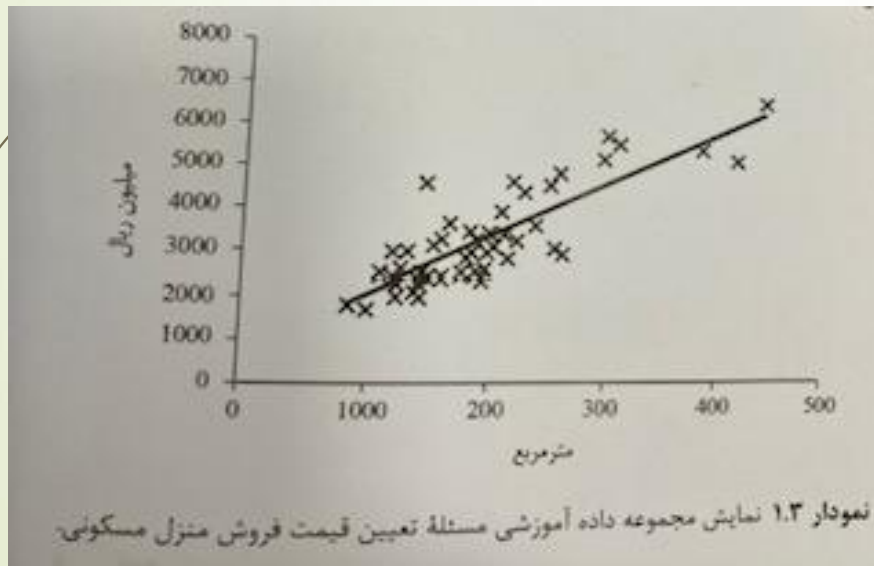


رگرسیون خطی یک متغیره

➤ به دنبال ترسیم خطی بین داده ها هستیم که با مجموعه داده ها آموزشی متناسب باشد.

➤ داده آموزشی: داده هایی که هدف از یادگیری آنها، پیش بینی است.



➤ m : تعداد داده های آموزشی

➤ X : متغیر یا ویژگی ورودی

➤ y : متغیر خروجی یا هدف

➤ (x, y) : یک داده آموزشی

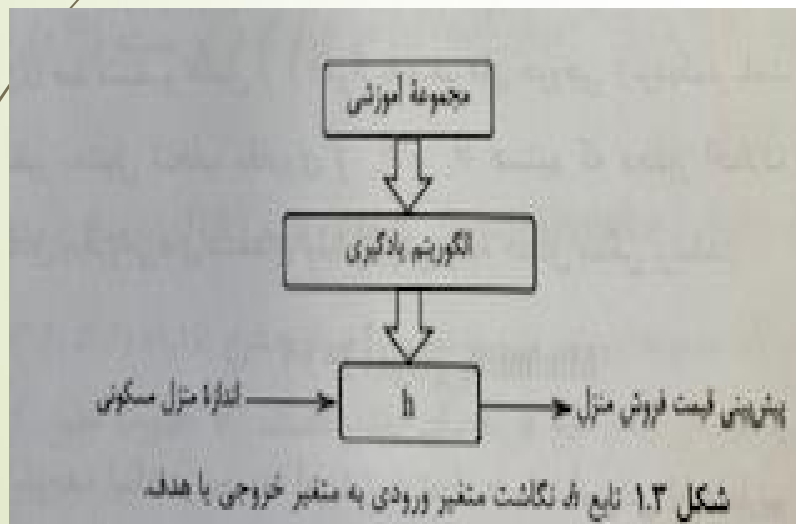
➤ (x_i, y_i) : i امین داده آموزشی

جدول ۱.۳ مجموعه آموزشی مسئله تعیین قیمت فروش منزل مسکونی

متر از خانه (مترمربع)	قیمت (میلیون ریال)
۲۱۰	۳۹۹۹
۱۶۰	۳۲۹۹
۲۴۰	۳۶۹۰
۱۴۱	۲۳۲۰
...	...

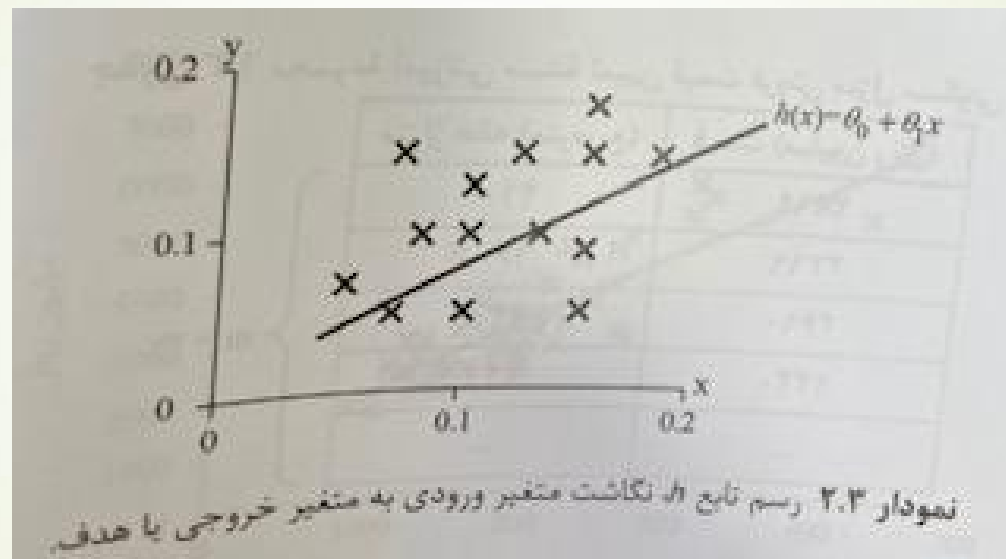
} $m = 47$

- داده های آموزشی به الگوریتم وارد می شود.
تابع فرض (Hypothesis) ایجاد می شود.
تابع h ورودی ها را به خروجی نگاشت می کند.
تعریف درست با پارمترهای مناسب مهمترین بخش در رگرسیون است.



