یادگیری نظارت شده: رگرسیون

سید ناصر رضوی n.razavi@tabrizu.ac.ir

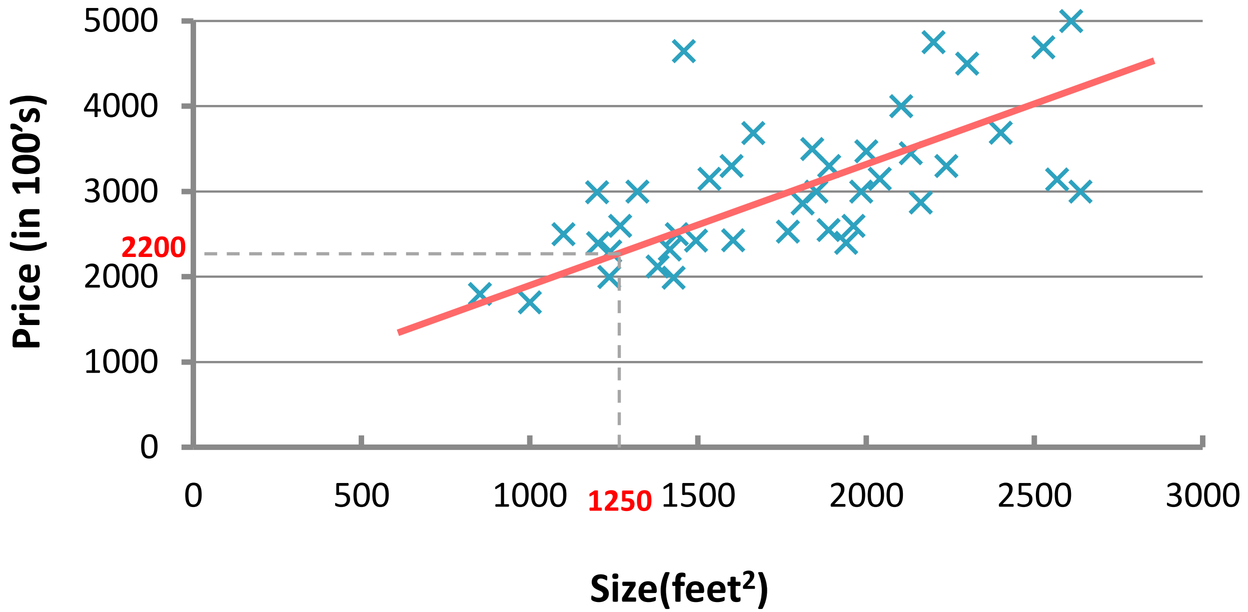
۱۳۹۷

فهرست مطالب

- رگرسیون.
  - رگرسیون خطی تک متغیره و چند متغیره
- گرادیان کاهشی.
- معادله نرمال.
- رگرسیون با وزن‌دهی محلی.
- تفسیر احتمالاتی رگرسیون.
- تخمین بیشترین درست‌نمایی.

رگرسیون خطی تک متغیره

قیمت‌گذاری خانه



- یادگیری نظارت شده.

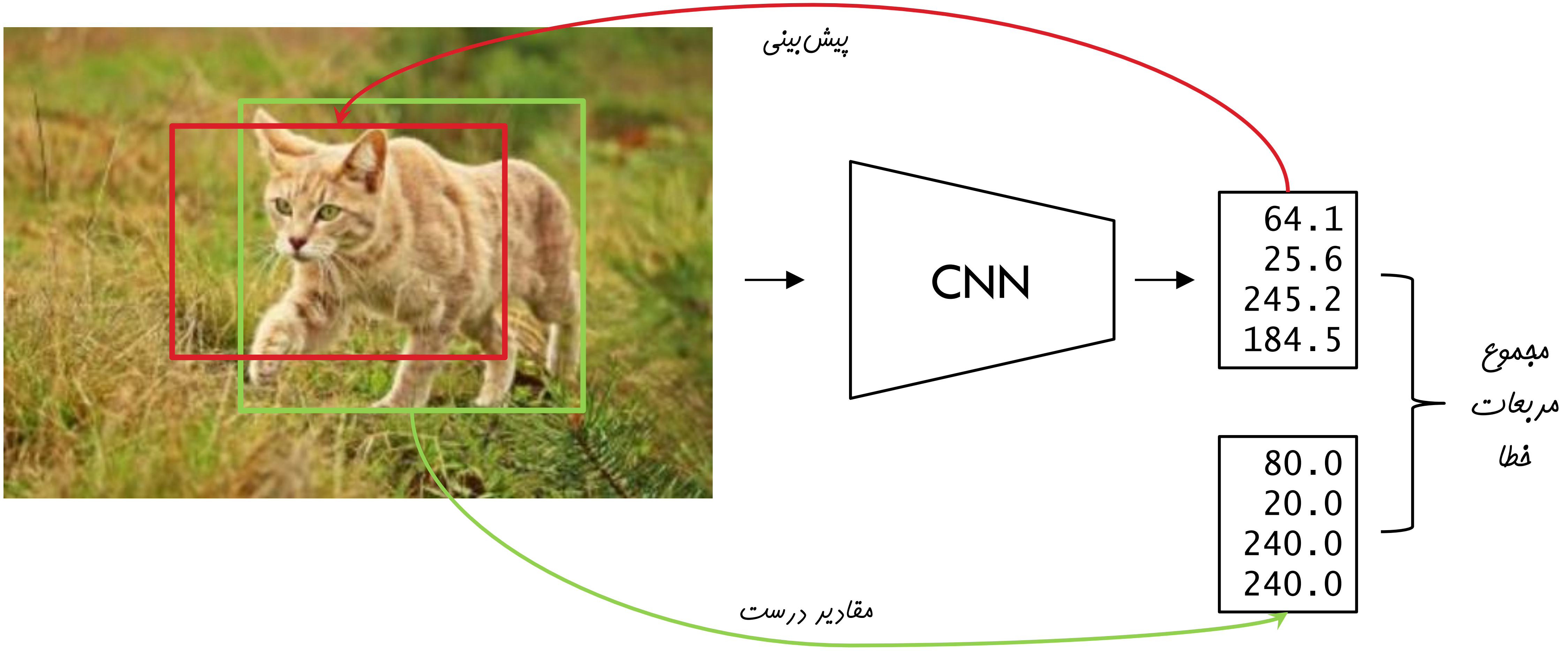
به ازای هر نمونه آموزشی، «پاسخ درست» داده شده است.

- رگرسیون.

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

رگرسیون: شناسایی اشیا

مکان‌یابی به عنوان رگرسیون



نمادگذاری

مجموعه آموزشی

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

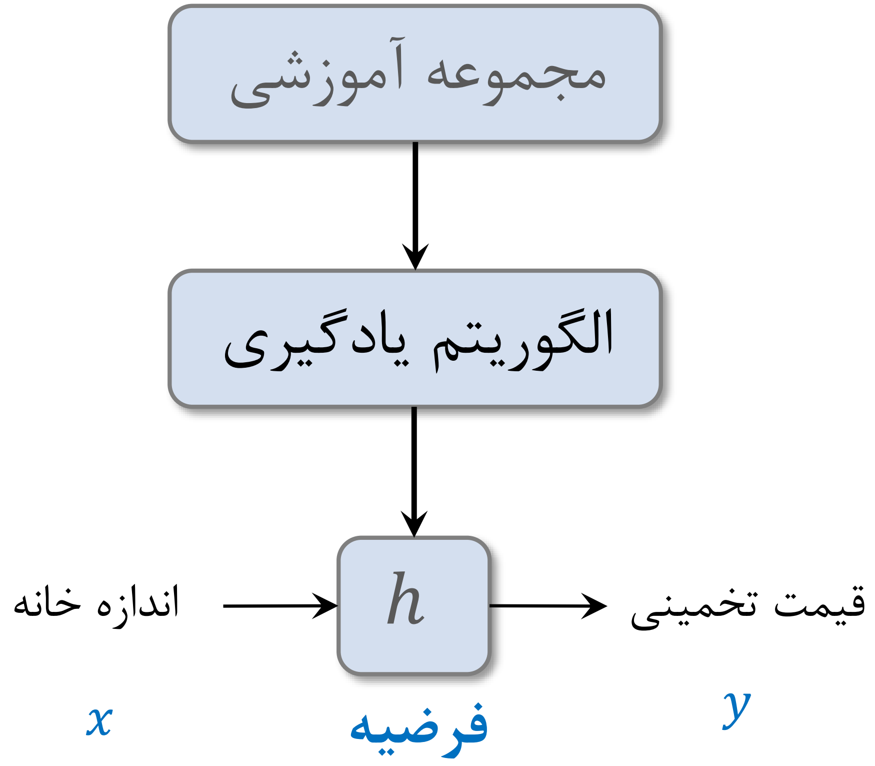
$m = ۴۷$

- نمادها.
  - $m =$ تعداد نمونه‌های آموزشی
  - $x =$ متغیر «ورودی»، ویژگی‌ها
  - $y =$ متغیر «خروجی»، متغیر «هدف»

(x,y): یک نمونه آموزشی

$(x^{(i)},y^{(i)})$: نمونه آموزشی i اُم

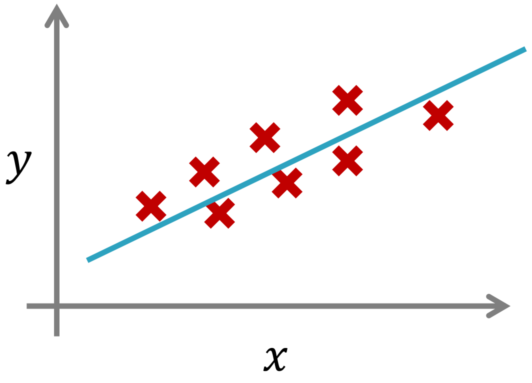
بازنمایی مدل



$$h : x \rightarrow y$$

نمایش فرضیه h.

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



رگرسیون خطی تک متغیره

تابع هزینه

ارزیابی فرضیه

مجموعه آموزشی

$m = ۴۷$

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

فرضیه:

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

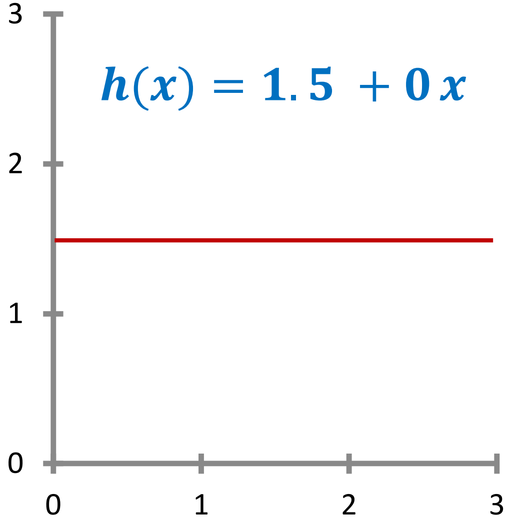
پارامترها:

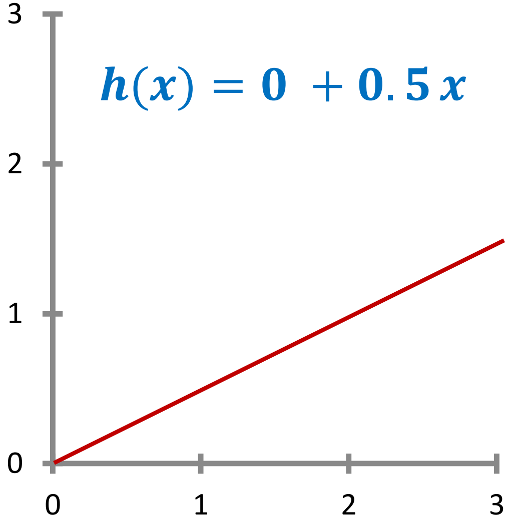
$$(\theta_0 , \theta_1)$$

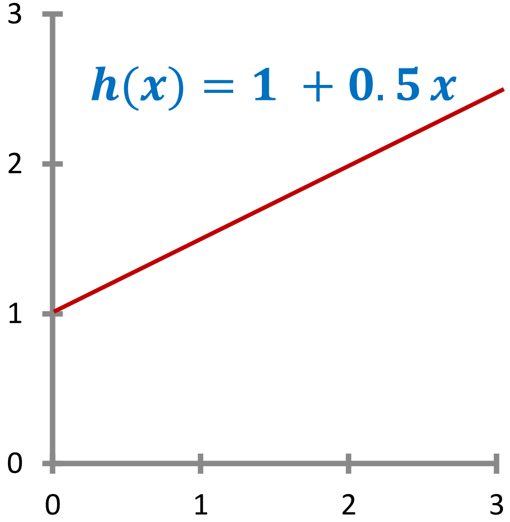
س. مقدار پارامترها را چگونه باید انتخاب نمود؟

ارزیابی فرضیه

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







$\theta_0 = 1.5$
$\theta_1 = 0$

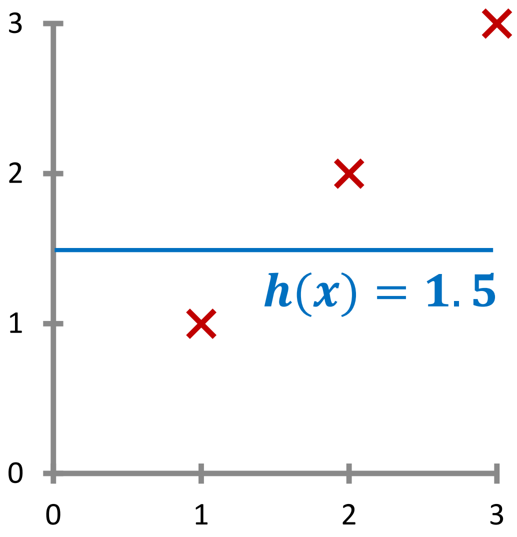
$\theta_0 = 0$
$\theta_1 = 0.5$

$\theta_0 = 1$
$\theta_1 = 0.5$

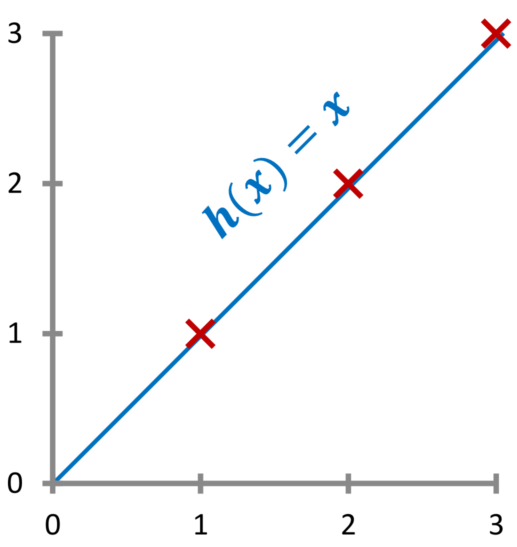
ارزیابی فرضیه

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

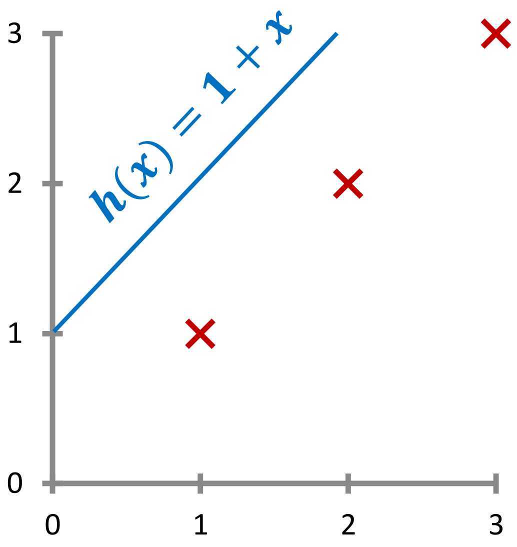
س. کدام فرضیه بهتر است؟



$\theta_0 = 1.5$
$\theta_1 = 0$



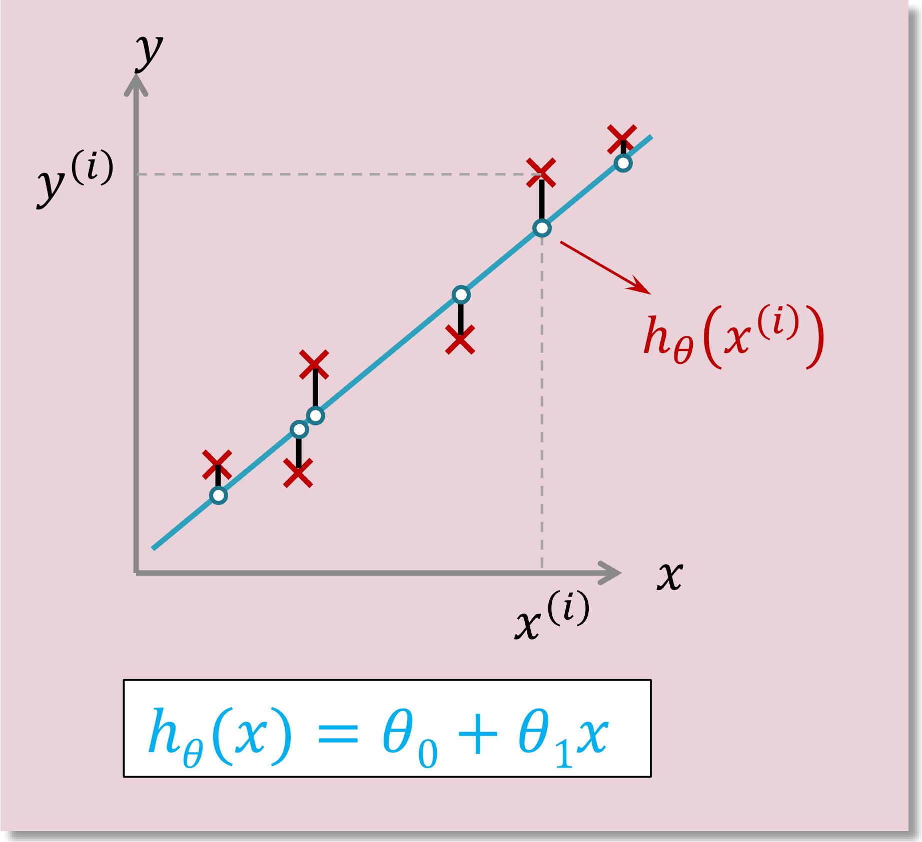
$\theta_0 = 0$
$\theta_1 = 1$



$\theta_0 = 1$
$\theta_1 = 1$

تابع هزینه

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



- تابع هزینه. مجموع مربعات خطا.

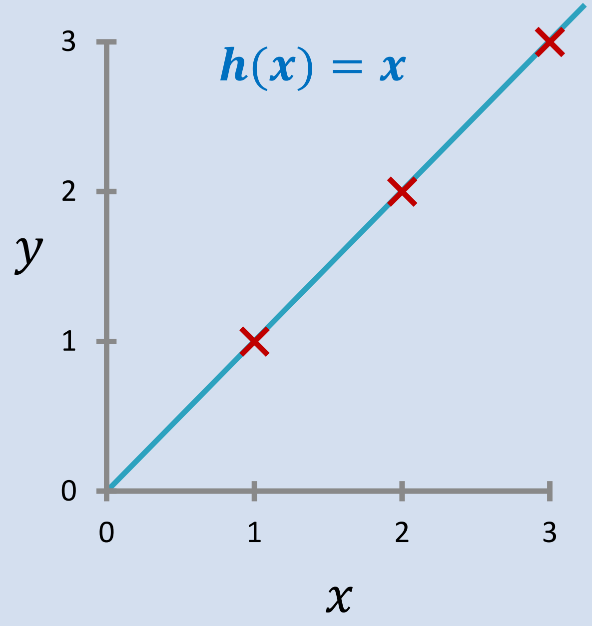
$$J(\theta_0, \theta_1) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- هدف.

