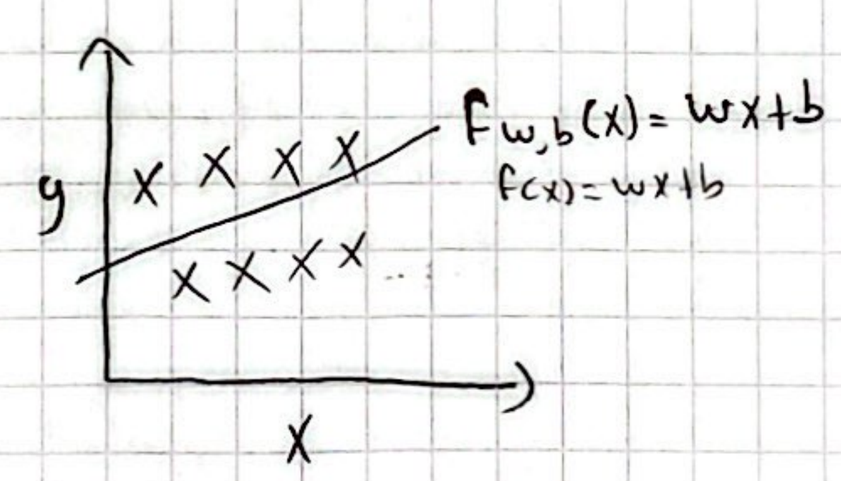
(x,y) = single training example

$(x^{(i)}, y^{(i)})$ = i^{th} training example



How to represent f?

$f_{w,b}(x) = wx+b$



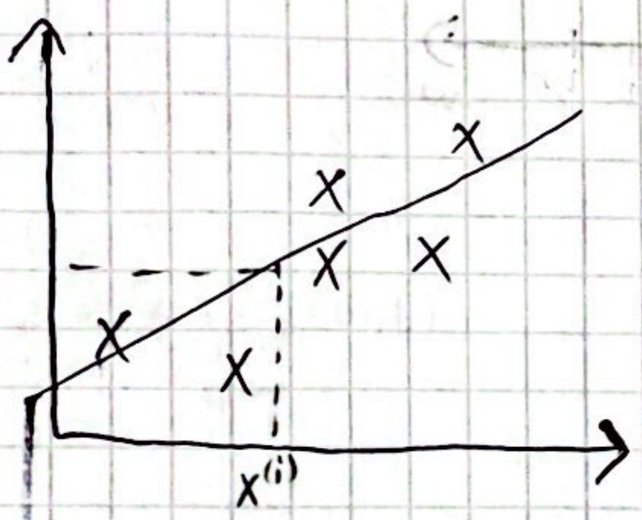
Linear regression with one variable

Univarite linear regression.

