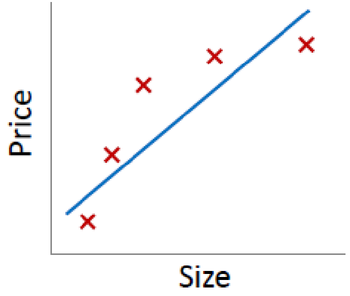
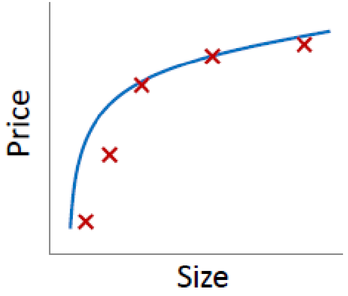
فصل پنجم : مثال(رگرسیون چند جمله ای)

$\theta_0 + \theta_1 x$



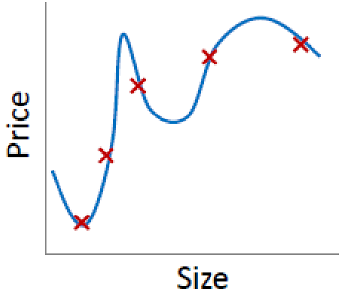
کم برازش (بایاس بالا)

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



مدل صحیح

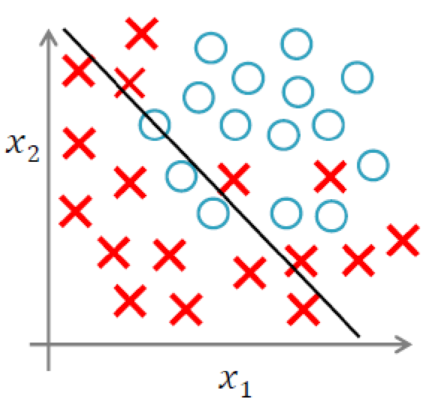
$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$



بیش برازش (واریانس بالا)

- بیش‌برازش. اگر تعداد ویژگی‌ها بسیار زیاد باشد، فرضیه یاد گرفته شده ممکن است داده‌های آموزشی را خیلی خوب یاد بگیرد، اما این امکان نیز وجود دارد که این فرضیه در پیش‌بینی داده‌های جدید شکست بخورد. [عدم قابلیت تعمیم]

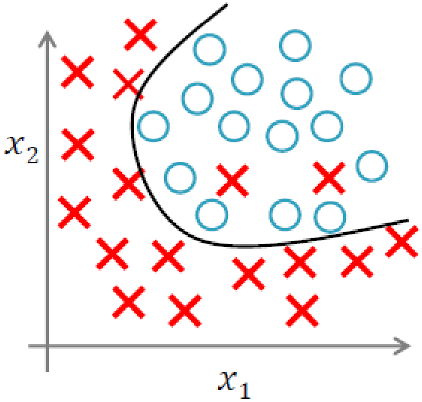
فصل پنجم : مثال(رگرسیون لجستیکی)



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

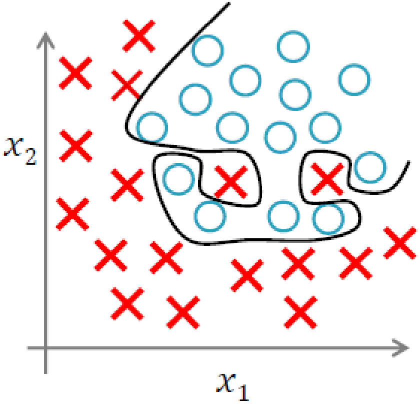
(g = sigmoid function)

کم‌برازش (بایاس بالا)



