

یادگیری نظارت شده: رگرسیون

سید ناصر رضوی n.razavi@tabrizu.ac.ir

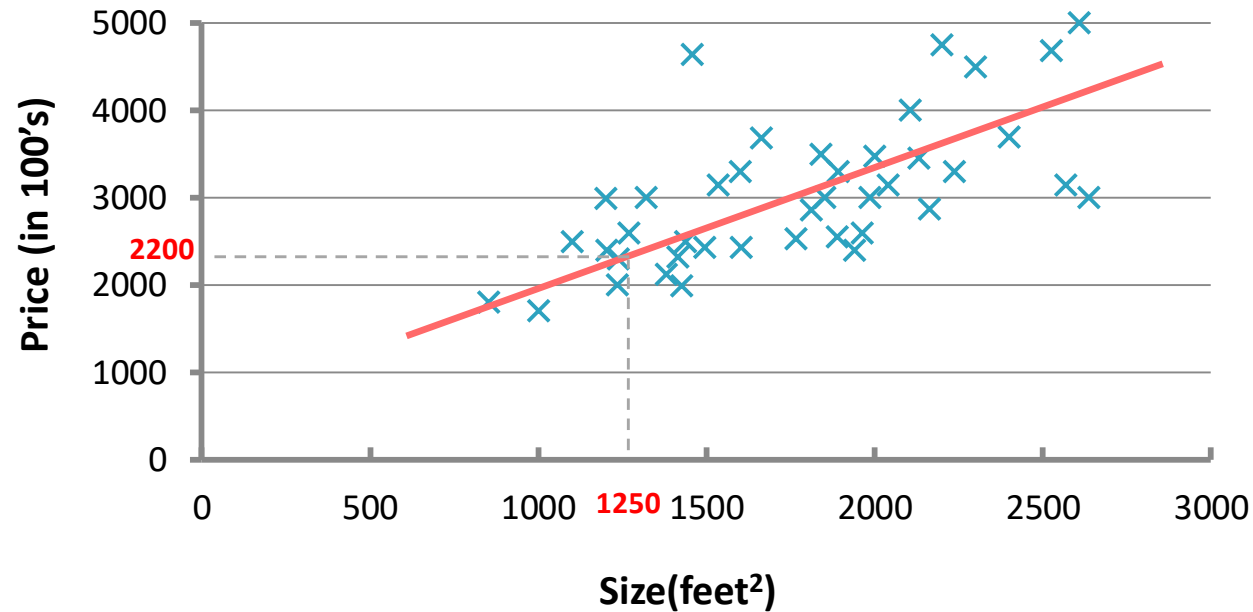
۱۳۹۵

فهرست مطالب

- رگرسیون.
- رگرسیون خطی تک متغیره و چند متغیره
- گرادیان کاهشی.
- معادله‌ی نرمال.
- رگرسیون با وزن دهی محلی.
- تفسیر احتمالاتی رگرسیون.
- تخمین بیشترین درست‌نمایی.

رگرسیون خطی تا متغیره

قیمت‌گذاری خانه



□ یادگیری نظارت شده.

به ازای هر نمونه‌ی آموزشی، «پاسخ درست» داده شده است.

□ رگرسیون.

پیش‌بینی کمیت‌هایی با مقادیر پیوسته. (مانند قیمت یک خانه)

نمادگذاری

۵

مجموعه‌ی آموزشی

متراف (فوت مربع) (x)	قیمت (در ۱۰۰۰ دلار) (y)
۲۱۰۴	۴۶۰
۱۴۱۶	۲۳۲
۱۵۳۴	۳۱۵
۸۵۲	۱۷۸
...	...

$m = ۴۷$

□ نمادها.

□ m = تعداد نمونه‌های آموزشی

□ x = متغیر «ورودی»، ویژگی‌ها

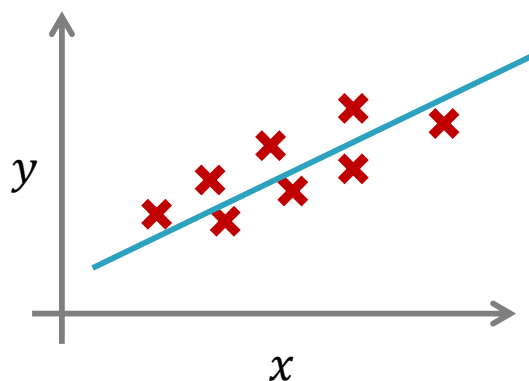
□ y = متغیر «خروجی»، متغیر «هدف»

(x, y) : یک نمونه‌ی آموزشی

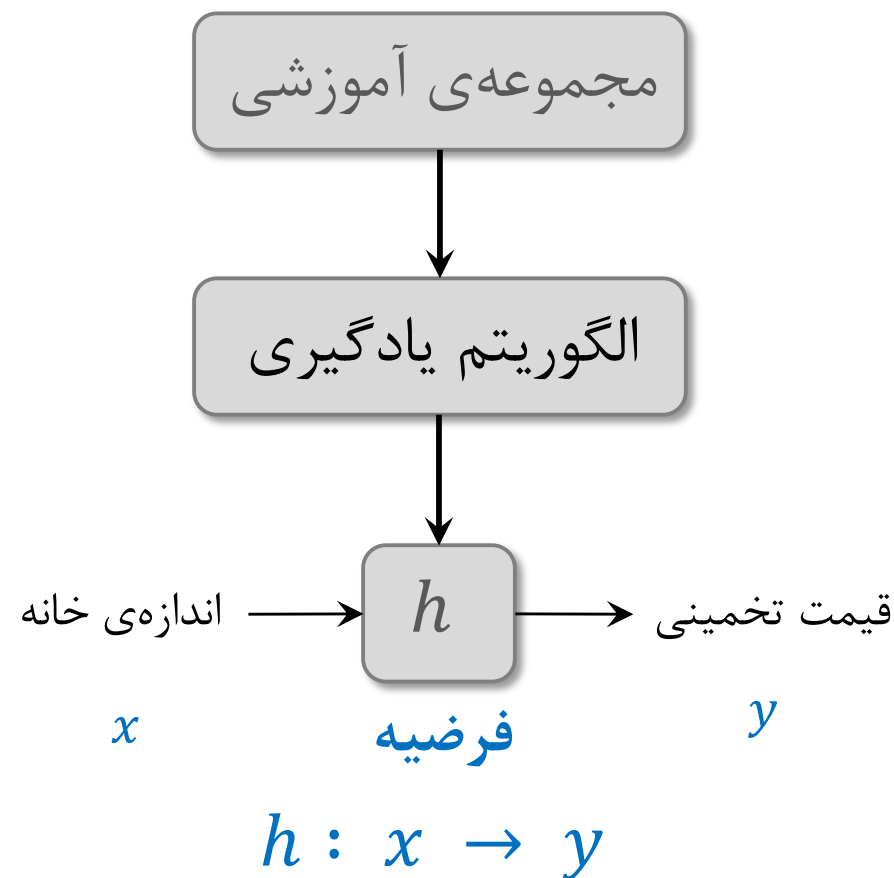
$(x^{(i)}, y^{(i)})$: نمونه‌ی آموزشی i ام

نمایش فرضیه h .

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



رگرسیون خطی تک متغیره

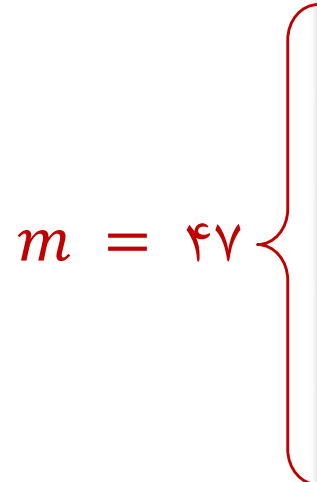


تابع هزینه

ارزیابی فرضیه

۸

مجموعه‌ی آموزشی



متراف (فوت مربع)	قیمت (در ۱۰۰۰ دلار)
۲۱۰۴	۴۶۰
۱۴۱۶	۲۳۲
۱۵۳۴	۳۱۵
۸۵۲	۱۷۸
...	...

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

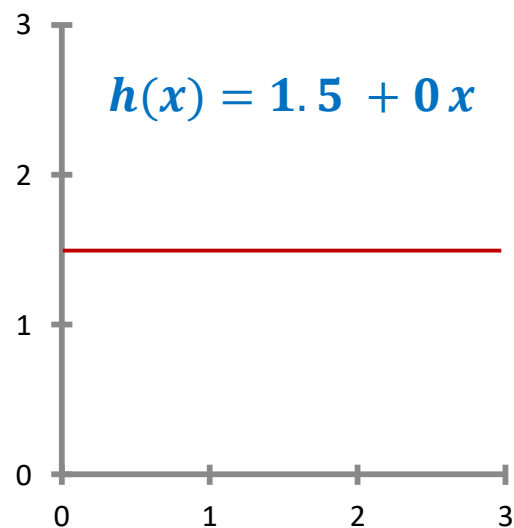
$$(\theta_0, \theta_1)$$

فرضیه:

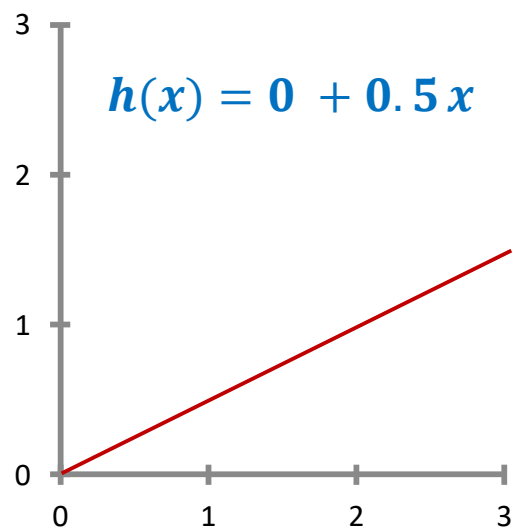
پارامترها:

س. مقدار پارامترها را چگونه باید انتخاب نمود؟

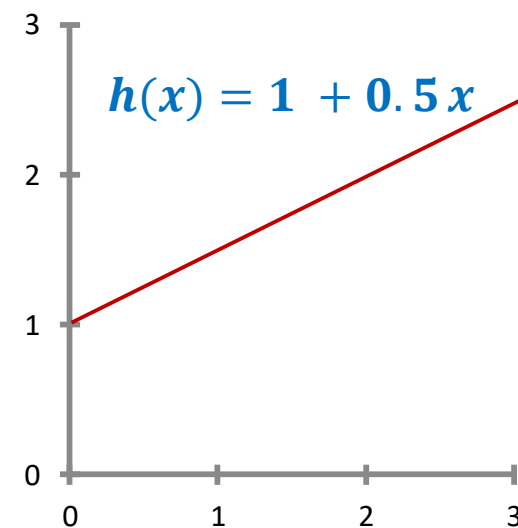
$$h_{\theta}(x) = \theta_0 + \theta_1 x$$



$$\begin{aligned}\theta_0 &= 1.5 \\ \theta_1 &= 0\end{aligned}$$



$$\begin{aligned}\theta_0 &= 0 \\ \theta_1 &= 0.5\end{aligned}$$



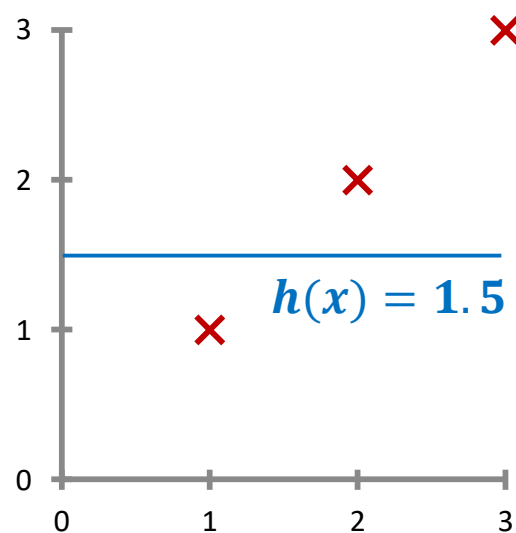
$$\begin{aligned}\theta_0 &= 1 \\ \theta_1 &= 0.5\end{aligned}$$

ارزیابی فرضیه

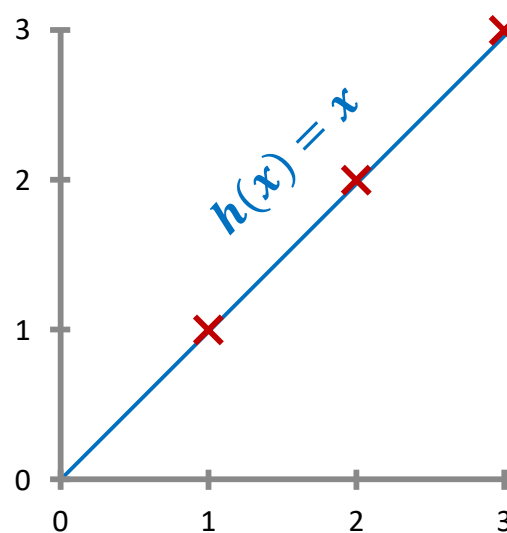
۱۰

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

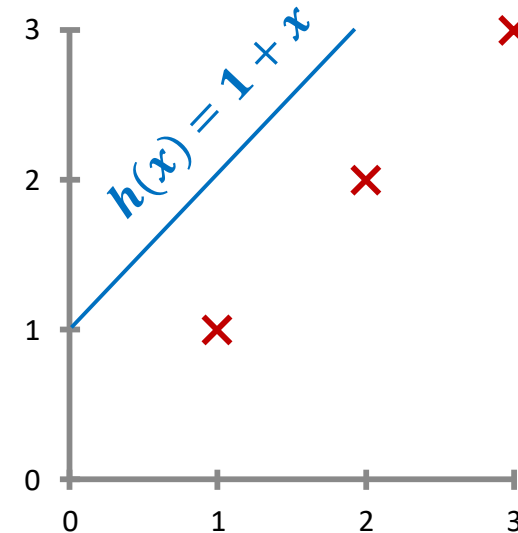
س. کدام فرضیه بهتر است؟



$$\begin{aligned}\theta_0 &= 1.5 \\ \theta_1 &= 0\end{aligned}$$



$$\begin{aligned}\theta_0 &= 0 \\ \theta_1 &= 1\end{aligned}$$



$$\begin{aligned}\theta_0 &= 1 \\ \theta_1 &= 1\end{aligned}$$

تابع هزینه

۱۱

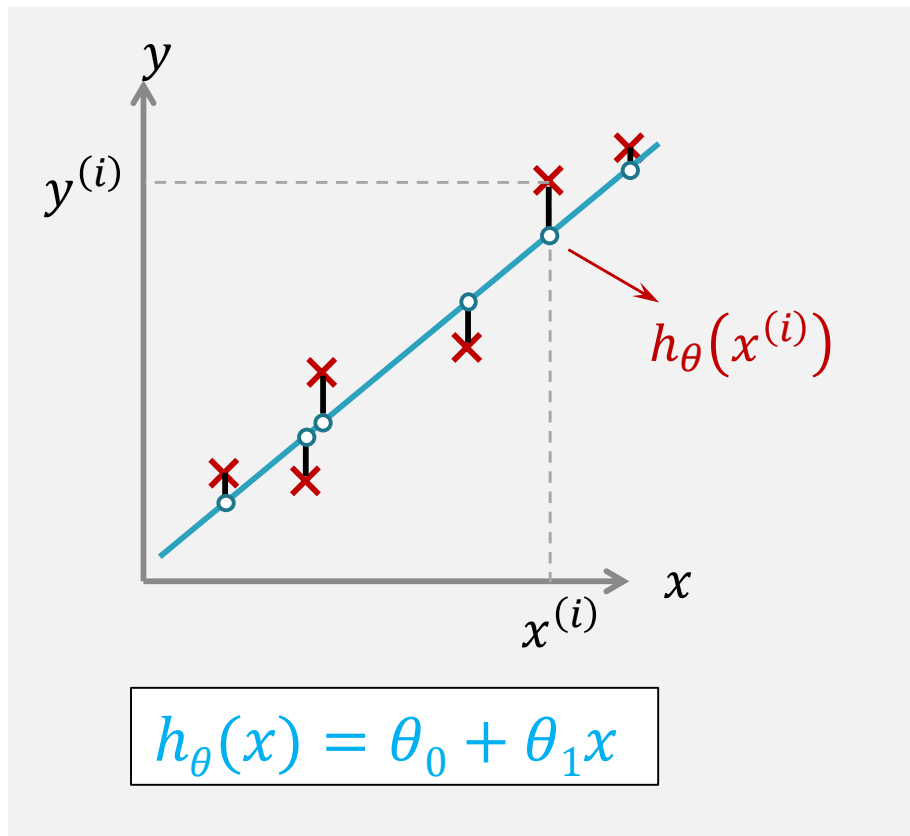
□ ایده. انتخاب پارامترها به گونه‌ای که به ازای هر نمونه‌ی آموزشی مانند (x, y) ، مقدار $h_\theta(x)$ تا حد ممکن به مقدار y نزدیک باشد.

