Machine Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

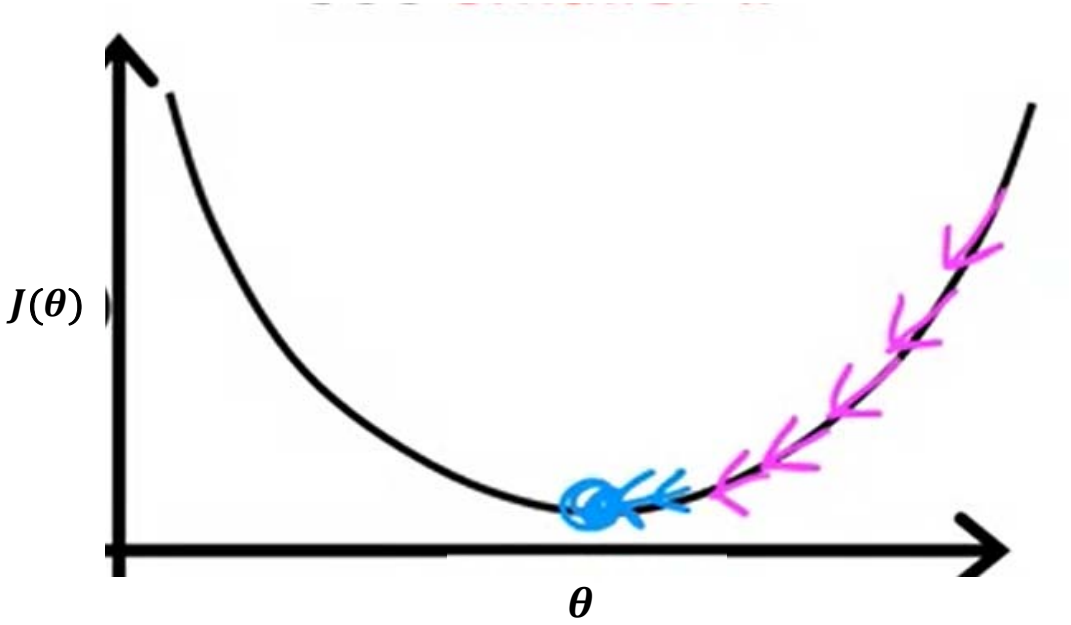
safayani.iut.ac.ir

https://www.aparat.com/mehran.safayani

https://github.com/safayani/machine_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

Gradient Descent



$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$\boldsymbol{\theta_1} = \boldsymbol{\theta_1} - \alpha \frac{d\boldsymbol{J}(\boldsymbol{\theta_1})}{d\boldsymbol{\theta_1}}$$

number	size	#bedrooms	# floors	Price(y)
1	100	2	1	10000
2	150	3	2	175000
…	…	…	…	…
m	…	…	…	…

n: #features = 3

m: #training data

x_i: i th data in training set

x_j^i: j th feature of i th data in training set

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

$$y = [y^1, y^2, \dots, y^m]^T \in R^{m*1}$$

$$X = [x^1, x^2, \dots, x^m]^T \in R^{m*(n+1)}$$

$$\vec{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix} \in R^{n+1} \quad , \quad \overrightarrow{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ . \\ . \\ \theta_n \end{bmatrix} \in R^{n+1} \Rightarrow h_\theta(x) = x^T\theta = \theta^T x$$

$x_0 = 1$

θ_0 ***is bias***

Cost function

$$J(\overrightarrow{\theta}) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

$$e^i = x^T \theta - y_i \longrightarrow e = X\theta - y \longrightarrow J(\theta) = \frac{1}{2m} e^T e$$

e , $X\theta$, $y \in R^m$

Gradient Descent

Repeat until convergence:

For j=0,…,n

$$\boldsymbol{\theta_j} = \boldsymbol{\theta_j} - \alpha \frac{d\boldsymbol{J(\theta_0, \theta_1, \ldots, \theta_n)}}{d\boldsymbol{\theta_j}}$$

$$\frac{d\boldsymbol{J(\theta_0, \theta_1)}}{d\boldsymbol{\theta_0}} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)$$

$$\frac{d\boldsymbol{J(\theta_0, \theta_1, \ldots, \theta_n)}}{d\boldsymbol{\theta_j}} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)\, x_j^i$$

(j=0,…,n , $x_0^i = 1$)

$$\frac{d\boldsymbol{J(\theta)}}{d\boldsymbol{\theta}} = \frac{1}{m} X^T \mathrm{e}$$

(n+1)*1

m*1

حجم محاسبات ضرب ماتریس

$A \in R^{a*b}$, $B \in R^{b*c}$ ⟹ $AB \in R^{a*c}$ $(2b-1)\,ac\ flops$

$O(abc)$

$Calculating: \ e = X\theta - y$

a=m
b=n+1
c=1

$m(2n+1)$ (ضرب و جمع)

m (تفریق)

2m(n+1) (مجموع) ⟹ O(mn)

$\frac{dJ(\theta)}{d\theta}: (2m-1)(n+1) + (n+1) = 2m(n+1)$ ⟹ O(mn) در مجموع

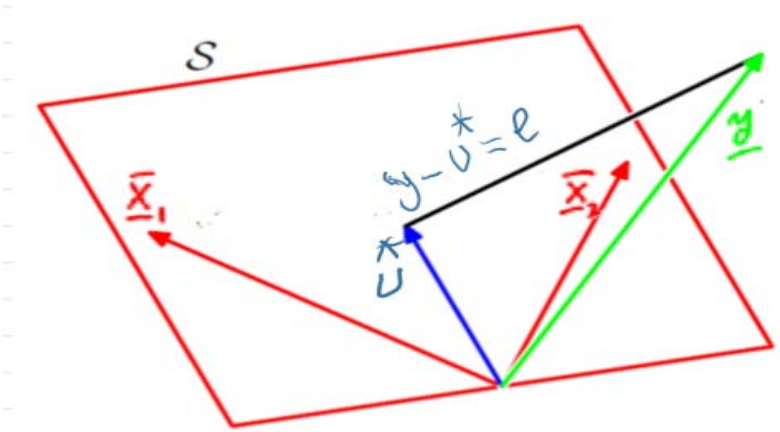
تقسیم بر m

$\frac{1}{m} X^T e$

a=n+1
b=m
c=1

مفهوم هندسی

$$\underset{W}{\text{Min}} \|y - XW\|_2 = \text{Min} \|e\|_2$$



تعداد ویژگی * تعداد داده

Span of X:

فضایی که توسط ستون های X پوشش داده می شود. هر بردار در این فضا به صورت U = XW نشان داده می شود. و به آن span(X) می گویند. U بهینه که به صورت U* نشان داده می شود برداری است که e = y – U* بر span(X) عمود باشد. یا به عبات دیگر U*ای باید انتخاب شود که برابر با نگاشت y در span(X) باشد.

