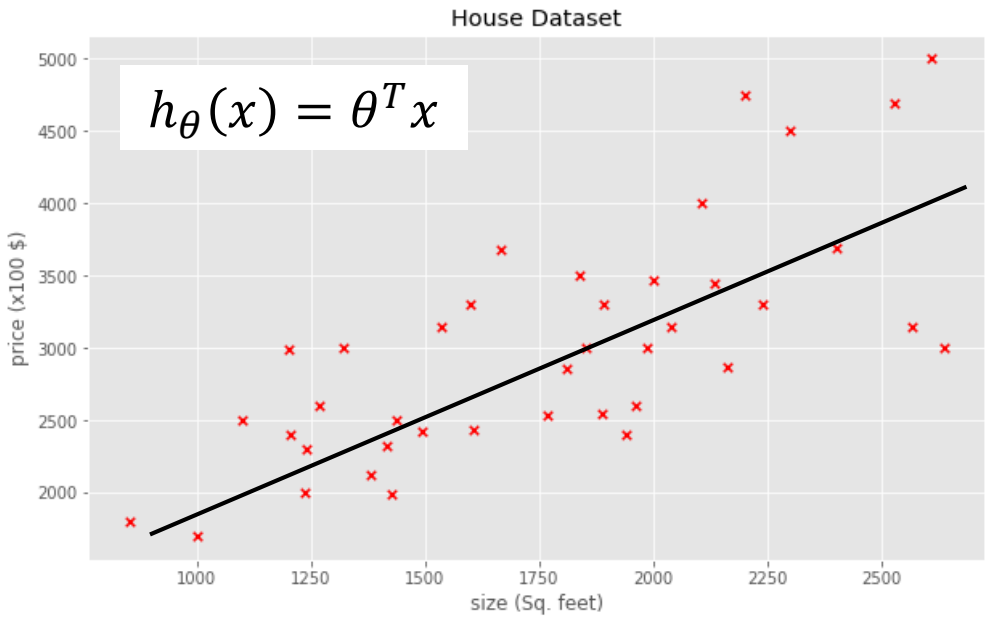
دسته‌بندی: رگرسیون لجیستیک

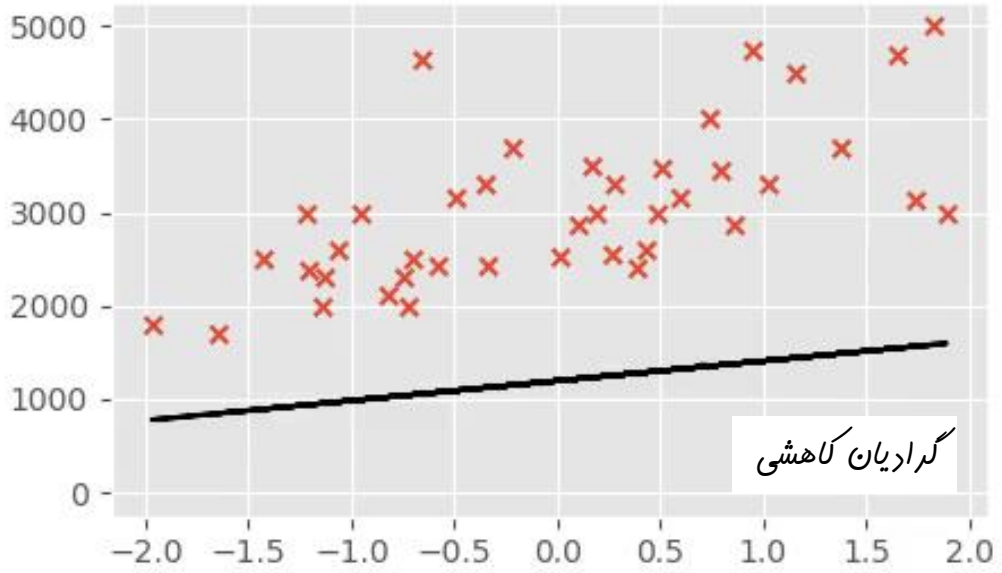
سید ناصر رضوی n.razavi@tabrizu.ac.ir

۱۳۹۷

یادآوری: رگرسیون

- هدف. تخمین یک کمیت پیوسته با توجه به مقادیر ویژگی‌ها.