تعریف تابع فرضیه خطی

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



در رگرسیون خطی یک متغیره به دنبال پارامترهای θ_0 و θ_1 بهینه هستیم.

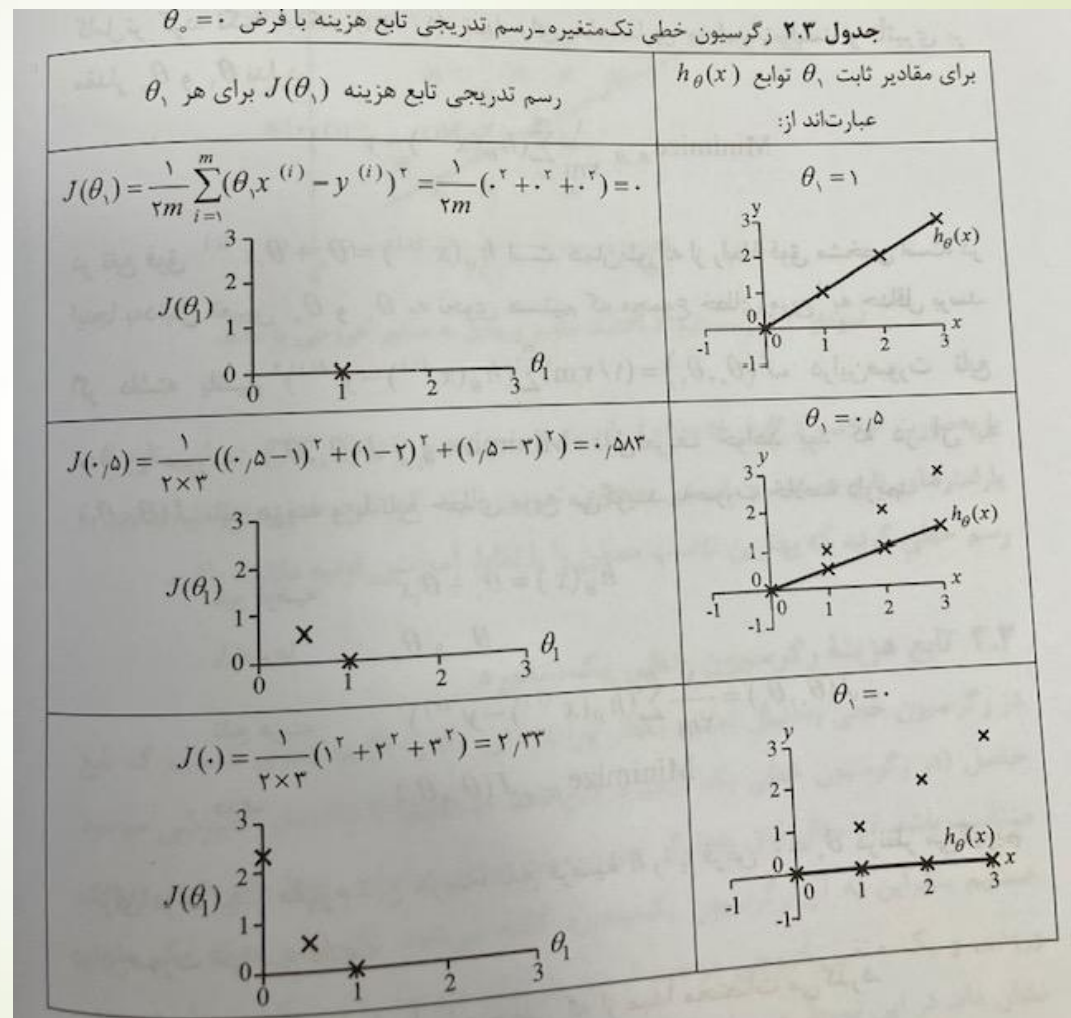
$$\text{Min } J(\theta_0, \theta_1) = (h_{\theta}(x) - y)^2$$

مقدار پارامتر θ_0 برابر صفر و هدف به دست آوردن حداقل $J(\theta_1)$ است.