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

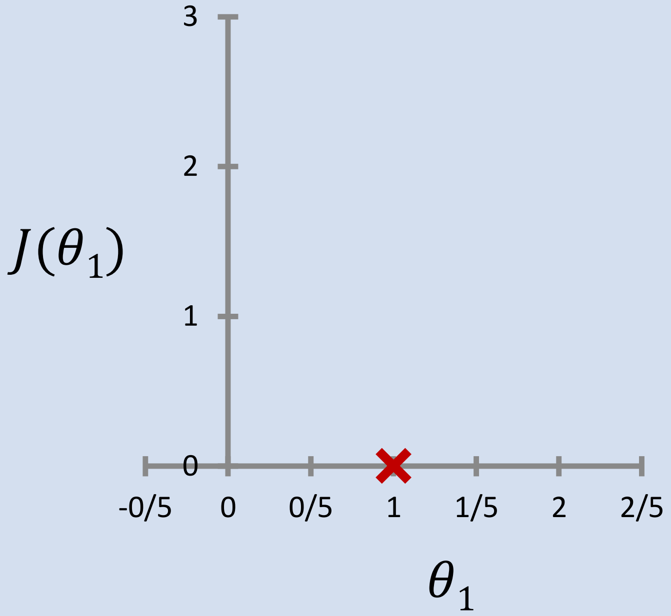
تابع هزینه ساده شده $(\theta_0 = 0)$

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



$\theta_1 = 1$

$J(\theta_1)$

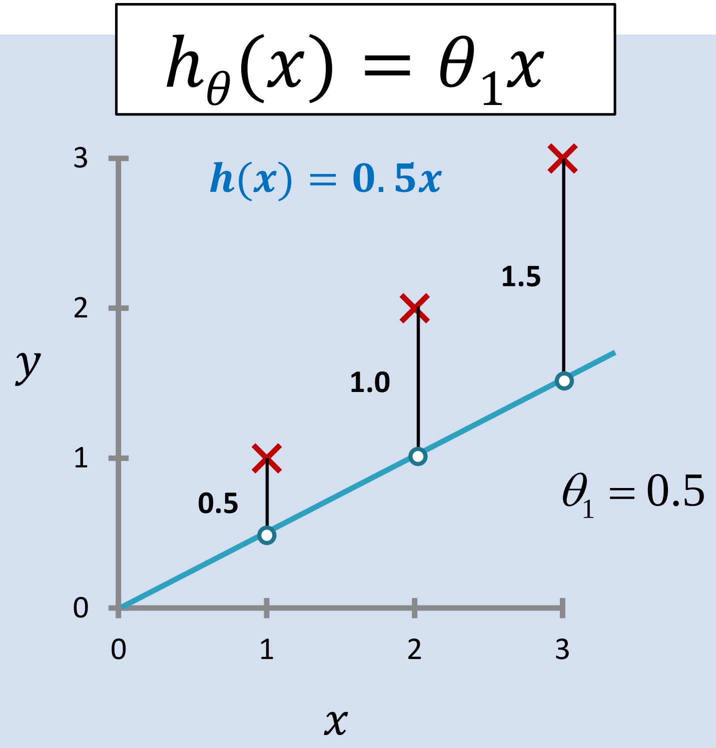


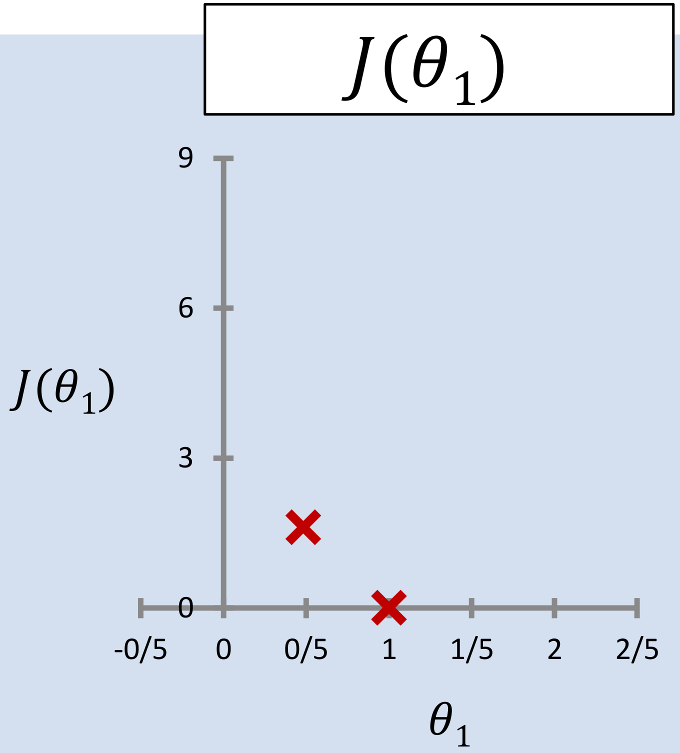
$$J(\theta_0, \theta_1) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{m}\left(x^{(i)} - y^{(i)}\right)^2 = \frac{1}{2}(0^2 + 0^2 + 0^2) = 0$$

$J(1) = 0$

تابع هزینه



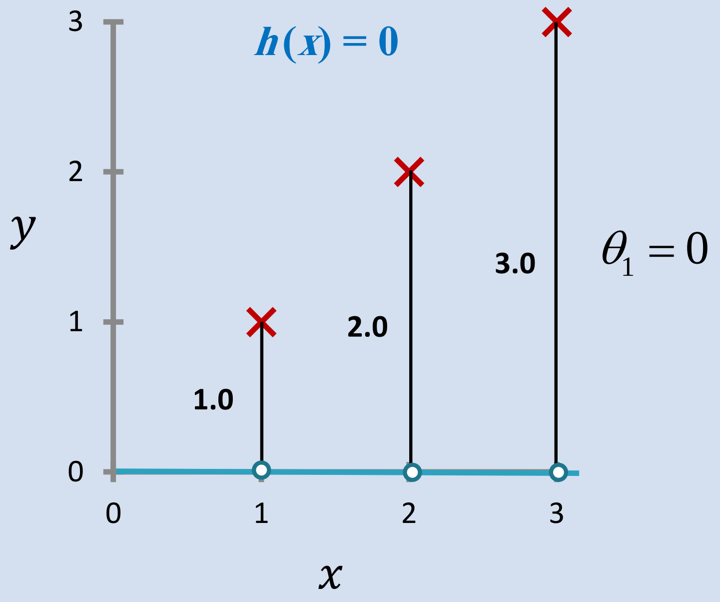


$$J(0.5) = \frac{1}{2}(0.5^2 + 1.0^2 + 1.5^2) = \frac{1}{2}(3.5) = 1.75$$

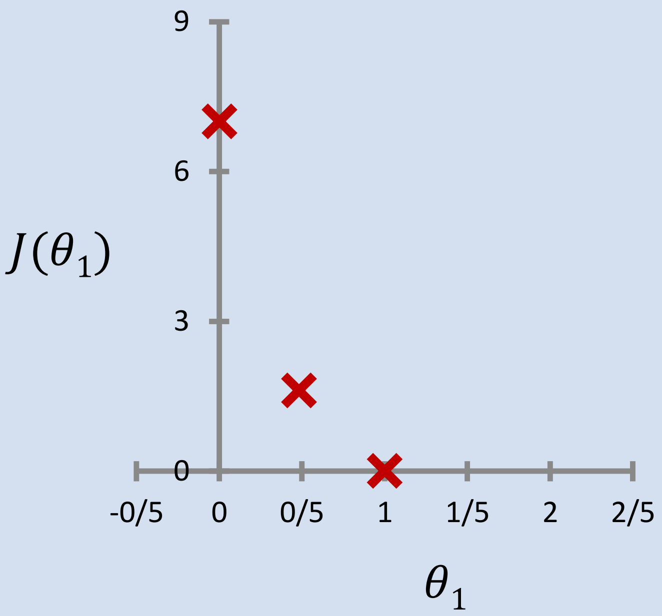
$$J(0.5) = 1.75$$

تابع هزینه

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



$J(\theta_1)$

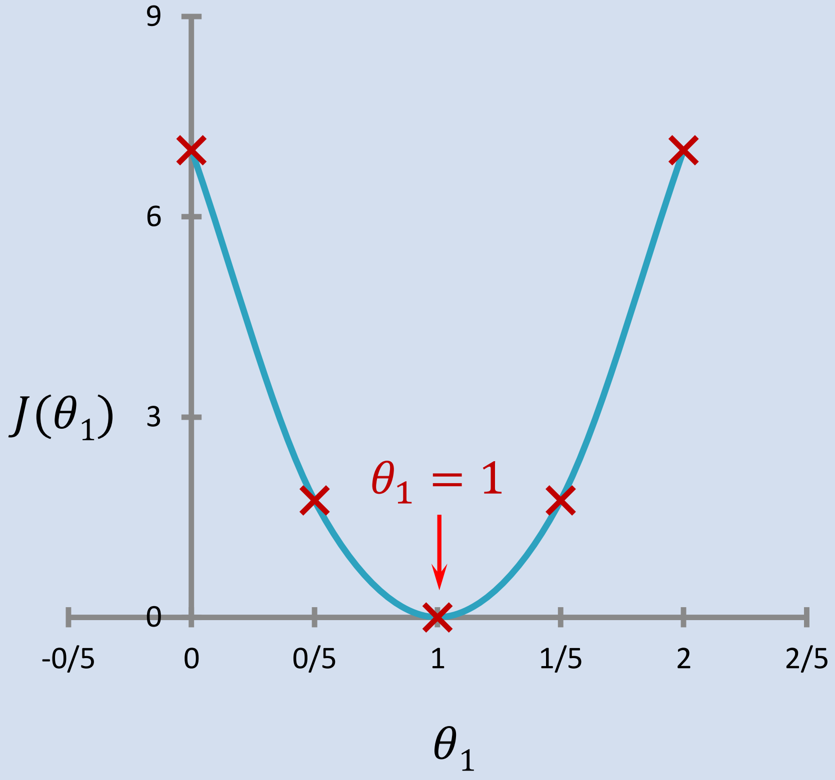


$J(0) = \frac{1}{2}(1.0^2 + 2.0^2 + 3.0^2) = \frac{1}{2}(14) = 7.0$

$J(0) = 7.0$

تابع هزینه

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



رگرسیون خطی تک متغیره

- فرضیه.

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

- پارامترها.

$$\theta_0, \theta_1$$

- تابع هزینه.

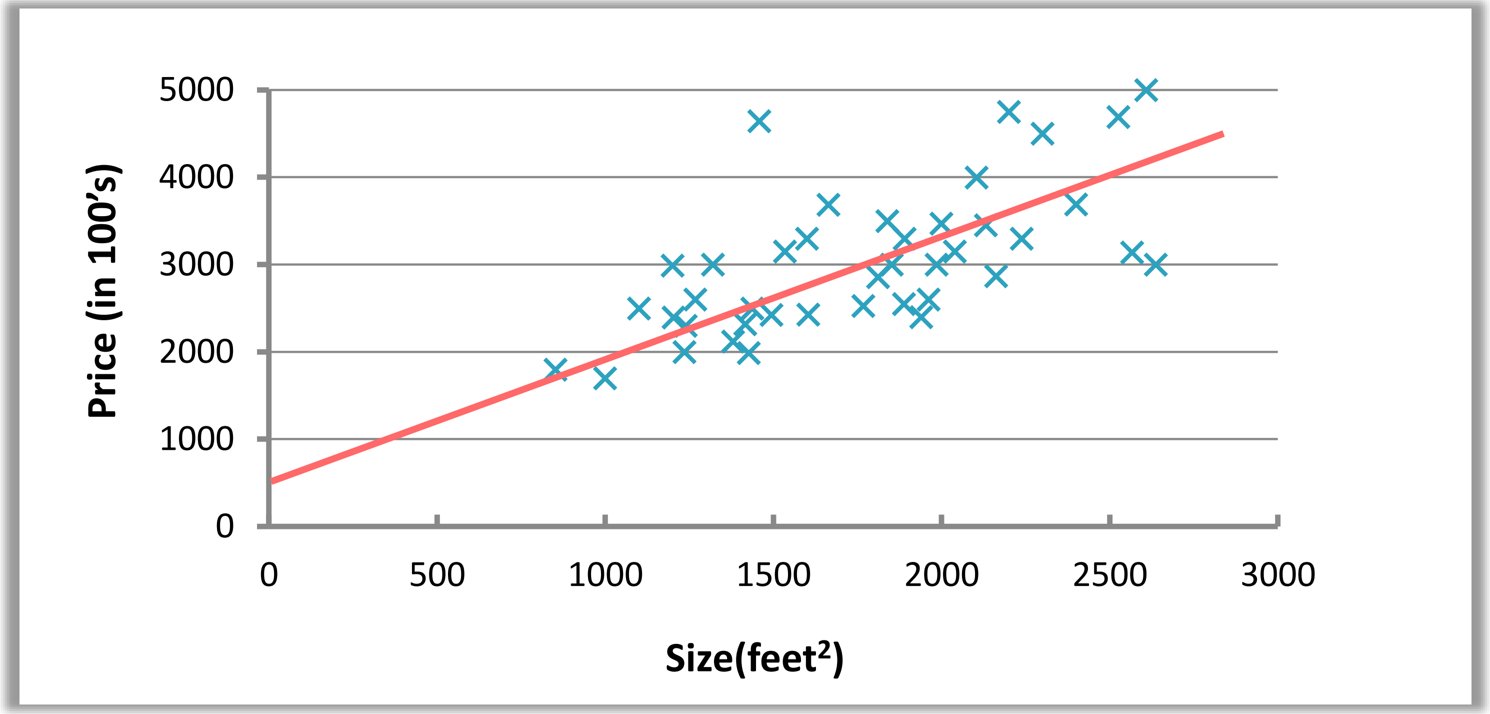
$$J(\theta_0, \theta_1) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- هدف.

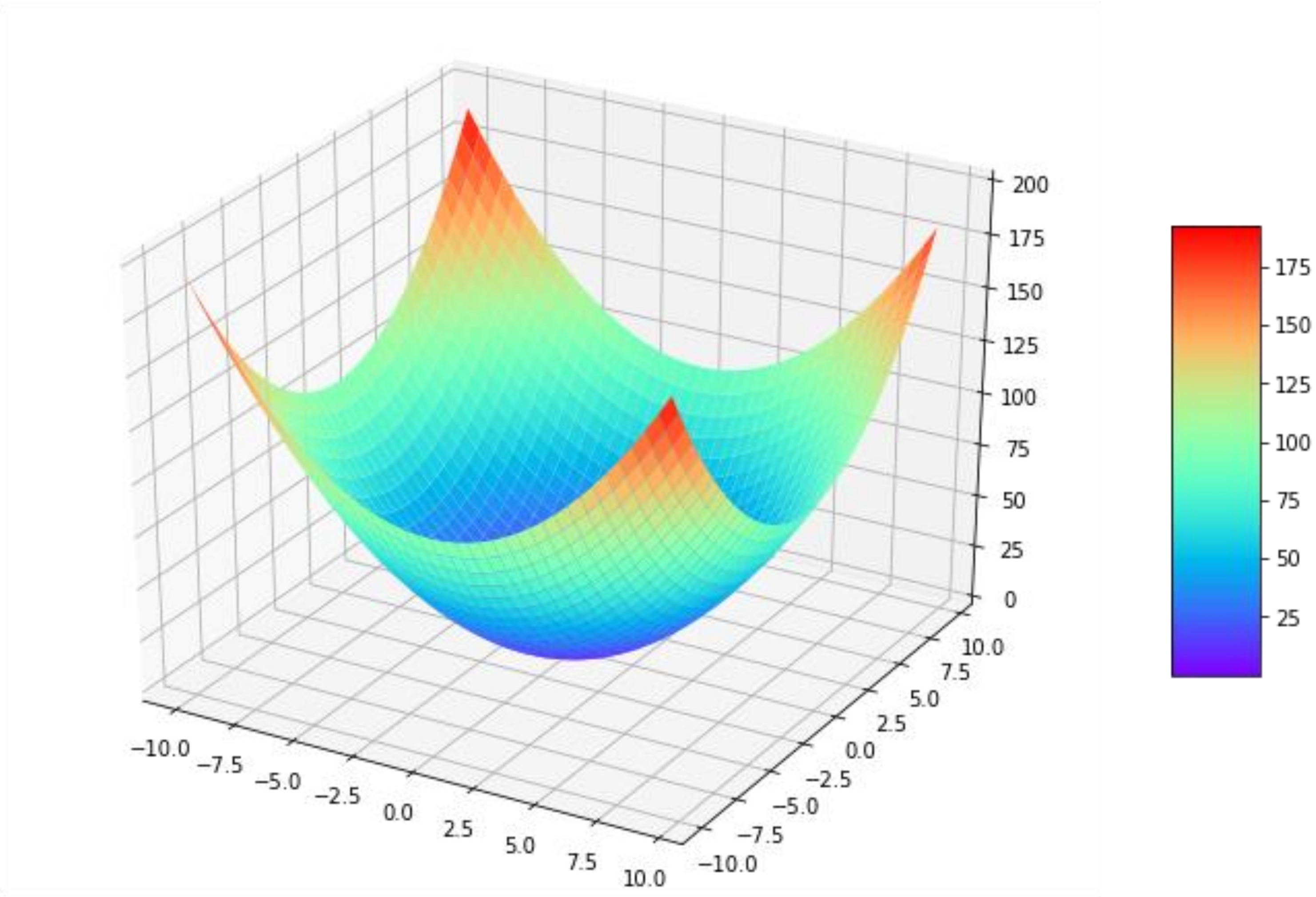
$$\underset{\theta_0, \theta_1}{\text{minimize}} \;\; J(\theta_0, \theta_1)$$