□ تابع هزینه. مجموع مربعات خطا.

$$J(\theta_0, \theta_1) = \frac{1}{2} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$$

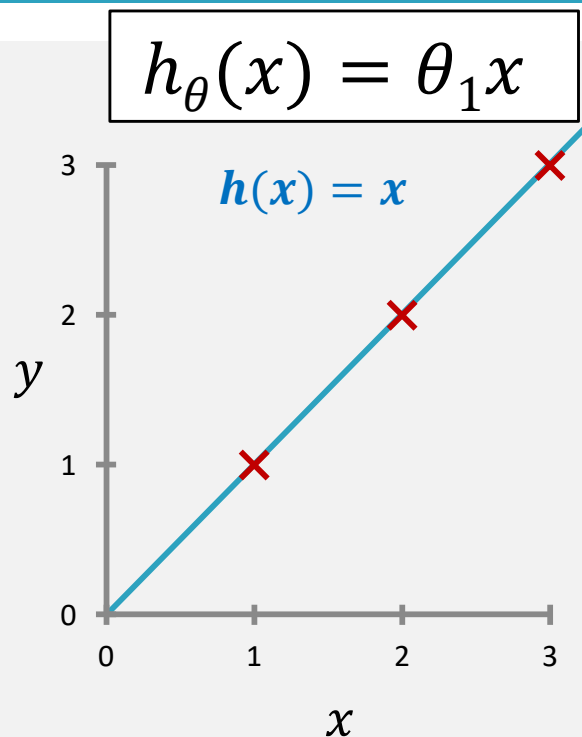
□ هدف.

$$\underset{\theta_0, \theta_1}{\text{minimize}} \quad J(\theta_0, \theta_1)$$

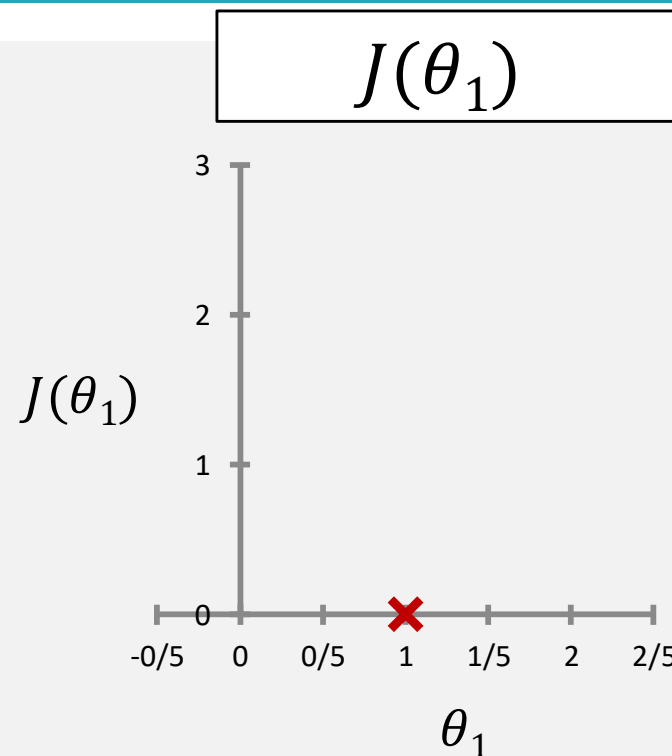


تابع هزینه ساده شده ($\theta_0 = 0$)

۱۲



$$\theta_1 = 1$$

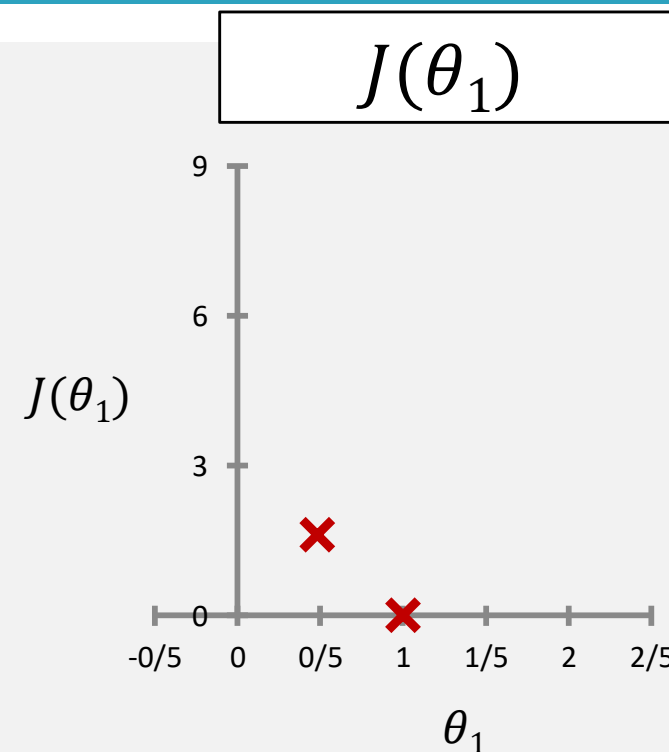
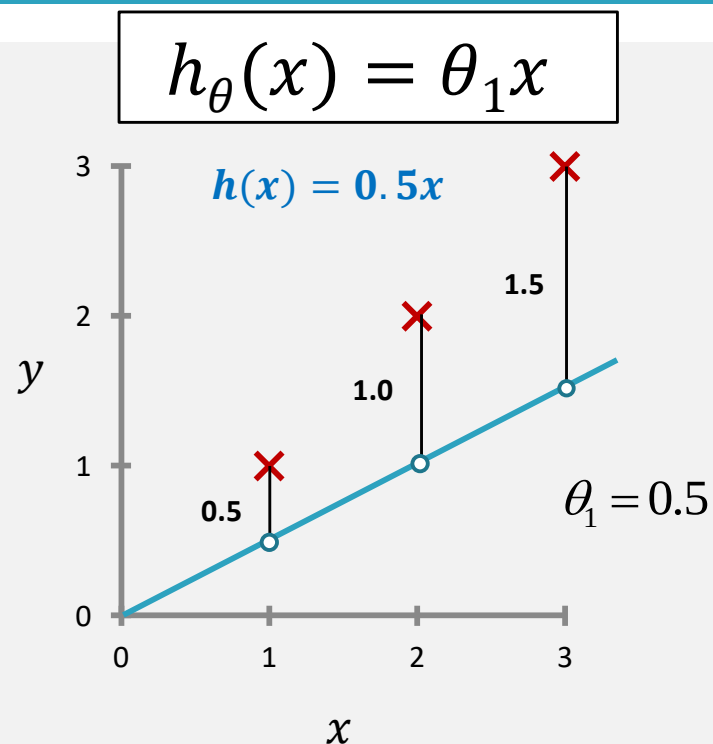


$$\begin{aligned} J(\theta_0, \theta_1) &= \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \\ &= \frac{1}{2} \sum_{i=1}^m (x^{(i)} - y^{(i)})^2 = \frac{1}{2} (0^2 + 0^2 + 0^2) = 0 \end{aligned}$$

$$J(1) = 0$$

تابع هزینه

۱۳

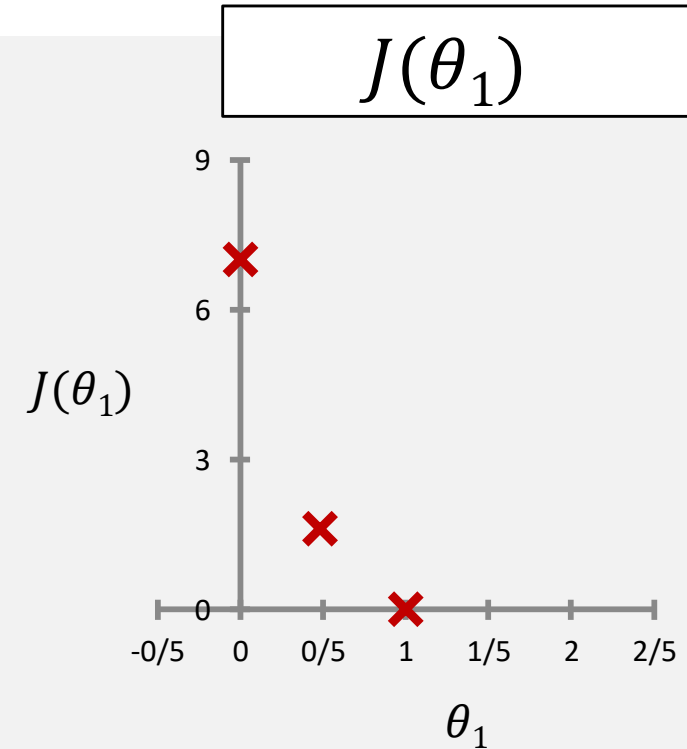
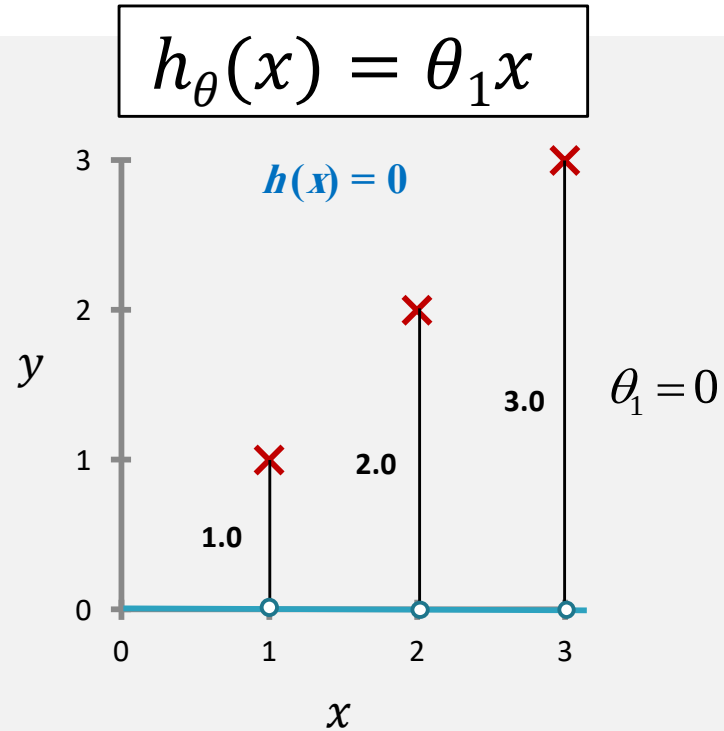


$$J(0.5) = \frac{1}{2}(0.5^2 + 1.0^2 + 1.5^2) = \frac{1}{2}(3.5) = 1.75$$

$$J(0.5) = 1.75$$

تابع هزینه

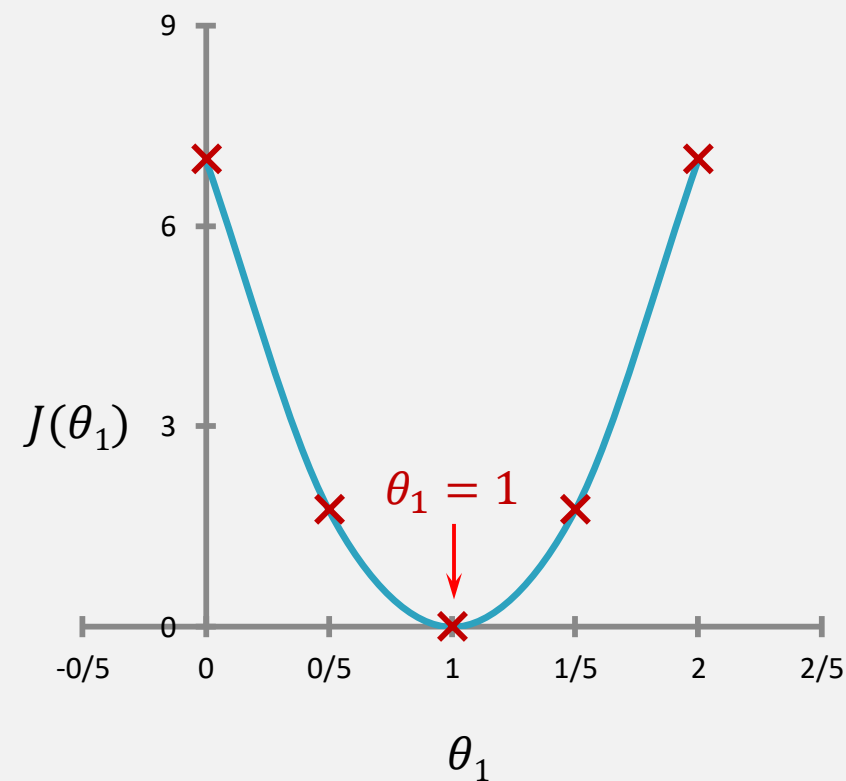
۱۴



$$J(0) = \frac{1}{2}(1.0^2 + 2.0^2 + 3.0^2) = \frac{1}{2}(14) = 7.0$$

$$J(0) = 7.0$$

$$\underset{\theta_1}{\text{minimize}} \quad J(\theta_1)$$



رگرسیون خطی تک متغیره

۱۶

□ فرضیه.

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

□ پارامترها.

$$\theta_0, \theta_1$$

□ تابع هزینه.

$$J(\theta_0, \theta_1) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

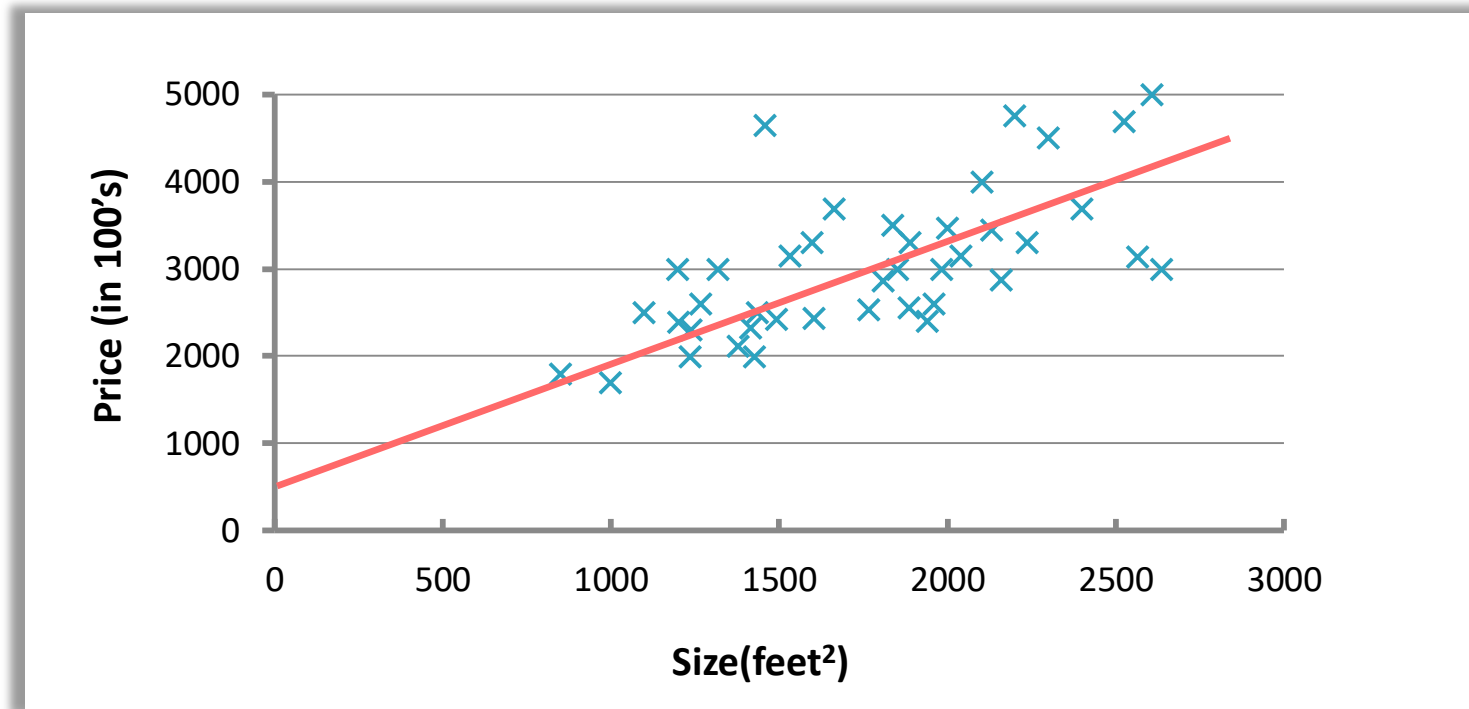
□ هدف.