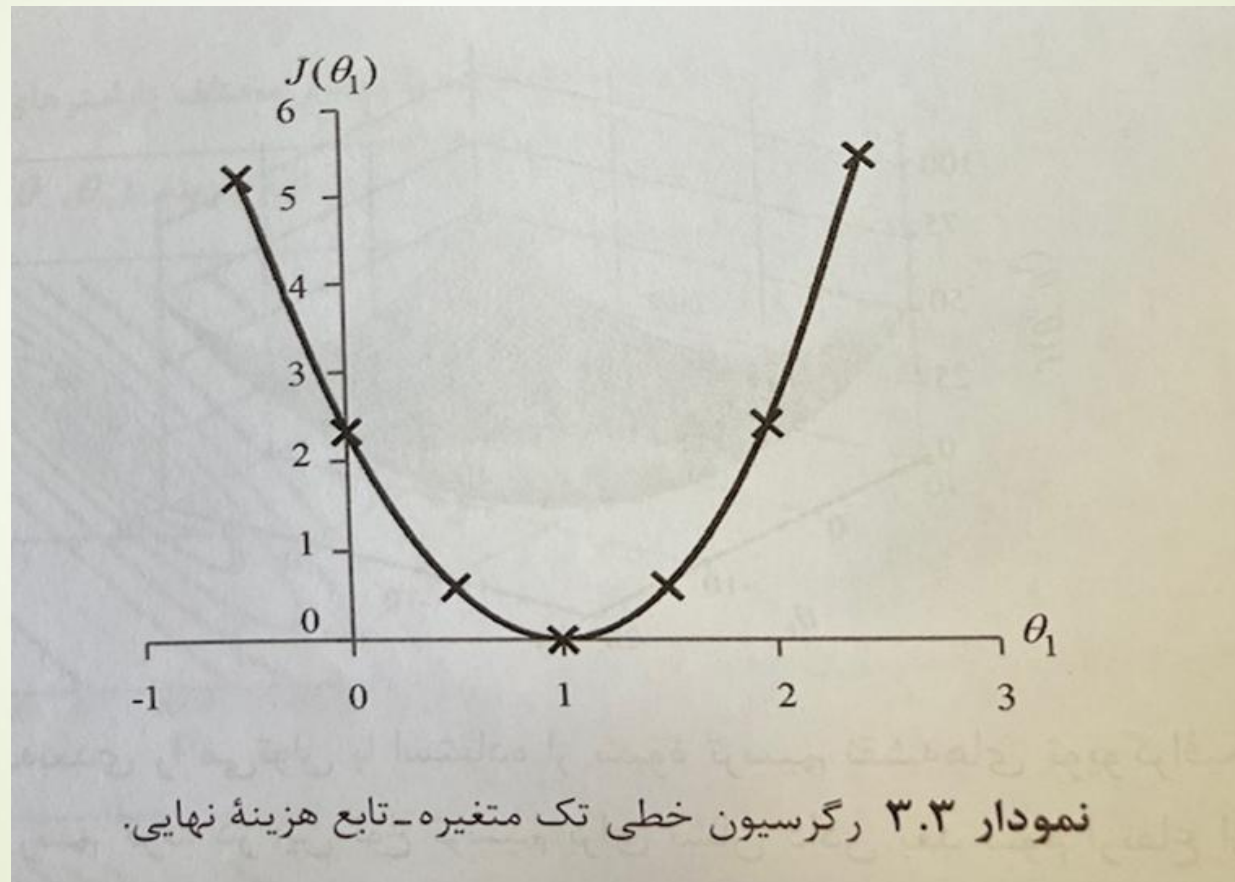
(1, 1)

(2, 2)

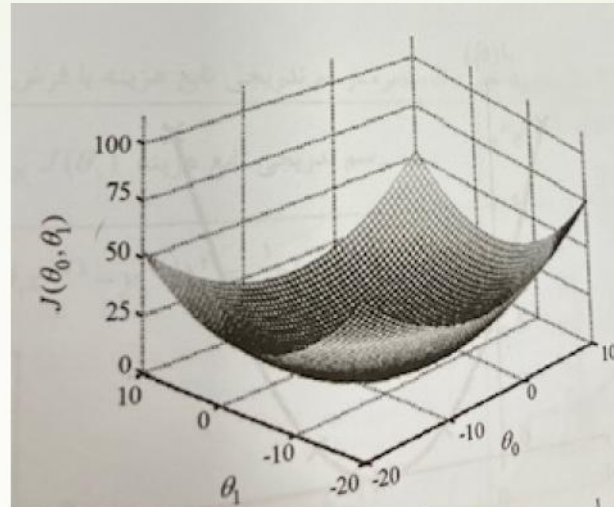
(3, 3)



مقدار پارامتر θ_0 برابر صفر و هدف به دست آوردن حداقل $J(\theta_1)$ است.

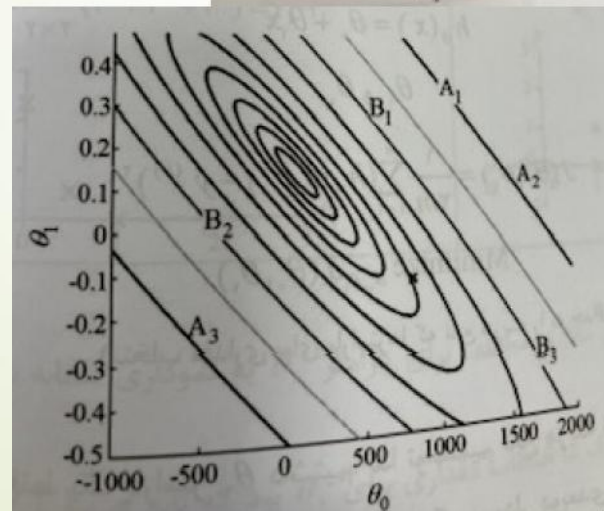


مقدار پارامتر θ_0 و θ_1 مخالف صفر و هدف به دست آوردن حداقل $J(\theta_0, \theta_1)$



نمودار سه بعدی می شود.

