



Machine Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir



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

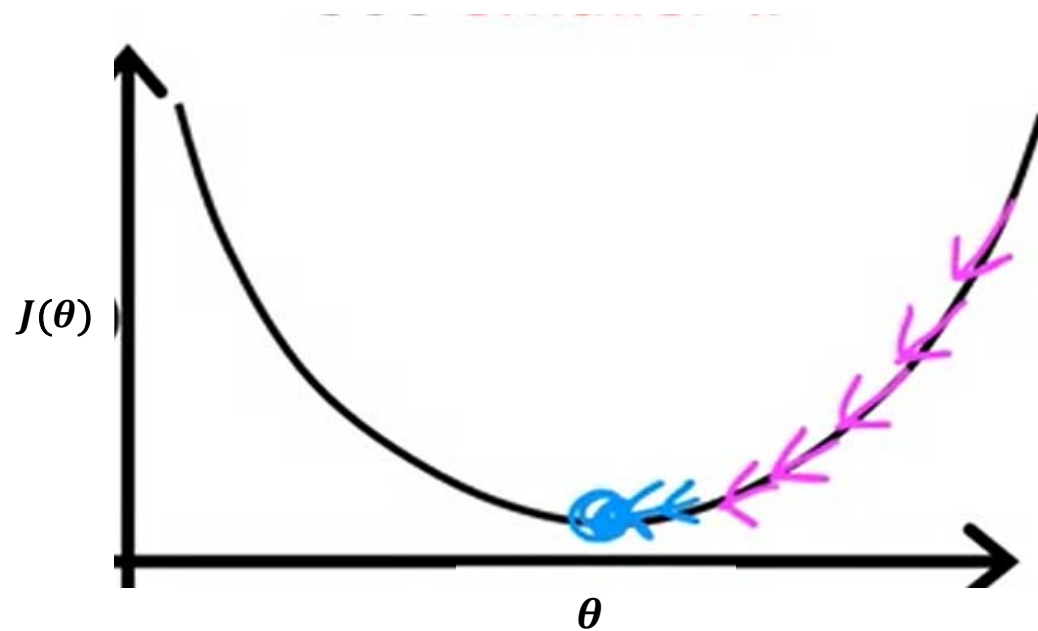


https://github.com/safayani/machine_learning_course



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

Gradient Descent



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

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

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

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

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

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

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

$x_0 = 1$ θ_0 is bias

Cost function

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

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

$$e, X\theta, y \in \mathbb{R}^m$$

Gradient Descent

Repeat until convergence: {

For $j=0,\dots,n$

$$\theta_j = \theta_j - \alpha \frac{dJ(\theta_0, \theta_1, \dots, \theta_n)}{d\theta_j}$$

}

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

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

($j=0,\dots,n$, $x_0^i = 1$)

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

حجم محاسبات ضرب ماتریس

$$A \in R^{a \times b}, \quad B \in R^{b \times c} \quad \longrightarrow \quad AB \in R^{a \times c} \quad (2b - 1) ac \text{ flops}$$

$O(abc)$

Calculating: $e = X\theta - y$

$$\begin{aligned} a &= m \\ b &= n+1 \\ c &= 1 \end{aligned}$$

$$\begin{array}{ll} m(2n + 1) & \text{(ضرب و جمع)} \\ m & \text{(تفریق)} \end{array}$$

$$\left. \begin{array}{l} m(2n + 1) \\ m \end{array} \right\} 2m(n+1) \text{ (مجموع)} \quad \longrightarrow \quad O(mn)$$

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

$$\begin{array}{c} \uparrow \\ \text{تقسیم بر } m \end{array}$$

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

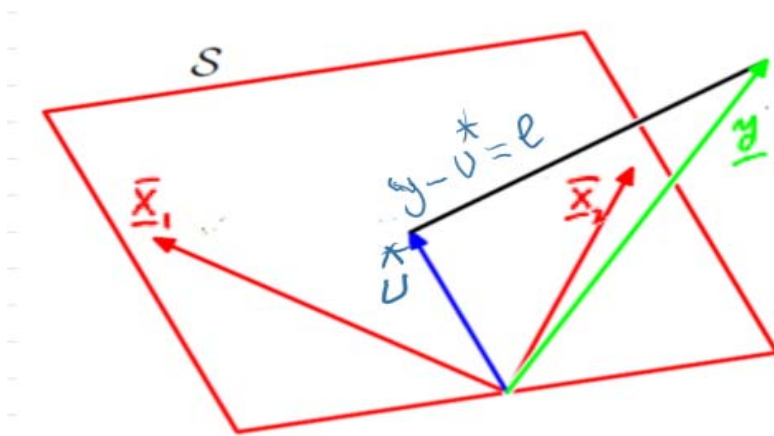
$$\begin{aligned} a &= n+1 \\ b &= m \\ c &= 1 \end{aligned}$$

مفهوم هندسی

$$\min_W \|y - XW\|_2 = \min \|e\|_2$$

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 $\text{span}(X)$ می گویند. U بهینه که به صورت U^* نشان داده می شود برداری است که $e = y - U^*$ بر $\text{span}(X)$ عمود باشد. یا به عبارت دیگر U^* ای باید انتخاب شود که برابر با نگاشت y در $\text{span}(X)$ باشد.



