

یادگیری ماشین با پایتون

فرزاد حسینیان

دانشگاه محقق اردبیلی

دانشکده علوم کامپیوتر

آبان 98

جلسه دوم

رگرسیون خطی چندمتغیره

رگرسون خطی چندمتغیره

No	Size(m*m) X_1	# Bedrooms X_2	# Floors X_3	Age X_4	Price(M Toman) y
1	50	1	1	7	80
2	53	1	1	10	75
3	55	1	2	5	78
4	60	2	1	3	82
...	...	...	...	...	...
15	145	2	2	1.5	200
16	150	1	2	12	150
17	200	2	3	2	320

Notations:

$n = \# \text{ features}$

$m = \# \text{ Training set Examples}$

$x^{(i)}$ = i^{th} example (vector)

$x_j^{(i)}$ = j^{th} input variable for the i^{th} example

No	Size(m*m) x_1	# Bedrooms x_2	# Floors x_3	Age x_4	Price(M Toman) y
1	50	1	1	7	80
2	53	1	1	10	75
3	55	1	2	5	78
4	60	2	1	3	82
...	...	...	...	...	...
15	145	2	2	1.5	200
16	150	1	2	12	150
17	200	2	3	2	320

$n = 4, m = 17$

$$x^{(2)} = \begin{bmatrix} 53 \\ 1 \\ 1 \\ 10 \end{bmatrix}, x^{(16)} = \begin{bmatrix} 150 \\ 1 \\ 2 \\ 12 \end{bmatrix}$$

$x_2^{(4)} = 2$ $x_4^{(17)} = 2$

تابع فرضیه رگرسیون خطی چندمتغیره

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

$$= \sum_{i=0}^n \theta_i x_i$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$h_{\theta}(x) = \theta^T x$$

تابع هزینه

الگوریتم گرادیان خطی

Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

(for $j = 1$ and $j = 0$)