$$\underset{\theta_0, \theta_1}{\text{minimize}} \quad J(\theta_0, \theta_1)$$

مثال: قیمت‌گذاری خانه

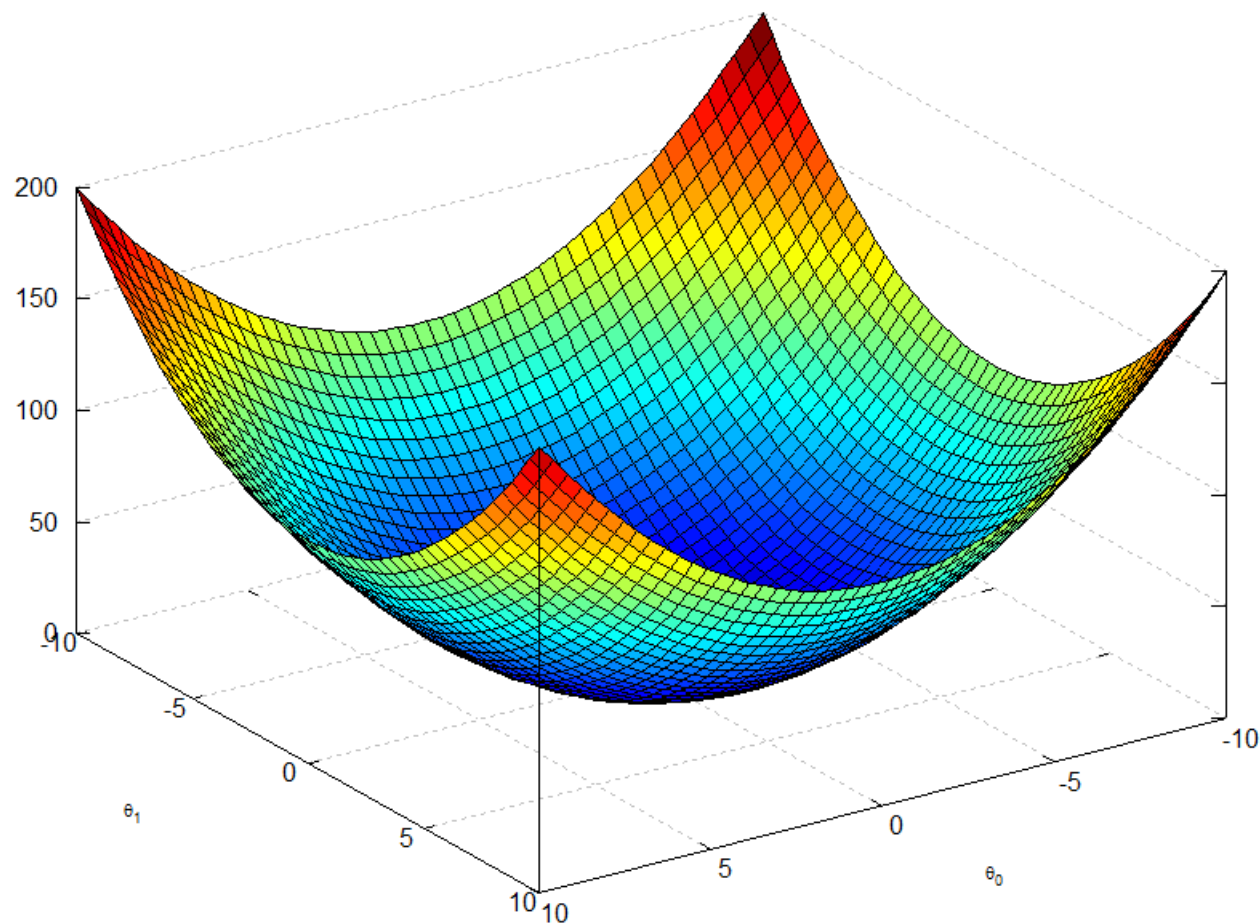
۱۷

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



تابع هزینه

۱۸

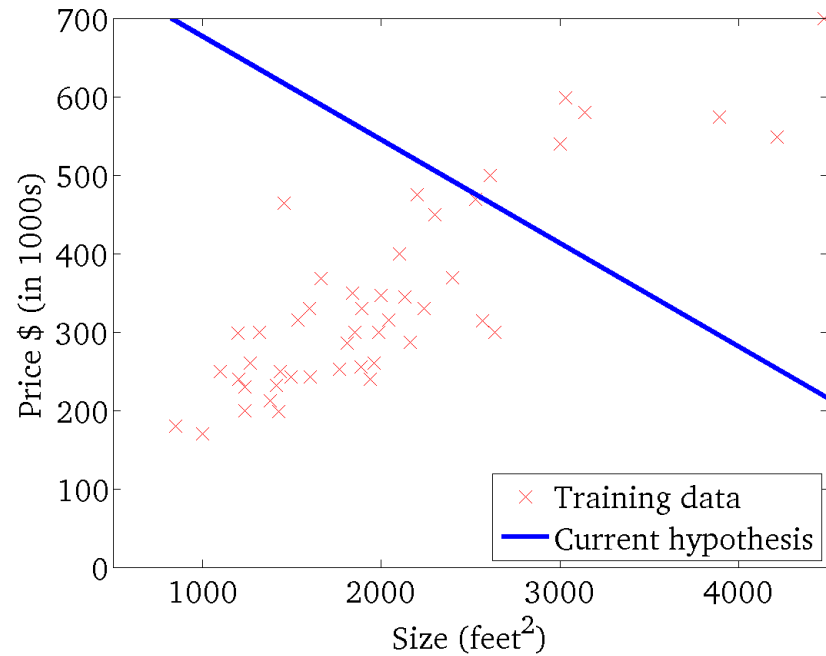


```
x = -10:0.5:10;  
y = -10:0.5:10;  
[X Y] = meshgrid(x, y);  
Z = X.^2 + Y.^2;  
surf(X, Y, Z);
```

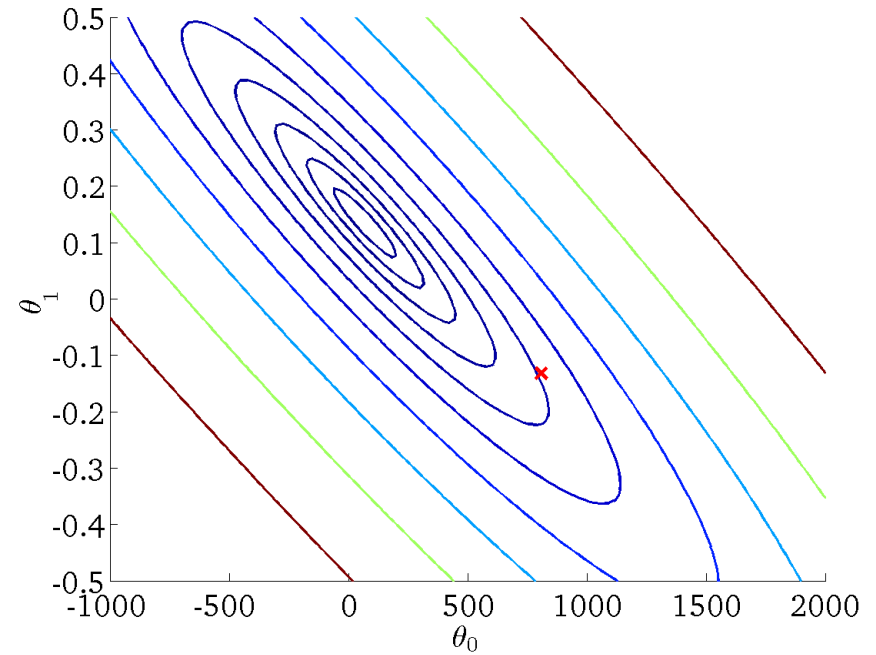
تابع هزینه: منحنی کانتور

۱۹

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

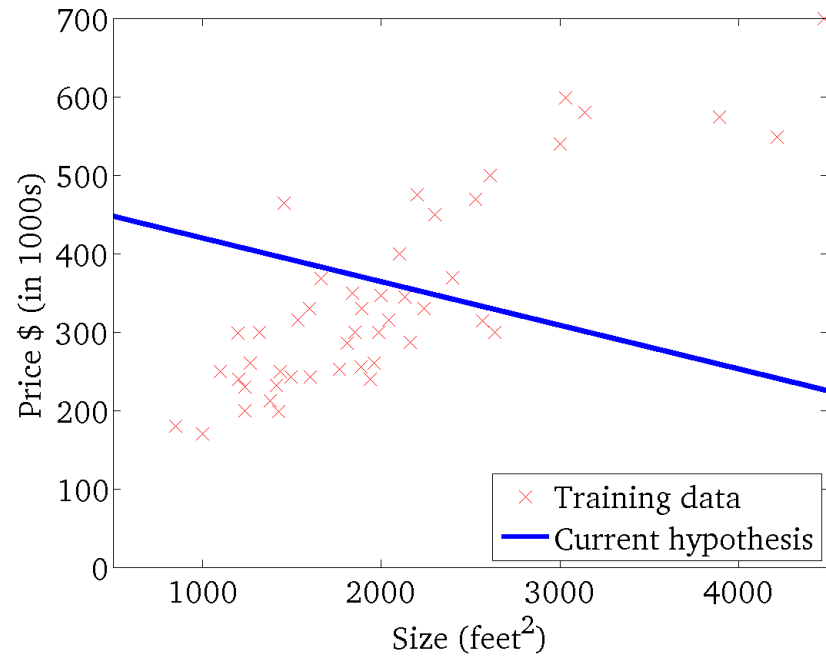


$$J(\theta_0, \theta_1)$$

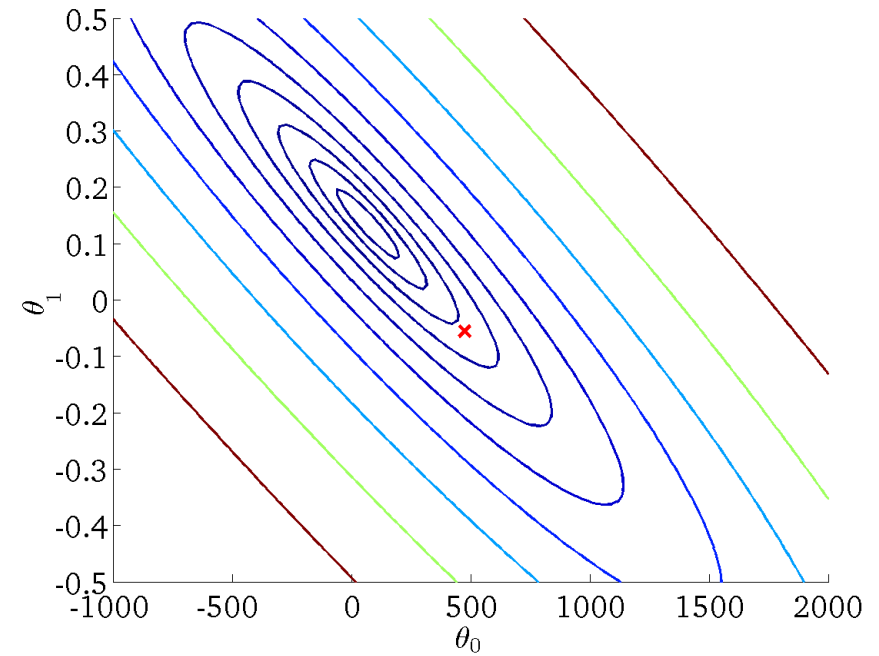


تابع هزینه: منحنی کانتور

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



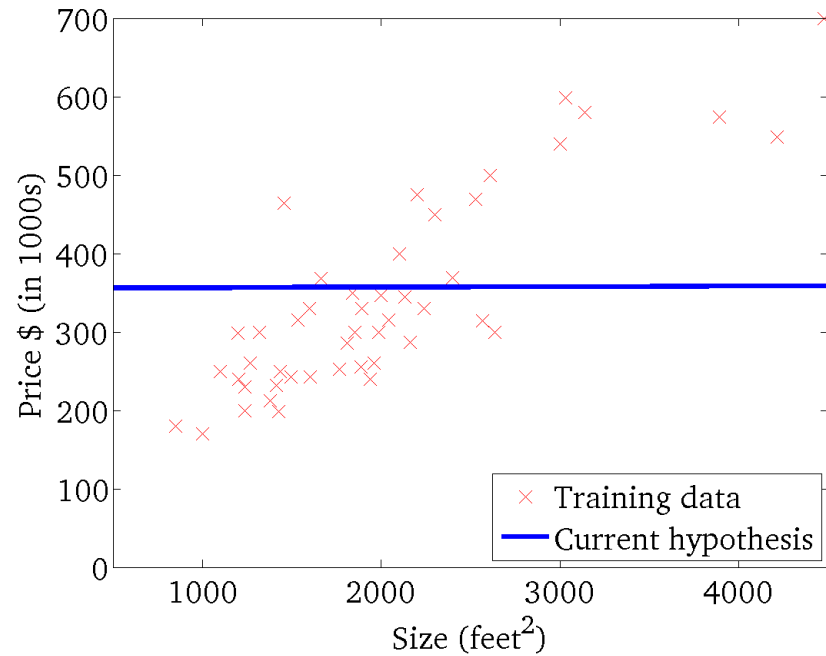
$$J(\theta_0, \theta_1)$$



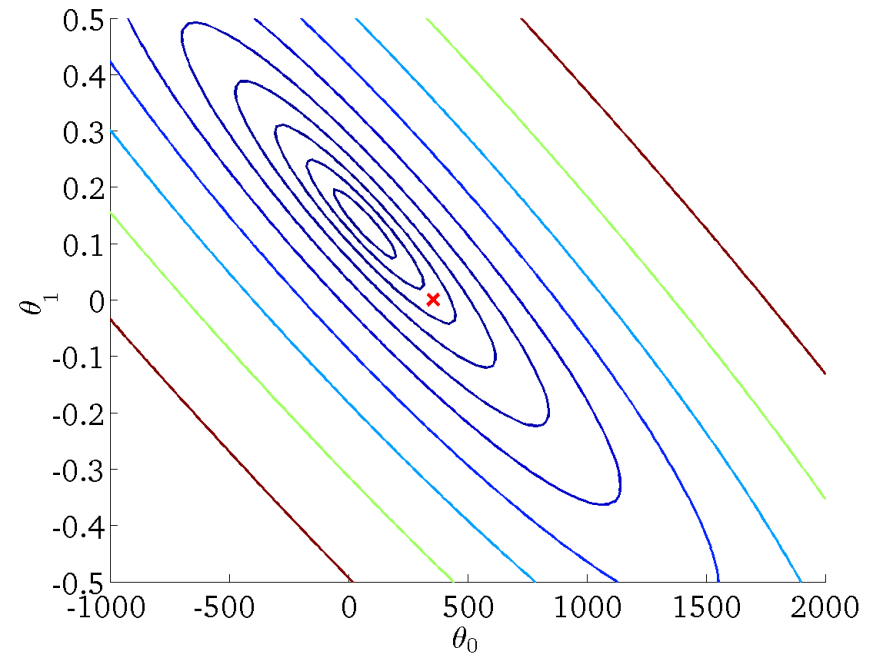
تابع هزینه: منحنی کانتور

۲۱

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



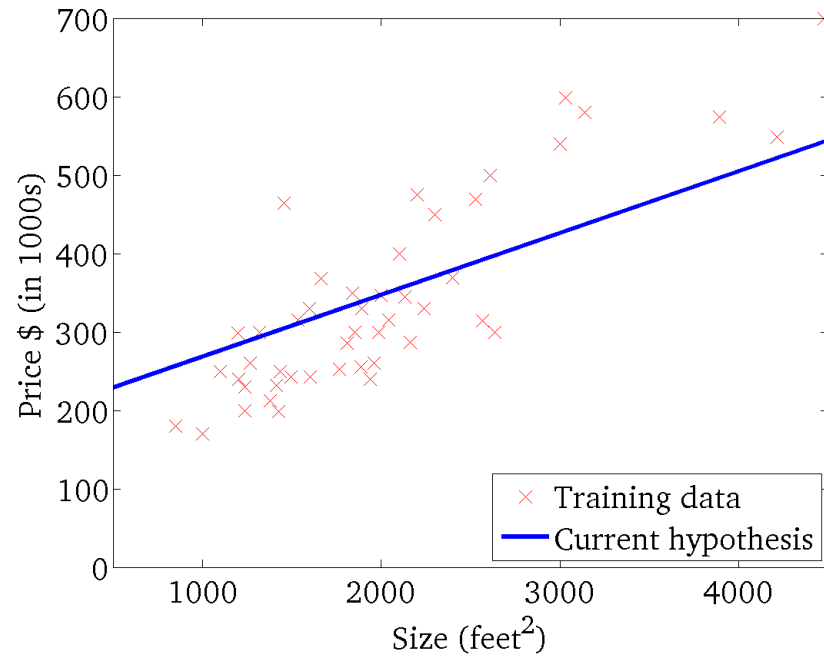
$$J(\theta_0, \theta_1)$$



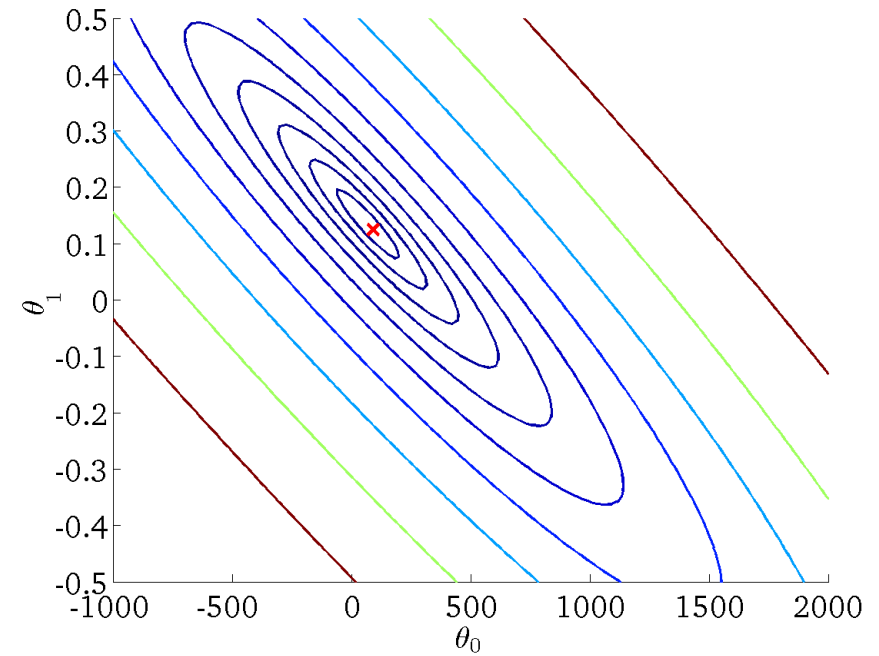
تابع هزینه: منحنی کانتور

۲۲

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



$$J(\theta_0, \theta_1)$$



گَرادیان کاهشی

□ تابع هزینه.

$$J(\theta_0, \theta_1)$$

□ هدف.

$$\underset{\theta_0, \theta_1}{\text{minimize}} J(\theta_0, \theta_1)$$

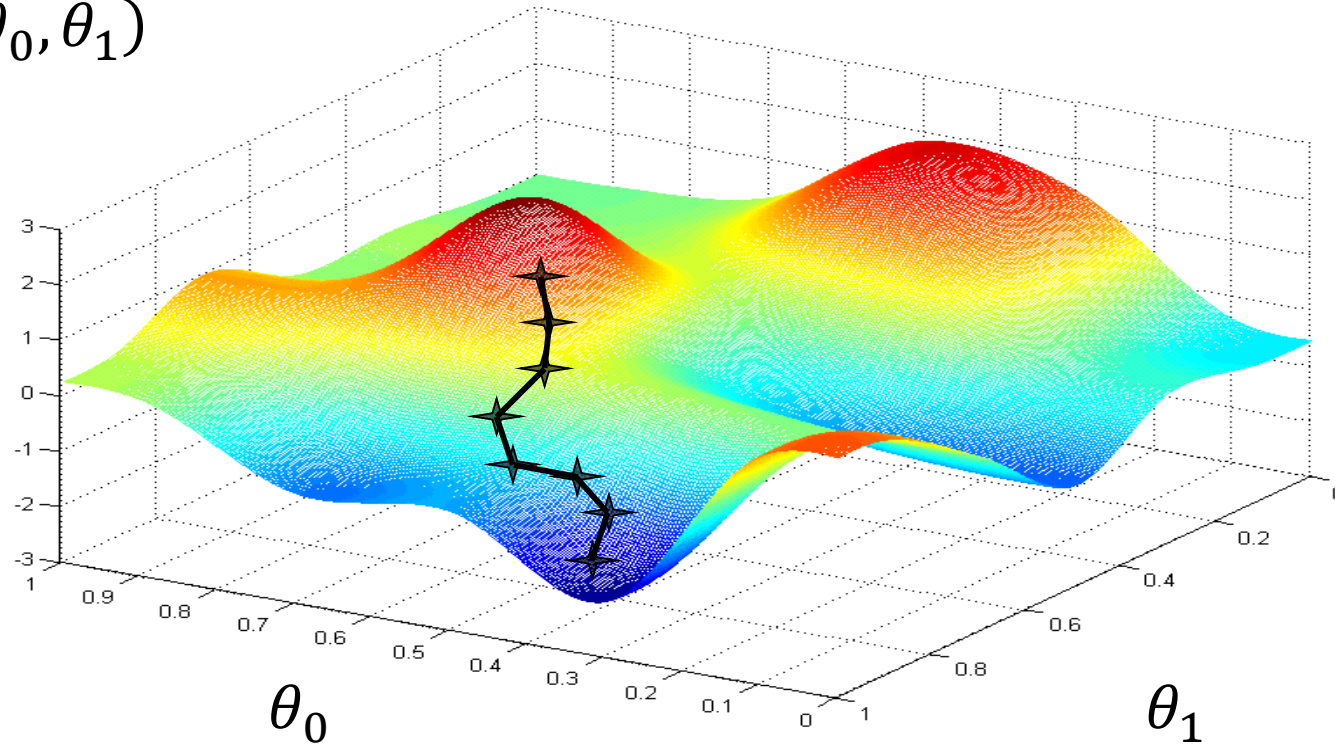
□ کلیات روش.