Cost Function

Training Set

Size in $feet^2$ (x)	price \$1000's (y)
2104	460
1416	232
1534	315
852	178



$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$

$f_{w,b} = wx+b$

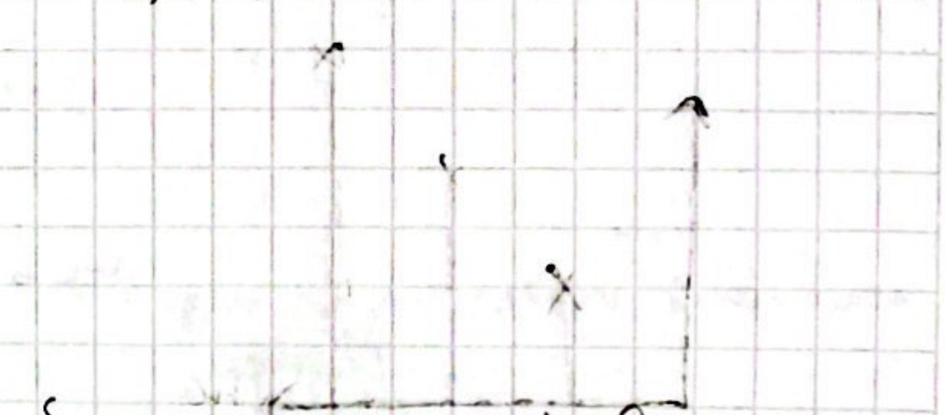
model: $f_{w,b}(x) = wx+b$

parameters: w, b

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

goal: minimize $J(w,b)$

$f_{w,b}(x) = wx+b$



Squared error cost function

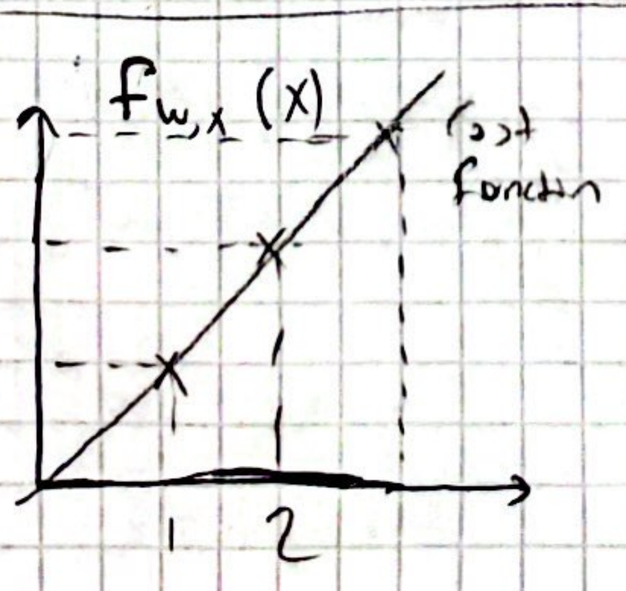
$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$

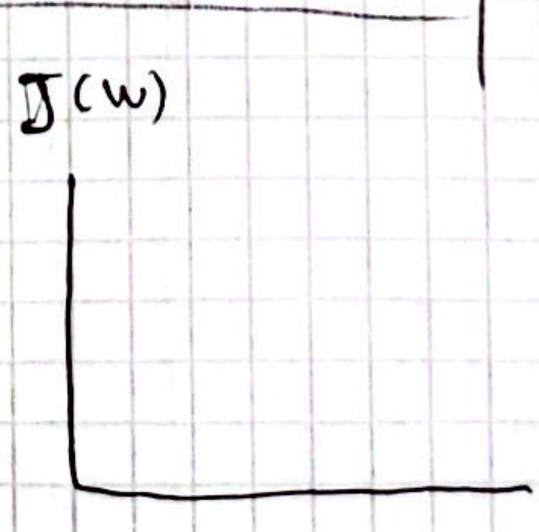
$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2$

Simplified

$f_w(x) = wx$

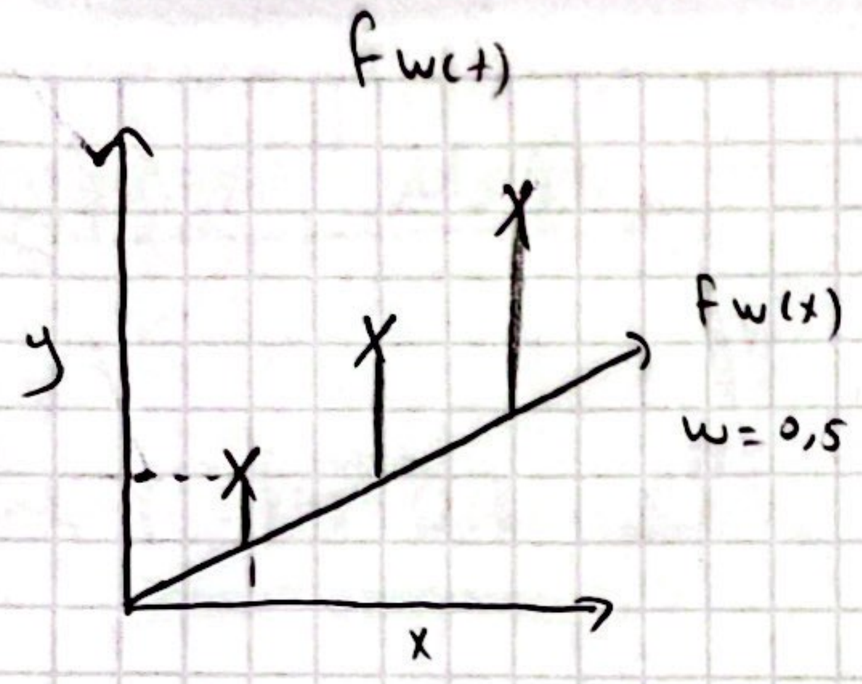
$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_w(x^{(i)}) - y^{(i)}\right)^2$