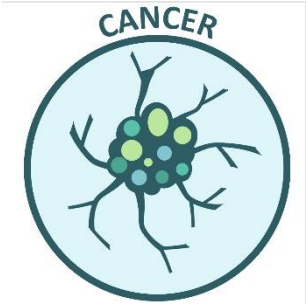
$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

تابع هزینه

کمینه‌سازی

دسته‌بندی

- ایمیل: هرزنامه (بله / خیر؟)
- تراکنش برخط: کلاهبرداری (بله / خیر؟)
- غده سرطانی: خوش‌خیم / بدخیم ؟



- در این مثال‌ها، متغیری که می‌خواهیم مقدارش را پیش‌بینی کنیم دارای دو مقدار است:

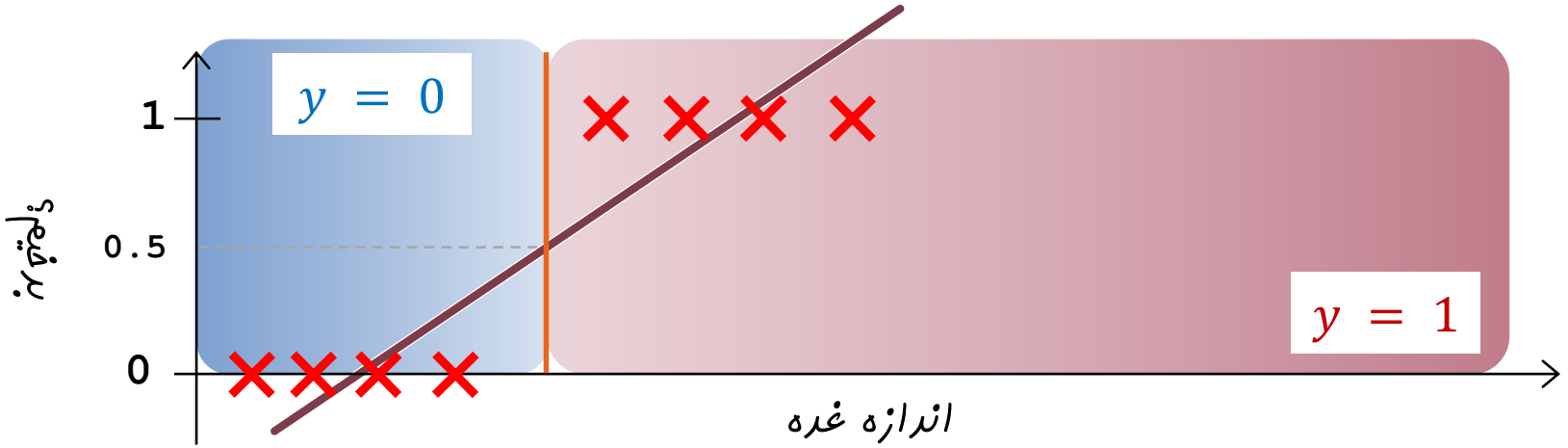
صفر: «کلاس منفی» (مانند غده خوش‌خیم)

یک: «کلاس مثبت» (مانند غده بدخیم)

$$y \in \{0,1\}$$

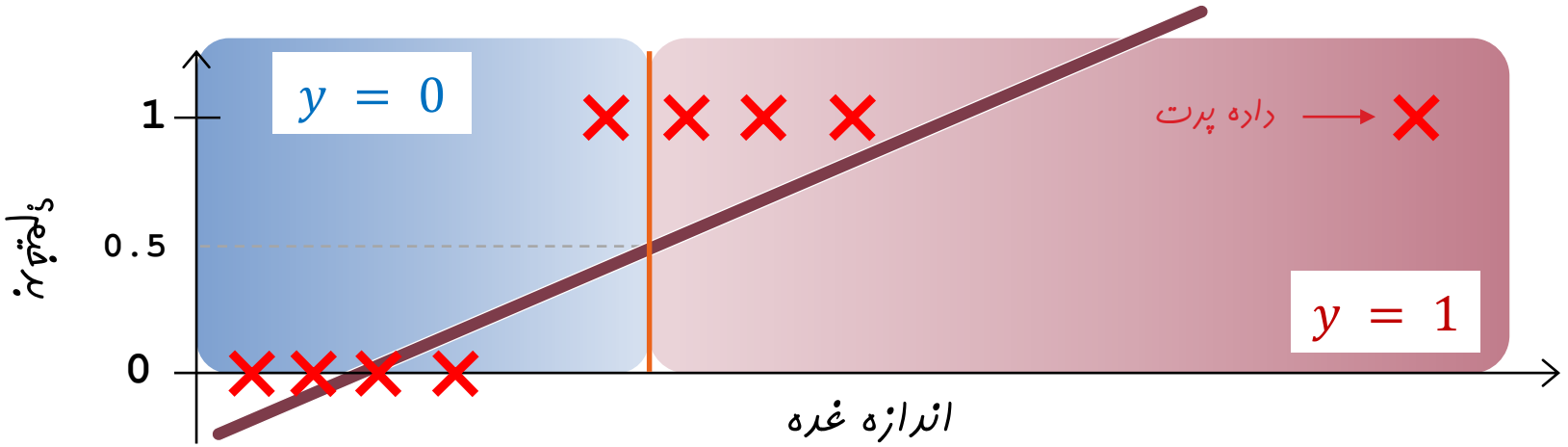
- دسته‌بندی. پیش‌بینی یک متغیر با مقادیر گسسته.
  - دسته‌بندی دودویی
  - دسته‌بندی چندکلاسی