$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2$

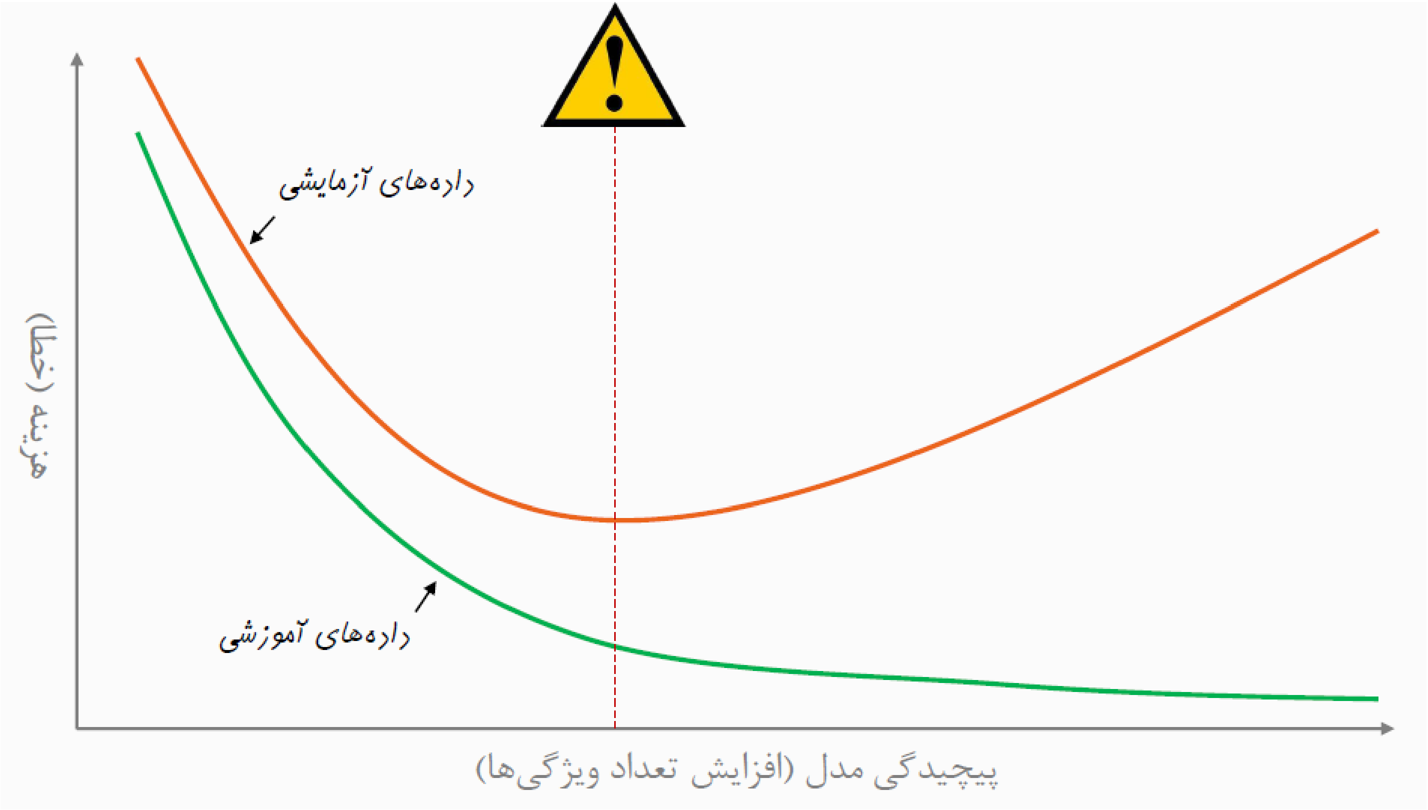
مدل صحیح



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2$$

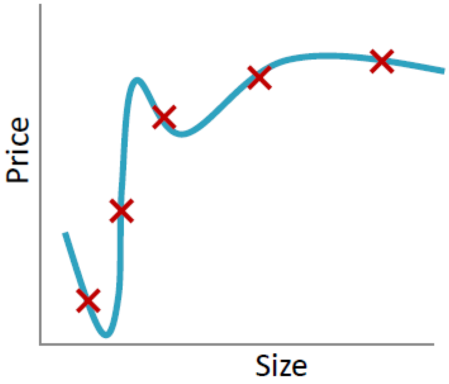
بیش‌برازش (واریانس بالا)

فصل پنجم : بیش برازش



فصل پنجم : برخورد با بیش برازش

- ویژگی‌ها.
  - x_1 : اندازه خانه
  - x_2 : تعداد اتاق‌ها
  - x_3 : تعداد طبقات
  - x_4 : قدمت
  - x_5 : اندازه آشپزخانه
  - x_6 : تعداد سرویس‌ها
  - ...
  - x_{100} : میانگین درآمد همسایه‌ها