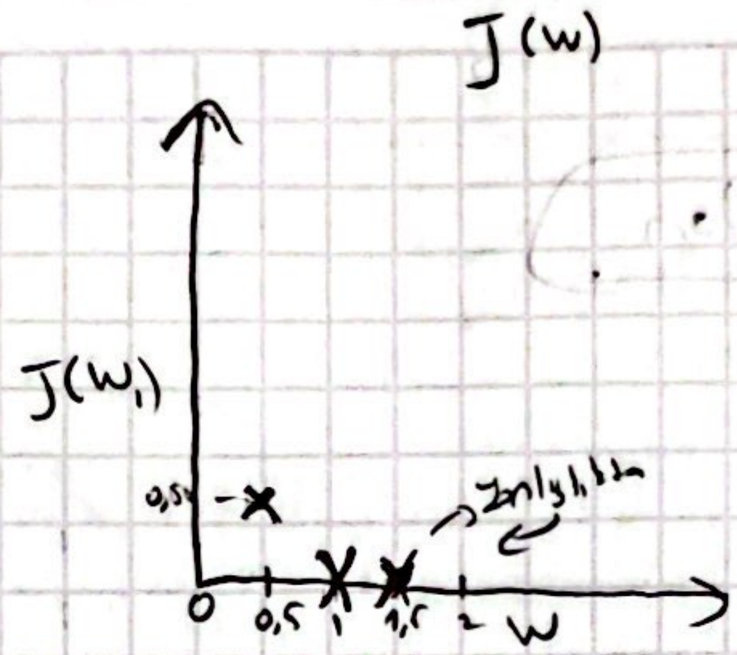




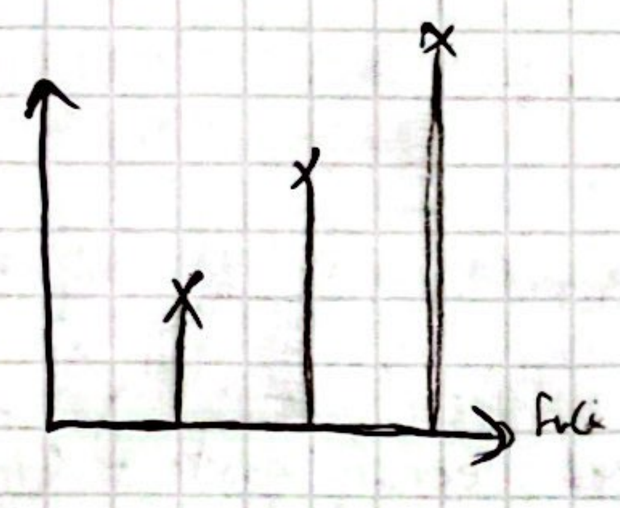
$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_w(x^{(i)}) - y^{(i)}\right)^2 = \frac{1}{2m} \sum_{i=1}^{m} \left(wx^{(i)} - y^{(i)}\right)^2$