دسته‌بندی



- قرار دادن یک آستانه بر روی خروجی دسته‌بند:
  - اگر $h_\theta(x) \geq 0.5$، آنگاه $y = 1$
  - اگر $h_\theta(x) < 0.5$، آنگاه $y = 0$

دسته‌بندی



- قرار دادن یک آستانه بر روی خروجی دسته‌بند:
  - اگر $h_\theta(x) \geq 0.5$، آنگاه $y = 1$
  - اگر $h_\theta(x) < 0.5$، آنگاه $y = 0$

دسته‌بندی

- در دسته‌بندی دودویی داریم:

$$y = 0 \quad \text{یا} \quad y = 1$$

- اما در رگرسیون ممکن است:

$$h_\theta(x) < 0 \quad \text{یا} \quad h_\theta(x) > 1$$

- رگرسیون لجستیکی (دسته‌بندی).

$$0 \leq h_\theta(x) \leq 1$$

بازنمایی فرضیه در رگرسیون لجستیک

بازنمایی فرضیه

- هدف.

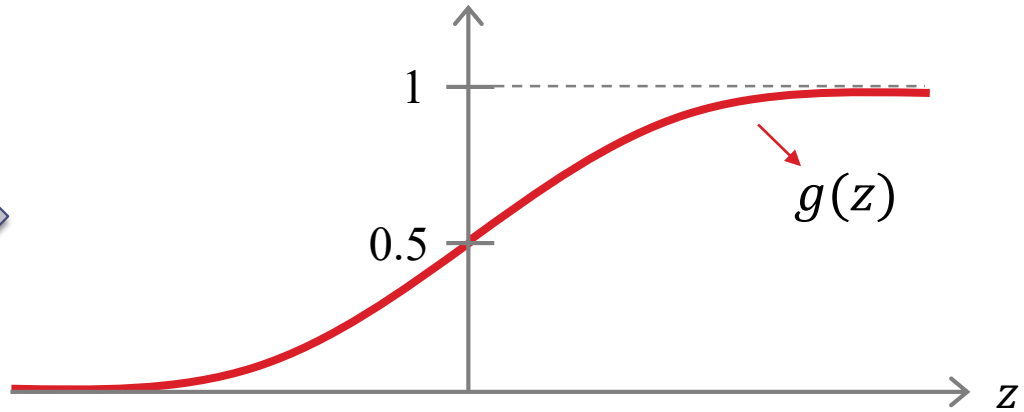
$$0 \leq h_\theta(x) \leq 1$$

- فرضیه.