مثال: قیمت‌گذاری خانه

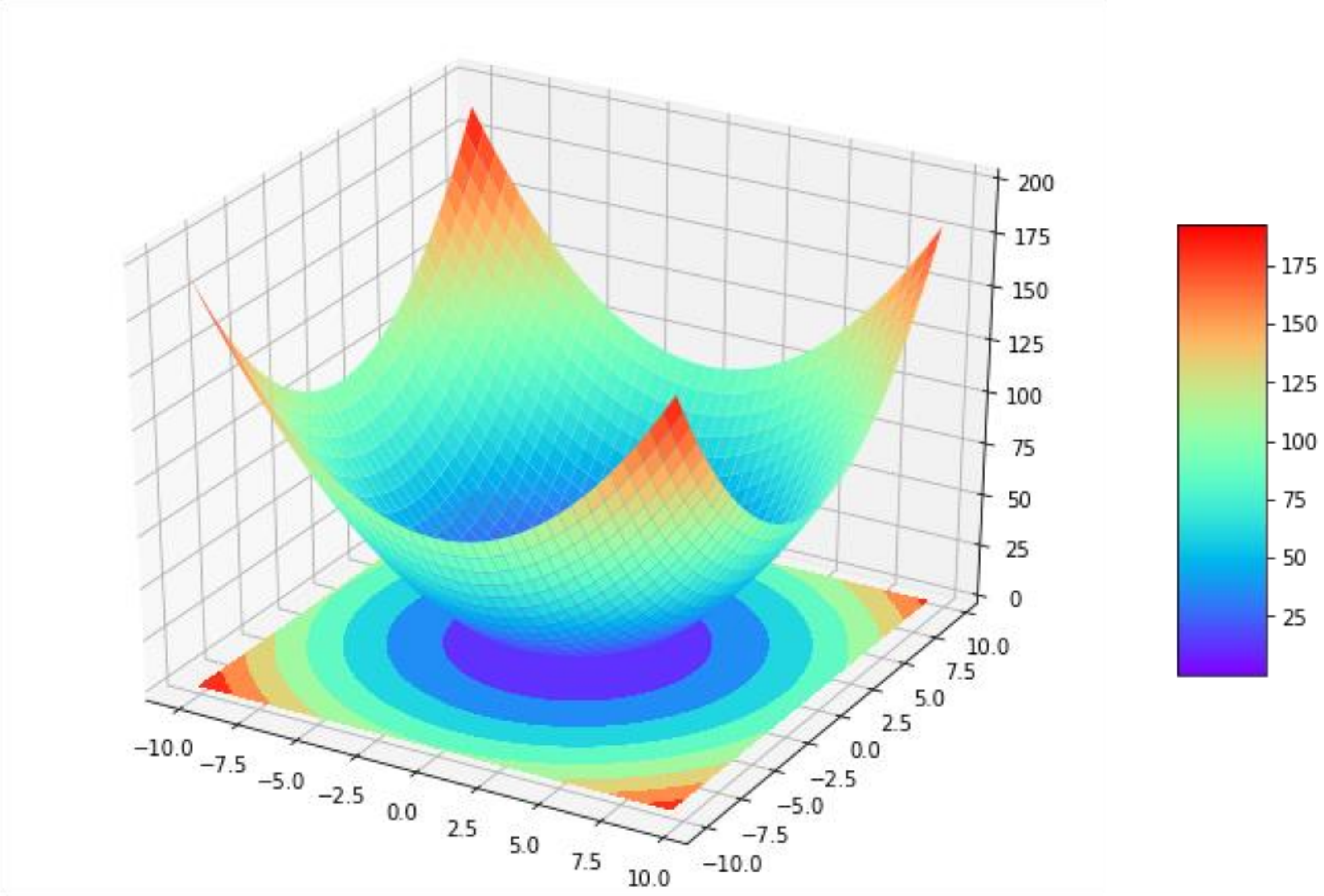
$$h_\theta(x) = \theta_0 + \theta_1 x$$



تابع هزینه


200
175
150
125
100
75
50
25
0
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
175
150
125
100
75
50
25

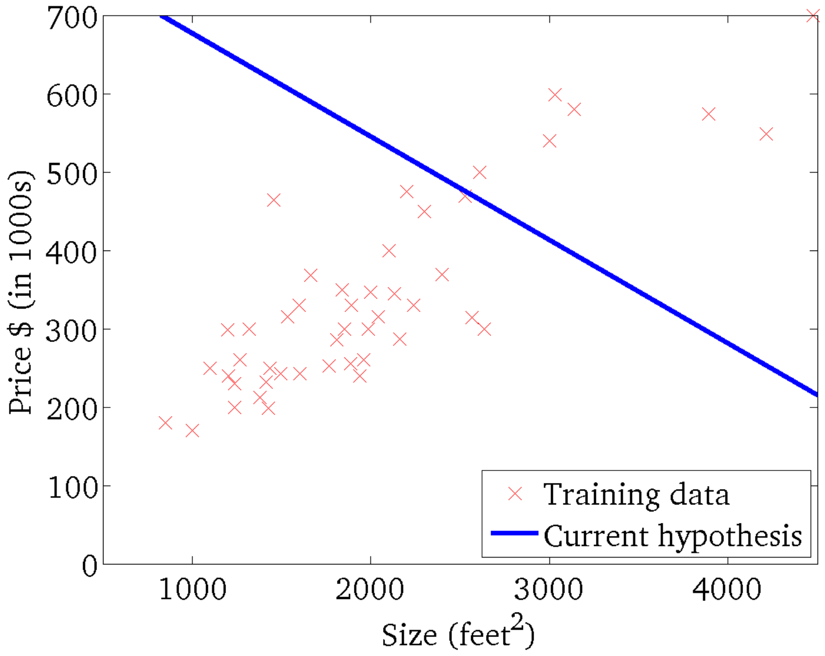
تابع هزینه

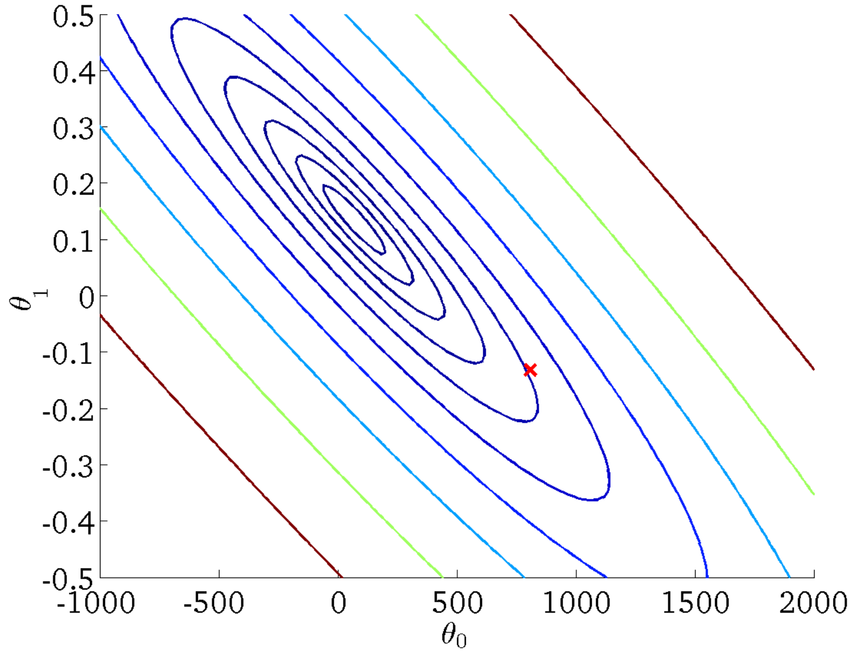

200
175
150
125
100
75
50
25
0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
175
150
125
100
75
50
25

تابع هزینه: منحنی کانتور

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

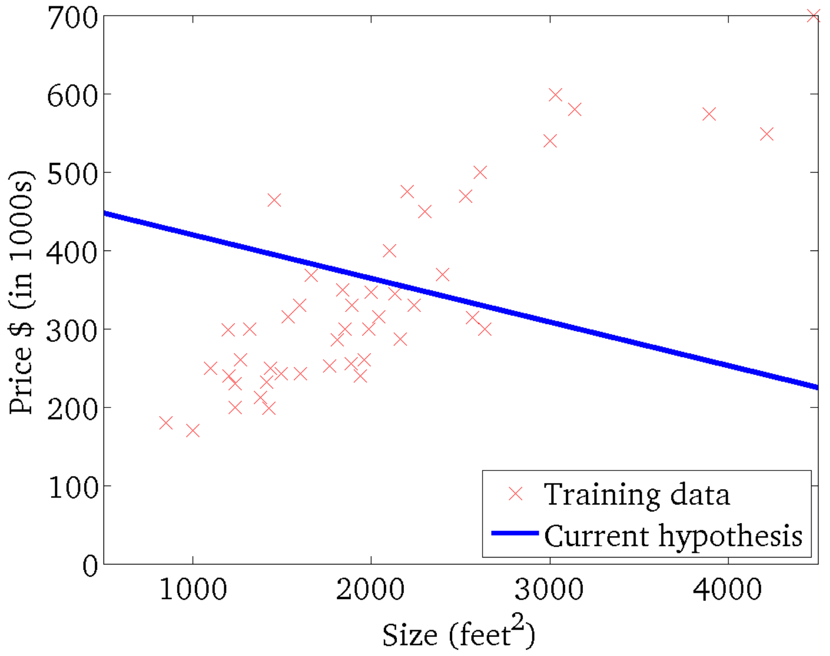
$J(\theta_0, \theta_1)$



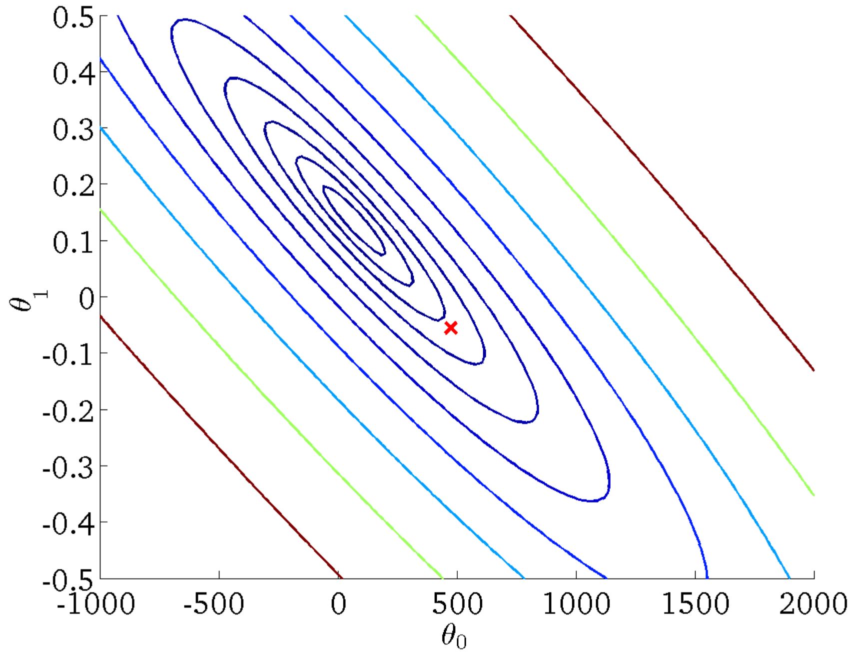


تابع هزینه: منحنی کانتور

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



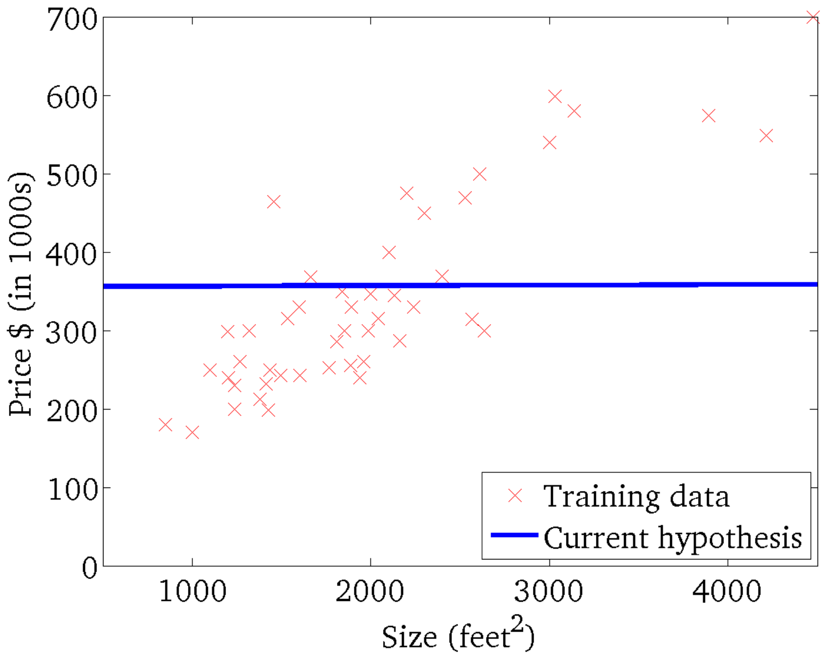
$J(\theta_0, \theta_1)$

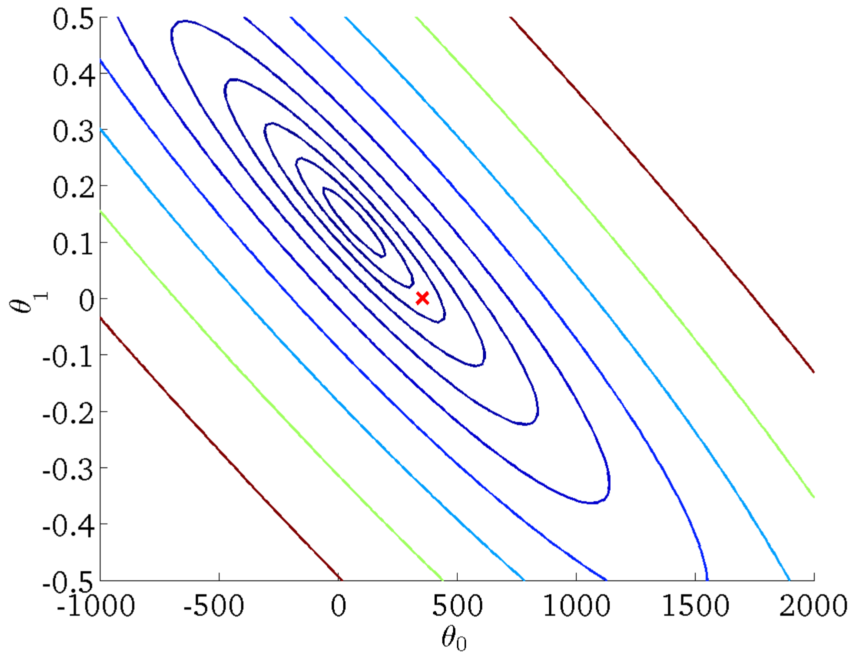


تابع هزینه: منحنی کانتور

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

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

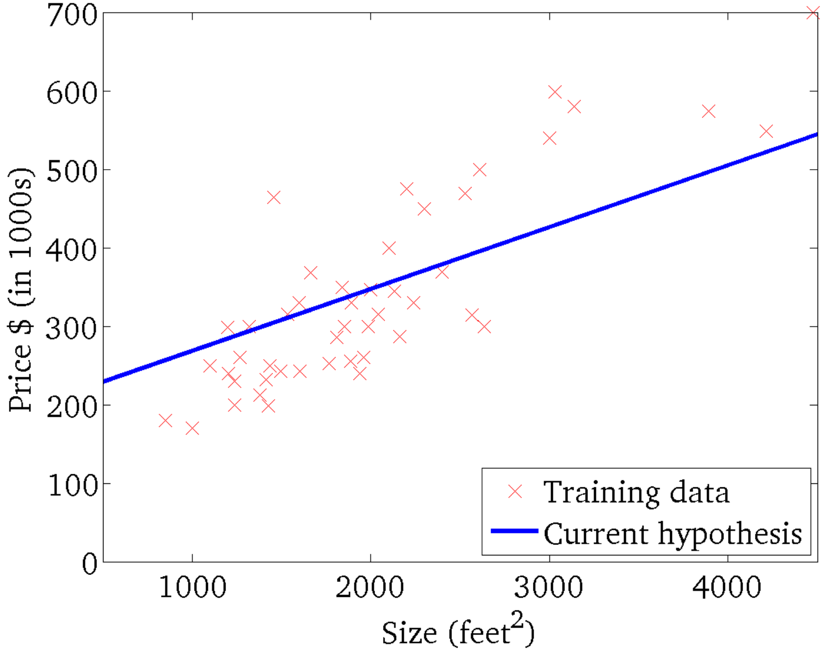


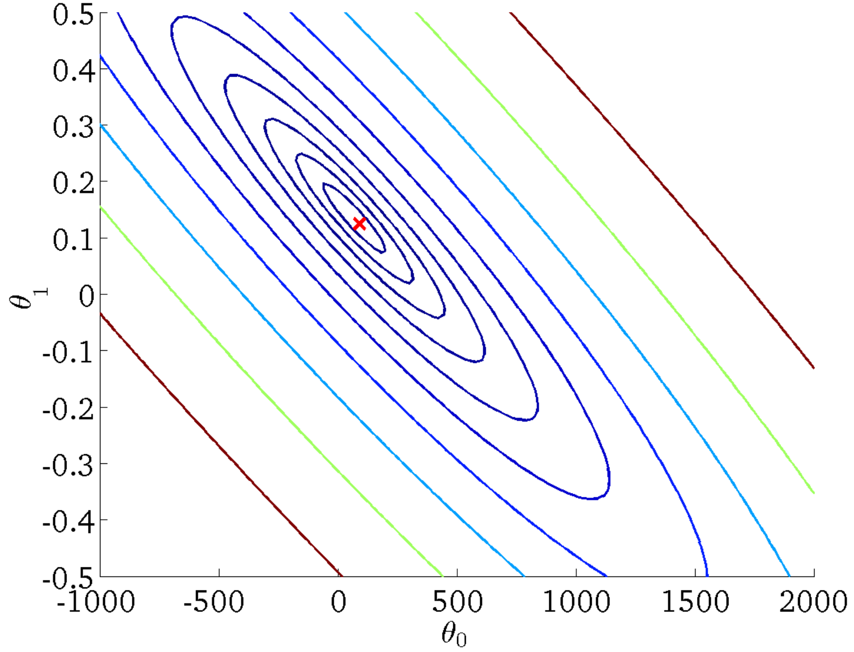


تابع هزینه: منحنی کانتور

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

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





گرادیان کاهشی

مسئله

- تابع هزینه.

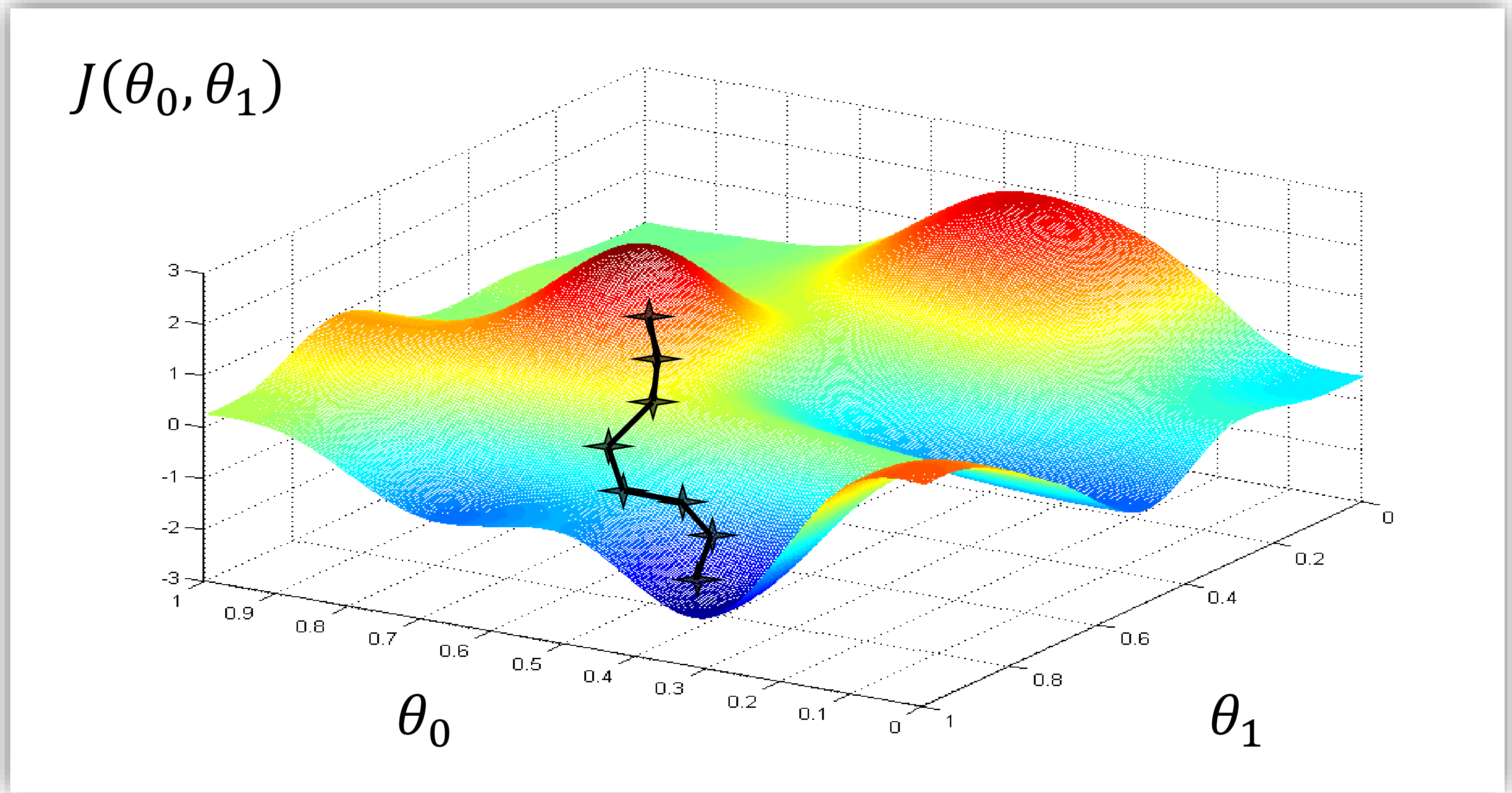
$$J(\theta_0, \theta_1)$$

- هدف.