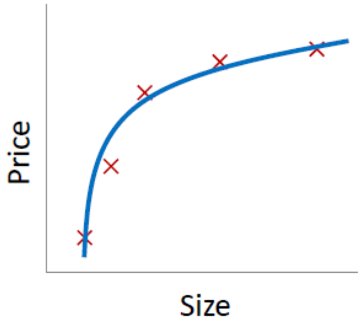


فصل پنجم : برخورد با بیش برازش

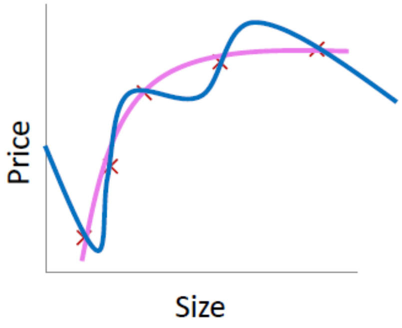
- راه‌حل‌های ممکن.
  - کاهش تعداد ویژگی‌ها.
    - به صورت دستی ویژگی‌های مهم‌تر را انتخاب و بقیه را حذف کن.
    - الگوریتم‌های انتخاب مدل [در ادامه]
  - تنظیم. (رگولاریزاسیون)
    - همه ویژگی‌ها را نگهدار، اما مقدار پارامترهای θ_j را کاهش بده.
    - زمانی که ویژگی‌های بسیاری داریم که هر کدام سهم اندکی در پیش‌بینی مقدار خروجی دارند، این روش به خوبی عمل می‌کند.

تابع هزینه

فصل پنجم : مفهوم تنظیم



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

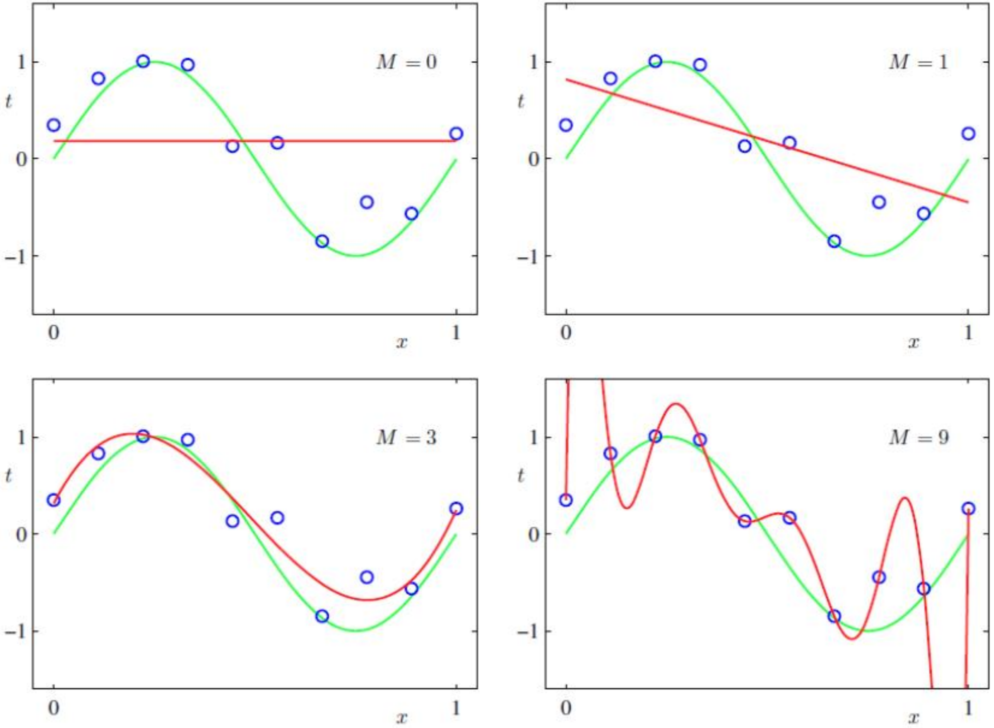


$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

- تابع هزینه.
  - می‌توان با جریمه کردن تابع هزینه، مقادیر پارامترهای θ۳ و θ۴ را بسیار کوچک نمود:

$$\min_\theta \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \underbrace{1000\theta_3^2}_{\approx 0} + \underbrace{1000\theta_4^2}_{\approx 0}$$