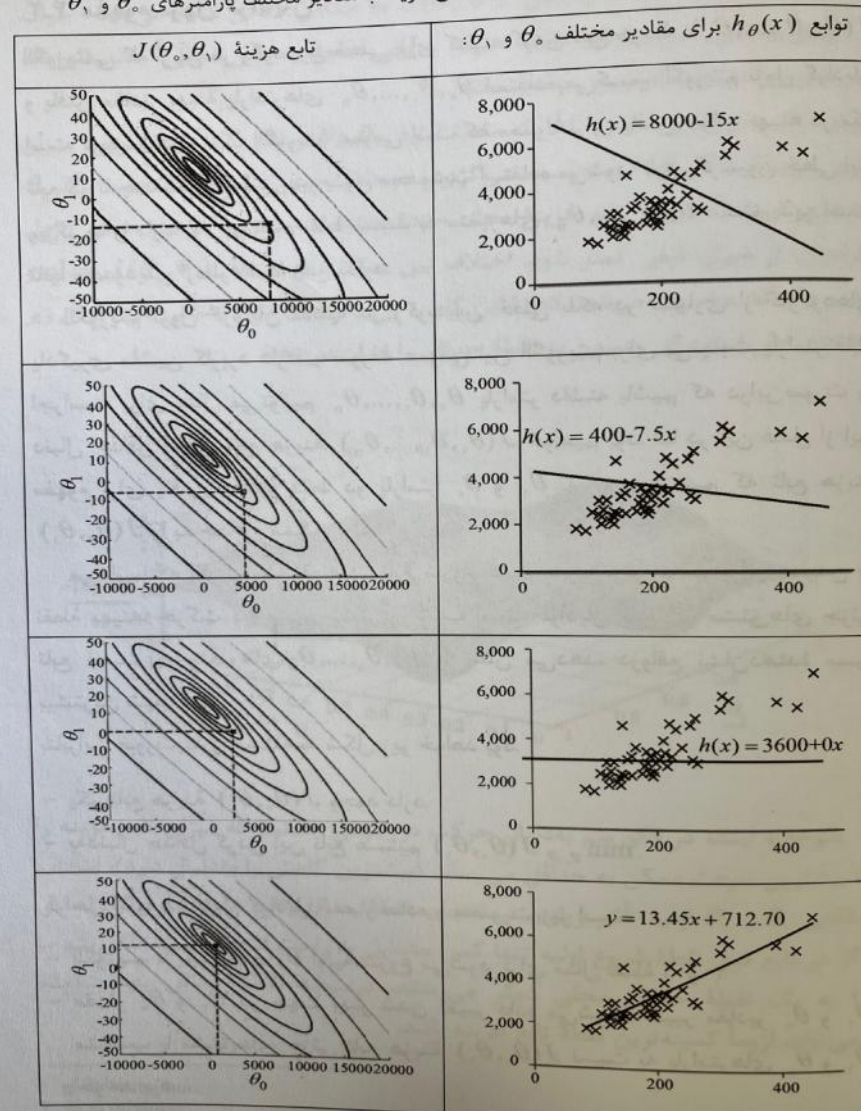
نقشه توپوگراف



جهت نشان دادن ارتفاع از حلقه استفاده می شود.

هر حلقه یک مقدار یکسان را نشان می دهد.

جدول ۳.۳ رگرسیون خطی تک متغیره-تابع هزینه با مقادیر مختلف پارامترهای θ_0 و θ_1

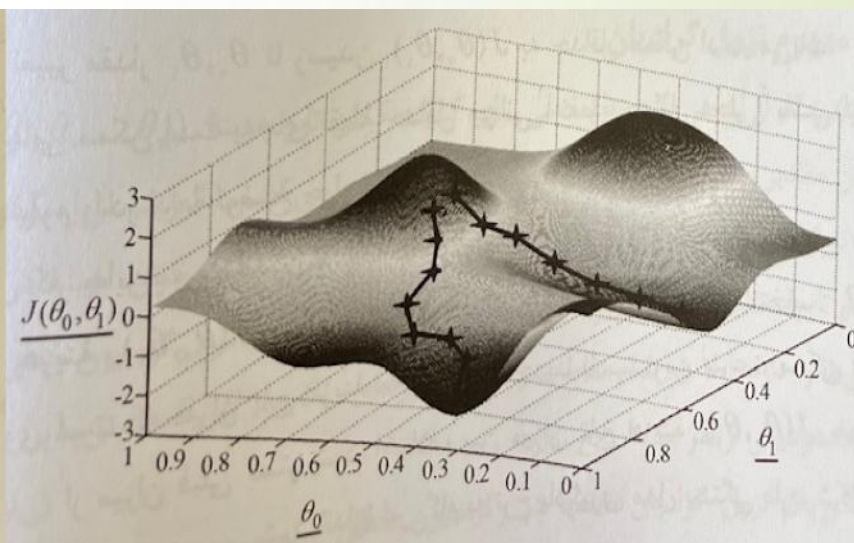
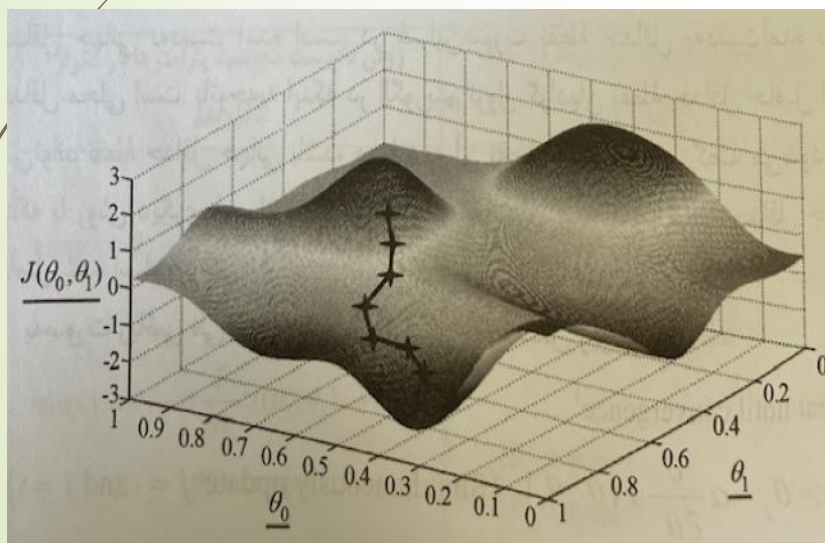


گرادیان

- در رگرسیون خطی، هدف به دست آوردن الگوریتمی است که توسط کامپیوتر مقادیر 0 و 1 را طوری تعیین کند که مقدار $J(0, 1)$ حداقل ممکن شود.
- الگوریتم نزول گرادیان
- جهت حداقل کردن مقدار تابع هزینه (J) و به دست آوردن پارامترهای استفاده می شود (در مورد تعداد ها محدودیت ندارد).
- تابع هزینه (J) باید در تمام نقاط نسبت به پارامترهای مشتق پذیر باشد.
- تابع هزینه (J) محدودیتی در دامنه تابع نداشته باشد.
- سریع ترین مسیر جهت رسیدن به نقطه بهینه حرکت در مسیر بیشترین شیب است.
- گرادیان جهت بیشترین تغییرات در تابع را نشان می دهد.
- گرادیان تابع مشتقات جزئی تابع نسبت به پارامترهای است.