- با یک مقدار اولیه‌ی تصادفی برای پارامترهای θ_0 و θ_1 شروع کن. [مثلاً مقدار صفر]
- مقدار پارامترها را به گونه‌ای تغییر بده که مقدار تابع هزینه $J(\theta_0, \theta_1)$ کاهش یابد.
- عمل بالا را آن قدر تکرار کن تا به یک مقدار کمینه برای تابع هزینه برسیم. [همگرایی]

گرادیان کاهشی: بهینه‌ی سراسری

۲۵

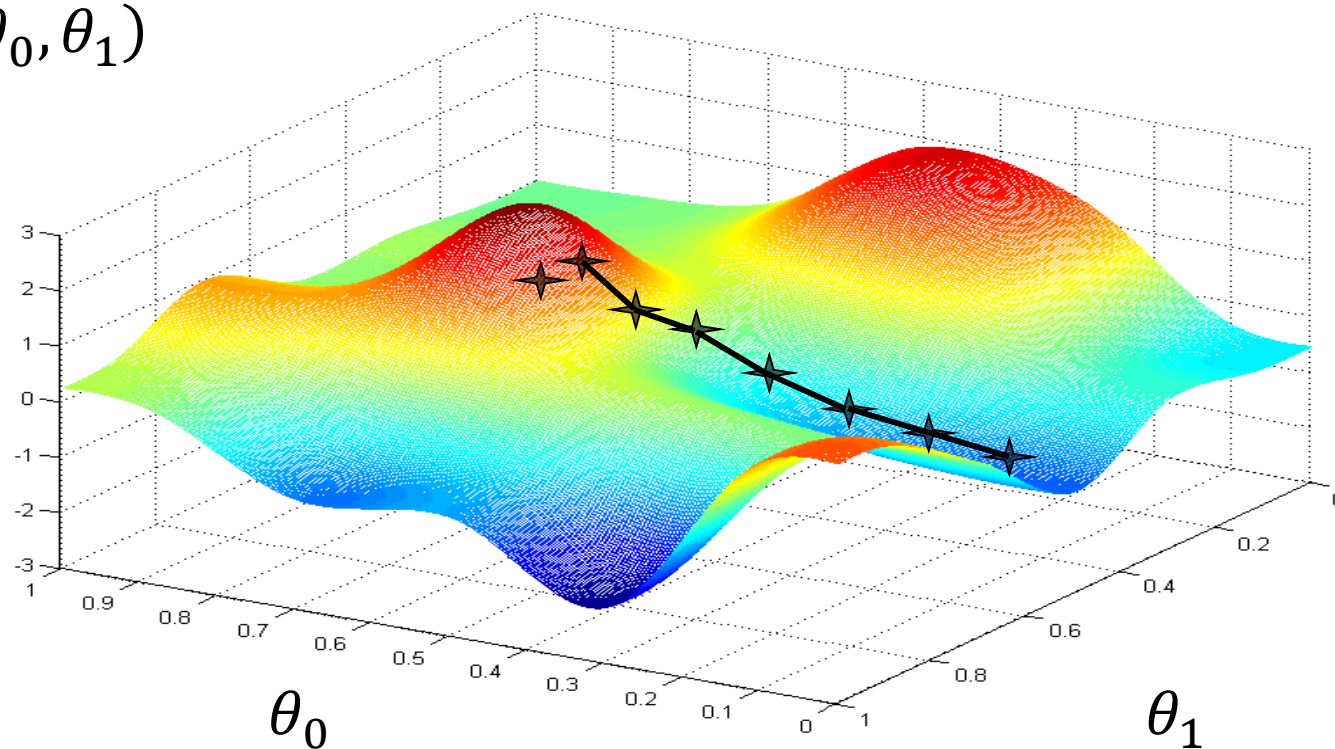
$J(\theta_0, \theta_1)$



گرادیان کاهشی: بهینه‌ی محلی

۲۶

$J(\theta_0, \theta_1)$



الگوریتم گرادیان کاهشی

۲۷

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1)$$

}

نرخ یادگیری

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

□ پیاده‌سازی درست. به روزرسانی مقدار پارامترها به طور همزمان

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

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

$$\theta_0 := \theta_0 + \Delta \theta_0$$

$$\theta_1 := \theta_1 + \Delta \theta_1$$

الگوریتم گرادیان کاهشی

۲۸

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1)$$

}

نرخ یادگیری

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

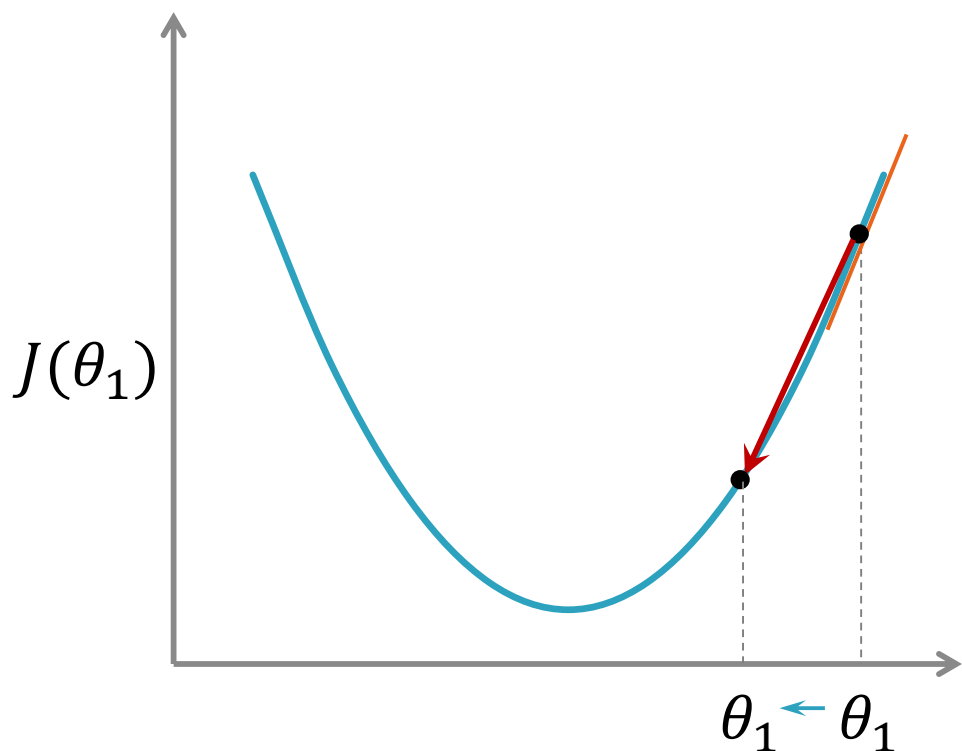
□ پیاده‌سازی نادرست. به روزرسانی مقدار پارامترها به طور ترتیبی

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

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

الگوریتم گرادیان کاهشی: گرادیان

۲۹



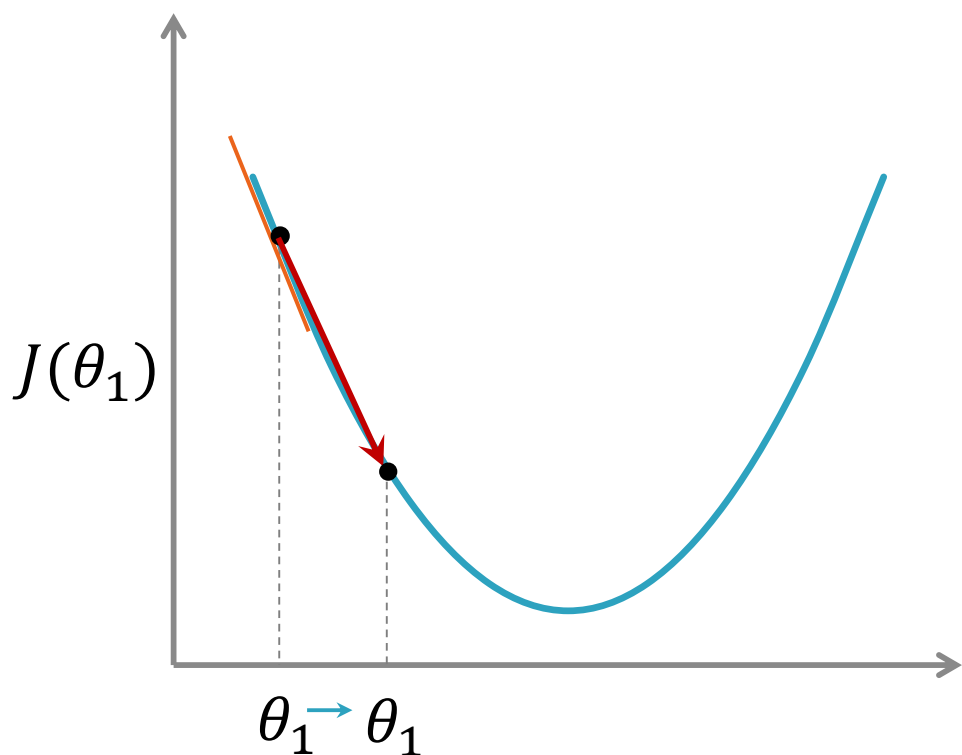
شیب مثبت

$$\theta_1 := \theta_1 - \alpha \boxed{\frac{\partial}{\partial \theta_1} J(\theta_1)}$$

$$\theta_1 := \theta_1 - \underbrace{\alpha}_{\geq 0} (\geq 0)$$

الگوریتم گرادیان کاهشی: گرادیان

۳۰



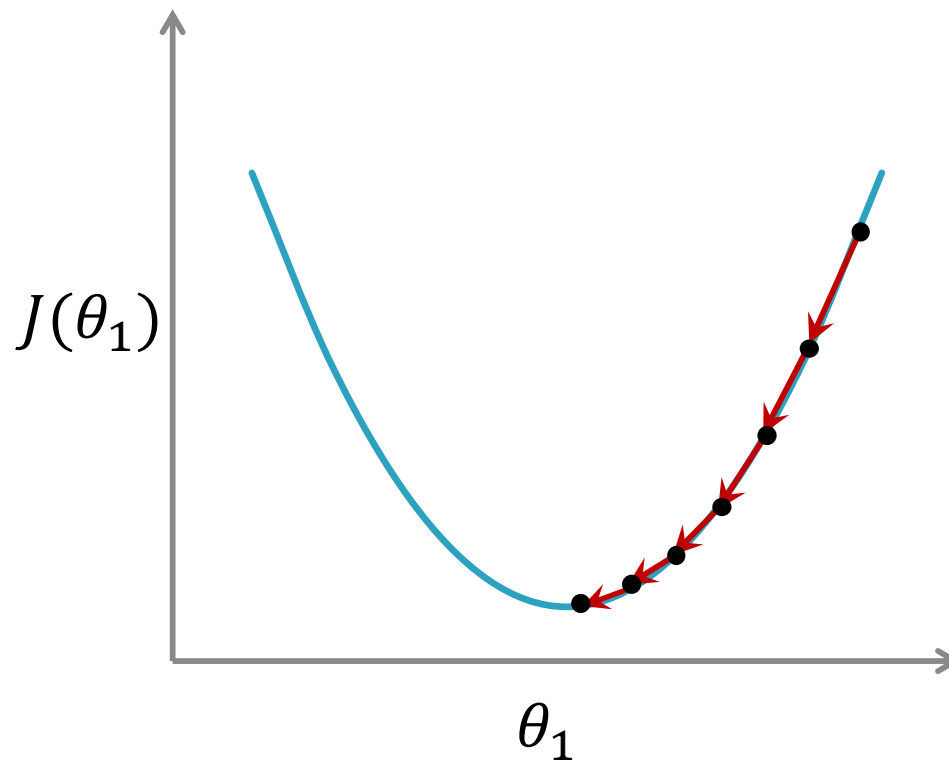
شیب منفی

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

$$\theta_1 := \theta_1 - \underbrace{\alpha(\leq 0)}_{\leq 0}$$

الگوریتم گرادیان کاهشی: نرخ یادگیری

□ اگر نرخ یادگیری بیش از حد کوچک باشد، گرادیان کاهشی به کندی همگرا خواهد شد.

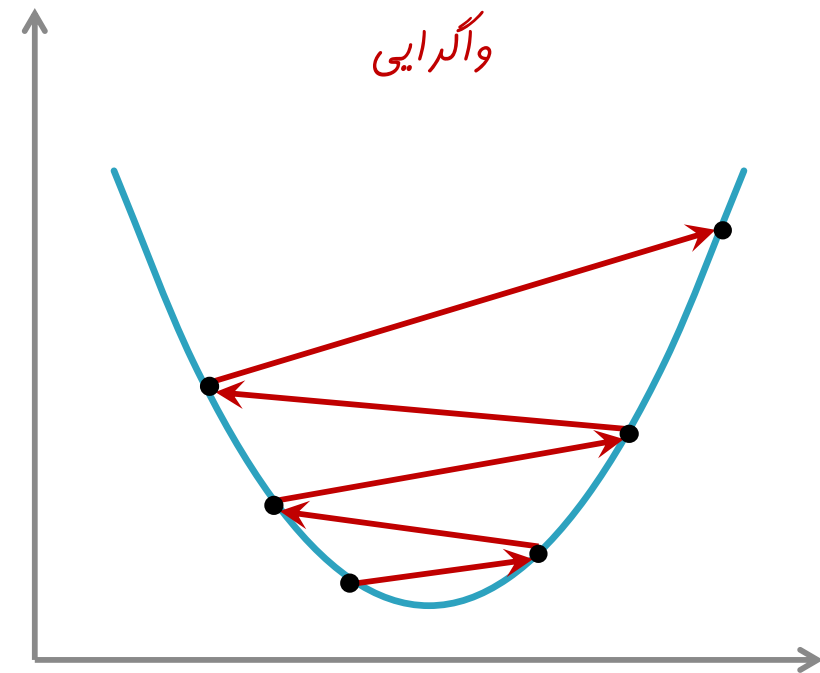
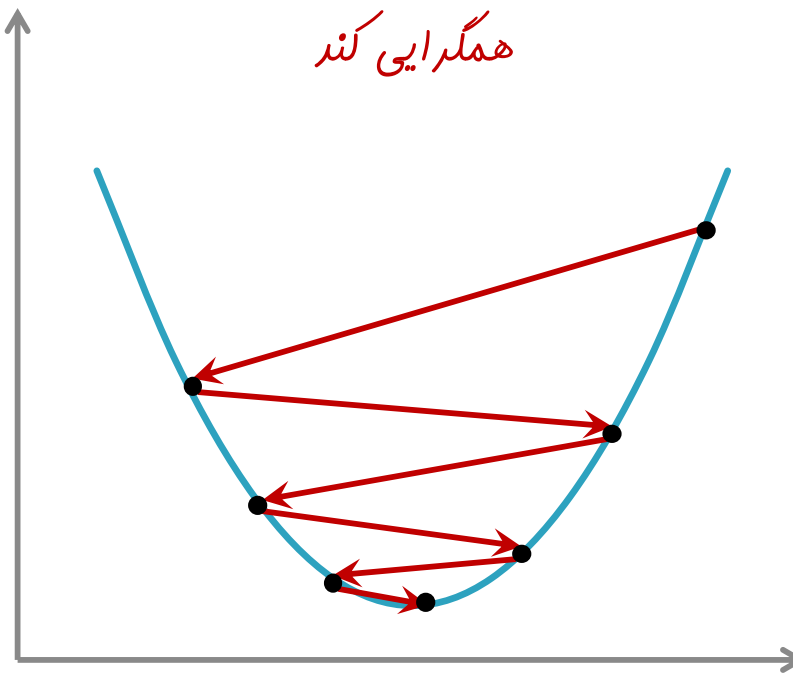