تعداد ویژگی * تعداد داده

Feature Scaling

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

$$-1 \leq x_j \leq 1$$

$$0 < x_1 < 1000$$

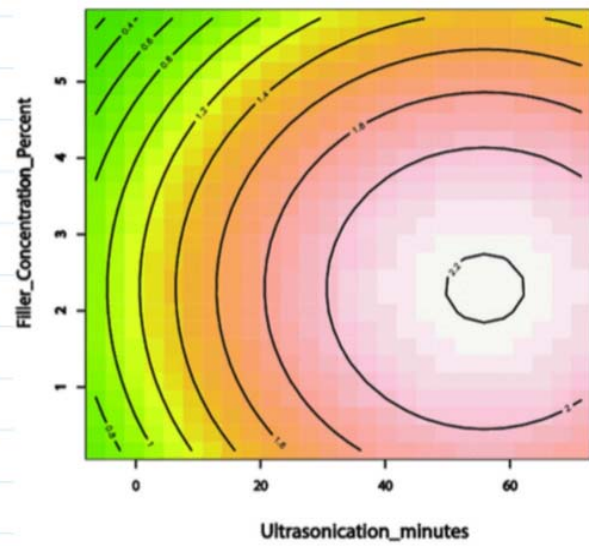
$$0 < x_2 < 5$$



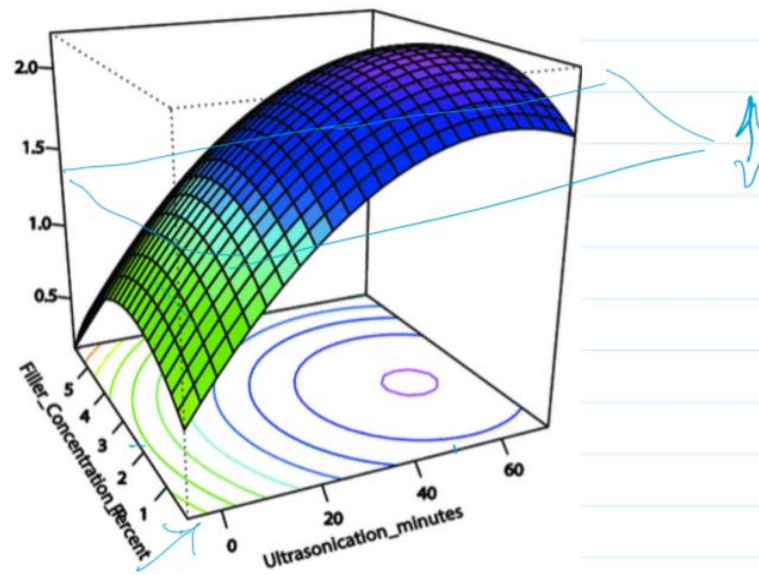
$$x_1: \frac{size}{1000}$$

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

a



b



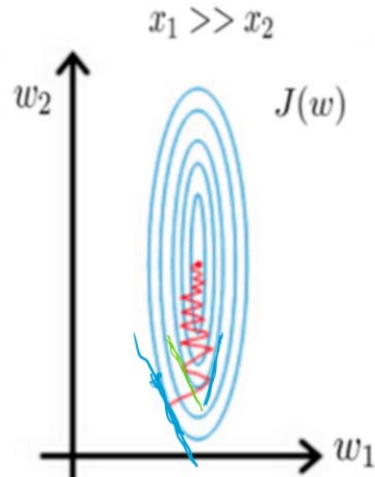
Contour Plot

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

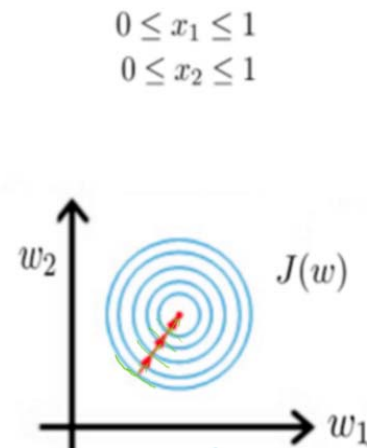
2a: قطر بزرگ

2b: قطر کوچک

Gradient descent
without scaling



Gradient descent
after scaling variables



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

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:

$$-100 \leq x_3 \leq 100 \quad \times$$

$$-0.001 \leq x_4 \leq 0.001 \quad \times$$

Feature Scaling

$$x_1^* = \frac{x_1 - \mu_1}{\text{standard_deviation}}$$


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

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

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

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

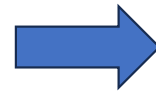
Creating New Features

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


قیمت خانه

طول خانه

عرض خانه

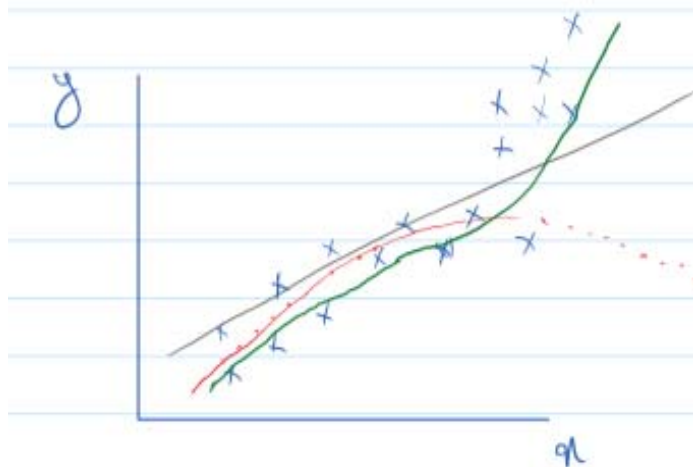


تعریف ویژگی جدید

$$x^* = x_1 * x_2$$

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

Creating New Features



We can use:

$$x, x^2, x^3, \sqrt{x}$$
$$\theta_0 + \theta_1 x + \theta_2 \sqrt{x}$$

درجه 2:

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

درجه 3:

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

Need scaling:

$x: 0, \dots, 1000$

$x^2: 0, \dots, 10^6$

$x^3: 0, \dots, 10^9$