مفهوم نزول گرادینان

- الگوریتم با یک مقدار θ_0 و θ_1 اولیه شروع می شود.
- مقادیر θ_0 و θ_1 متناسب با مشتق های جزئی تابع هزینه J تغییر داده می شوند.
- تغییر مقادیر θ_0 و θ_1 تا رسیدن به مقدار بهینه ادامه می یابد.
- مقدار نهایی ممکن است به جای حداقل جهانی به حداقل محلی همگرا شود.

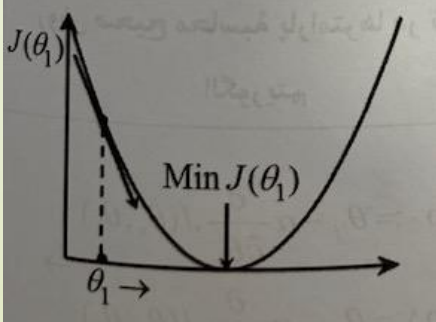
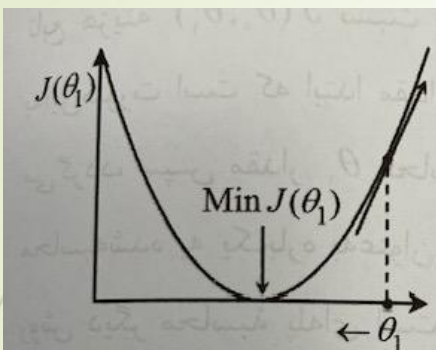


Repeat until convergence

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (j=0 \text{ و } j=1 \text{ به روز رسانی همزمان})$$

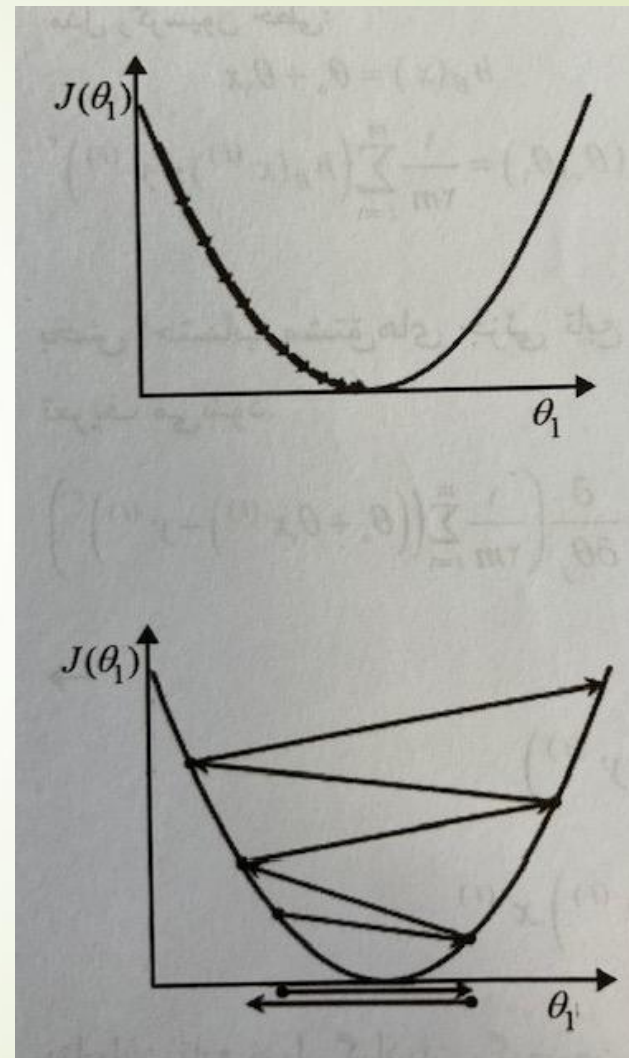
روش نادرست محاسبه پارامترها در تکرار الگوریتم	روش صحیح محاسبه پارامترها در تکرار الگوریتم
$\text{temp} \cdot := \theta_j - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \rightarrow$ $\theta_0 := \text{temp} \cdot$ $\text{temp} \setminus := \theta_j - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \rightarrow$ $\theta_1 := \text{temp} \setminus$	$\left\{ \begin{array}{l} \text{temp} \cdot := \theta_j - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\ \text{temp} \setminus := \theta_j - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \end{array} \right. \rightarrow$ $\rightarrow \left\{ \begin{array}{l} \theta_0 := \text{temp} \cdot \\ \theta_1 := \text{temp} \setminus \end{array} \right.$

α نرخ یادگیری است که طول گام
های الگوریتم در هر تکرار است.