Feature Scaling

$x_1^i, x_2^i, \ldots, x_n^i$

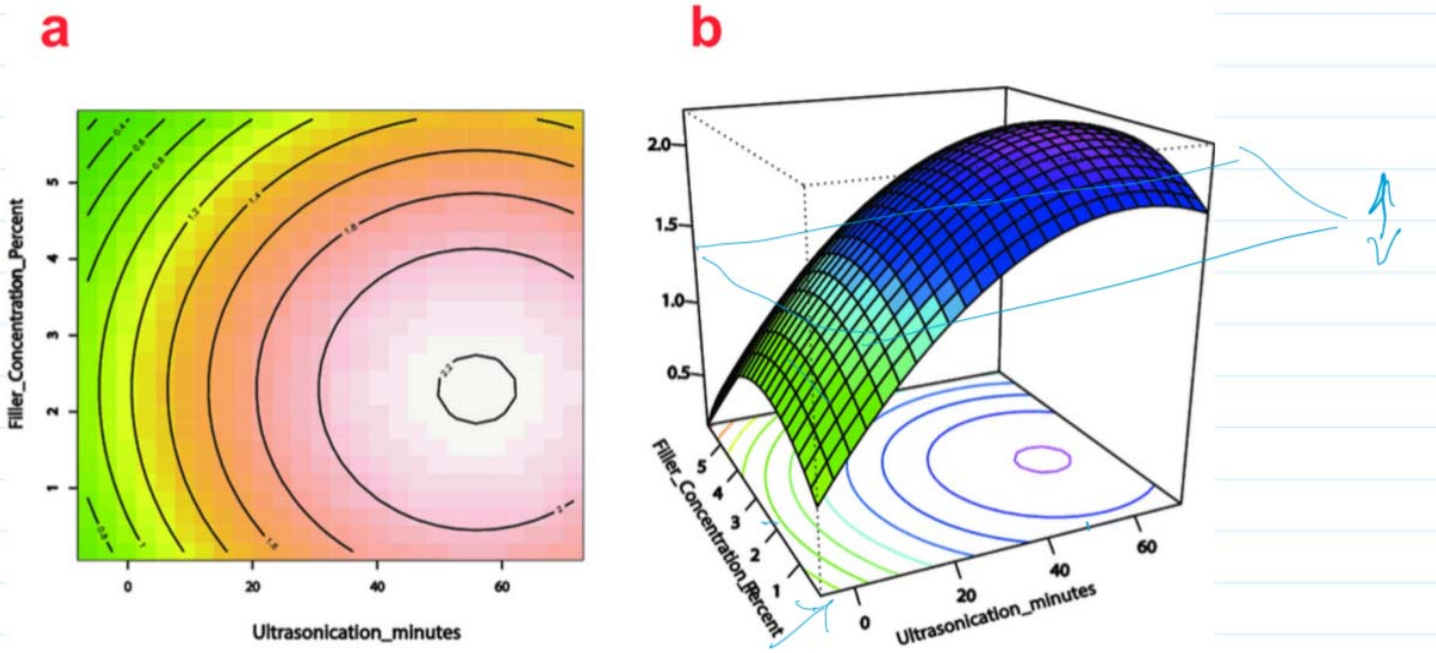
$-1 \leq x_j \leq 1$

$0 < x_1 < 1000$

$0 < x_2 < 5$

$x_1: \frac{size}{1000}$

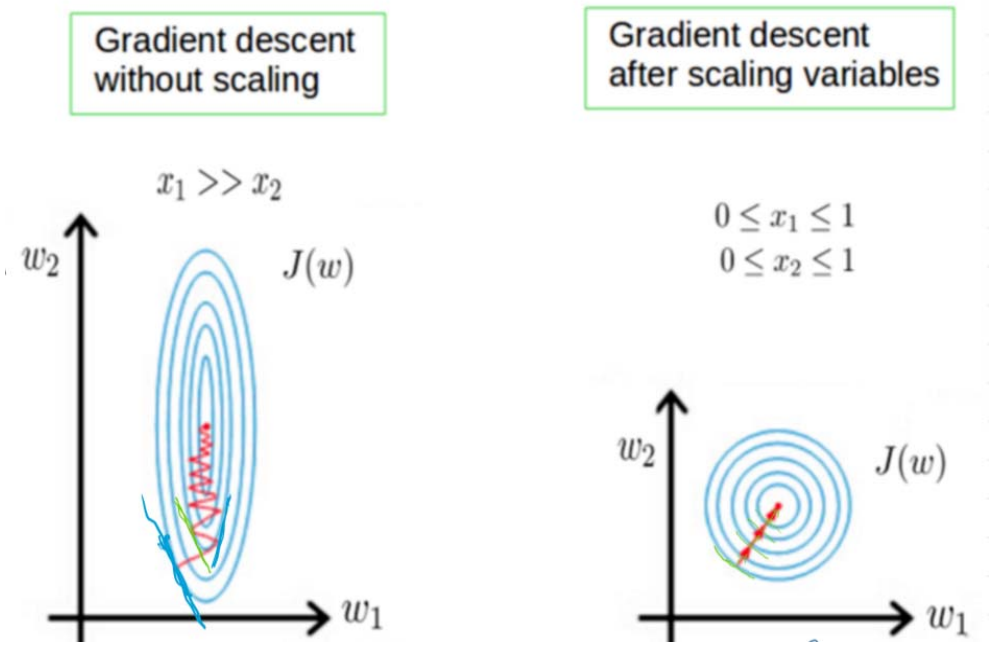
$x_2: \frac{\#bedrooms}{5}$


a
b
Filler_Concentration_Percent
Ultrasonication_minutes
0
20
40
60
1
2
3
4
5
2.0
1.5
1.0
0.5

Contour Plot

$$\frac{w_1^2}{b^2} + \frac{w_2^2}{a^2} = 1$$

2a: قطر بزرگ
2b: قطر کوچک



$$\frac{w_1^2}{a^2} + \frac{w_2^2}{a^2} = 1$$

Feature Scaling

Scaled features:

- $0 \leq x_1 \leq 3$ ✓
- $-3 \leq x_1 \leq 3$ ✓
- $-2 \leq x_2 \leq 0.5$ ✓
- $-\frac{1}{3} \leq x_2 \leq \frac{1}{3}$ ✓

Need scaling:

$-100 \leq x_3 \leq 100$ ✗

$-0.001 \leq x_4 \leq 0.001$ ✗

Feature Scaling

$$x_1^* = \frac{x_1 - \mu_1}{standard\_deviation}$$