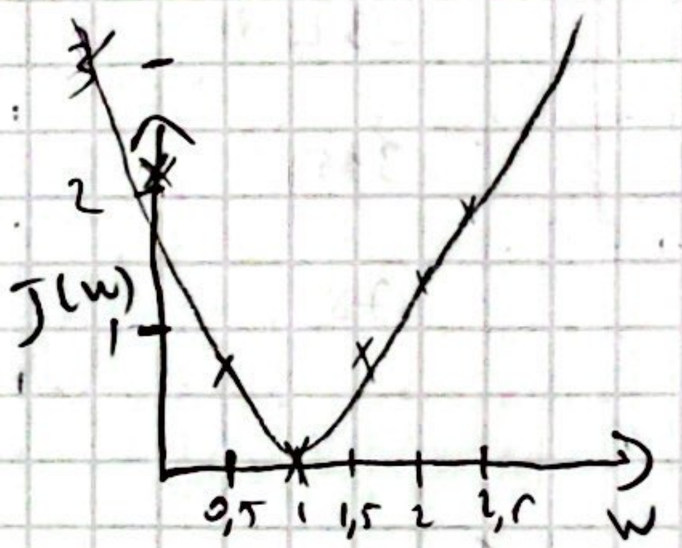
$= \frac{1}{2}(0+0+0) = 0$





$$J(0,5) = \frac{1}{2m}\left[(0,5-1)^2 + (1-2)^2 + (1,5-3)^2\right] = \frac{3.5}{6} \approx 0,58$$





$$J(0) = \frac{1}{2m}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right] \approx 2,3$$

goal of linear regression : minimize $J(w)$ $_w$

general case : minimize $J(w,b)$ $_{w,b}$

Visualizing the Cost Function

Gradient Descent (Gradyan iniş:)

Have some function $J(w,b)$ for linear regression or any function.

Want $\min_{w,b} J(w,b)$ $\quad\quad$ $\min_{w_1,w_2 \cdots w_n,b} J(w_1,w_2,\cdots w_n,b)$

Outline:

- Start with some w,b (set $w=0, b=0$)
- Keep changing w,b to reduce $J(w,b)$
- Until we settle at or near a minimum

* minimum noktası birden fazla olabilir.

$$W = w - \alpha \frac{dJ(w,b)}{dw}$$

α = learning rate
0-1 arasında olur

α büyükse doğru inen adımlar büyük

α küçükse " " " küçük

$\frac{dJ(w,b)}{dw}$ ⇒ adım atarken hangi yöne atacağımıza karar verme

$$b = b - \alpha \frac{dJ(w,b)}{db}$$

w-b değerlerini aynı anda güncellemeliyiz.

Correct: Simultaneous update

$$tmp\_w = w - \alpha \frac{dJ(w,b)}{dw}$$
$$tmp\_b = w - \alpha \frac{dJ(w,b)}{db}$$

$w = tmp\_w$

$b = tmp\_b$

Gradient Descent Intuition

(Gradyn iş sezgisi)