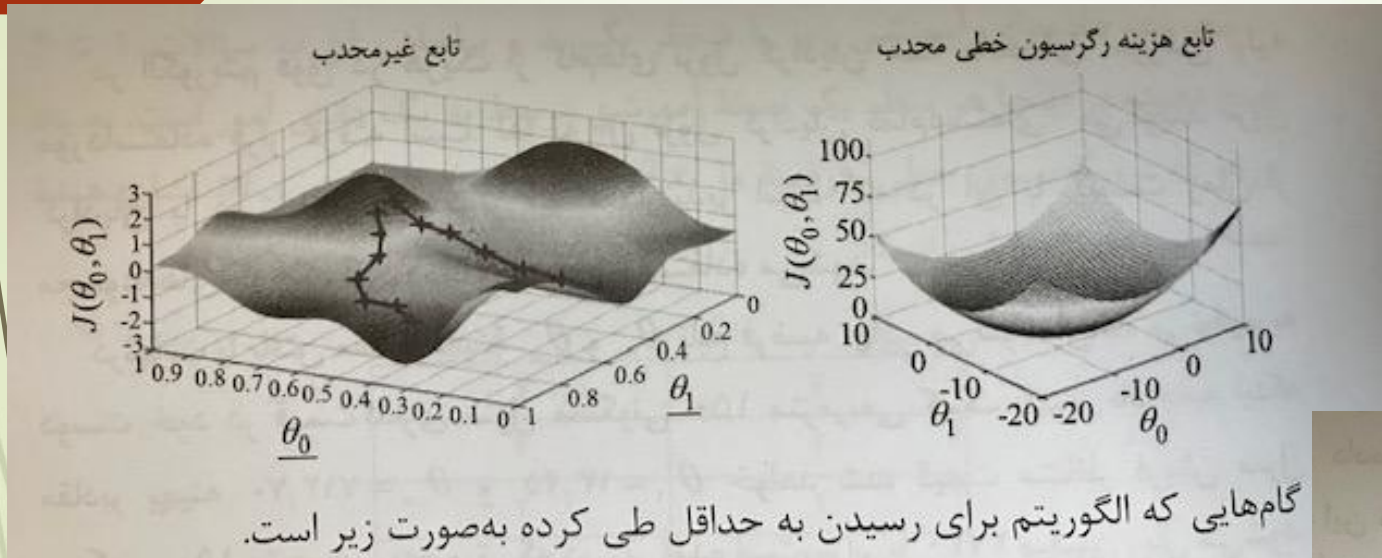


α کوچک

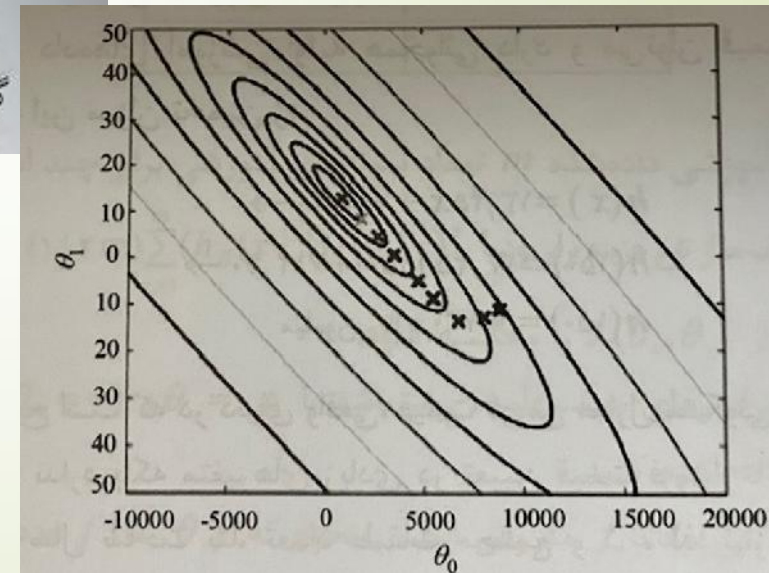
α بزرگ



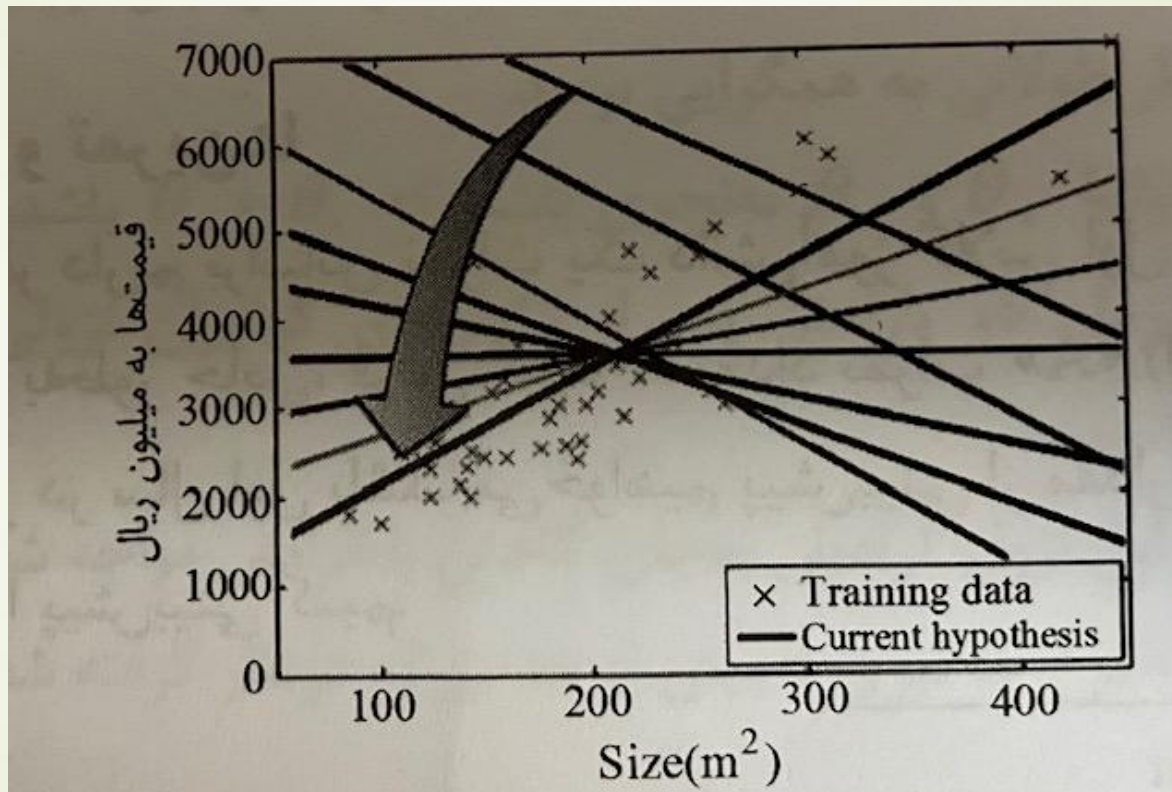
مشکل نزول گرادیان:
گیر افتادن در مینیموم محلی به جای مینیموم عمومی



در رگرسیون خطی مینیموم محلی وجود ندارد.