$$h_\theta(x) = g(\theta^T x)$$

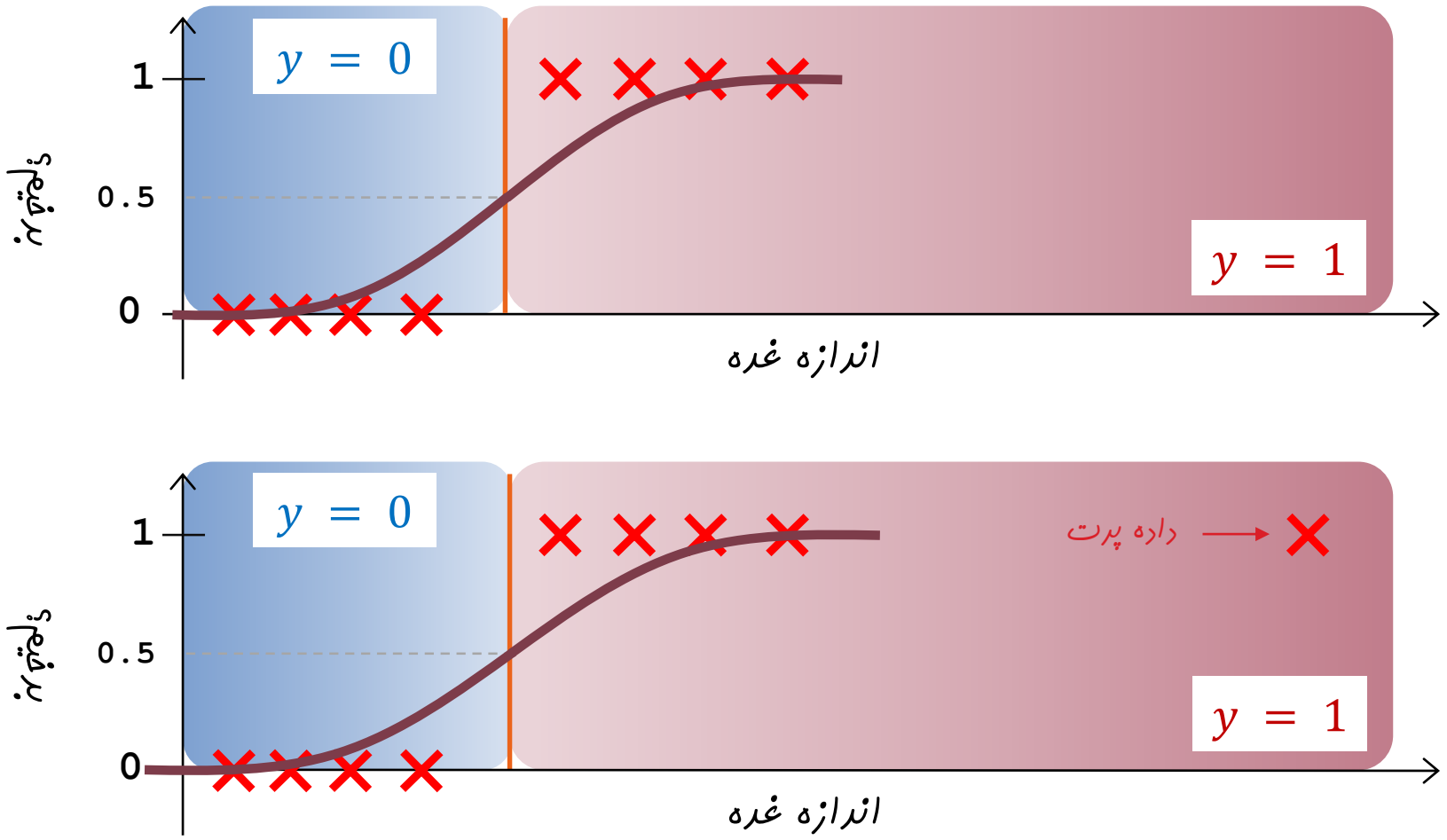
$$g(z) = \frac{1}{1 + e^{-z}}$$

تابع سیگموید
(لجستیک)



$$0 \leq g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \leq 1$$

رگرسیون لجستیک و دسته‌بندی



فرضیه

- تفسیر خروجی فرضیه.

«احتمال این که ورودی x به دسته $y = 1$ تعلق داشته باشد»

- مثال. اگر داشته باشیم: $h_\theta(x) = 0.7$ $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ tumorSize \end{bmatrix}$,

در این صورت، به احتمال ۷۰ درصد، این غده سرطانی بدخیم است.

$$p(y = 1|x; \theta) = h_\theta(x)$$

$$p(y = 0|x; \theta) = 1 - p(y = 1|x; \theta) = 1 - h_\theta(x)$$

فرضیه

- تفسیر احتمالاتی فرضیه.

$$p(y=1|x;\theta) = h_\theta(x)$$

$$p(y=0|x;\theta) = 1 - h_\theta(x)$$

$$p(y|x;\theta) = h_\theta(x)^y \left(1 - h_\theta(x)\right)^{1-y}$$

- تابع درست‌نمایی.