$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)$$

الگوریتم گرادیان کاهشی: نرخ یادگیری

۳۲

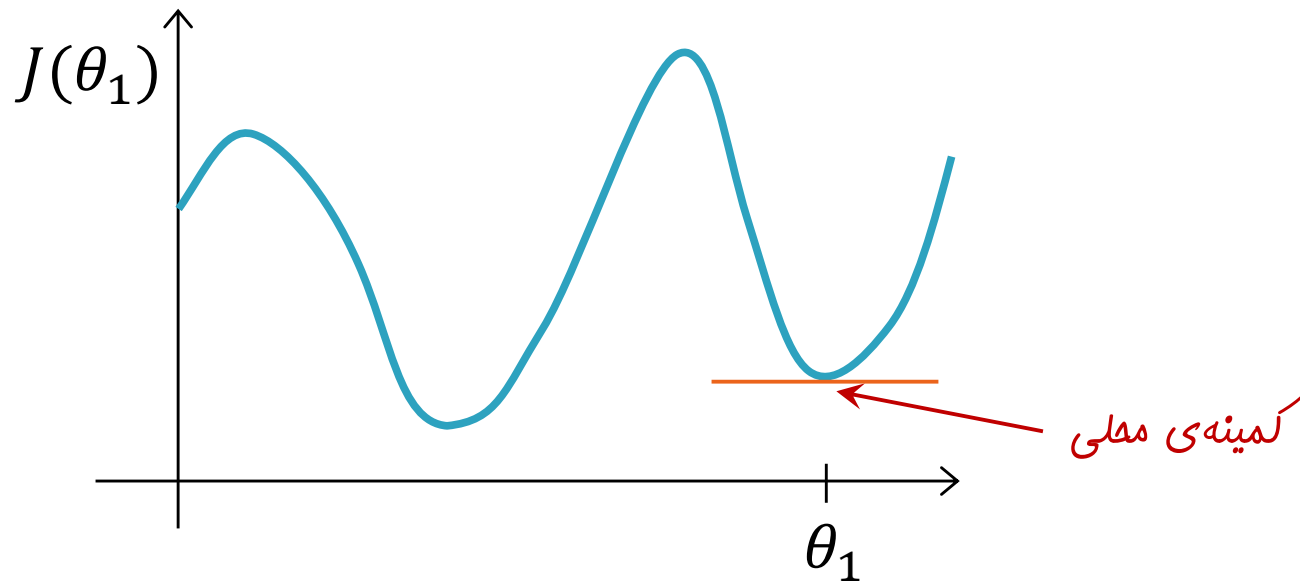
□ اگر نرخ یادگیری بیش از حد بزرگ باشد، گرادیان کاهشی ممکن است به کندی همگرا شود و یا حتی واگرا شود.



الگوریتم گرادیان کاهشی: همگرایی

۳۳

□ همگرایی. زمانی که مقدار پارامتر θ_1 در یک **کمینه محلی** قرار بگیرد.



$$\theta_1 := \theta_1 - \underbrace{\alpha \frac{\partial}{\partial \theta_1} J(\theta_1)}_{\text{صفر}}$$

کاربرد گرادیان کاهش در رگرسیون خطی



گرادیان کاهشی و رگرسیون خطی

۳۵

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1)$$

}

□ رگرسیون خطی.


$$J(\theta_0, \theta_1) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

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

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_j} \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 = \frac{\partial}{\partial \theta_j} \frac{1}{2} \sum_{i=1}^m ((\theta_0 + \theta_1 x^{(i)}) - y^{(i)})^2$$

$$j = 0 \Rightarrow \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$j = 1 \Rightarrow \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

گرادیان کاهشی و رگرسیون خطی

۳۷

repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$\theta_1 := \theta_1 - \alpha \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

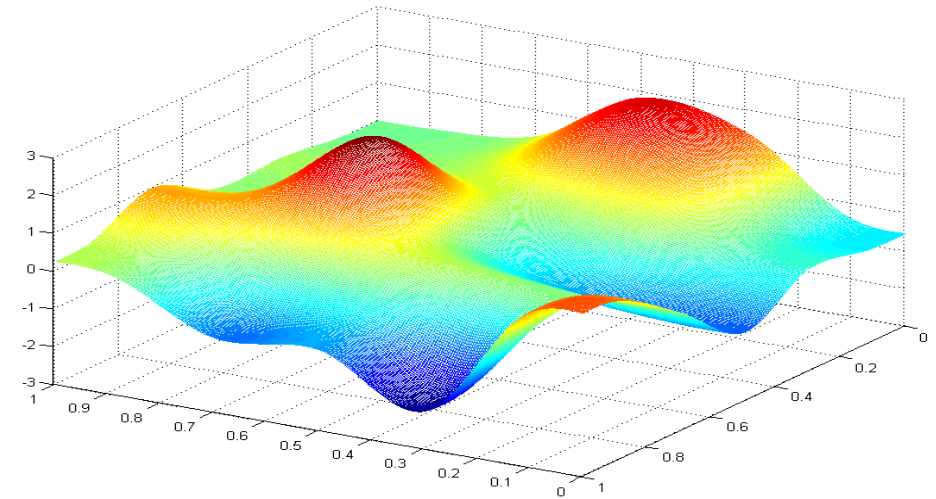
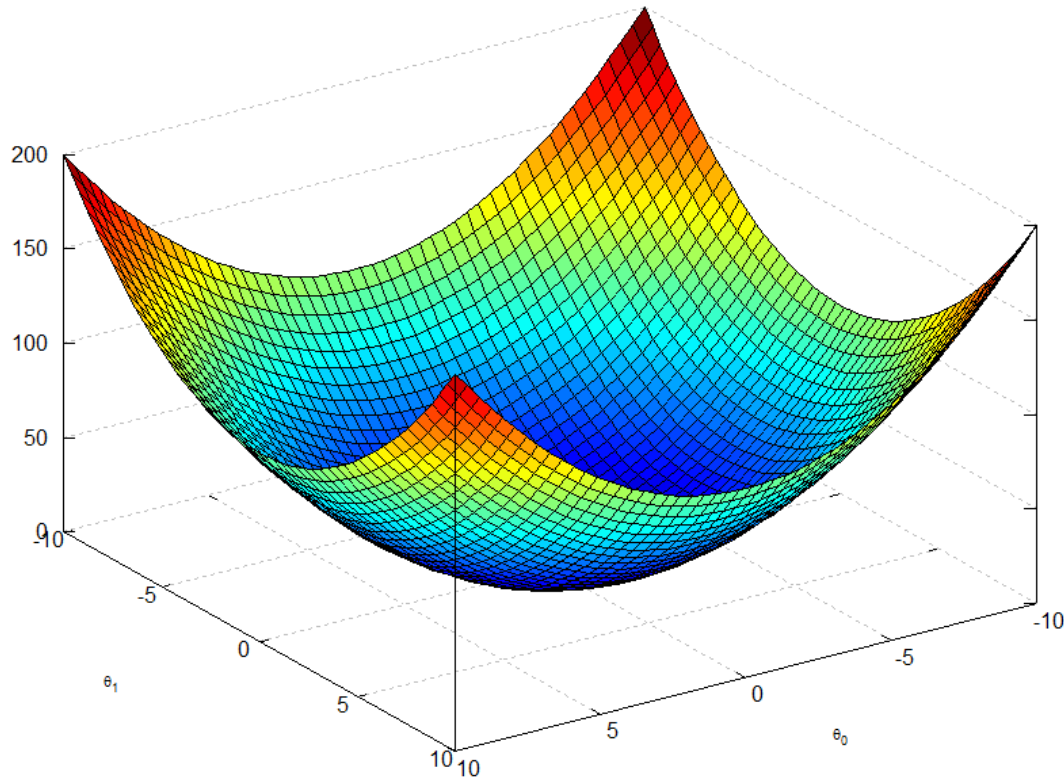
}

به روزرسانی همزمان

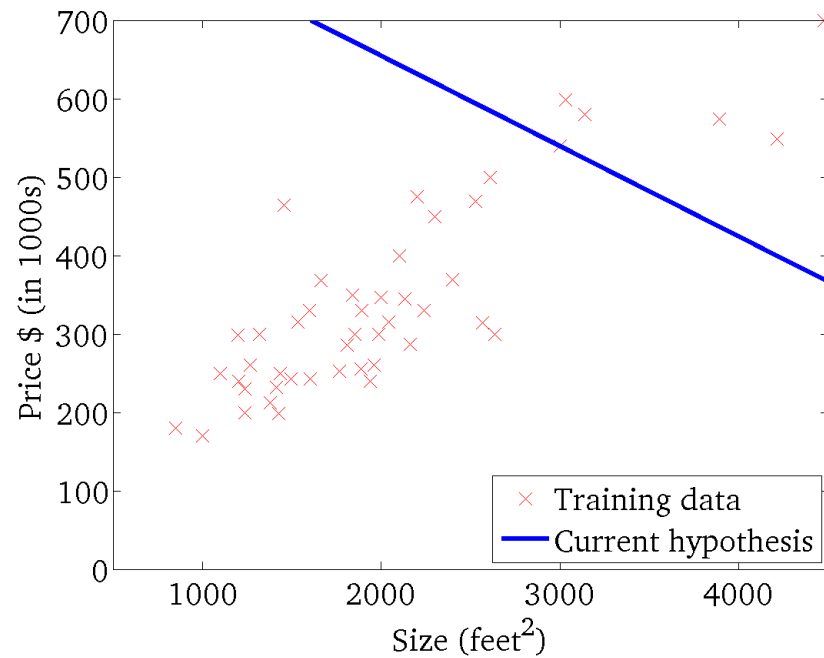
گرادیان کاهش‌ی و رگرسیون خطی

۳۸

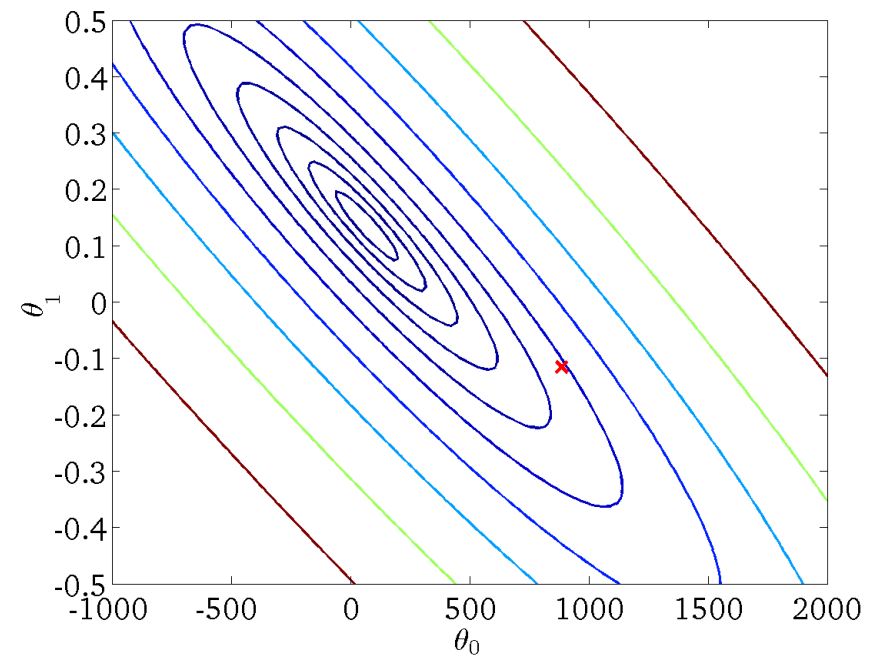
□ توجه. در رگرسیون خطی تابع هزینه یک تابع کوژ است و در نتیجه گرادیان کاهش‌ی در صورت همگرایی لزوماً در بهینه‌ی سراسری همگرا می‌شود.