حرکت تابع فرض اولیه به سمت تابع یادگیری بهینه



رگرسیون خطی چند متغیره

متراژ خانه (مترمربع)	تعداد اتاق ها	تعداد طبقات	سن بنا به سال	قیمت (میلیون ریال)
x_1	x_2	x_3	x_4	y
۲۱۰	۵	۱	۴۵	۳,۹۹۹
۱۶۰	۳	۲	۴۰	۳,۲۹۹
۲۴۰	۳	۲	۳۰	۳,۳۶۰
۱۴۱	۲	۱	۳۶	۲,۳۲۰

n تعداد ویژگی ها

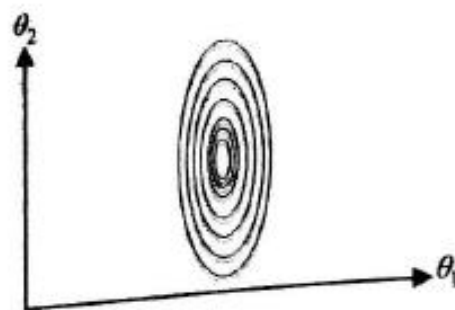
m تعداد داده های آموزشی

y متغیر خروجی یا هدف

x^i بردار ویژگی های داده ام

x_j^i مقدار ویژگی j ام در داده آموزشی ام

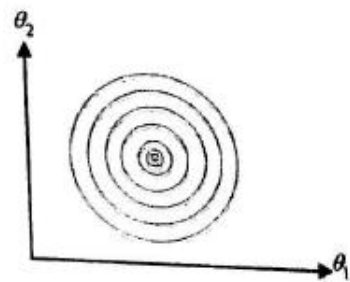
هم مقیاس نبودن متغیرها



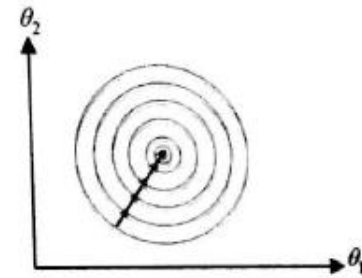
نمودار ۱.۵ تابع هزینه متغیرهای با مقیاس متفاوت.



هم مقیاس کردن متغیرها



نمودار ۳.۵ تابع هزینه متغیرهای هم مقیاس



$$S_j = \max x_j^i - \min x_j^i \quad 1 \leq i \leq m$$

$$u_j = \sum_{i=1}^m x_j^i / m$$

$$x_j^i = \frac{x_j^i}{S_j}$$

$$x_j^i = \frac{x_j^i - u_j}{S_j}$$

$$X_{(n+1) \times 1} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \theta_{(n+1) \times 1} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$\hat{y}(x) = \theta^T X = [\theta_0 \ \theta_1 \ \theta_n]^T \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

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

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

تابع فرض

$$\theta_0, \theta_1, \dots, \theta_n$$

پارامترهای مسئله

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

تابع هزینه

$$\text{Min } J(\theta_0, \theta_1, \dots, \theta_n)$$

تابع هدف

الگوریتم نزول گرادیان

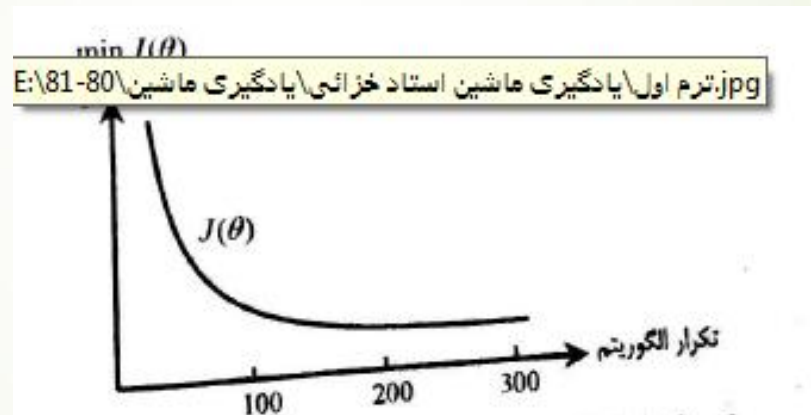
Repeat Until Convergence {

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

}Simultaneously Update for every $j=0, 1, 2, \dots, n$

ترسیم تابع هزینه جهت فهم عملکرد صحیح الگوریتم

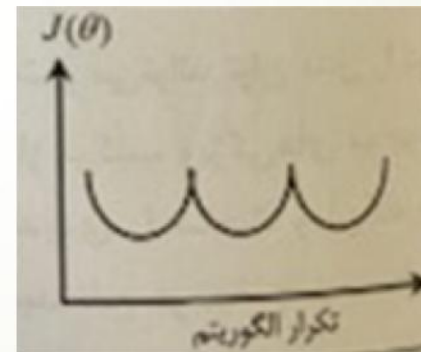
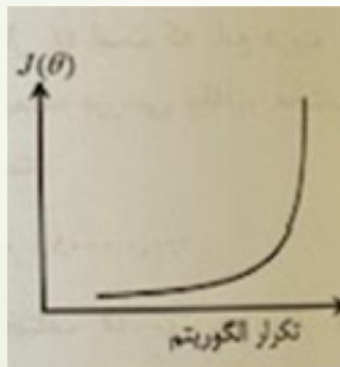
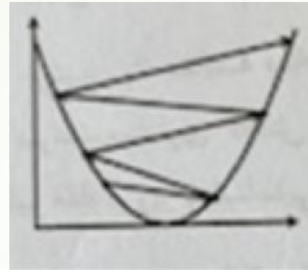
نمودار یک تابع هزینه نرمال نسبت به تعداد تکرارها