$$L(\theta) = p(Y|X;\theta) = \prod_{i=1}^{m} p\left(y^{(i)}|x^{(i)};\theta\right)$$

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

تخمین بیشترین درست‌نمایی

- لگاریتم تابع درست‌نمایی.

$$l(\theta) = \log L(\theta) = \log \prod_{i=1}^{m} h_\theta\left(x^{(i)}\right)^{y^{(i)}} \left(1 - h_\theta\left(x^{(i)}\right)\right)^{1-y^{(i)}}$$

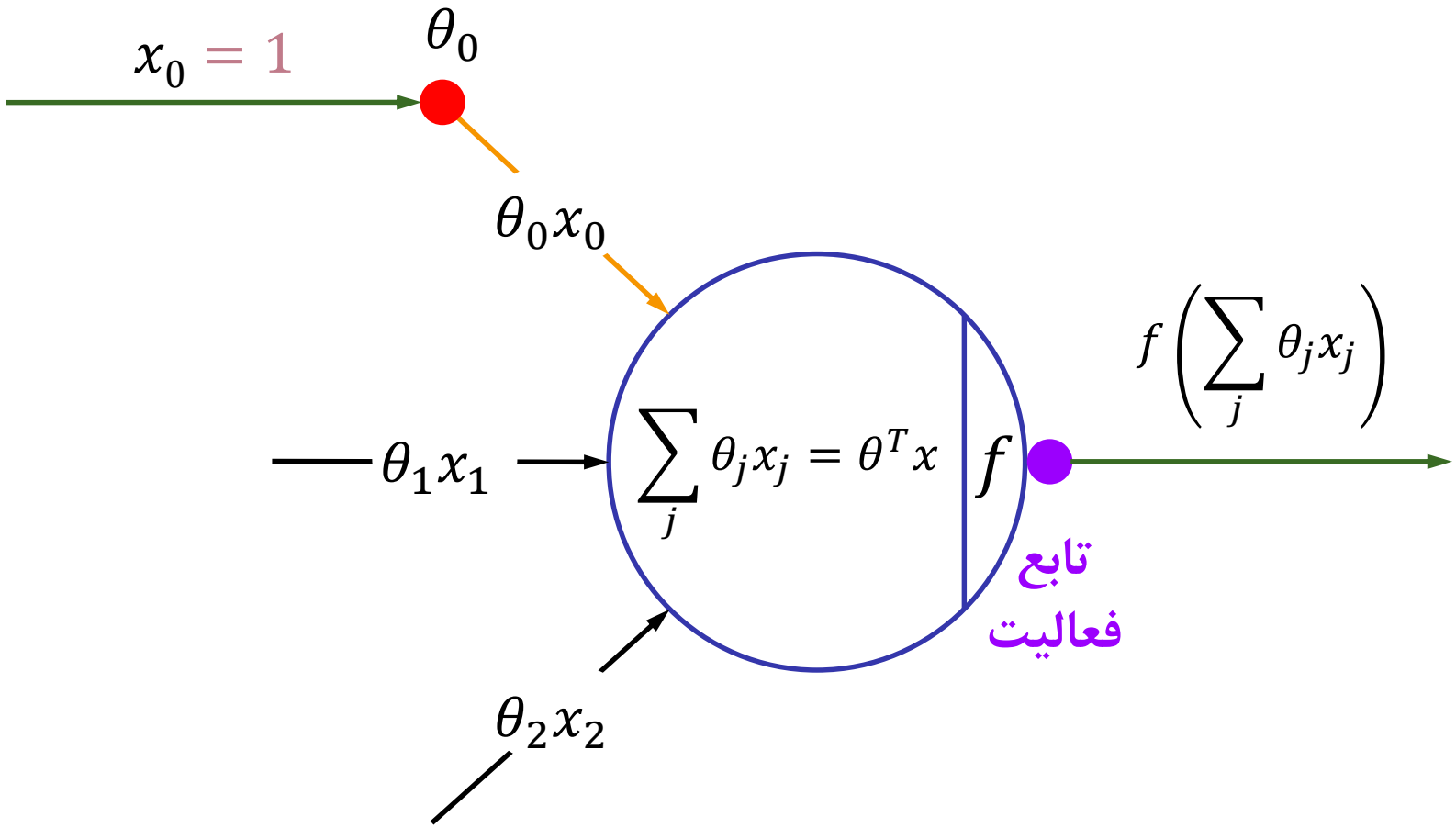
$$= \sum_{i=1}^{m} \log\left(h_\theta\left(x^{(i)}\right)^{y^{(i)}} \left(1 - h_\theta\left(x^{(i)}\right)\right)^{1-y^{(i)}}\right)$$