فصل پنجم : مفهوم تنظیم



	$M = 0$	$M = 1$	$M = 6$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

فصل پنجم : تنظیم

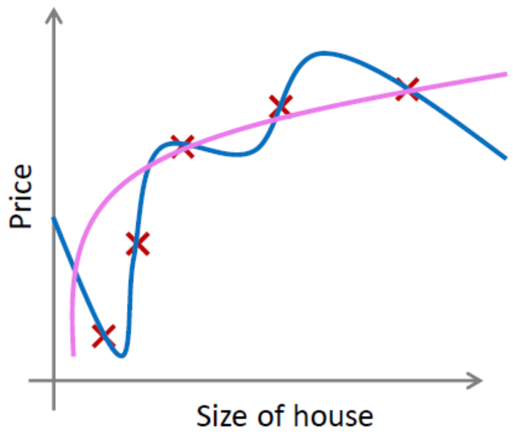
- تنظیم. استفاده از مقادیر کوچک برای پارامترهای «تتا»
  - ایجاد فرضیه‌های «ساده‌تر»
  - «اصل تراش اوکام»: تراشیدن اجزای غیر ضروری از مدل
  - کاهش خطر بیش‌برازش

- مثال.
  - ویژگی‌ها: x_1، x_2، ... و x_{100}
  - پارامترها: θ_0، θ_1، θ_2، ... و θ_{100}

$$J(\theta) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2\right]$$

فصل پنجم : تنظیم

- تنظیم.



ضریب تنظیم

$$J(\theta) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2\right]$$

$$\min_\theta J(\theta)$$

فصل پنجم : پرسش کلاسی

- در رگرسیون خطی تنظیم شده، مقادیر پارامترها به گونه‌ای انتخاب می‌شوند که مقدار تابع هزینه کمینه گردد.

$$J(\theta) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2\right]$$

س. اگر ضریب تنظیم λ را با یک مقدار بسیار بزرگ مقداردهی کنیم (مثلاً $۱۰^{۱۰}$)، در این صورت چه خواهد شد؟

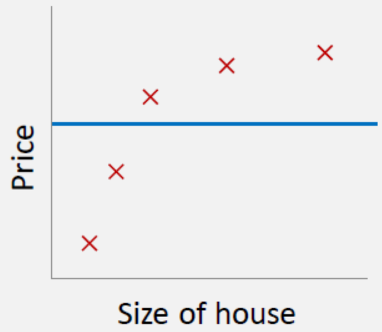
- الگوریتم به خوبی کار می‌کند و مقدار بزرگ λ آسیبی به آن وارد نمی‌کند.
- الگوریتم در برخورد با بیش‌برازش شکست می‌خورد.
- الگوریتم با کم‌برازش پایان می‌پذیرد.
- الگوریتم گرادیان کاهشی همگرا نمی‌شود.

فصل پنجم : پرسش کلاسی

- در رگرسیون خطی تنظیم شده، مقادیر پارامترها به گونه‌ای انتخاب می‌شوند که مقدار تابع هزینه کمینه گردد.

$$J(\theta) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda \sum_{j=1}^{n} \theta_j^2\right]$$

س. اگر ضریب تنظیم λ را با یک مقدار بسیار بزرگ مقداردهی کنیم (مثلاً ۱۰^۱۰)، در این صورت چه خواهد شد؟



$$\theta_0 + \underset{\approx 0}{\cancel{\theta_1}} x + \underset{\approx 0}{\cancel{\theta_2}} x^2 + \underset{\approx 0}{\cancel{\theta_3}} x^3 + \underset{\approx 0}{\cancel{\theta_4}} x^4$$

رگرسیون خطی تنظیم شده

فصل پنجم : رگرسیون خطی تنظیم شده

- تابع هزینه.