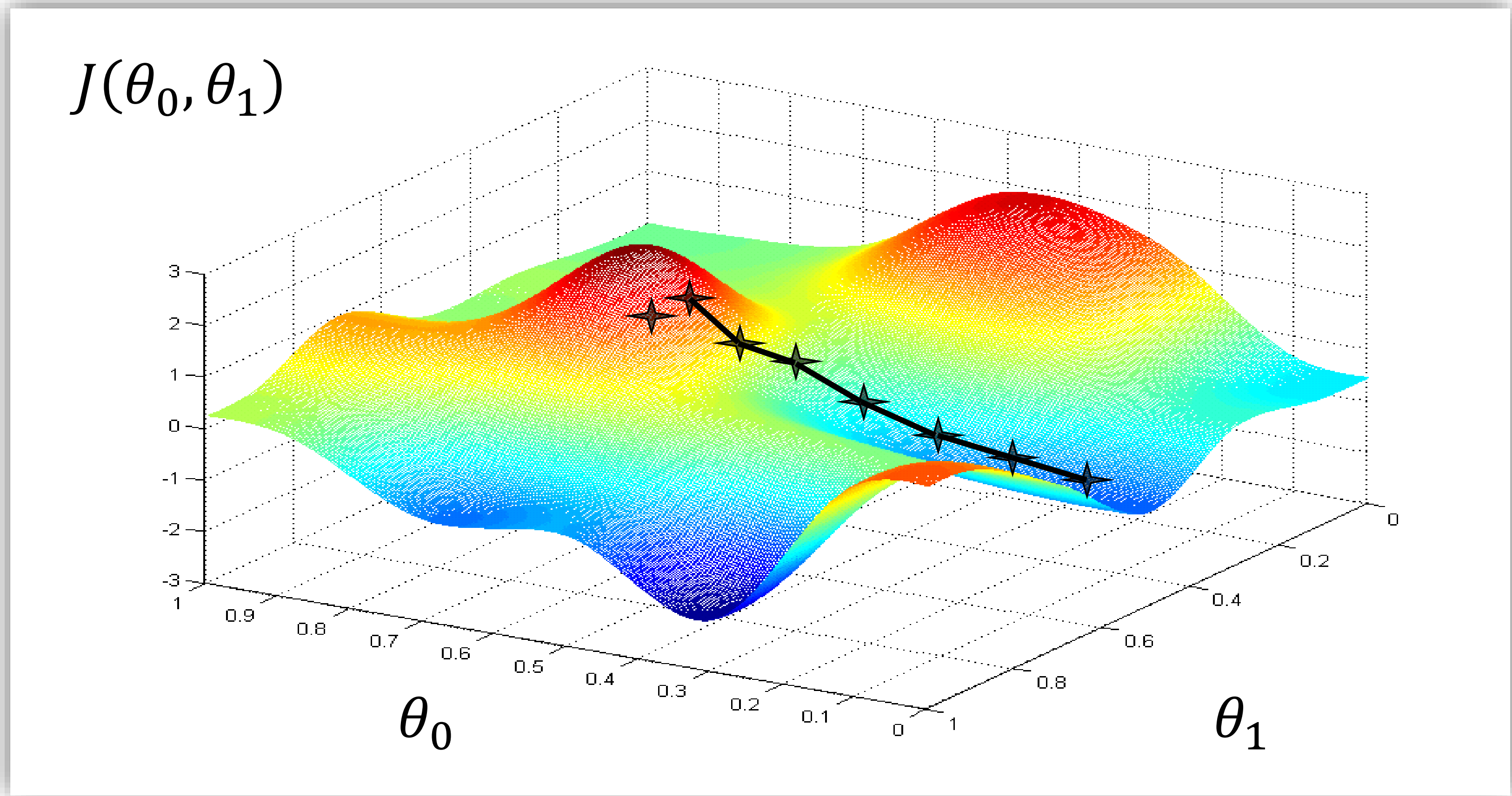
$$\underset{\theta_0,\theta_1}{\text{minimize}} \;\; J(\theta_0, \theta_1)$$

- کلیات روش.
  - با یک مقدار اولیه 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: بهینه سراسری



گرادیان کاهشی: بهینه محلی



الگوریتم گرادیان کاهشی

repeat until convergence {

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

}

نرخ یادگیری

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

- 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

repeat until convergence {

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

}

نرخ یادگیری

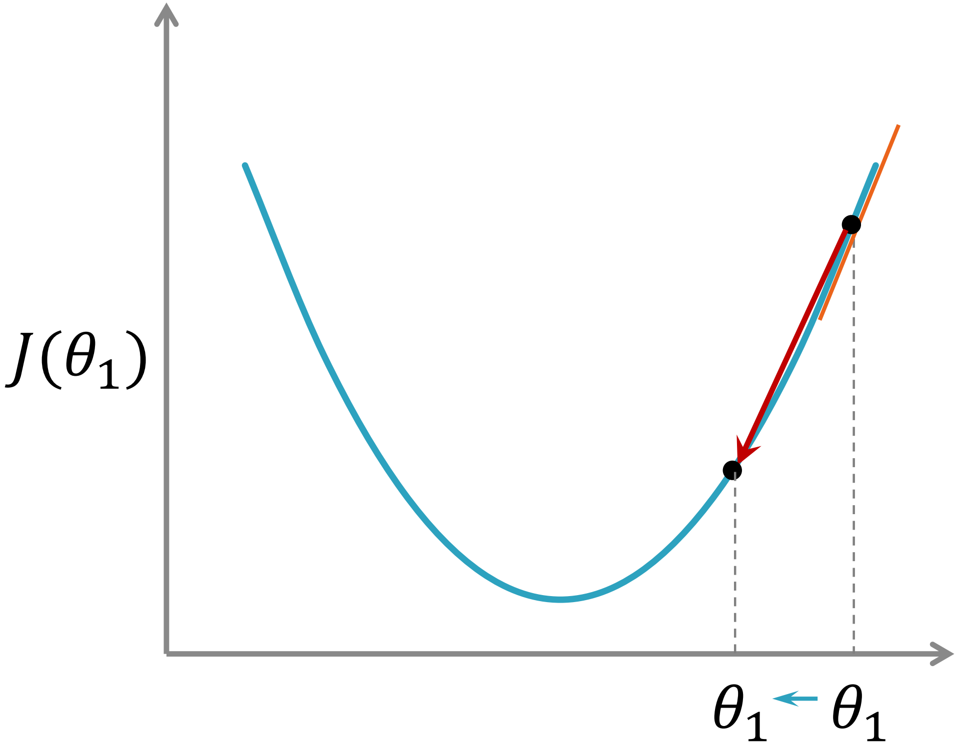
$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x}$$

- 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

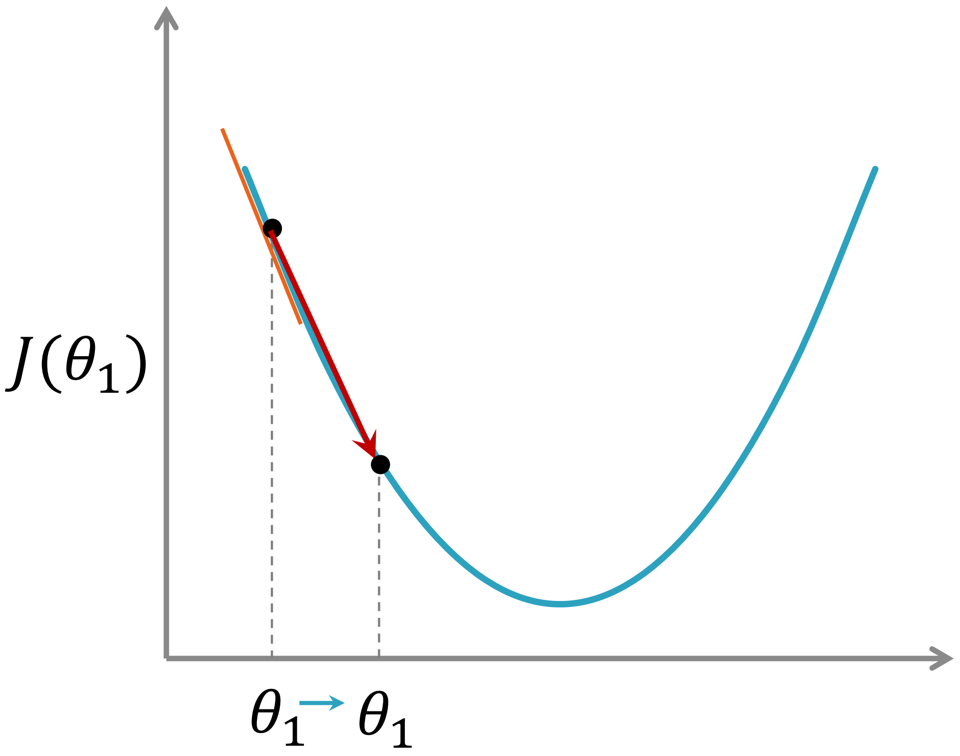


شیب مثبت

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

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

الگوریتم گرادیان کاهشی: گرادیان



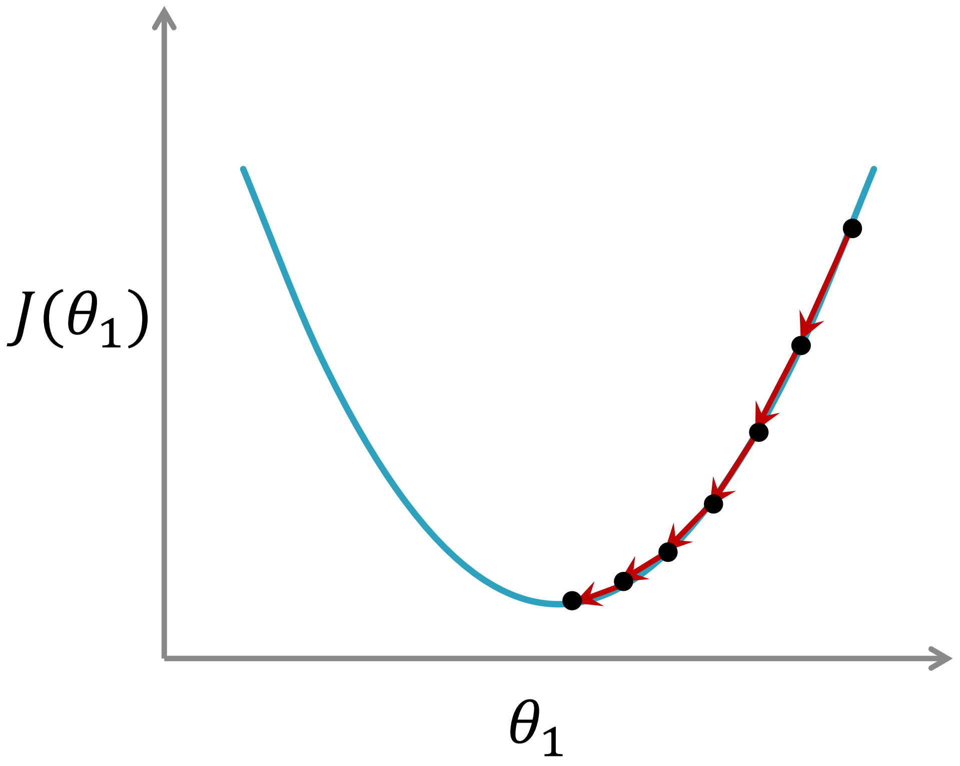
شیب منفی

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

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

الگوریتم گرادیان کاهشی: نرخ یادگیری

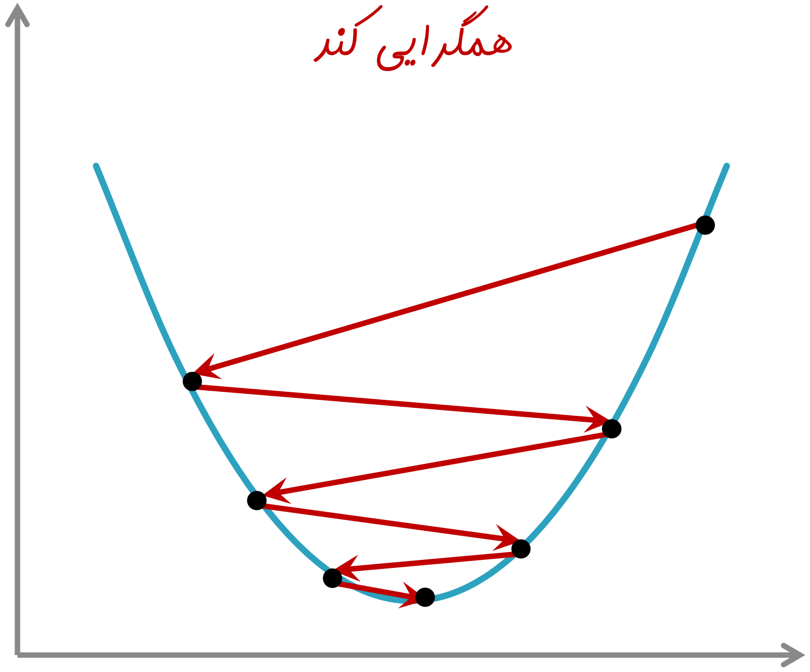
- اگر نرخ یادگیری بیش از حد کوچک باشد، گرادیان کاهشی به کندی همگرا خواهد شد.

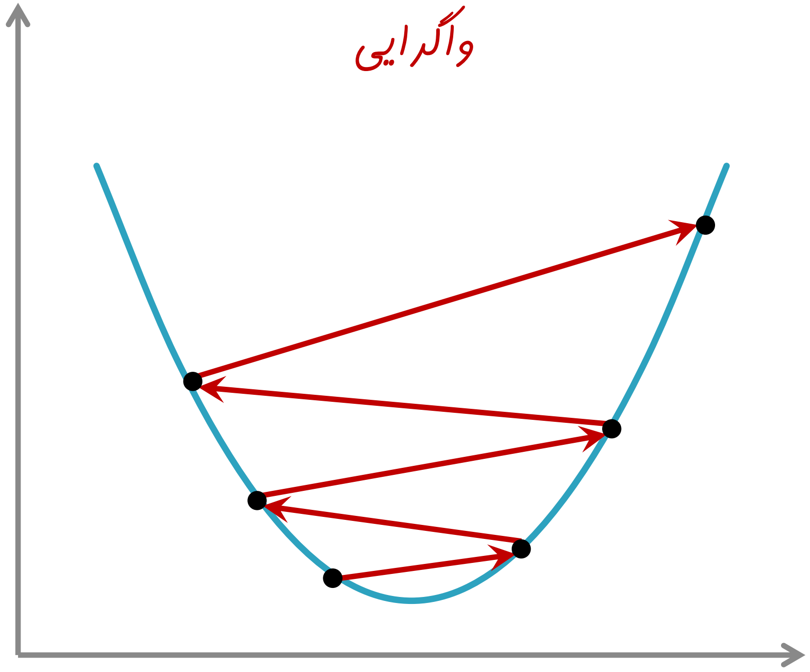


$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)$$

الگوریتم گرادیان کاهشی: نرخ یادگیری

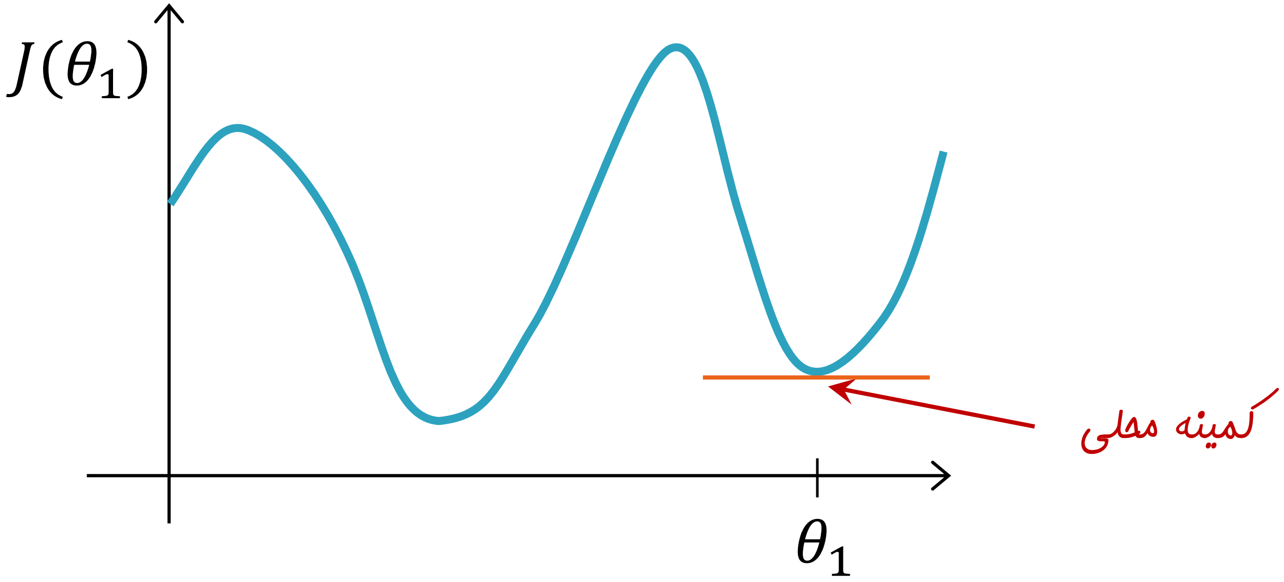
- اگر نرخ یادگیری بیش از حد بزرگ باشد، گرادیان کاهشی ممکن است به کندی همگرا شود و یا حتی واگرا شود.





الگوریتم گرادیان کاهش: همگرایی

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



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

کاربرد گرادیان کاهشی در رگرسیون خطی

گرادیان کاهشی و رگرسیون خطی

repeat until convergence {

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

}

- رگرسیون خطی.