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



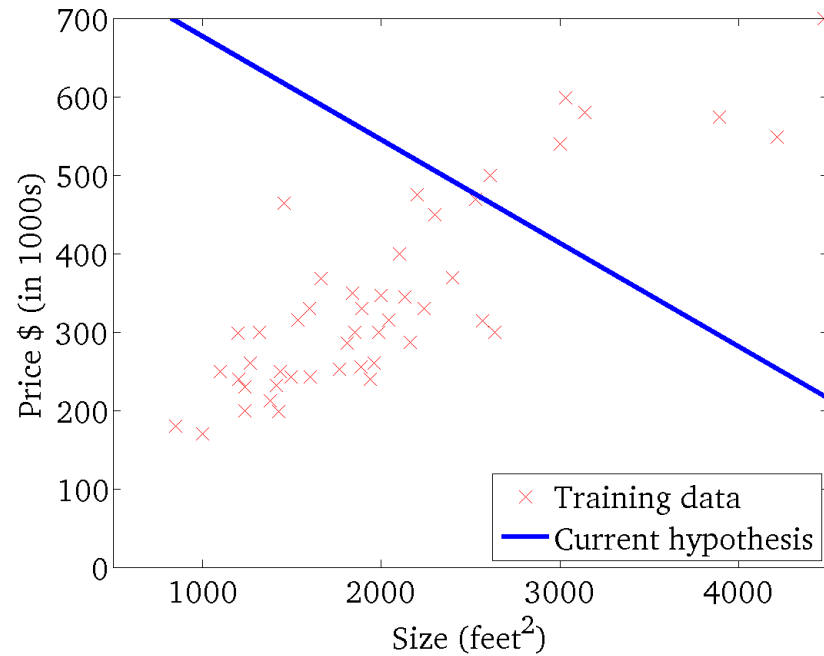
$$J(\theta_0, \theta_1)$$



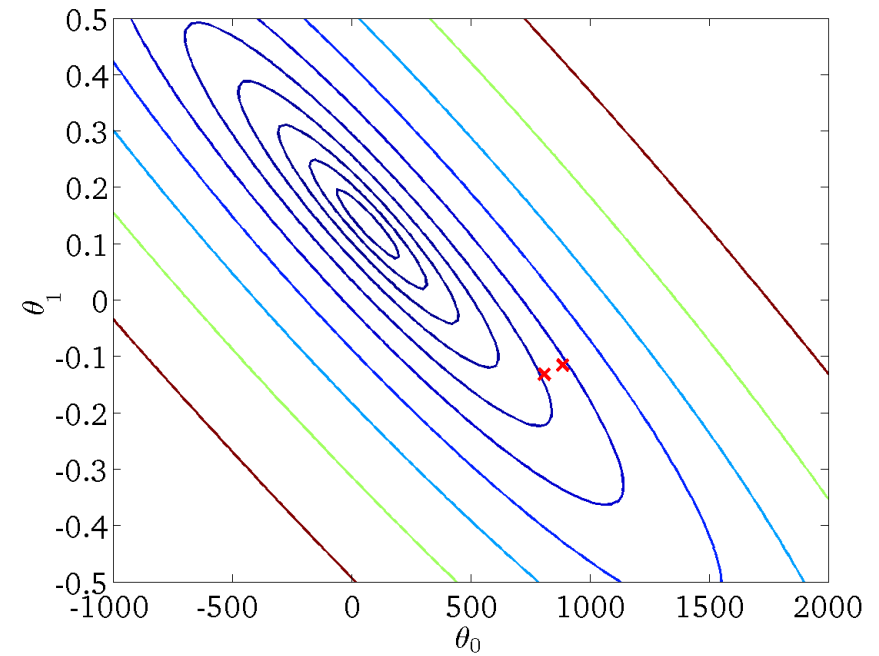
گرادیان کاهشی

۴۰

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



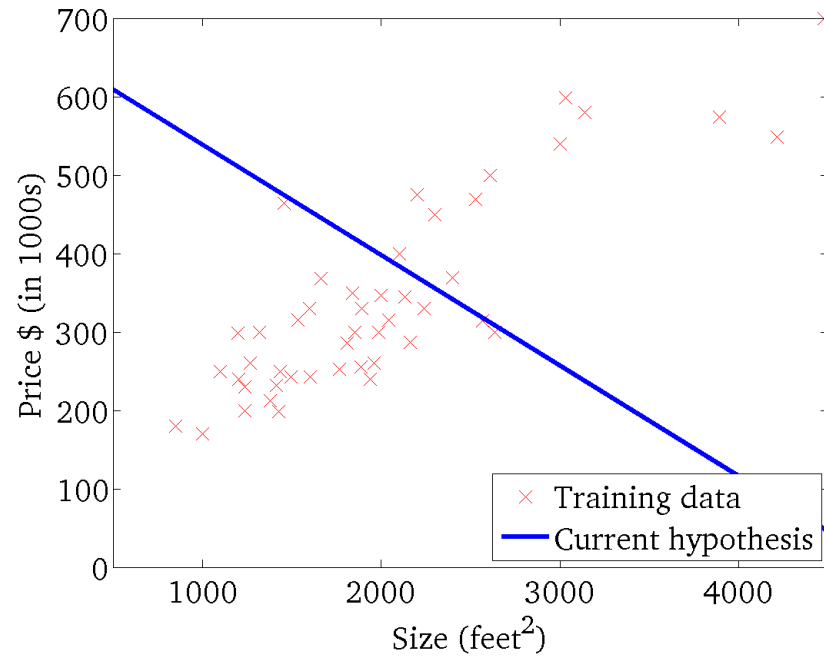
$$J(\theta_0, \theta_1)$$



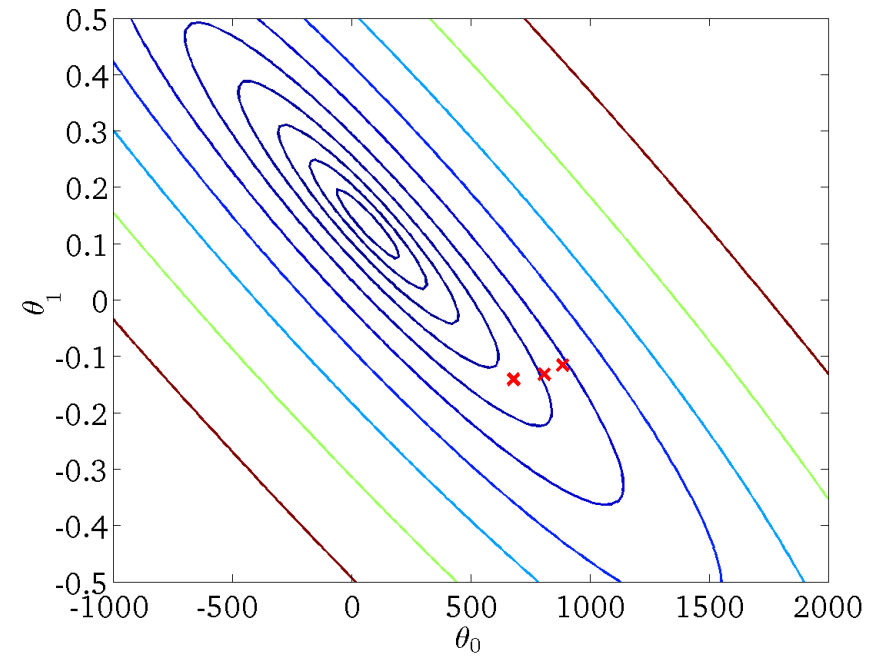
گرادیان کاهشی

۴۱

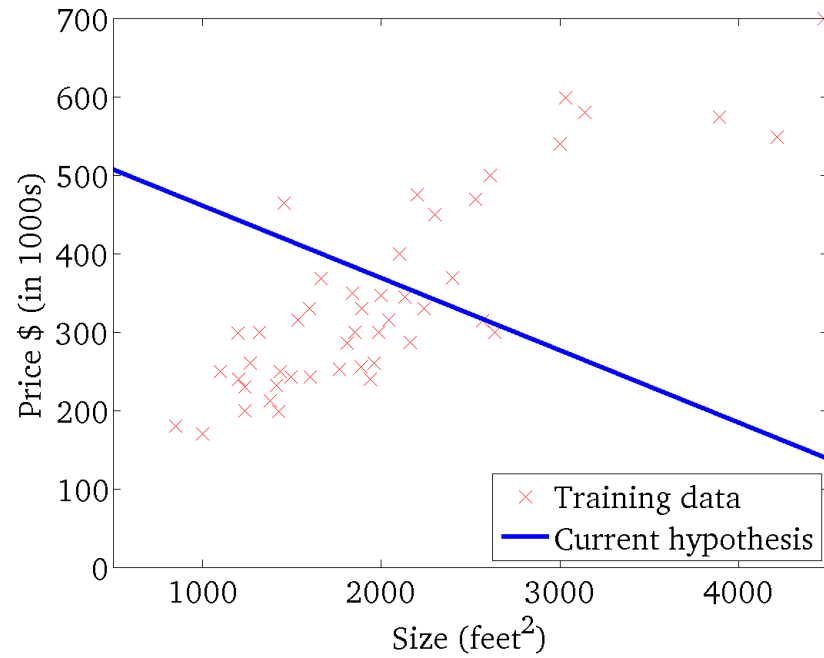
$$h_{\theta}(x) = \theta_0 + \theta_1 x$$



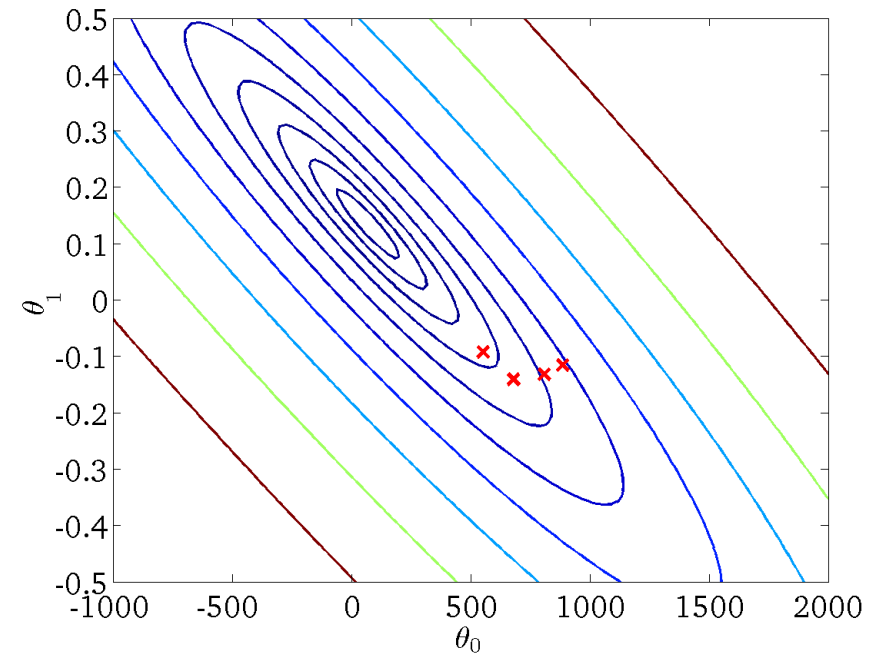
$$J(\theta_0, \theta_1)$$



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



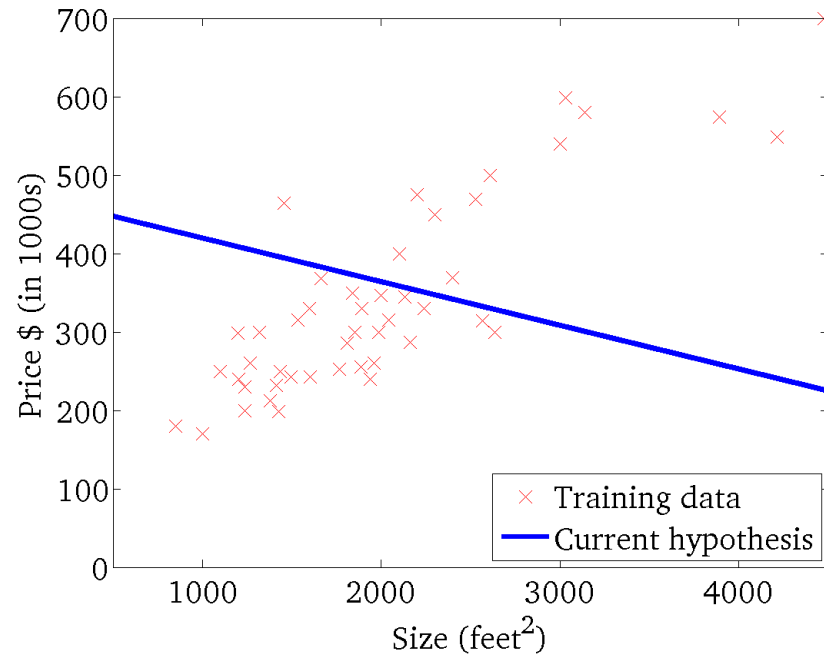
$$J(\theta_0, \theta_1)$$



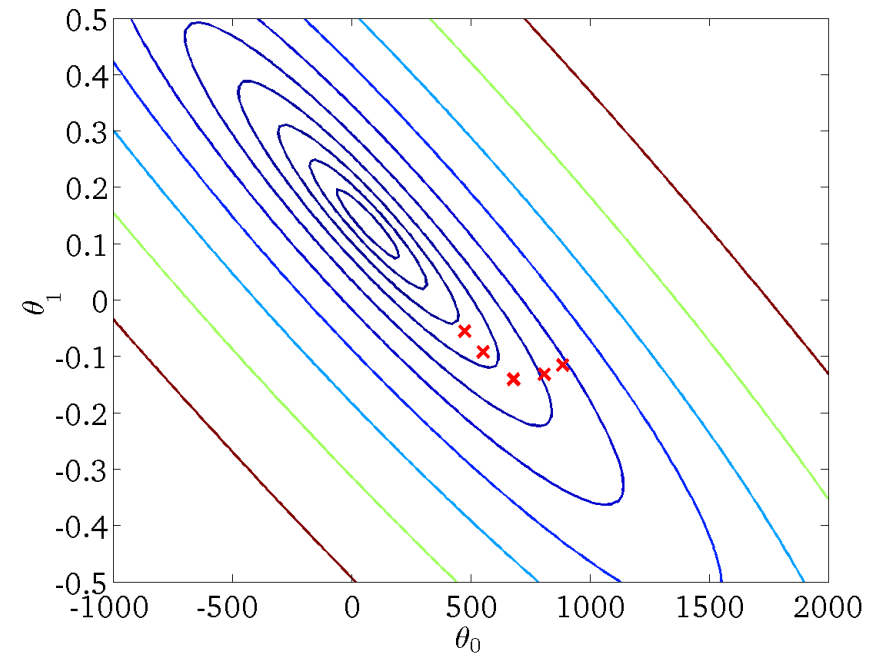
گرادیان کاهشی

۴۳

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



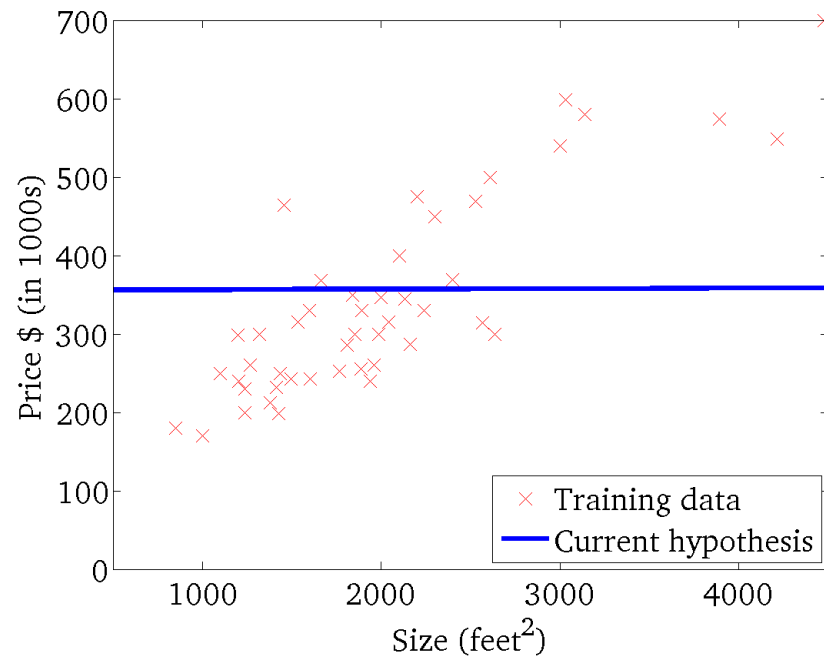
$$J(\theta_0, \theta_1)$$



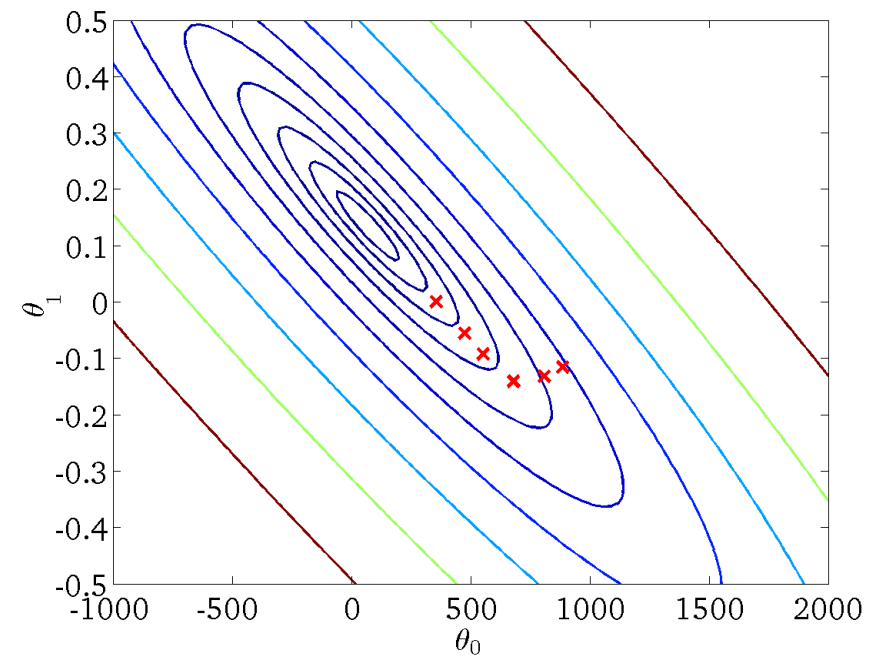
گرادیان کاهشی

۴۴

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



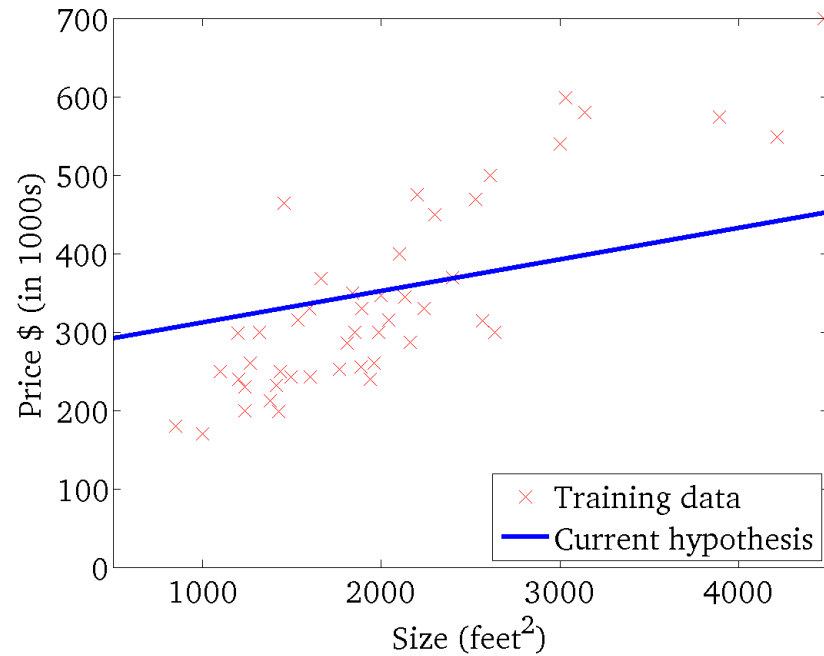
$$J(\theta_0, \theta_1)$$



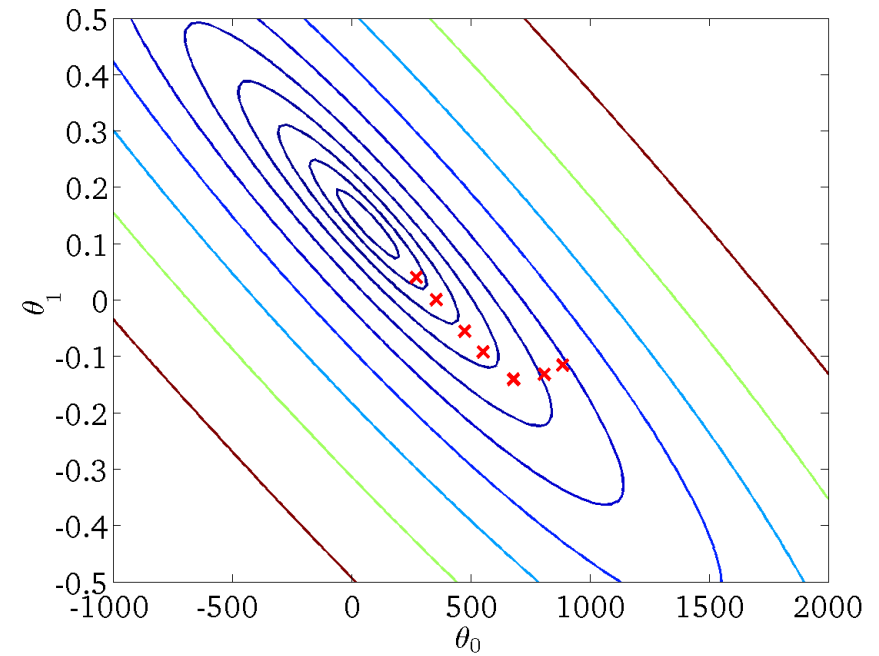
گرادیان کاهشی

۴۵

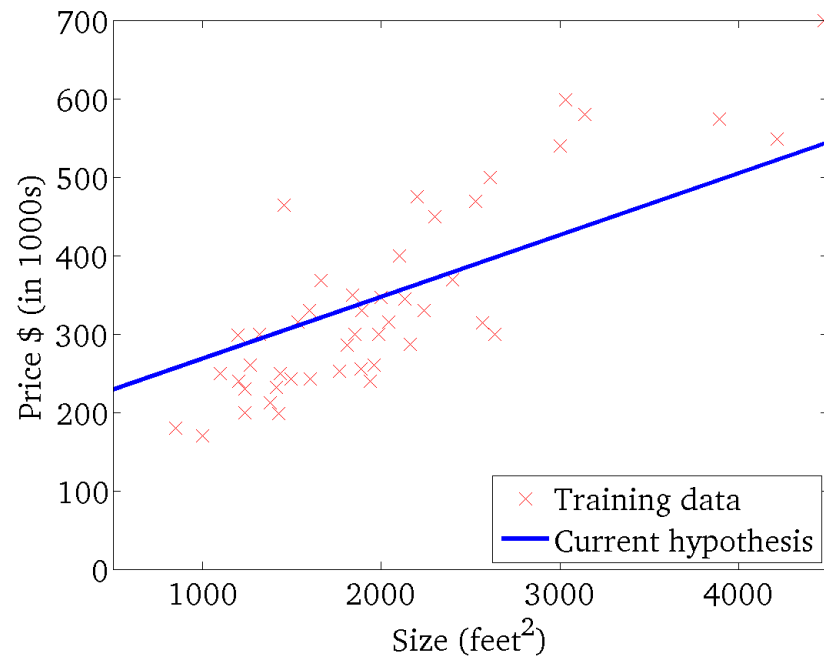
$$h_{\theta}(x) = \theta_0 + \theta_1 x$$