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

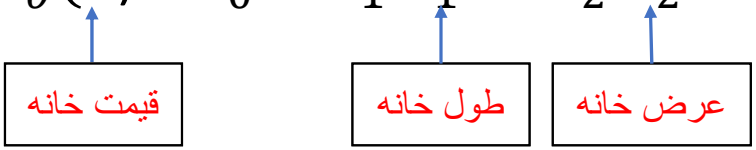
$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} x_j^i$$

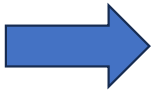
$$bedroom^* = \frac{bedroom - 2.5}{5}$$

$$size^* = \frac{size - 300}{2000}$$

Creating New Features

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

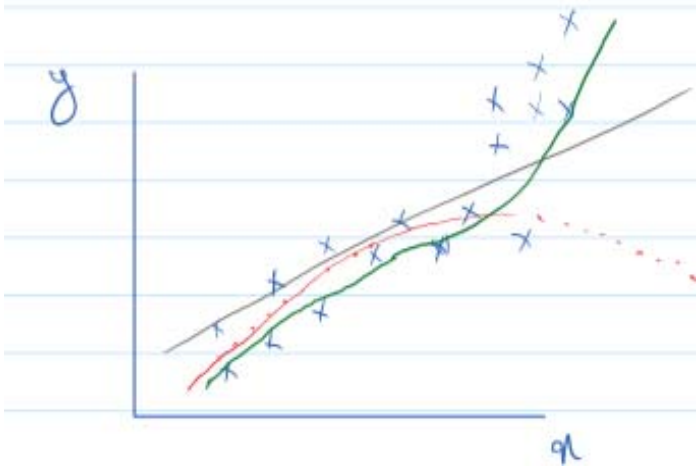




$$x^* = x_1 * x_2$$

$$h_\theta(x) = \theta_0 + \theta_1 x^*$$

Creating New Features



We can use:

$x , x^2 , x^3 , \sqrt{x}$

$\theta_0 + \theta_1 x + \theta_2 \sqrt{x}$

$\theta_0 + \theta_1 x + \theta_2 x^2$

درجه 3:

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$

Need scaling:

x: 0,…,1000

x^2: 0,…, 10^6

x^3: 0,…, 10^9