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

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

گرادیان تابع هزینه

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

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

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

گرادیان کاهشی و رگرسیون خطی

repeat until convergence {

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

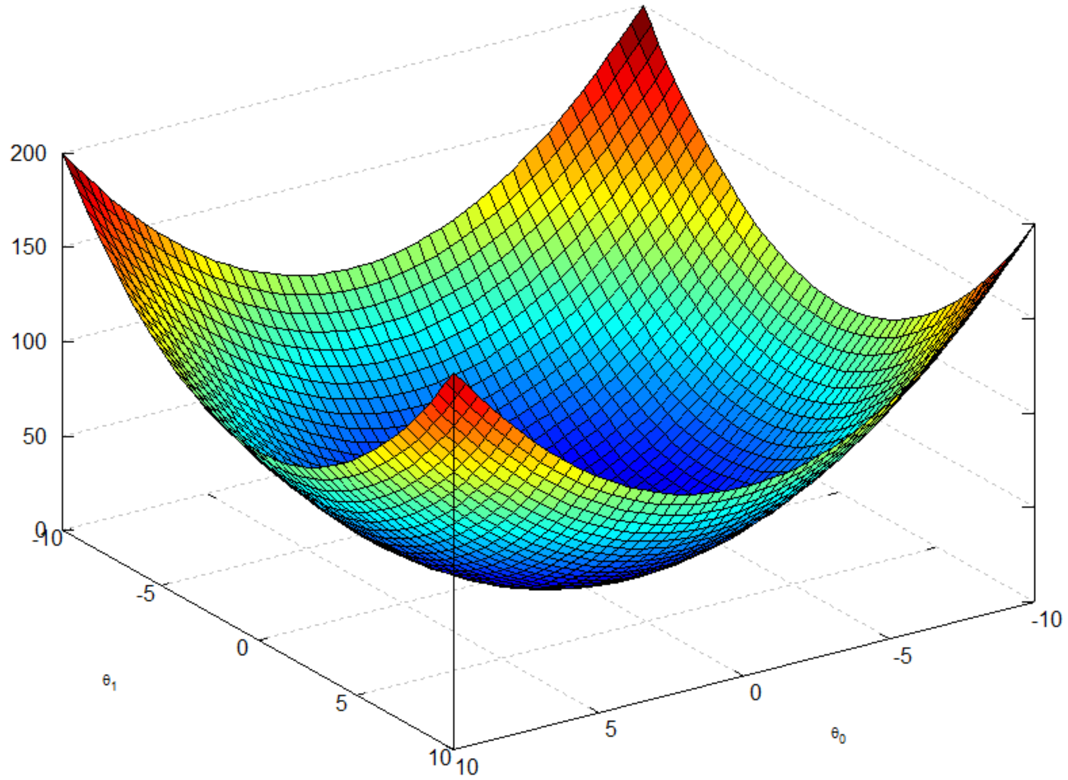
$$\theta_1 := \theta_1 - \alpha \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) \cdot x^{(i)}$$

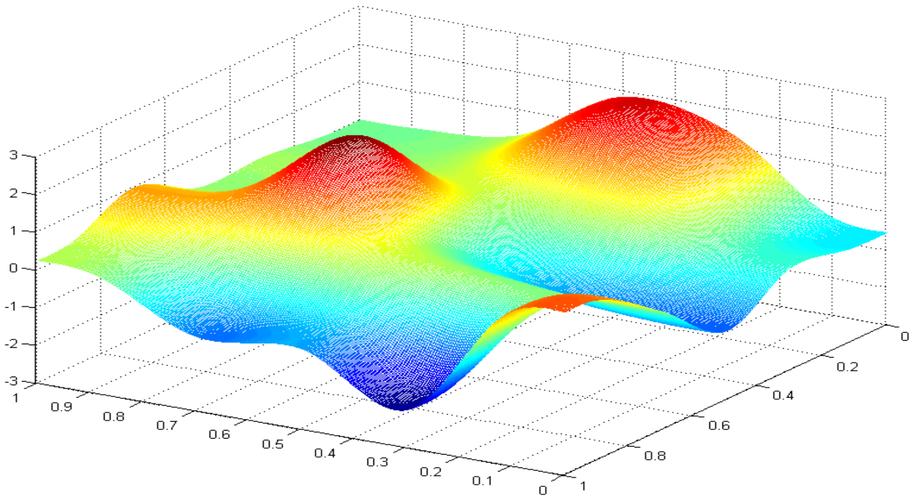
به روزرسانی همزمان

}

گرادیان کاهشی و رگرسیون خطی

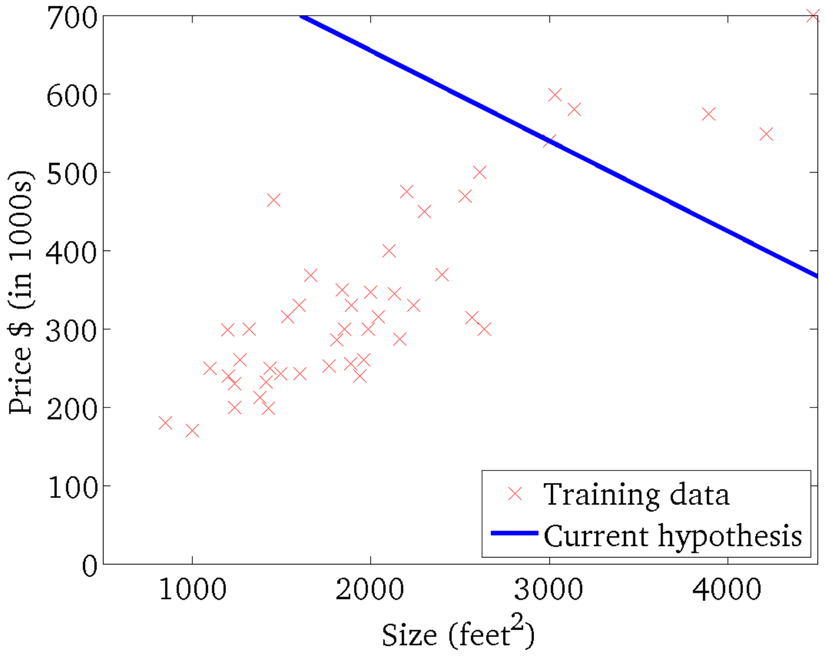
- توجه. در رگرسیون خطی تابع هزینه یک تابع کوژ است و در نتیجه گرادیان کاهشی در صورت همگرایی لزوماً در بهینه‌ی سراسری همگرا می‌شود.