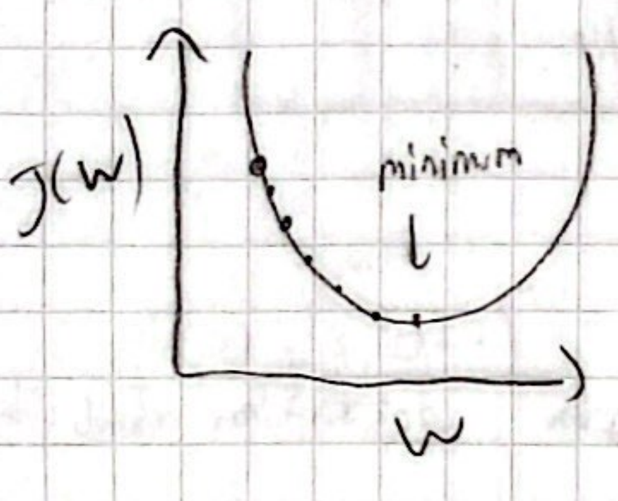
Learning Rate

$$W = W - \alpha \frac{d}{dw} J(w,b)$$

If α is too small

Gradient descent may be slow

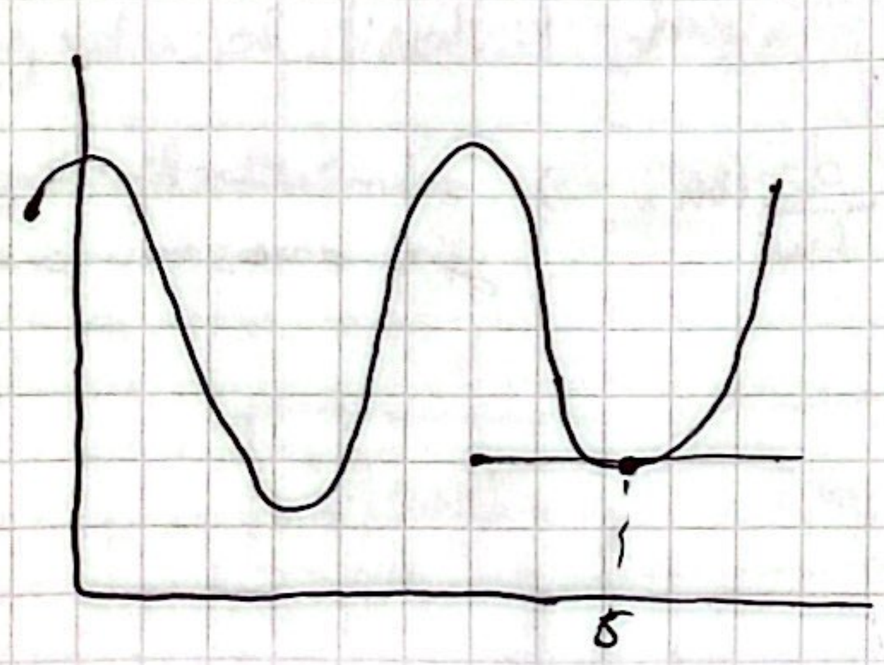
(yavaşlayabilir)



If α is too large

Gradient descent may:

- Overshoot (hedefi aşar), never reach minimum
- Fail to converge (yakınsayamaz), diverge (ıraksar)



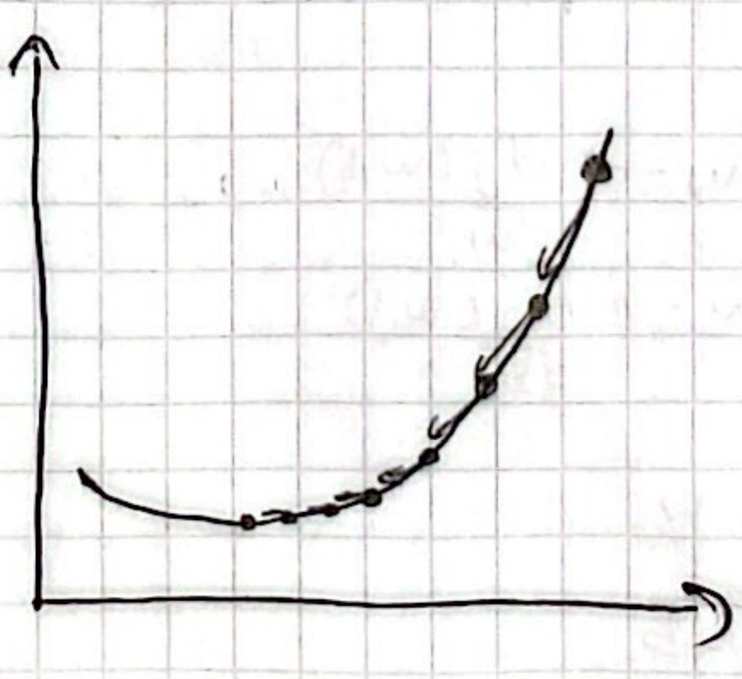
W değeri yerel minimum noktasına geldiğinde

W değeri güncellenmez. $W = W - \alpha \cdot 0 = W$

Yerel minimuma yaklaştıkça

→ Türev azalır

→ Adım büyüklüğü azalır



☆ α azaltmadan minimum'a ulaşabiliriz

Doğrusal Regresyon için Dereceli Azalma

linear Regression model

$$f_{w,b}(x) = wx + b$$

Cost function

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

gradient descent algorithm

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

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

repeat until convergence {

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

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

global minimum : yerel minimumların en küçüğü

Convex Function : Kase şeklinde fonk.

"Batch" gradient descent (Toplu derecli azalma

" Her dereceli azalma adımında tüm data seti kullanılır