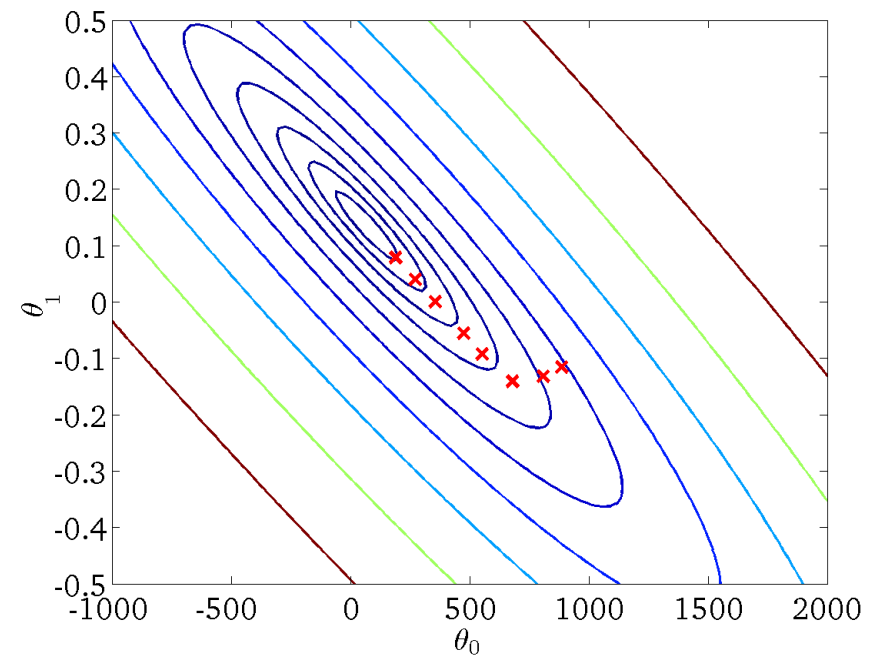
$$J(\theta_0, \theta_1)$$



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



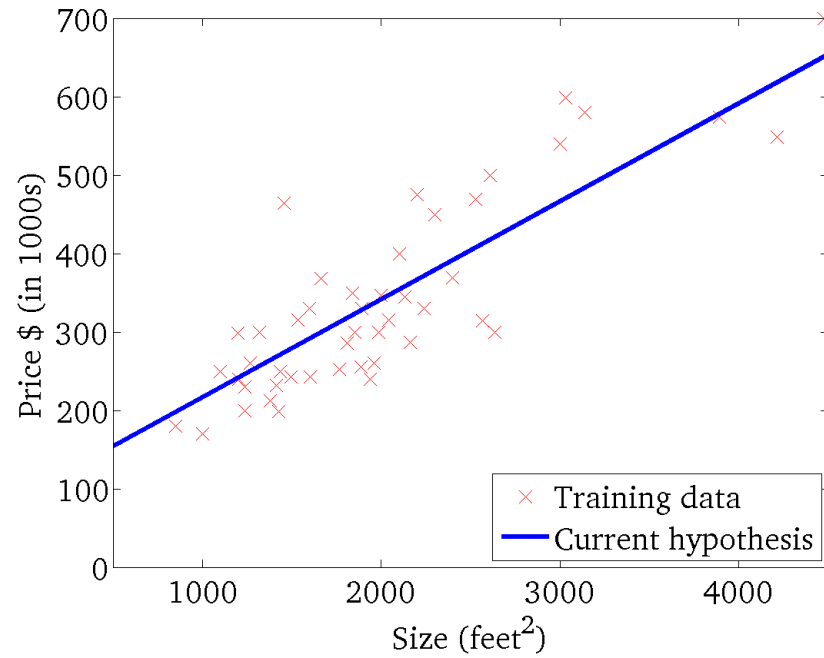
$$J(\theta_0, \theta_1)$$



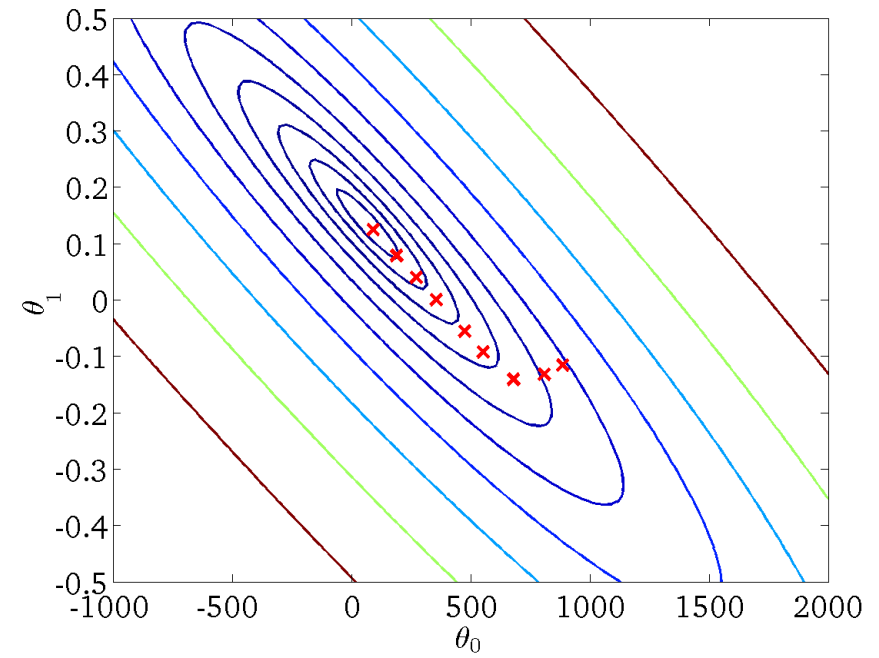
گرادیان کاهشی

۴۷

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



$$J(\theta_0, \theta_1)$$



گرادیان کاهشی دسته‌ای

۴۸

□ **گرادیان کاهشی دسته‌ای.** در هر تکرار الگوریتم، از تمام نمونه‌های آموزشی برای به‌روزرسانی مقدار پارامترها استفاده می‌شود.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (\text{for } j = 0 \text{ and } j = 1)$$

$$j = 0 \Rightarrow \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$j = 1 \Rightarrow \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

رگرسيون خطى چند متغيره

رگرسیون تک متغیره (یک ویژگی)

۵۰

□ رگرسیون خطی با یک ویژگی.

متر از (فوت مربع)	قیمت (۱۰۰۰ دلار)
x	y
۲۱۰۴	۴۶۰
۱۴۱۶	۲۳۲
۱۵۳۴	۳۱۵
۸۵۲	۱۷۸
...	...

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

رگرسیون چند متغیره (چند ویژگی)

۵۱

□ رگرسیون خطی با چند ویژگی.

مترآژ (فوت مربع) x_1	تعداد اتاق خوابها x_2	تعداد طبقات x_3	سن خانه (سال) x_4	قیمت (۱۰۰۰ دلار) y
۲۱۰۴	۵	۱	۴۵	۴۶۰
۱۴۱۶	۳	۲	۴۰	۲۳۲
۱۵۳۴	۳	۲	۳۰	۳۱۵
۸۵۲	۲	۱	۳۶	۱۷۸
...	...	...	...	...

رگرسیون چند متغیره (چند ویژگی)

۵۲

مترائز (فوت مربع) x_1	تعداد اتاق خواب ها x_2	تعداد طبقات x_3	سن خانه (سال) x_4	قیمت (۱۰۰۰ دلار) y
۲۱۰۴	۵	۱	۴۵	۴۶۰
۱۴۱۶	۳	۲	۴۰	۲۳۲
۱۵۳۴	۳	۲	۳۰	۳۱۵
۸۵۲	۲	۱	۳۶	۱۷۸
...	...	...	...	...

تعداد نمونه های آموزشی

m

n

□ نمادها.

n □

$x^{(i)}$ □

$x_j^{(i)}$ □

تعداد ویژگی ها

ورودی ها در i امین نمونه آموزشی

مقدار ویژگی j ام در i امین نمونه آموزشی

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix}$$

□ رگرسیون خطی تک متغیره.

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

□ رگرسیون خطی چند متغیره.

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

□ برای سادگی، تعریف می کنیم $x_0 = 1$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$



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

گرادیان کاهشی در رگرسیون خطی چند متغیره

گرادیان کاهشی

۵۵

□ فرضیه.

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

□ پارامترها.

$$\theta = (\theta_0, \theta_1, \dots, \theta_n)$$

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

□ تابع هزینه.

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

}

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

ترندهای عملی در رگرسیون چند متغیره

مقیاس بندی ویژگی ها

تعیین نرخ یادگیری

مقیاس بندی ویژگی‌ها (نرمال سازی)

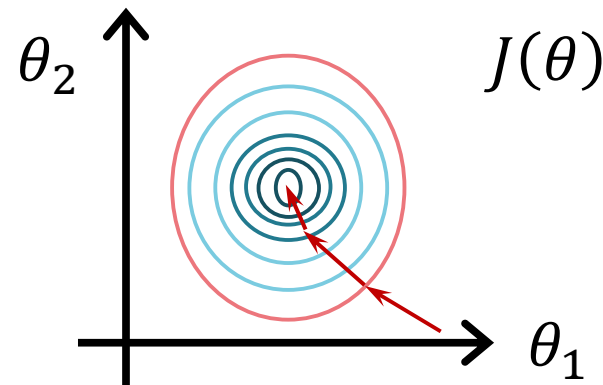
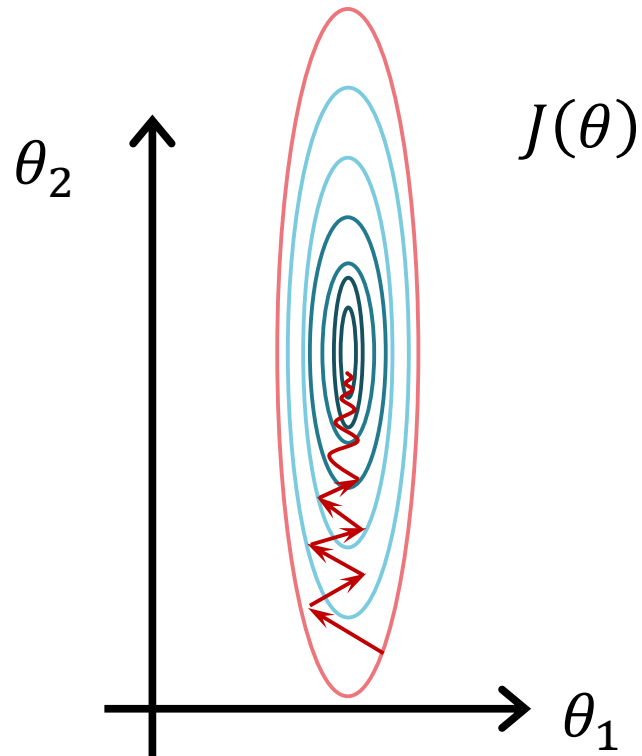
۵۷

- ایده. اطمینان از این که مقادیر ویژگی‌ها در یک مقیاس مشابه قرار دارند.
- هدف. افزایش سرعت همگرایی در گرادیان کاهشی.

□ مثال.

□ x_1 : اندازه خانه (۰ تا ۲۰۰۰)

□ x_2 : تعداد اتاق خواب‌ها (۱ تا ۵)



مقیاس بندی ویژگی‌ها

۵۸

□ مقیاس بندی. مقدار هر ویژگی در ضریب کوچکی از بازه ۱- تا ۱+ قرار دارد.

□ مثال.

$$x_1 = \frac{size - 1000}{2000} \quad -0.5 \leq x_1 \leq 0.5$$

$$x_2 = \frac{\# \text{ bedrooms} - 2}{5} \quad -0.5 \leq x_2 \leq 0.5$$

□ نرمال سازی میانگین.

$$x_i = \frac{x_i - \mu_i}{s_i}$$

میانگین μ_i

انحراف معیار s_i

گرادیان کاهش

۵۹

□ گرادیان کاهش.

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

□ س. چگونه می‌توان مطمئن شد گرادیان کاهش به درستی عمل می‌کند؟

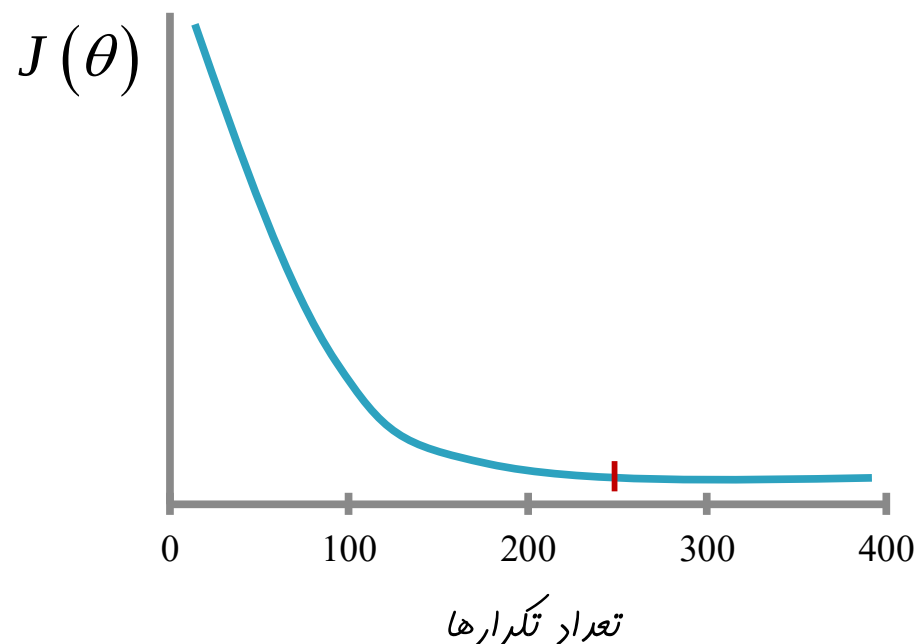
□ س. مقدار مناسب برای نرخ یادگیری چیست؟

عملکرد صحیح برای گرادیان کاهشی

۶۰

□ آزمایش همگرایی.

□ اگر مقدار $J(\theta)$ در یک تکرار به اندازه‌ای کمتر از 10^{-3} تغییر کند، همگرایی رخ داده است.

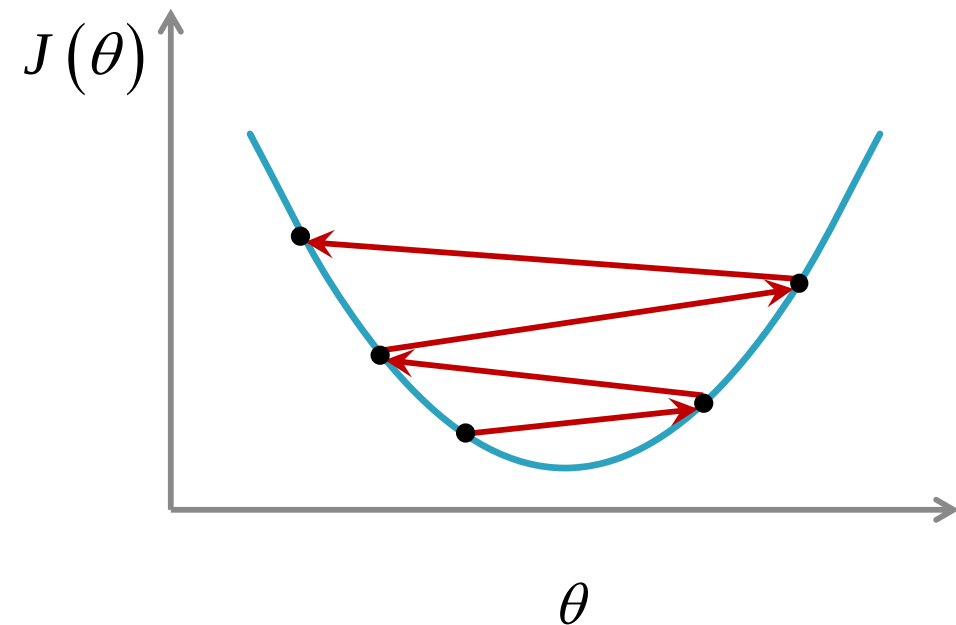
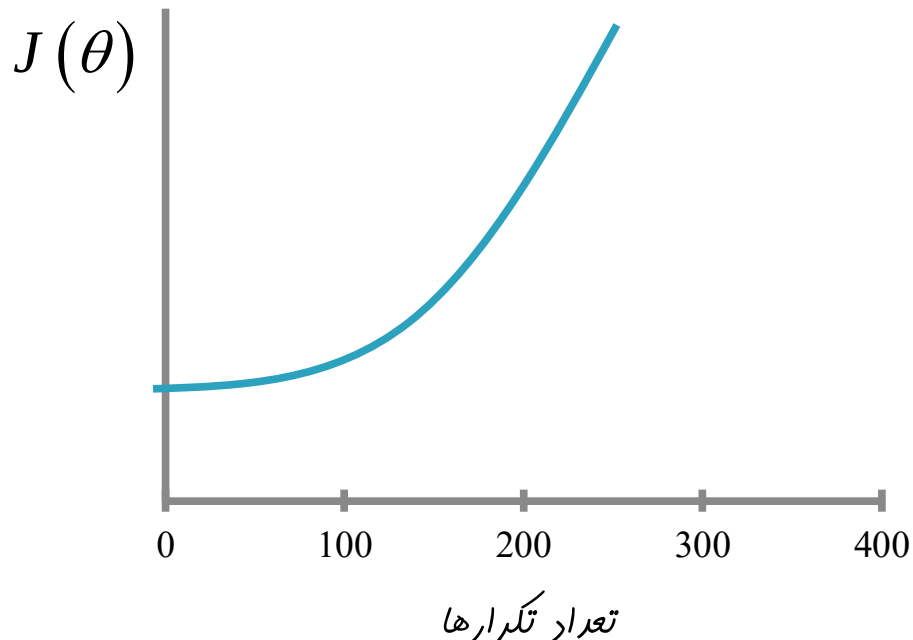


نشانه‌های عملکرد نادرست گرادینان کاهششی

۶۱

□ عدم همگرایی.