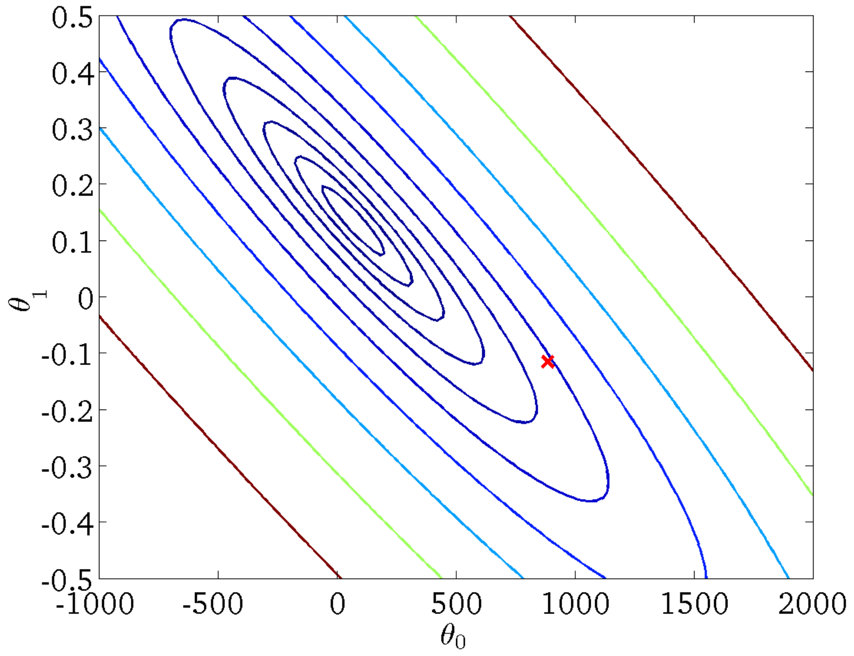


گرادیان کاهشی

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



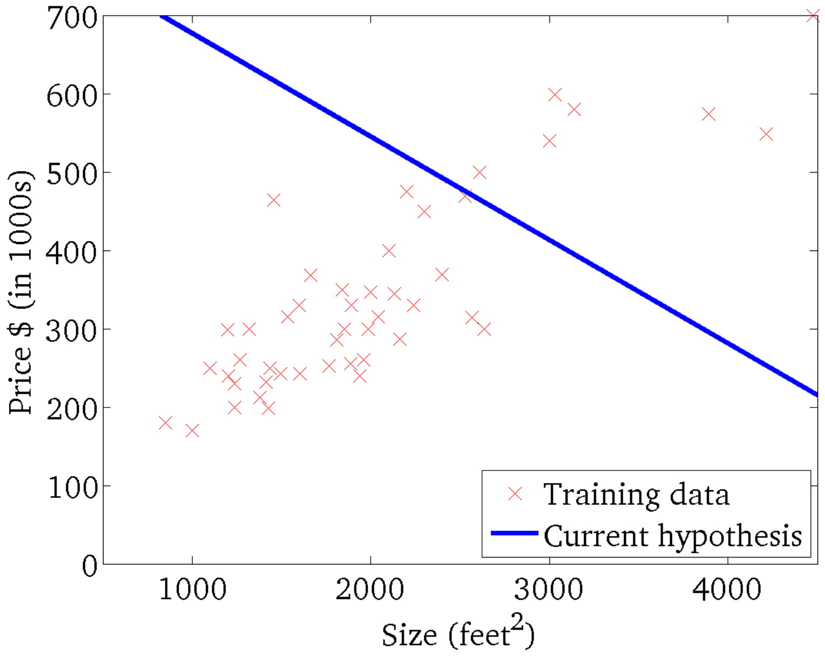
$J(\theta_0, \theta_1)$

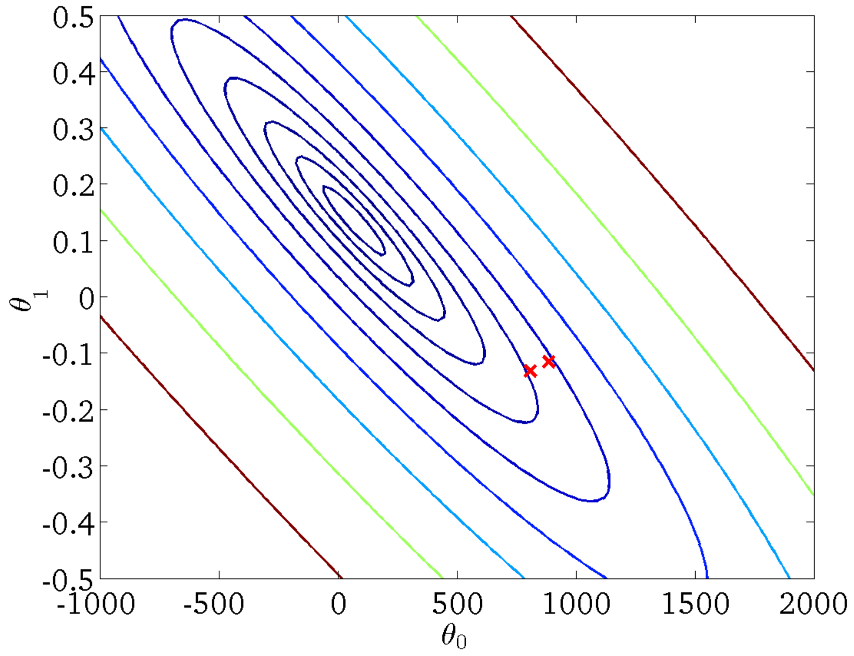


گرادیان کاهشی

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

$J(\theta_0, \theta_1)$

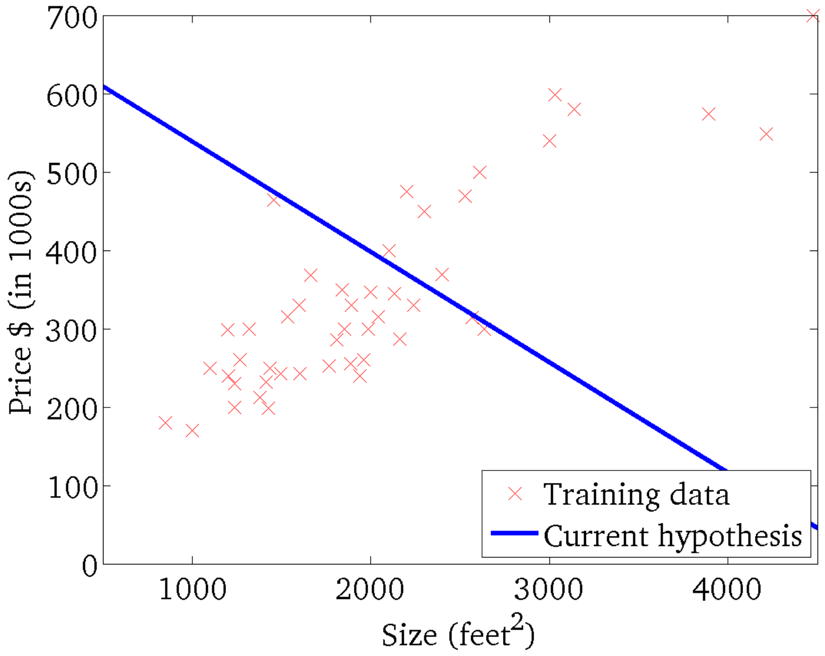


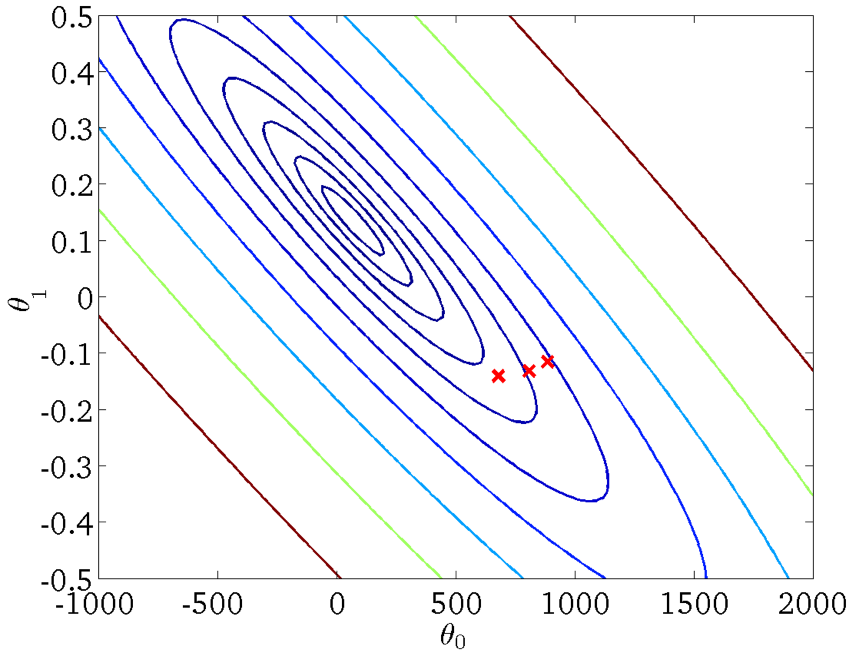


گرادیان کاهشی

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

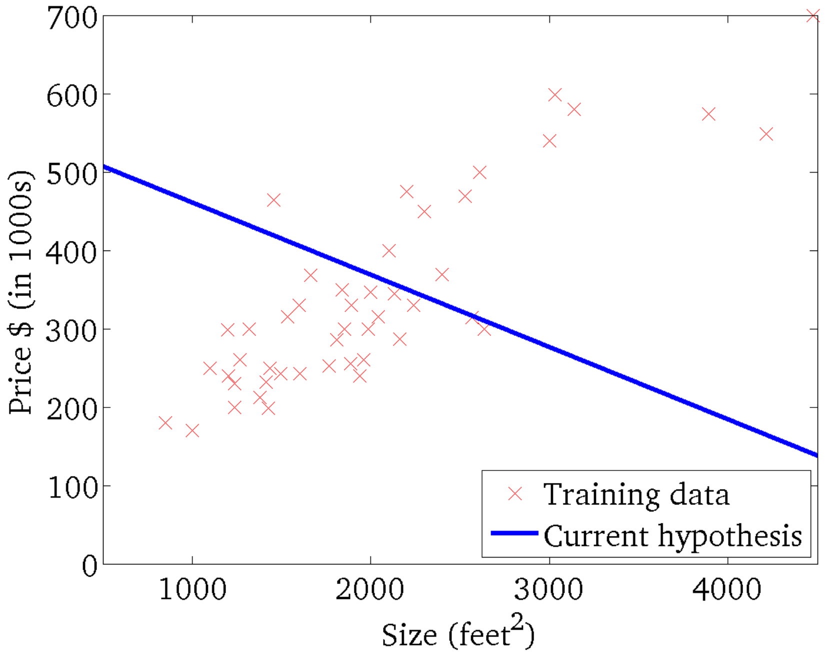
$J(\theta_0, \theta_1)$



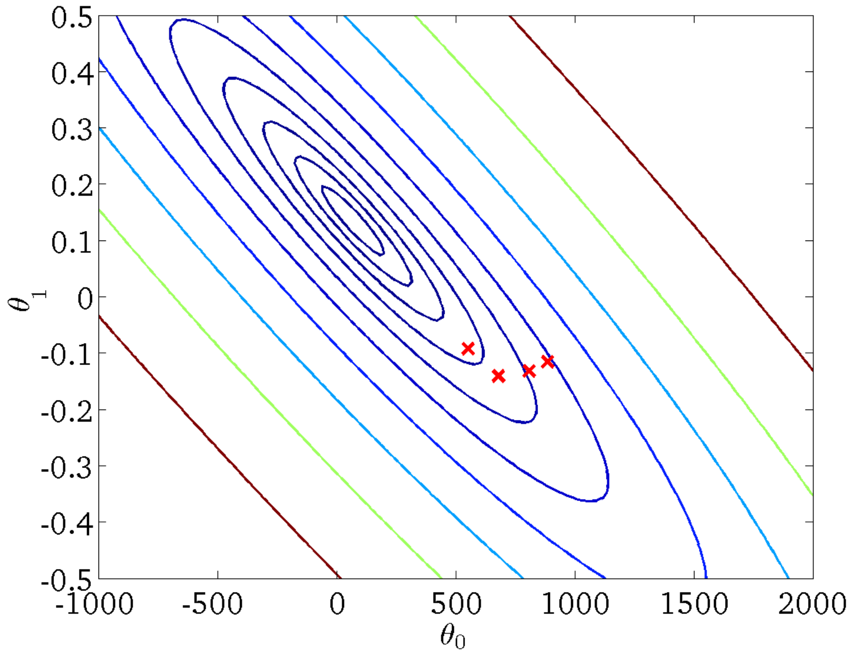


گرادیان کاهشی

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

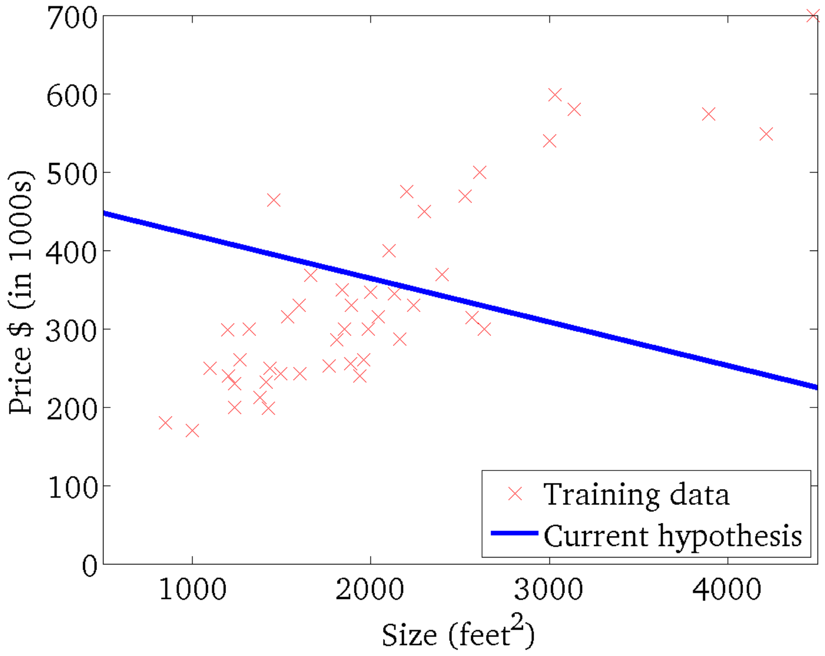


$$J(\theta_0, \theta_1)$$

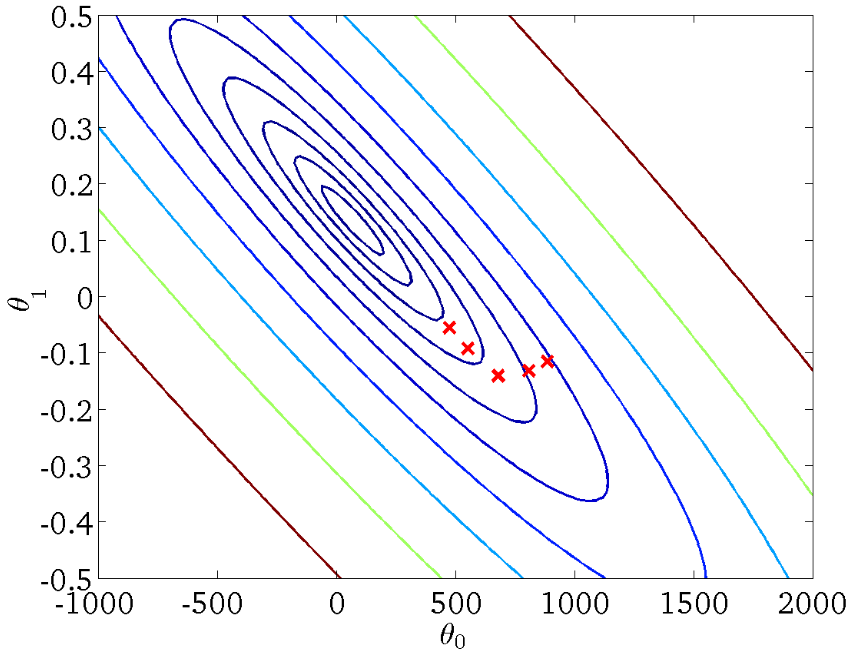


گرادیان کاهشی

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



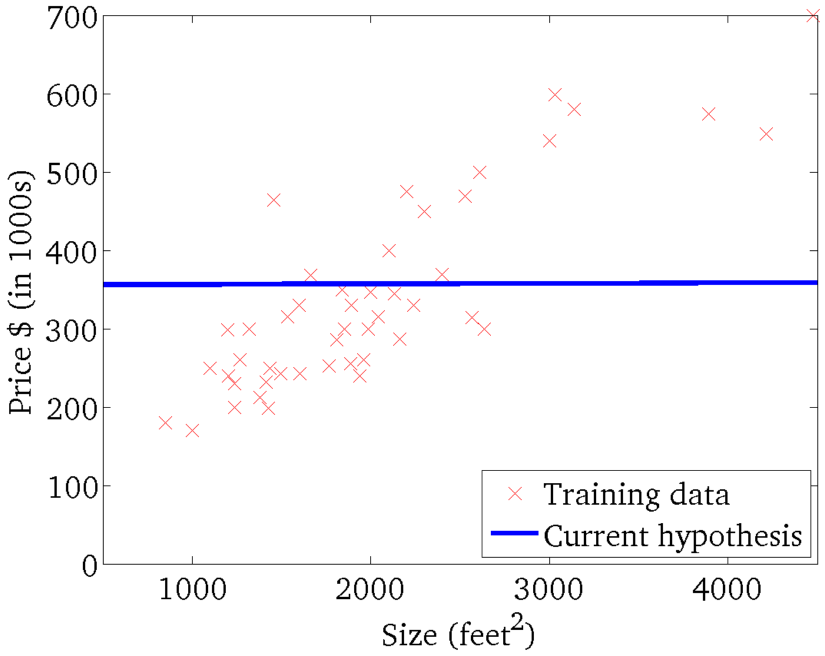
$J(\theta_0, \theta_1)$

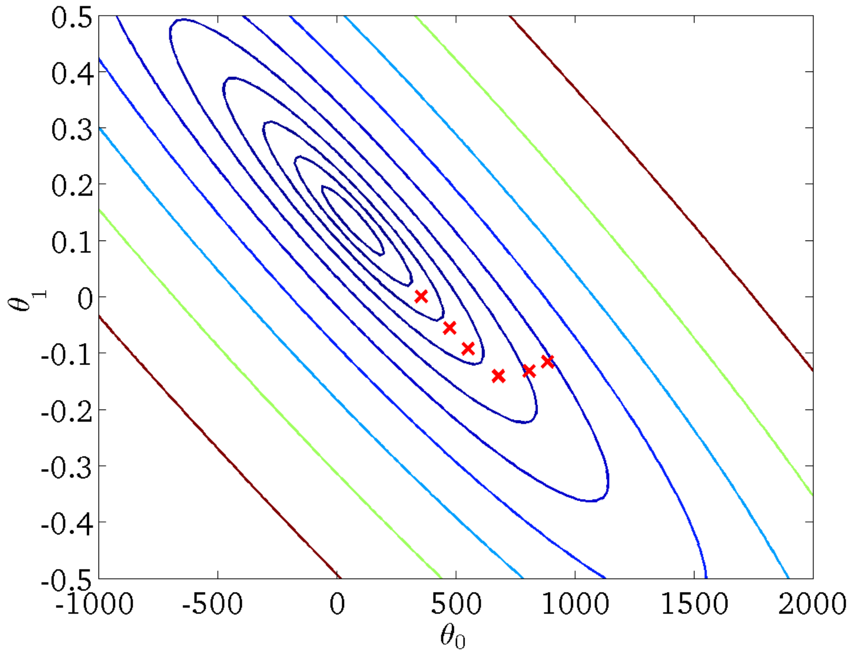


گرادیان کاهشی

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

$J(\theta_0, \theta_1)$

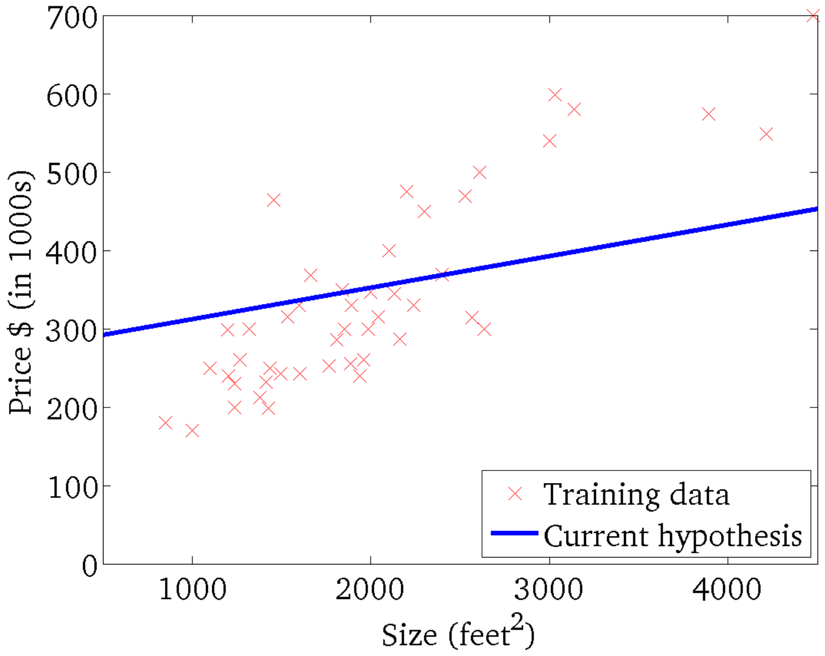


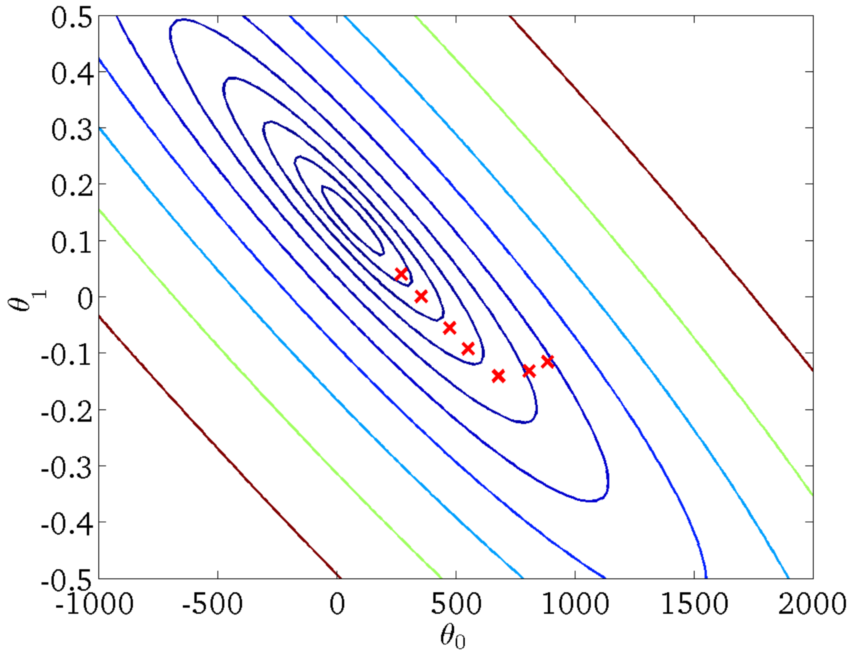


گرادیان کاهشی

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

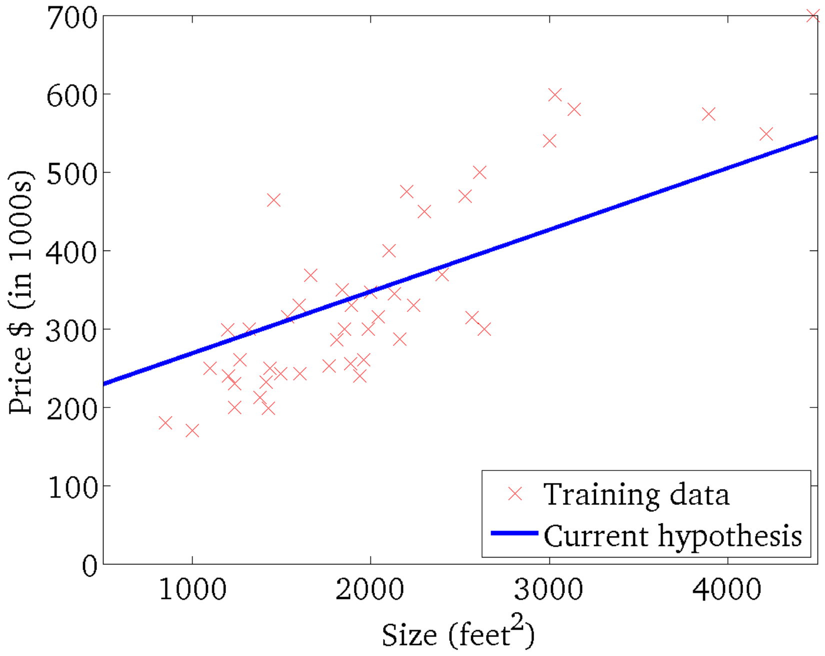
$J(\theta_0, \theta_1)$



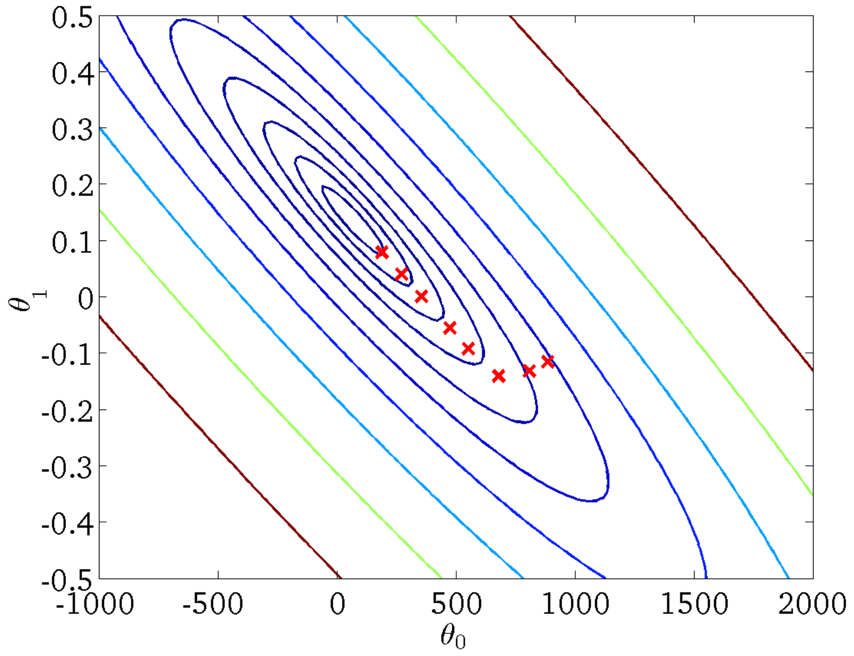


گرادیان کاهشی

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



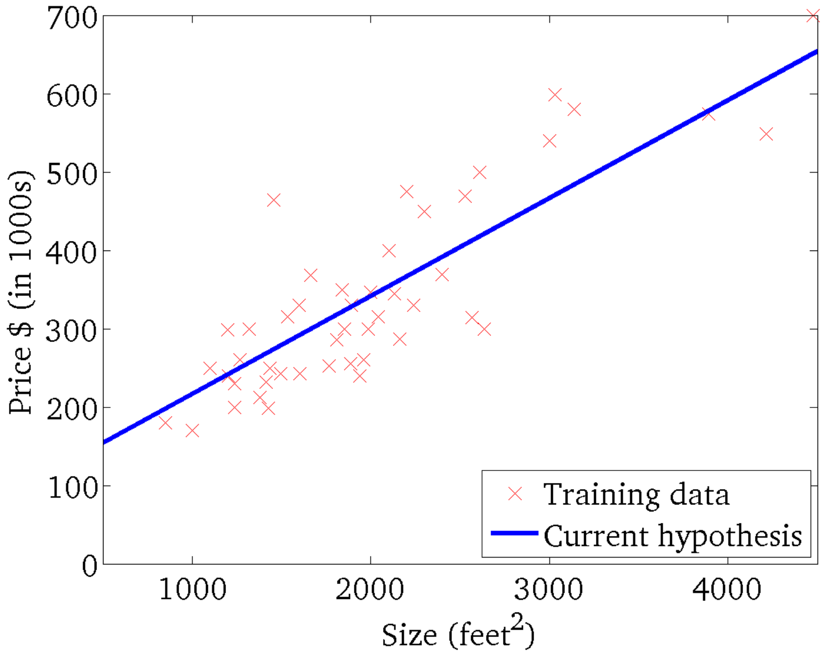
$J(\theta_0, \theta_1)$

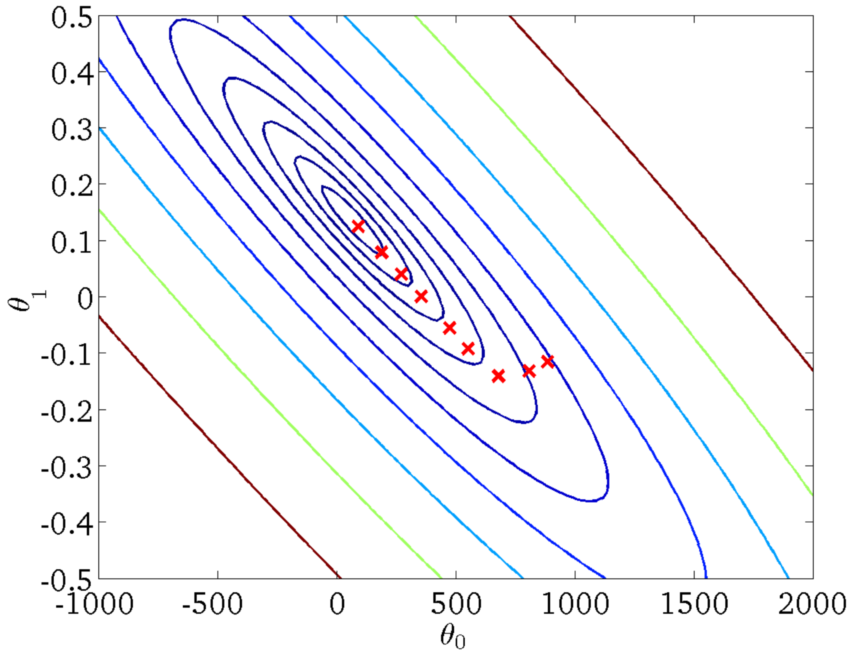


گرادیان کاهشی

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

$J(\theta_0, \theta_1)$





گرادیان کاهشی دسته‌ای

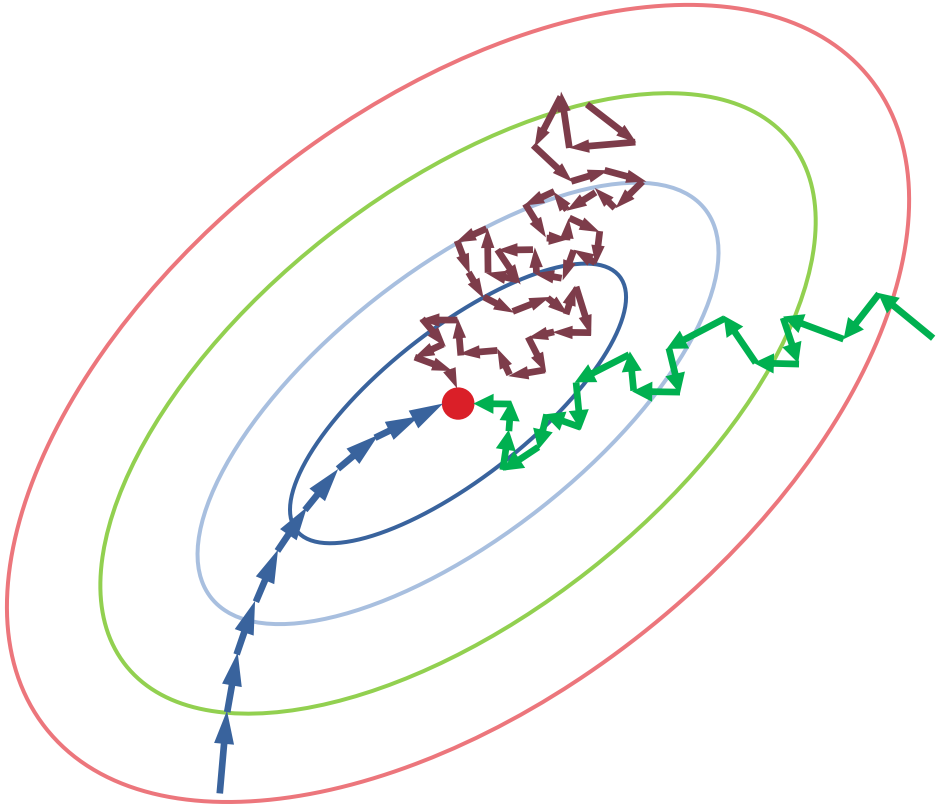
- گرادیان کاهشی دسته‌ای. در هر تکرار الگوریتم، از تمام نمونه‌های آموزشی برای به روز رسانی مقدار پارامترها استفاده می‌شود.