$$J(\theta) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right]$$

$$\min_\theta J(\theta)$$

فصل پنجم : گرادیان کاهشی (بدون تنظیم)

- بدون استفاده از تنظیم.

repeat until convergence {

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} \qquad (j = 0,1,2, \dots, n)$$

}

فصل پنجم : گرادیان کاهشی (با تنظیم)

- با استفاده از تنظیم.

repeat until convergence {

$$\theta_0 = \theta_0 - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_0^{(i)}$$

$$\theta_j = \theta_j - \alpha \left[\sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} + \lambda\theta_j\right] \qquad (j = 1,2, \dots, n)$$

}

$$\theta_j = \theta_j \underbrace{(1 - \alpha\lambda)}_{<1} - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)}$$

فصل پنجم : معادله نرمال

$$X = \begin{bmatrix} (x^{(1)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix} \qquad y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$$\boxed{X\theta = y}$$

عدم استفاده از تنظیم برای پارامتر θ_0

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \right)^{-1} X^T y$$

فصل پنجم : معکوس ناپذیری

- فرض کنید $n < m$:

$$\theta = \underbrace{(X^T X)}_{\text{معکوس ناپذیر}}{}^{-1} X^T y$$

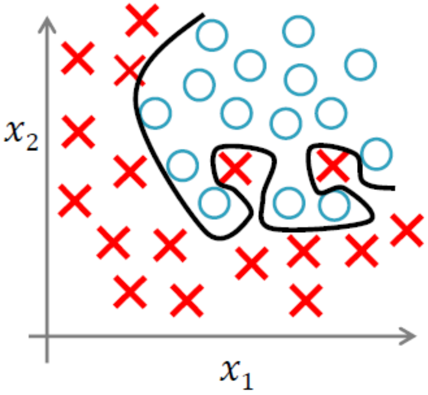
- اگر $\lambda > 0$:

$$\theta = \overbrace{\left(X^T X + \lambda \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \right)}^{\text{معکوس پذیر}}{}^{-1} X^T y$$

رگرسیون لجستیکی تنظیم شده

فصل پنجم : رگرسیون لجستیکی

- فرضیه.

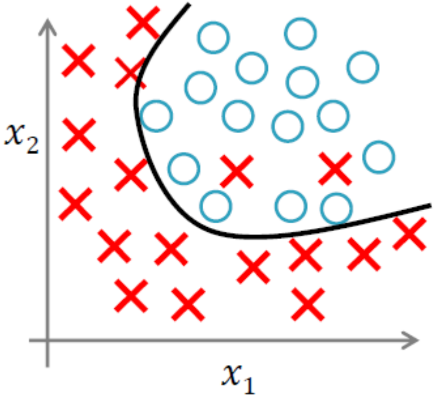


$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \cdots)$$

- تابع هزینه.

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

فصل پنجم : رگرسیون لجستیکی



- فرضیه.

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \cdots)$$

- تابع هزینه.

$$J(\theta) = -\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right) + \frac{\lambda}{2}\sum_{i=1}^{m} \theta_j^2$$

فصل پنجم : گرادیان کاهشی

- با استفاده از تنظیم.

repeat until convergence {

$$\theta_0 = \theta_0 - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_0^{(i)}$$

$$\theta_j = \theta_j - \alpha \left[\sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} + \lambda\theta_j\right] \qquad (j = 1,2, \dots, n)$$

}

$$h_\theta\left(x^{(i)}\right) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$

فصل پنجم : بهینه سازی پیشرفته

```
function [jVal, gradient] = costFunction(theta)

    jVal = [code to compute J(θ) ];
```

$$J(\theta) = \left[-\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right)\right] + \frac{\lambda}{2}\sum_{j=1}^{n}\theta_j^2$$

```
    gradient(1)      = [code to compute ∂/∂θ0 J(θ) ];
```

$$\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x_0^{(i)}$$

```
    gradient(2)      = [code to compute ∂/∂θ1 J(θ) ];
```

$$\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x_1^{(i)} + \lambda\theta_1$$

```
    ...

    gradient(n+1)    = [code to compute ∂/∂θn J(θ) ];
```

با تشکر از توجه شما