با افزایش تعداد تکرارها تابع هزینه کاهش و از یک تعداد تکرار به بعد همگرا شده و هزینه دیگر کاهش نمی یابد.

مقادیر ناصحیح پارامتر α

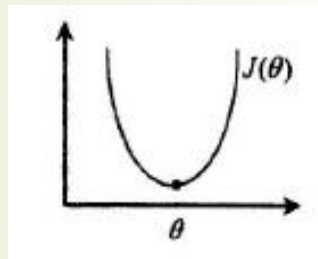
بهترین راه مقایسه مقادیر مختلف α
و انتخاب مقدار مناسب α



افزایش تابع هزینه و عملکرد اشتباه نزول گرادینت
مقدار α کاهش یابد

حل تحلیلی مدل

مشتق از تابع هزینه بر حسب پارمترها و برابر صفر قرار دادن حاصل



$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \dots, \theta_n) = 0 \quad \text{for every } j$$

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

$$\theta = \text{inv}(X' * X) * X' * y$$

$$\theta = \text{pinv}(X' * X) * X' * y$$

عدم نیاز به پارامتر α برخلاف الگوریتم نزول گرادینت
عدم نیاز به تکرار برخلاف الگوریتم نزول گرادینت
برای ویژگیهای زیاد (n) سرعت کاهش می یابد برخلاف الگوریتم نزول گرادینت
مشکل در صورت صفر بودن دترمینان ماتریس ویژگی ها

متراژ خانه (مترمربع)	تعداد اتاق ها	تعداد طبقات	سن بنا به سال	قیمت (میلیون ریال)
x_1	x_2	x_3	x_4	y
۲۱۰	۵	۱	۴۵	۳,۹۹۹
۱۶۰	۳	۲	۴۰	۳,۲۹۹
۲۴۰	۳	۲	۳۰	۳,۳۶۰
۱۴۱	۲	۱	۳۶	۲,۳۲۰
...

n تعداد ویژگی ها

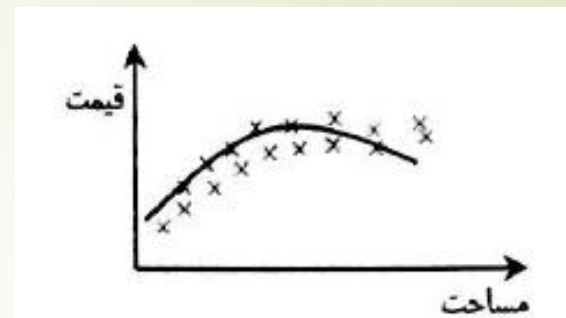
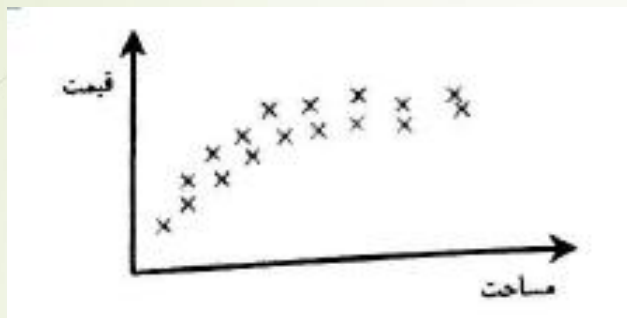
m تعداد داده های آموزشی

y متغیر خروجی یا هدف

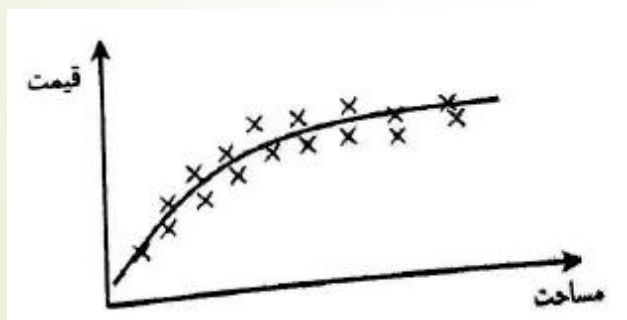
x^i بردار ویژگی های داده ام

x_j^i مقدار ویژگی j ام در داده آموزشی ام

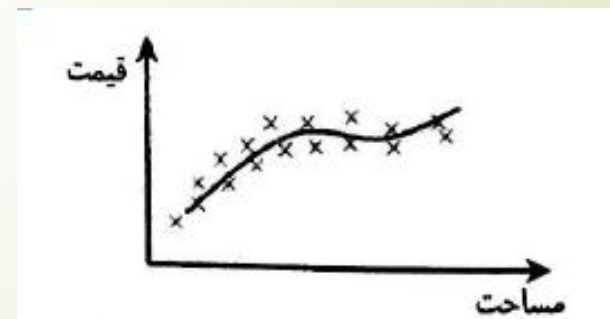
انتخاب صحیح ویژگی ها



$$\hat{y}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



$$\hat{y}(x) = \theta_0 + \theta_1 x + \theta_2 \sqrt{x}$$



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