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

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

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

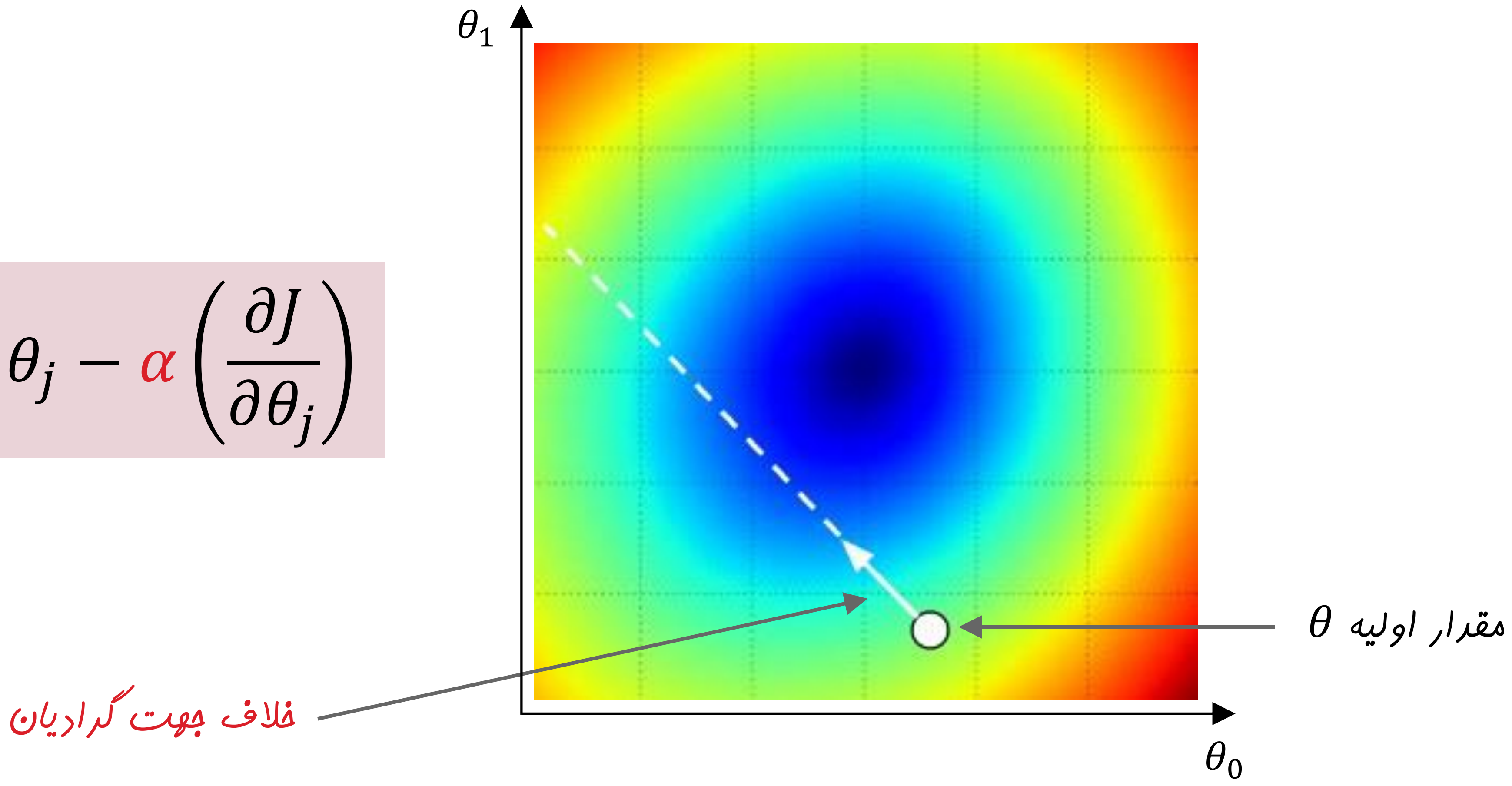
انواع گرادیان کاهشی



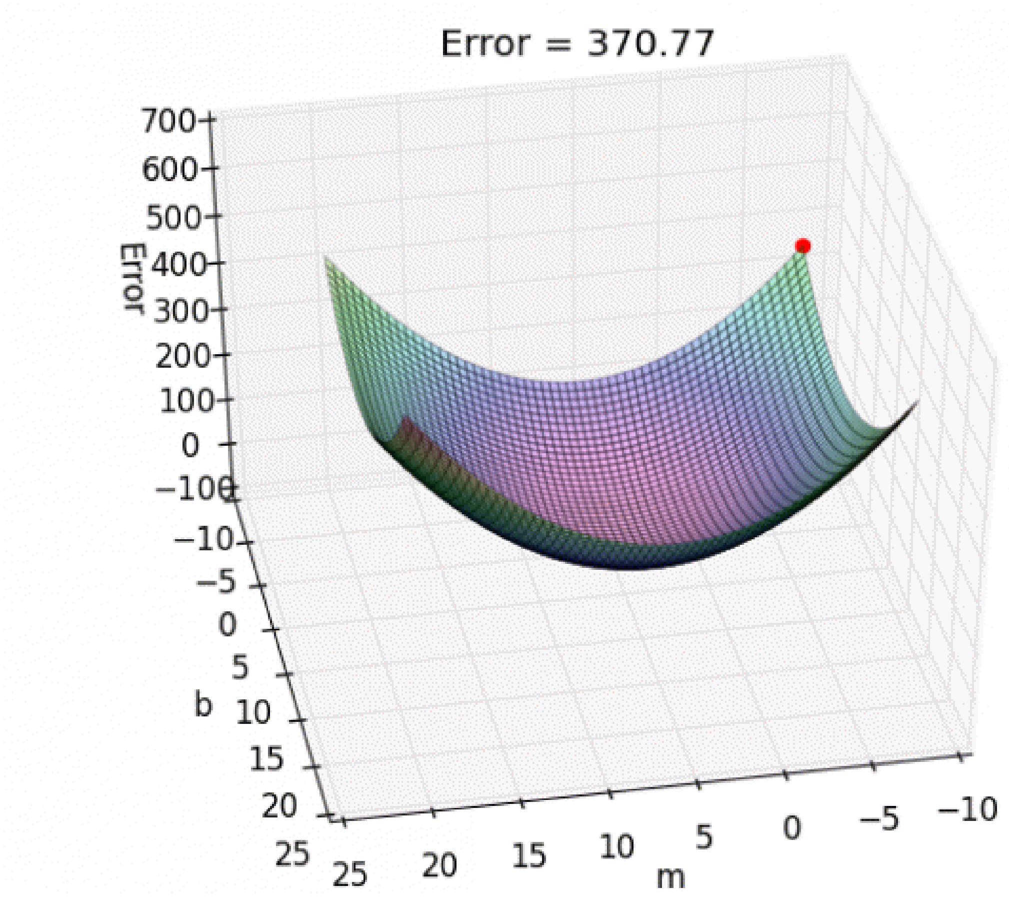
- گرادیان کاهشی با دسته‌های کامل
  به روز رسانی در هر تکرار با استفاده از تمام نمونه‌ها
- گرادیان کاهشی اتفاقی (آنلاین)
  به روز رسانی در هر تکرار با استفاده از یک نمونه تصادفی
- گرادیان کاهشی با دسته‌های کوچک
  به روز رسانی در هر تکرار با استفاده از یک دسته کوچک تصادفی از نمونه‌ها

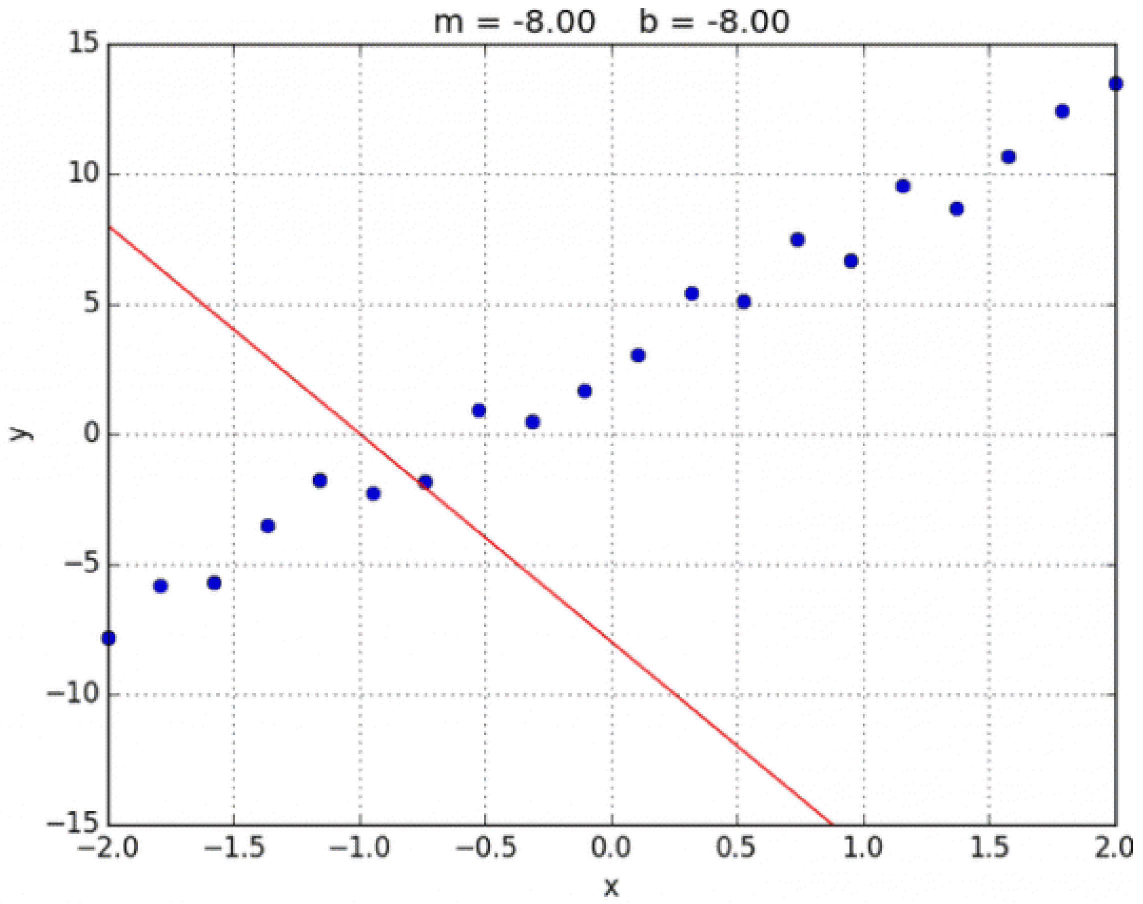
یادآوری: الگوریتم گرادیان کاهشی

$$\theta_j = \theta_j - \alpha\left(\frac{\partial J}{\partial \theta_j}\right)$$



یادآوری: الگوریتم گرادیان کاهشی





رگرسیون خطی چند متغیره

رگرسیون تک متغیره (یک ویژگی)

- رگرسیون خطی با یک ویژگی.

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

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

رگرسیون چند متغیره (چند ویژگی)

- رگرسیون خطی با چند ویژگی.

متراژ (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
…	…	…	…	…

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

رگرسیون چند متغیره (چند ویژگی)

متراژ (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
...	...	...	...	...

m تعداد نمونه‌های آموزشی

n

- نمادها.
  - n تعداد ویژگی‌ها
  - $x^{(i)}$ ورودی‌ها در i امین نمونه آموزشی
  - $x_j^{(i)}$ مقدار ویژگی j ام در i امین نمونه آموزشی

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix} \begin{array}{l} \leftarrow x_1^{(2)} \\ \\ \leftarrow x_3^{(2)} \\ \\ \end{array}$$

فرضیه

- رگرسیون خطی تک متغیره.

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

- رگرسیون خطی چند متغیره.

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

- برای سادگی، تعریف می‌کنیم $x_0 = 1$

$$x = \begin{bmatrix} x_0 = 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \Rightarrow \boxed{h_\theta(x) = \theta^T x}$$

گرادیان کاهشی در رگرسیون خطی چند متغیره

گرادیان کاهشی

- فرضیه.

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

- پارامترها.

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

- تابع هزینه.

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

repeat until convergence {

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

}

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

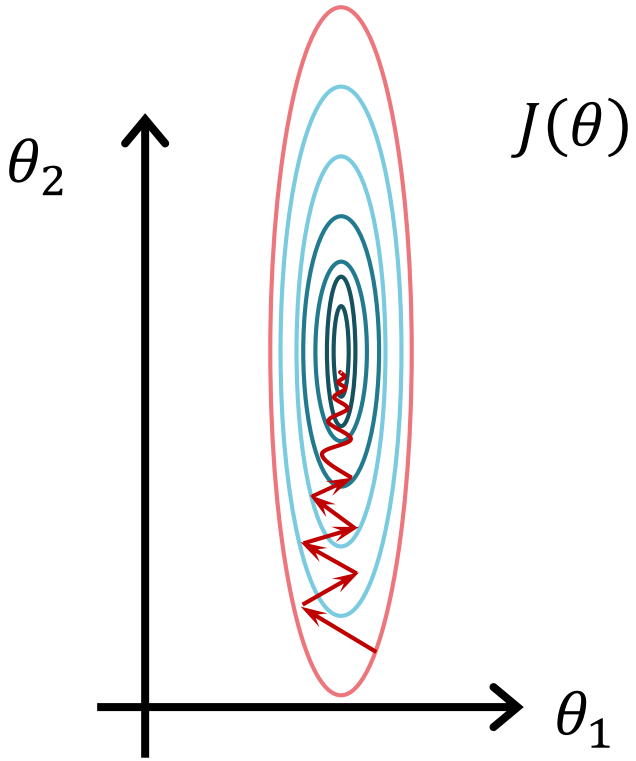
ترفندهای عملی در رگرسیون چند متغیره

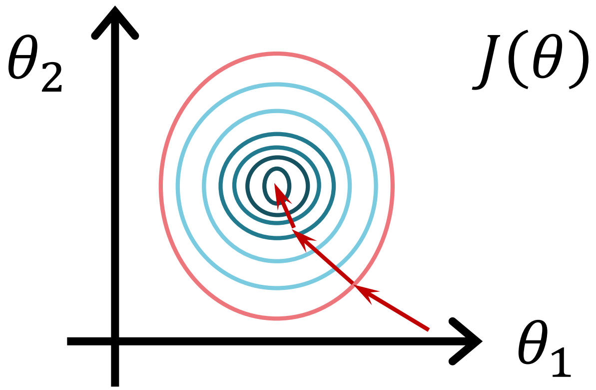
مقیاس‌بندی ویژگی‌ها

تعیین نرخ یادگیری

مقیاس‌بندی ویژگی‌ها (نرمال‌سازی)

- 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 خواب ها (۱ تا ۵)





مقیاس‌بندی ویژگی‌ها

- مقیاس‌بندی. مقدار هر ویژگی در ضریب کوچکی از بازه ۱- تا ۱+ قرار دارد.

- مثال.

$$x_1 = \frac{size - 1000}{2000} \qquad -0.5 \le x_1 \le 0.5$$

$$x_2 = \frac{\#\, bedrooms - 2}{5} \qquad -0.5 \le x_2 \le 0.5$$

- نرمال‌سازی میانگین.

$$x_j = \frac{x_j - \mu_j}{\sigma_j}$$

میانگین

انحراف معیار

گرادیان کاهشی

- گرادیان کاهشی.

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

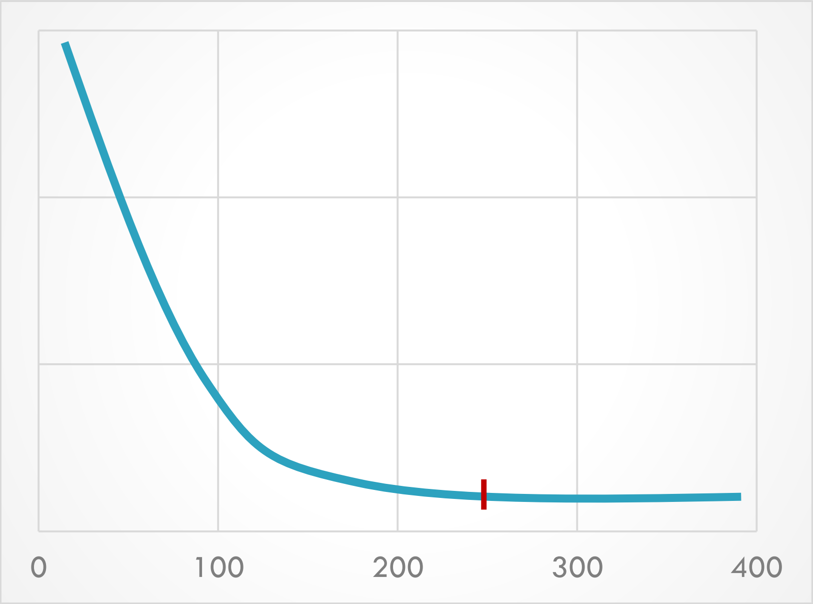
- س. چگونه می‌توان مطمئن شد گرادیان کاهشی به درستی عمل می‌کند؟

- س. مقدار مناسب برای نرخ یادگیری چیست؟

عملکرد صحیح برای گرادیان کاهشی

- آزمایش همگرایی.
  - اگر مقدار $J(\theta)$ در یک تکرار به اندازه‌ای کمتر از $۱۰^{-۳}$ تغییر کند، همگرایی رخ داده است.