}

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

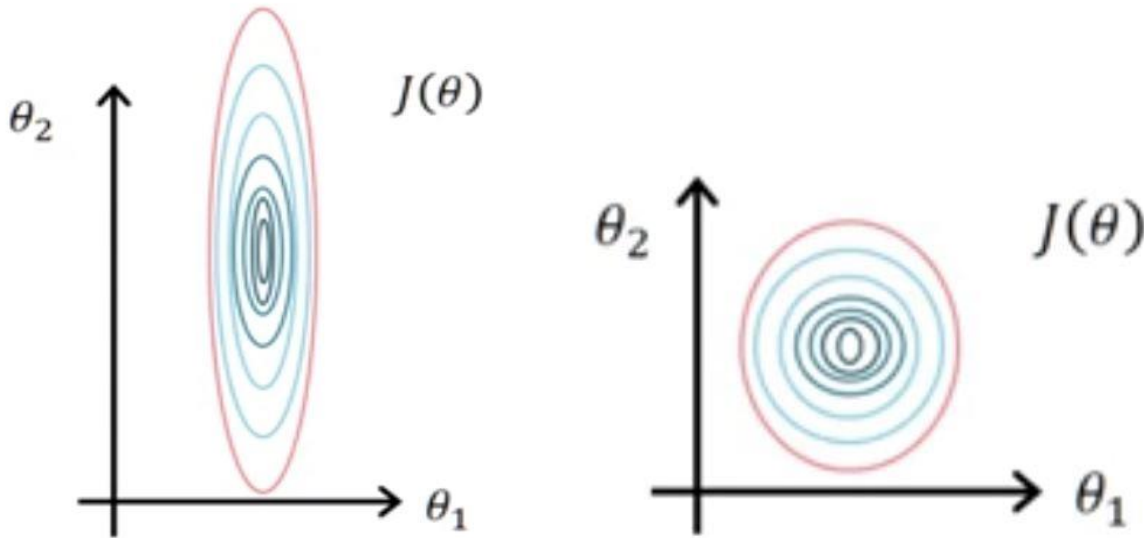
↑
 $\theta^T x$

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

کد نویسی پایتون

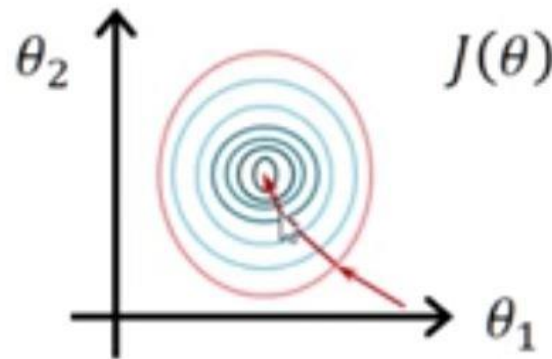
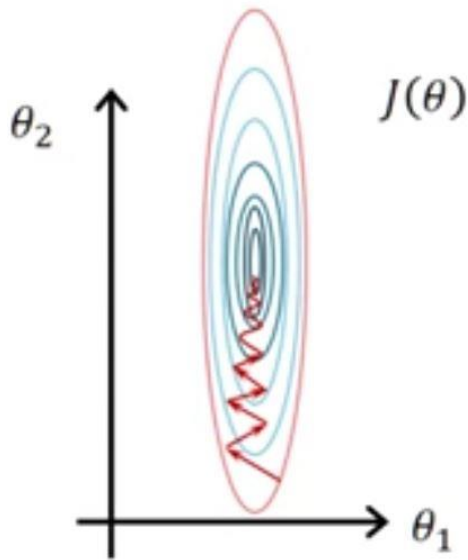
Feature scaling

ایده: اطمینان از اینکه ویژگی‌ها در مقیاس مشابه قرار دارند
دلیل: همگرایی سریعتر



Feature scaling

ایده: اطمینان از اینکه ویژگی ها در مقیاس مشابه قرار دارند
دلیل: همگرایی سریعتر

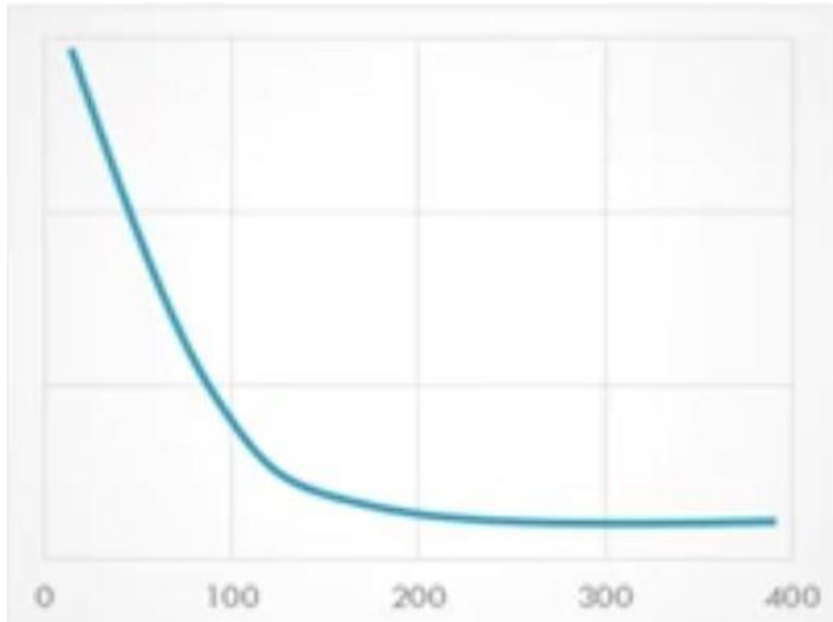


Mean Normalization:

Replace each x_i with $\frac{x_i - \mu_i}{s_i}$
(Not for x_0)