□ راه حل. از مقادیر کوچک‌تری برای نرخ یادگیری استفاده کن، اما اگر نرخ یادگیری بیش از حد کوچک باشد همگرایی بسیار کند خواهد بود.



□ نرخ یادگیری.

□ بسیار کوچک: همگرایی بسیار کند

□ بسیار بزرگ: همگرایی کند یا عدم همگرایی

□ انتخاب نرخ یادگیری.

□ به منظور انتخاب یک مقدار مناسب برای نرخ یادگیری، مقادیر زیر را امتحان کنید:

..., 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1.0, ...

رگرسیون چندجمله‌ای



قیمت‌گذاری یک خانه: انتخاب ویژگی‌ها

۶۴

$$h_{\theta}(x) = \theta_0 + \theta_1 \times \underbrace{(\text{frontage})}_{x_1} + \theta_2 \times \underbrace{(\text{depth})}_{x_2}$$

$$\text{Area} = \text{frontage} \times \text{depth}$$

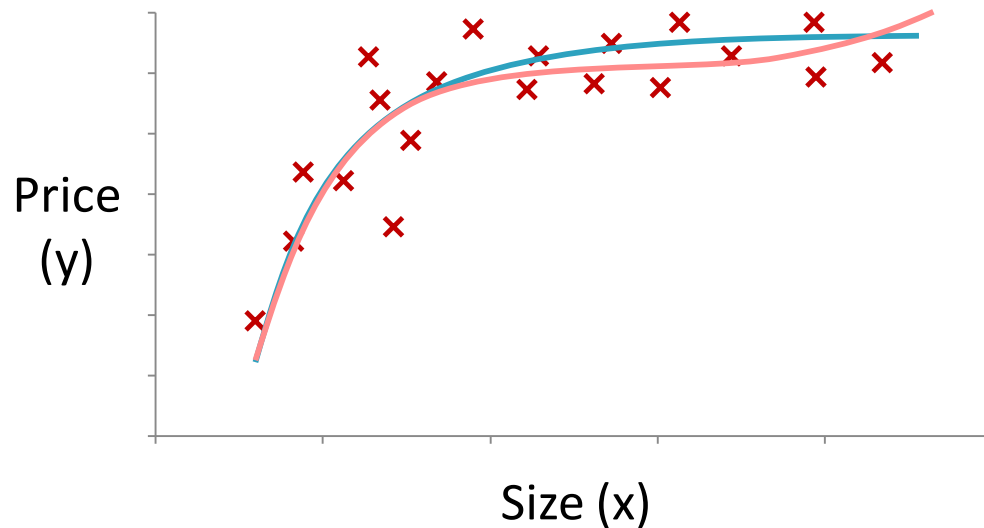
$$h_{\theta}(x) = \theta_0 + \theta_1 \times (\text{Area})$$



رگرسیون چند جمله‌ای

۶۵

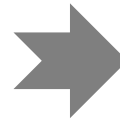
□ رگرسیون چند جمله‌ای.



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

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

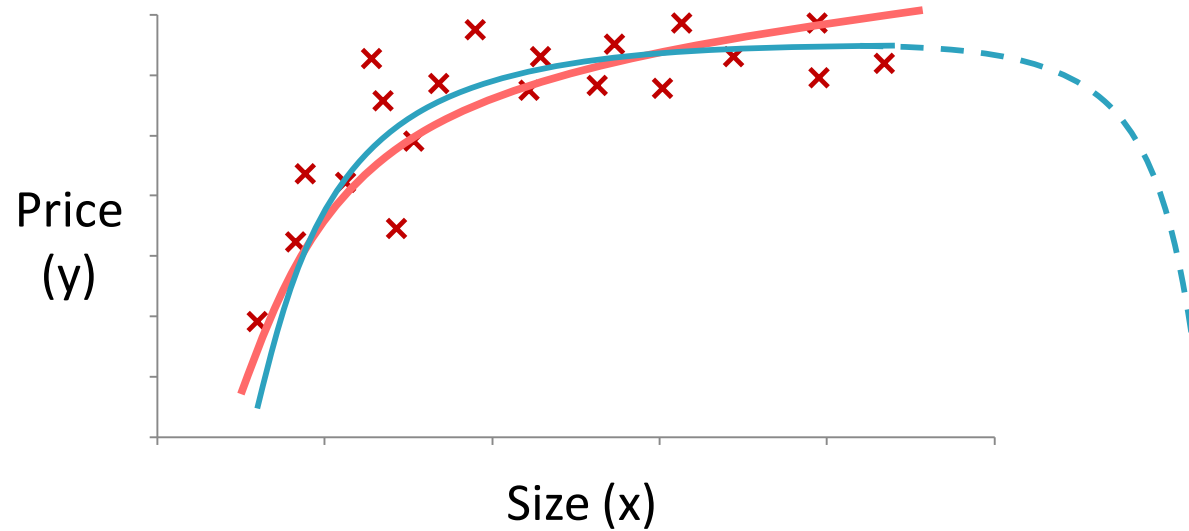
$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$



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

انتخاب ویژگی‌ها

۶۶



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

$$h_{\theta}(x) = \theta_0 + \theta_1(\text{size}) + \theta_2(\sqrt{\text{size}})$$

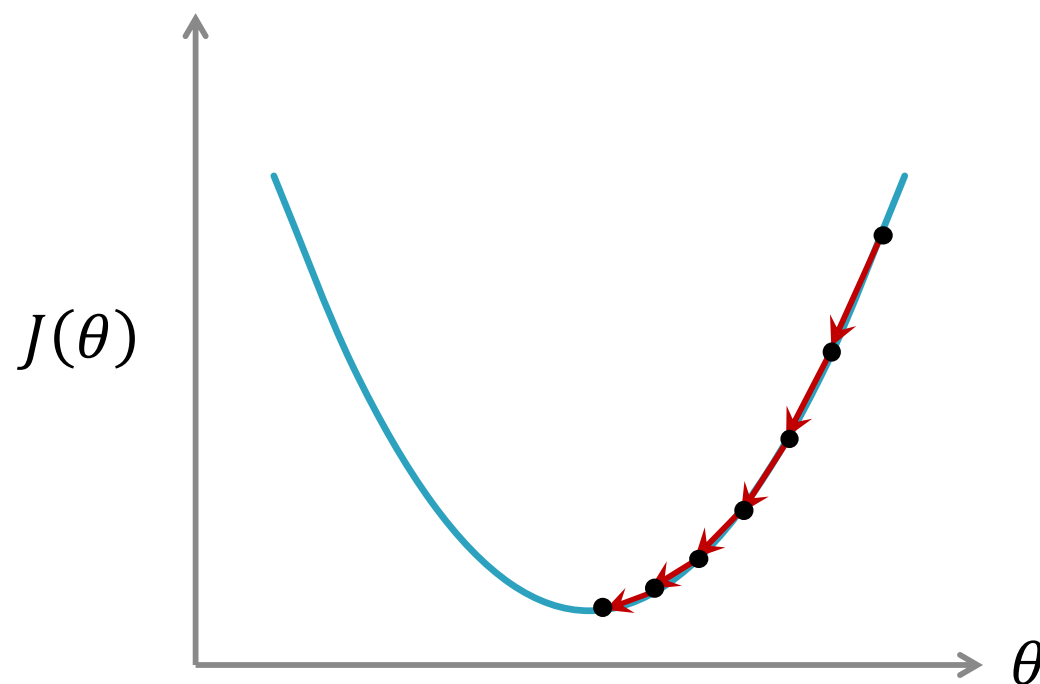
رگرسیون خطی: معادله نرمال



گرادیان کاهش و معادله نرمال

۶۸

□ گرادیان کاهش.



$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

□ معادله نرمال. یک روش تحلیلی به منظور تعیین مقدار پارامترها.

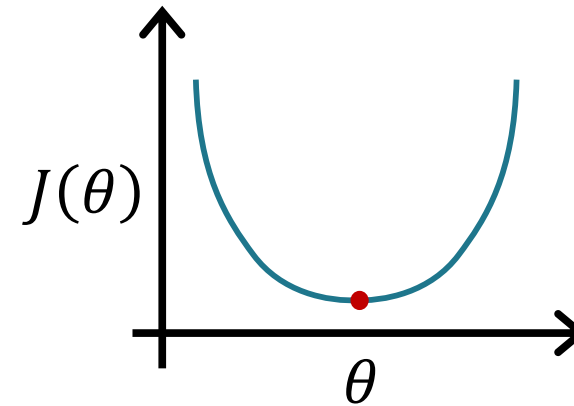
معادله‌ی نرمال

۶۹

□ معادله‌ی نرمال.

$$\theta \in \mathbb{R}: J(\theta) = a\theta^2 + b\theta + c$$

$$\frac{d}{d\theta}J(\theta) \stackrel{\text{def}}{=} 0$$



$$\theta \in \mathbb{R}^{n+1}: J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$$\frac{\partial}{\partial \theta_j} J(\theta) \stackrel{\text{def}}{=} 0 \quad (j = 0, 1, 2, \dots, n)$$

معادله‌ی نرمال: مثال ($m = 4$)

۷۰

	Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
x_0	x_1	x_2	x_3	x_4	y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$$X\theta = y$$

معادله‌ی نرمال: مثال ($m = 5$)

۷۱

	Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
x_0	x_1	x_2	x_3	x_4	y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178
1	3000	4	1	38	540

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \\ 1 & 3000 & 4 & 1 & 38 \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \\ 540 \end{bmatrix}$$

$$X\theta = y$$

معادله‌ی نرمال: حالت کلی

۷۲

□ m نمونه آموزشی؛ n ویژگی

$$X\theta = y$$

$$x^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \in \mathbb{R}^{n+1} \quad X = \begin{bmatrix} \text{---} (x^{(1)})^T \text{---} \\ \text{---} (x^{(2)})^T \text{---} \\ \text{---} (x^{(3)})^T \text{---} \\ \vdots \\ \text{---} (x^{(m)})^T \text{---} \end{bmatrix} \in \mathbb{R}^{m \times (n+1)} \quad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{bmatrix} \in \mathbb{R}^m$$

ماتریس طراحی

معادله‌ی نرمال

۷۳

□ حل دستگاه معادلات خطی.

$$X\theta = y$$

$$X^T X \theta = X^T y$$

← معادله‌ی نرمال

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

← ماتریس شبه معکوس X^+

Octave:

```
theta = pinv(X' * X) * X' * y
```

گرادیان کاهشی و معادله‌ی نرمال

معادله‌ی نرمال

- عدم نیاز به انتخاب α
- عدم نیاز به تکرار
- به دلیل نیاز به محاسبه‌ی معکوس ماتریس $X^T X$ ، برای n های بزرگ کند است.

$$n < 10000$$

گرادیان کاهشی

- نیاز به انتخاب α
- نیاز به تکرارهای زیاد
- حتی برای مقادیر بسیار بزرگ n به خوبی کار می‌کند.

$$n \geq 10000$$

معادله‌ی نرمال و معکوس ناپذیری $X^T X$

۷۶

□ معادله‌ی نرمال.

$$\theta = (X^T X)^{-1} X^T y$$

□ س. اما اگر $X^T X$ معکوس پذیر نباشد چه می‌شود؟

Octave: `theta = pinv(X' * X) * X' * y`

شبه معکوس

علل معکوس ناپذیری $X^T X$

۷۷

□ افزونگی ویژگی‌ها (وابستگی خطی).

$$x_1 = \text{size (feet}^2\text{)}$$

$$x_2 = \text{size (m}^2\text{)}$$

$$x_1 = (3.28)^2 x_2$$

□ تعداد بسیار زیاد ویژگی‌ها (مثلاً $n \geq m$).

□ راه حل. حذف برخی از ویژگی‌ها با استفاده از **تنظیم** [در ادامه]

رگرسیون با وزن‌دهی محلی



روش‌های یادگیری پارامتری و غیرپارامتری

۷۹

□ روش‌های یادگیری پارامتری.

- یک مجموعه‌ی ثابت از پارامترها وجود دارد.
- برای پیش‌بینی داده‌های جدید نیازی به مجموعه‌ی آموزشی نداریم.
- مثال: رگرسیون و رگرسیون لجستیک [در ادامه].

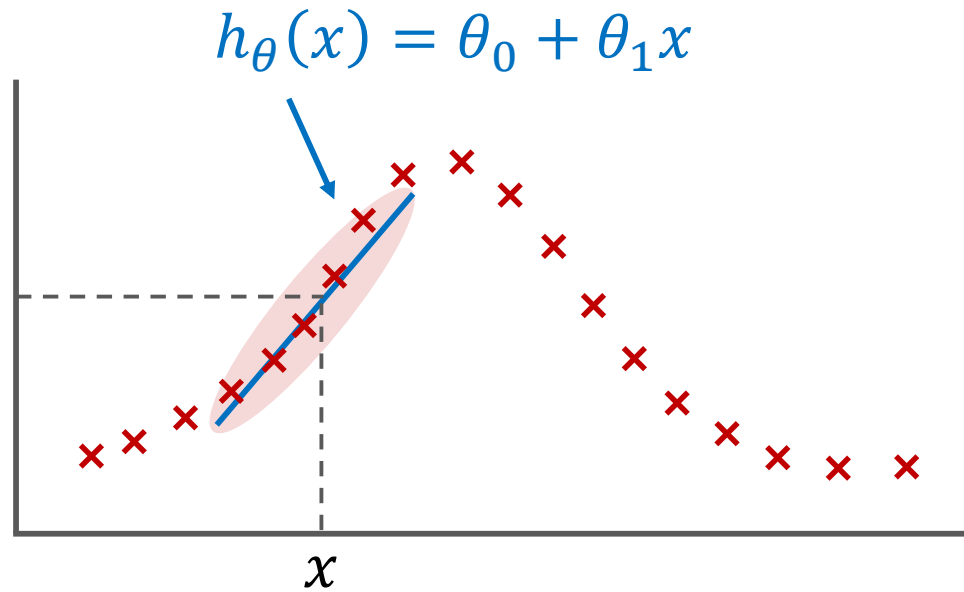
□ روش‌های یادگیری غیرپارامتری.

- تعداد پارامترها با افزایش اندازه‌ی مجموعه‌ی آموزشی (به صورت خطی) افزایش می‌یابد.
- به منظور پیش‌بینی داده‌های جدید به تمام مجموعه‌ی آموزشی نیاز داریم.
- مثال: رگرسیون با وزن‌دهی محلی [صفحه‌ی بعد].

رگرسیون با وزن‌دهی محلی

۸۰

□ ایده. دادن اهمیت بیشتر به داده‌های نزدیک‌تر.



$$w^{(i)} = \exp\left(-\frac{(x^{(i)} - x)^2}{2\tau^2}\right)$$

پهنای باند

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

رگرسیون: تفسیر احتمالاتی

رگرسیون: تفسیر احتمالاتی

۸۲

□ مدل.

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

خطا

□ خطا.

□ در نظر گرفتن تأثیر عوامل دیگر

■ مانند ویژگی‌های در نظر گرفته نشده

□ در نظر گرفتن تأثیر نویز.

$$\epsilon^{(i)} \sim N(0, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right)$$

□ خطاها **مستقل** هستند و از یک **توزیع یکسان** گاوسی پیروی می‌کنند. $[iid]$

$$y^{(i)} \sim N(\theta^T x^{(i)}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

تخمین بیشترین درست‌نمایی

۸۳

□ تابع درست‌نمایی. احتمال مشاهده‌ی داده‌های آموزشی به عنوان تابعی از پارامترهای θ

$$L(\theta) = p(Y|X; \theta) = \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

□ لگاریتم تابع درست‌نمایی.

$$\begin{aligned} l(\theta) = \ln L(\theta) &= \ln \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right) \\ &= \sum_{i=1}^m \ln \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right) \right] \\ &= m \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum_{i=1}^m (y^{(i)} - \theta^T x^{(i)})^2 \end{aligned}$$

تخمین بیشترین درست‌نمایی

۸۴

□ تخمین بیشترین درست‌نمایی.

□ انتخاب یک مقدار برای پارامتر θ به گونه‌ای که $l(\theta)$ بیشینه گردد.

$$\begin{aligned}\theta_{MLE} &= \arg \max_{\theta} l(\theta) \\ &= \arg \max_{\theta} m \ln \frac{1}{\sigma \sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum_{i=1}^m (y^{(i)} - \theta^T x^{(i)})^2 \\ &= \arg \max_{\theta} -\frac{1}{2} \sum_{i=1}^m (y^{(i)} - \theta^T x^{(i)})^2 \\ &= \arg \min_{\theta} \frac{1}{2} \sum_{i=1}^m (y^{(i)} - \theta^T x^{(i)})^2 \quad \leftarrow \text{تابع هزینه}\end{aligned}$$