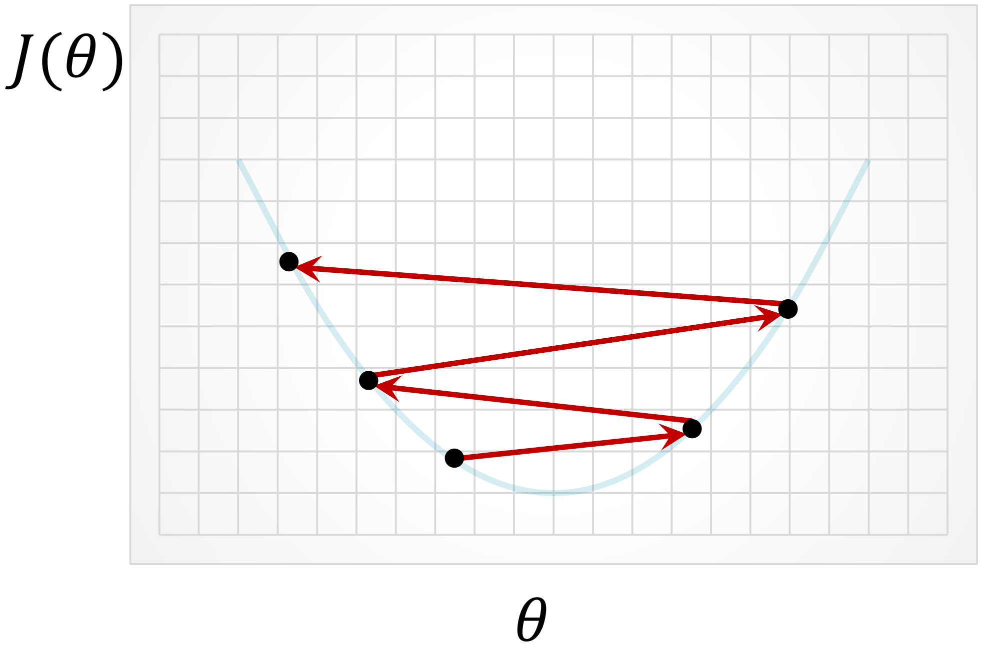
$J(\theta)$

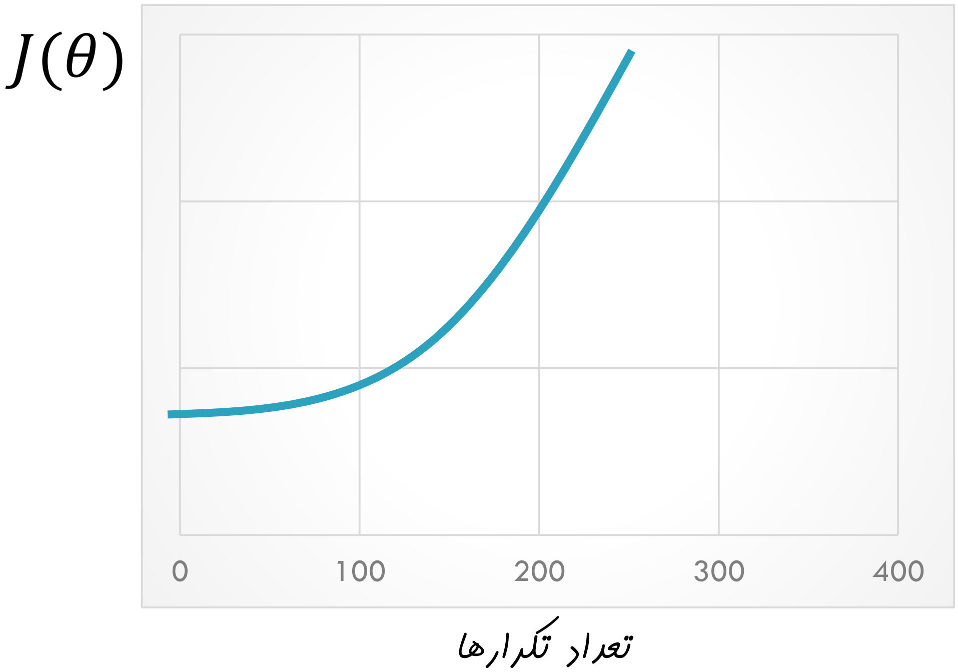


تعداد تکرارها

نشانه‌های عملکرد نادرست گرادیان کاهشی

- عدم همگرایی.
- راه‌حل. از مقادیر کوچک‌تری برای نرخ یادگیری استفاده کن، اما اگر نرخ یادگیری بیش از حد کوچک باشد همگرایی بسیار کند خواهد بود.





خلاصه

- 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:

$$\ldots,\quad 0.001,\quad 0.003,\quad 0.01,\quad 0.03,\quad .01,\quad 0.3,\quad 1.0,\quad \ldots$$

رگرسیون چندجمله‌ای

قیمت‌گذاری یک خانه: انتخاب ویژگی‌ها

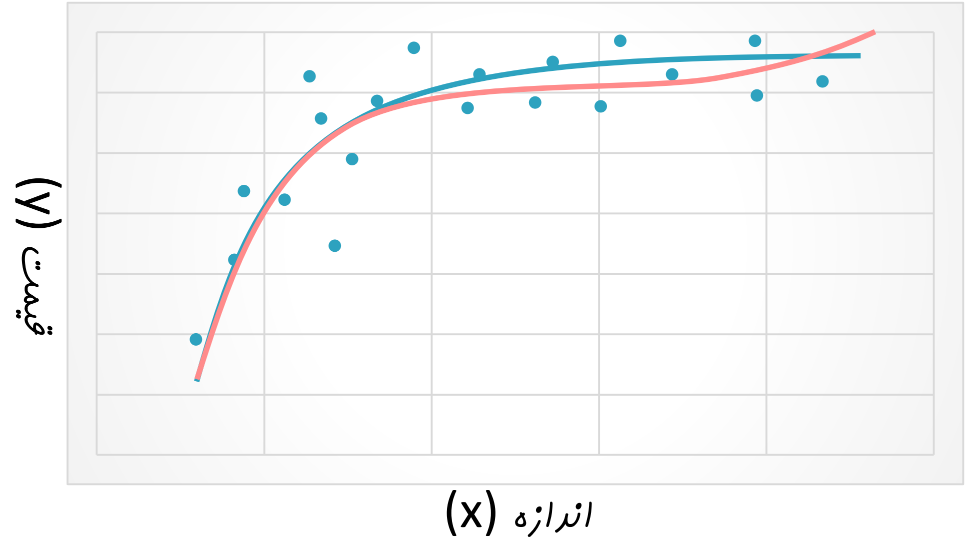
$$h_\theta(x) = \theta_0 + \theta_1 \times \underbrace{(\text{frontage})}_{x_1} + \theta_2 \times \underbrace{(\text{depth})}_{x_2}$$

$$Area = frontage \times depth$$

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

رگرسیون چند جمله‌ای

- رگرسیون چند جمله‌ای.



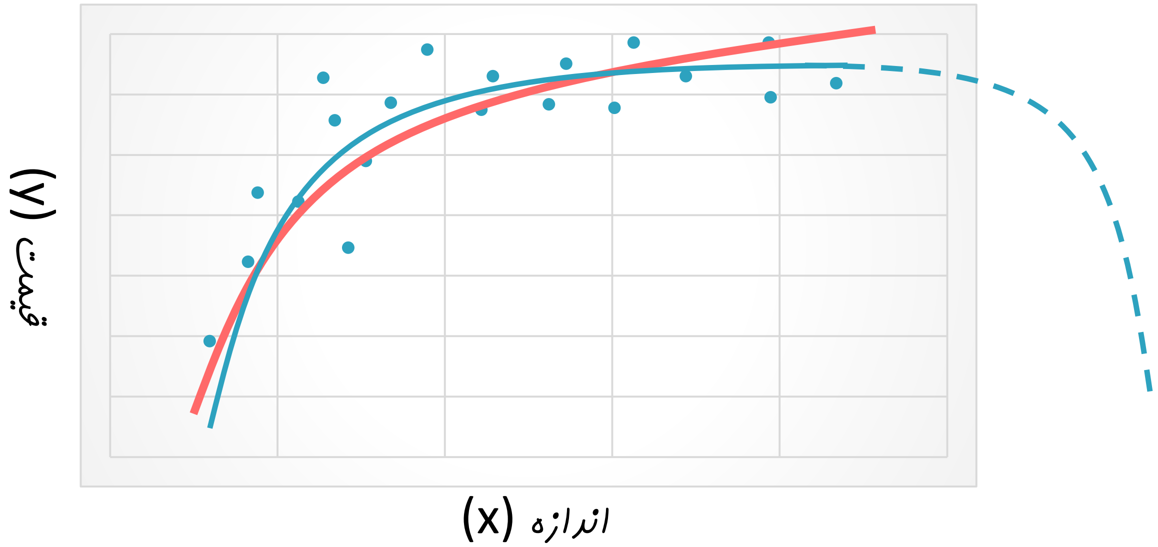
$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

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

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

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

انتخاب ویژگی‌ها



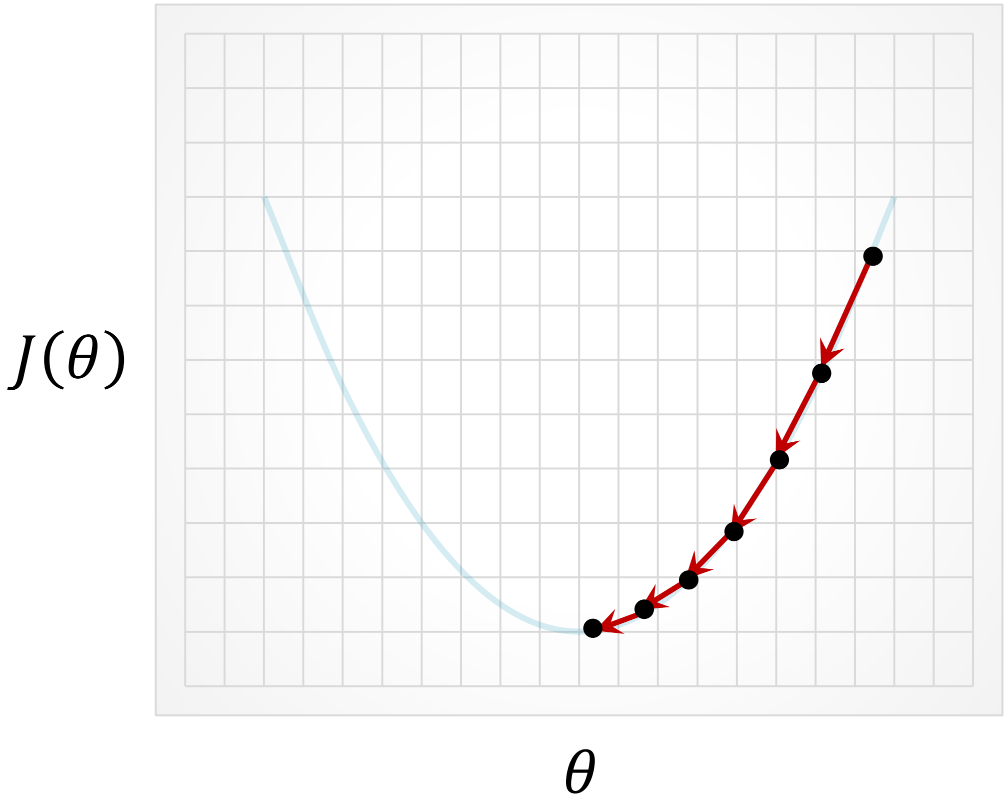
$$h_\theta(x) = \theta_0 + \theta_1(size) + \theta_2(size)^2$$

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

رگرسیون خطی: معادله نرمال

گرادیان کاهشی و معادله نرمال

- گرادیان کاهشی.



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

قاعده به روز رسانی

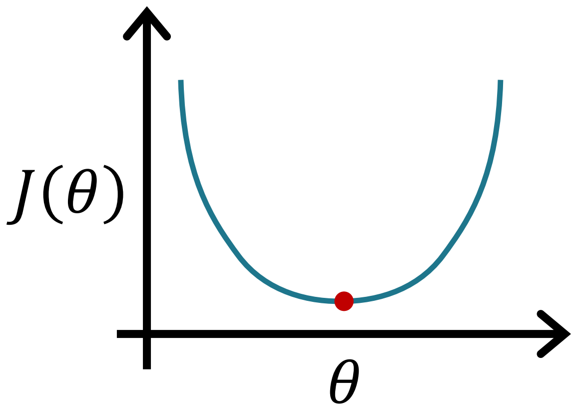
- معادله نرمال. یک روش تحلیلی به منظور تعیین مقدار پارامترها.

معادله نرمال

- معادله نرمال.

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

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



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

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

معادله نرمال: مثال (m = 4)

x_0	Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) 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

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$$X\theta = y$$

معادله نرمال: مثال (m = 5)

x_0	Size (feet^2) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) 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

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \\ 1 & 3000 & 4 & 1 & 38 \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \\ 540 \end{bmatrix}$$

$$X\theta = y$$

معادله نرمال: حالت کلی

- m نمونه آموزشی؛ n ویژگی

$$X\theta = y$$

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

ماتریس طراحی

معادله نرمال

- حل دستگاه معادلات خطی.

$$X\theta = y$$

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

معادله نرمال

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

Python:

```
theta = pinv(X.T @ X) @ X.T @ y
```

گرادیان کاهشی و معادله نرمال

گرادیان کاهشی	معادله نرمال
نیاز به انتخاب α	عدم نیاز به انتخاب α
نیاز به تکرارهای زیاد	عدم نیاز به تکرار
حتی برای مقادیر بسیار بزرگ n به خوبی کار می‌کند.	به دلیل نیاز به محاسبه معکوس ماتریس X^TX، برای n های بزرگ کند است.
$n \geq 10000$	$n < 10000$

معادله نرمال و معکوس ناپذیری X^TX

- معادله نرمال.

$$\theta = (X^TX)^{-1}X^Ty$$

- س. اما اگر X^TX معکوس‌پذیر نباشد چه می‌شود؟

Python:

```
theta = pinv(X.T @ X) @ X.T @ y
```

شبه معکوس

علل معکوس‌ناپذیری X^TX

- افزونگی ویژگی‌ها. [وابستگی خطی]

$$x_1 = size\ (feet^2)$$

$$x_2 = size\ (m^2)$$

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

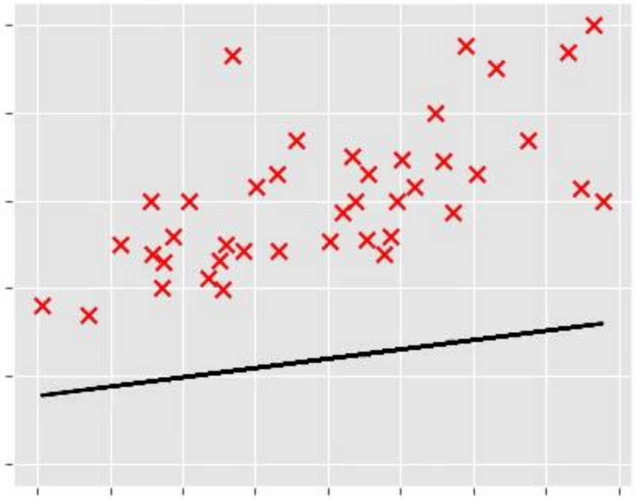
- تعداد بسیار زیاد ویژگی‌ها. $[n \geq m]$
  - راه‌حل. حذف برخی از ویژگی‌ها با استفاده از **تنظیم** [در ادامه]

رگرسیون با وزن‌دهی محلی

روش‌های یادگیری پارامتری و غیرپارامتری

- روش‌های یادگیری پارامتری.
  - یک مجموعه ثابت از پارامترها وجود دارد.
  - برای پیش‌بینی داده‌های جدید، نیازی به مجموعه آموزشی نداریم.
  - مثال: رگرسیون، و رگرسیون لجستیک، شبکه‌های عصبی.

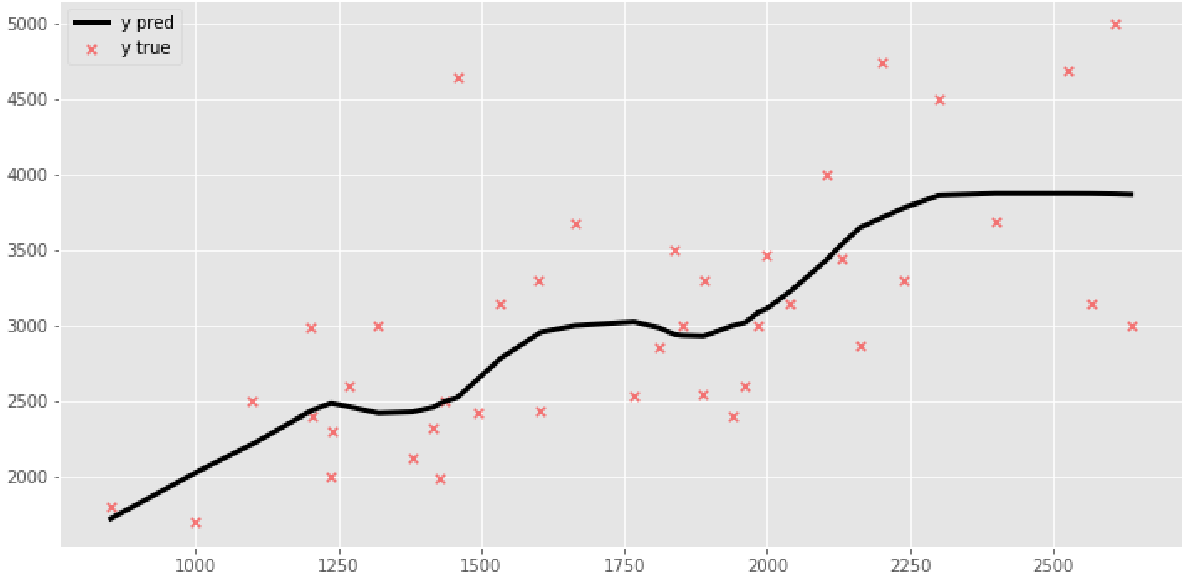
$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$



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

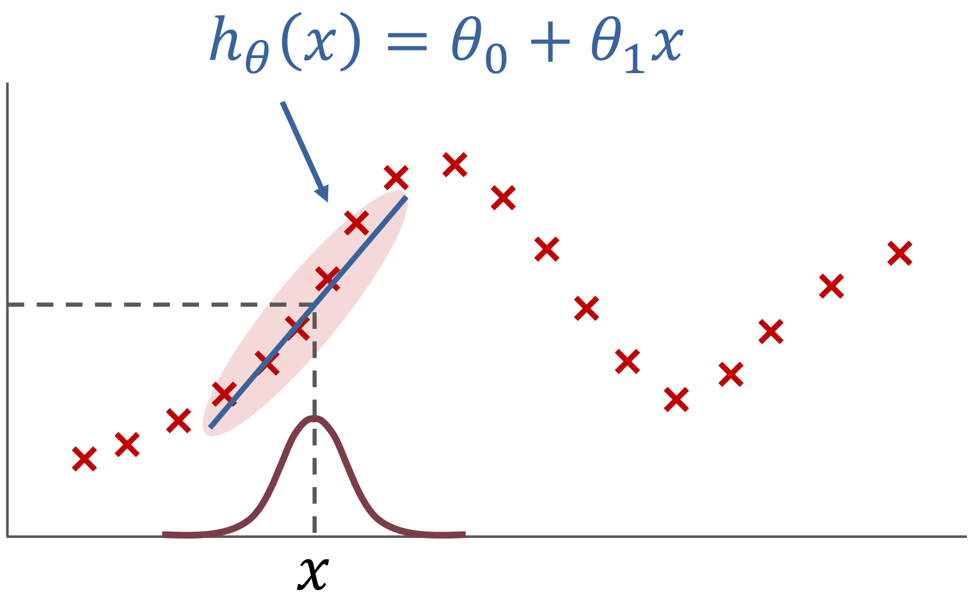
روش‌های یادگیری پارامتری و غیرپارامتری

- روش‌های یادگیری غیرپارامتری.
  - تعداد پارامترها با افزایش اندازه مجموعه آموزشی (به صورت خطی) افزایش می‌یابد.
  - به منظور پیش‌بینی داده‌های جدید، به تمام مجموعه آموزشی نیاز داریم.
  - مثال: رگرسیون با وزن‌دهی محلی [صفحه بعد].



رگرسیون با وزن‌دهی محلی

- ایده. دادن اهمیت بیشتر به داده‌های نزدیک‌تر.



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

پهنای باند

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

رگرسیون: تفسیر احتمالاتی

رگرسیون: تفسیر احتمالاتی

- مدل.

خطا

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

- خطا.
  - در نظر گرفتن تأثیر عوامل دیگر
    - مانند ویژگی‌های درنظر 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. $[iid]$

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

تخمین بیشترین درست‌نمایی

- تابع درست‌نمایی. احتمال مشاهده داده‌های آموزشی به عنوان تابعی از پارامترهای θ

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

- لگاریتم تابع درست‌نمایی.

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

تخمین بیشترین درست‌نمایی

- تخمین بیشترین درست‌نمایی.
  - انتخاب یک مقدار برای پارامتر θ به گونه‌ای که $l(\theta)$ بیشینه گردد.

$$\theta_{MLE} = \arg\max_{\theta} l(\theta)$$
$$= \arg\max_{\theta} \; m \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2$$
$$= \arg\max_{\theta} \; -\frac{1}{2}\sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2$$
$$= \arg\min_{\theta} \; \frac{1}{2}\sum_{i=1}^{m}\left(y^{(i)} - \theta^T x^{(i)}\right)^2 \longleftarrow \text{تابع هزینه}$$