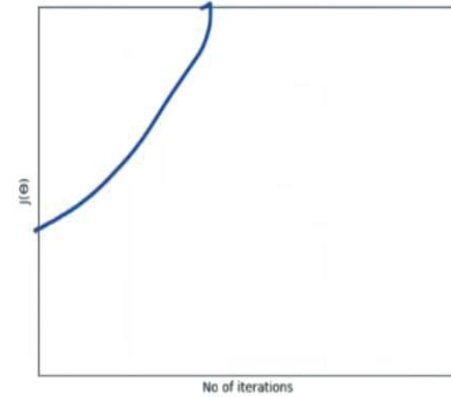
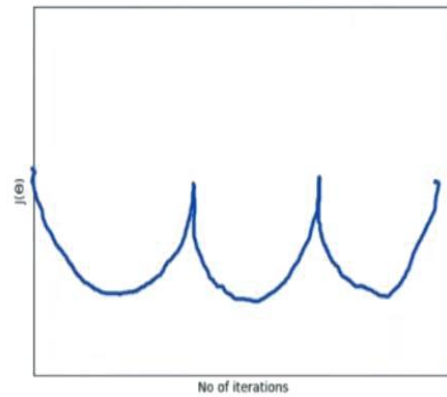
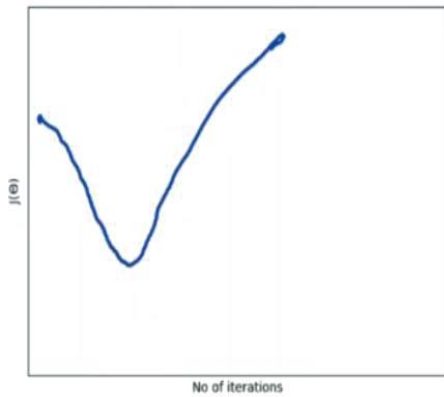
نمودار همگرایی صحیح

هزینه



تعداد تکرار

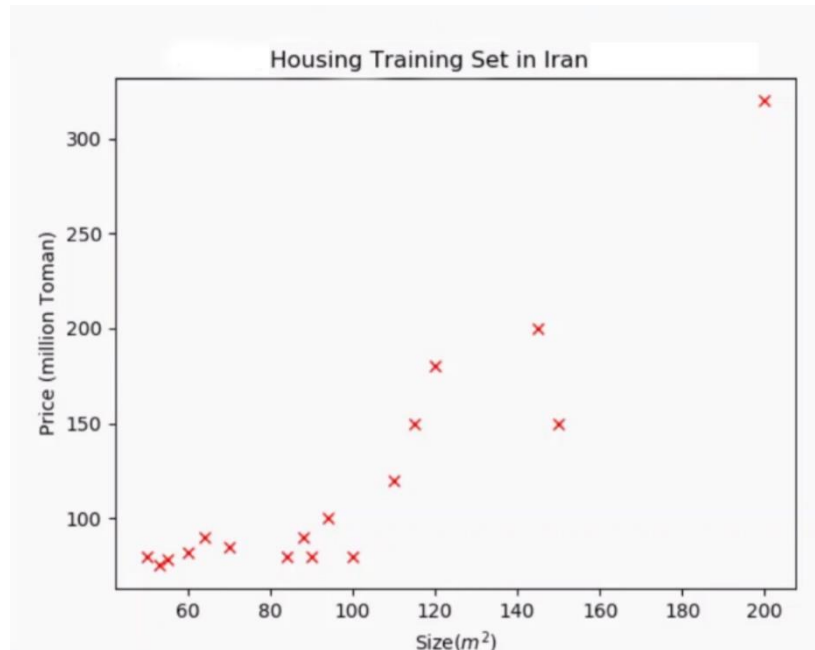
آلفا به اندازه کافی کوچک نیست

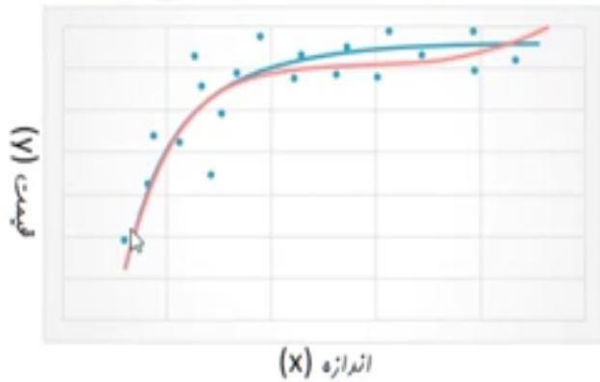


انتخاب آلفا

..., 0.001, 0.003, 0.01, 0.03, .01, 0.3, 1.0, ...

Polynomial regression





رگرسیون چند جمله‌ای.

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$



$$\begin{aligned} x_1 &= (Size)^1 & 1 \leq x_1 \leq 10^3 \\ x_2 &= (Size)^2 & 1 \leq x_2 \leq 10^6 \\ x_3 &= (Size)^3 & 1 \leq x_3 \leq 10^9 \end{aligned}$$