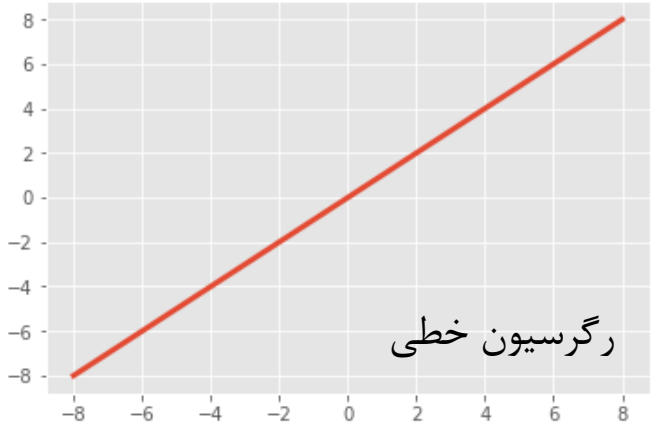
$$= \sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

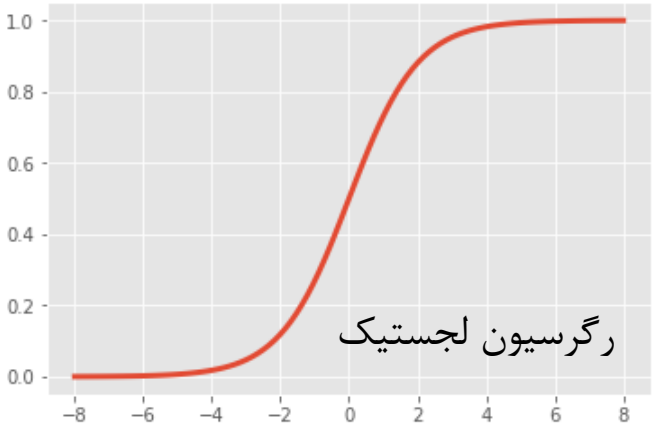
- تابع هزینه.

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

رگرسیون خطی و رگرسیون لجستیک







مرز تصمیم‌گیری

مرز تصمیم‌گیری

- رگرسیون لجستیک.

$$h_\theta(x) = g(\theta^T x)$$

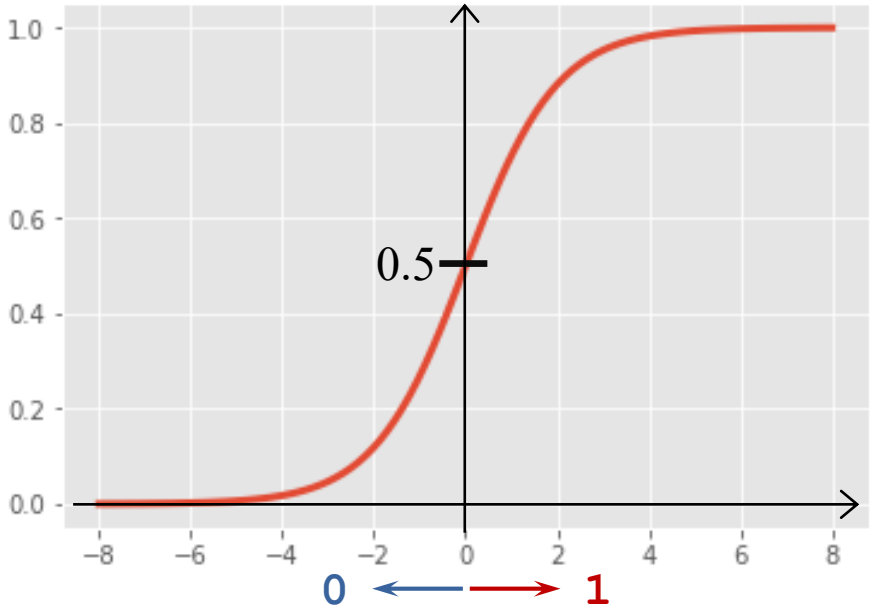
- قرار دادن یک آستانه بر روی خروجی دسته‌بند:

$$y = 1:\ h_\theta(x) \geq 0.5 \Rightarrow \theta^T x \geq 0$$

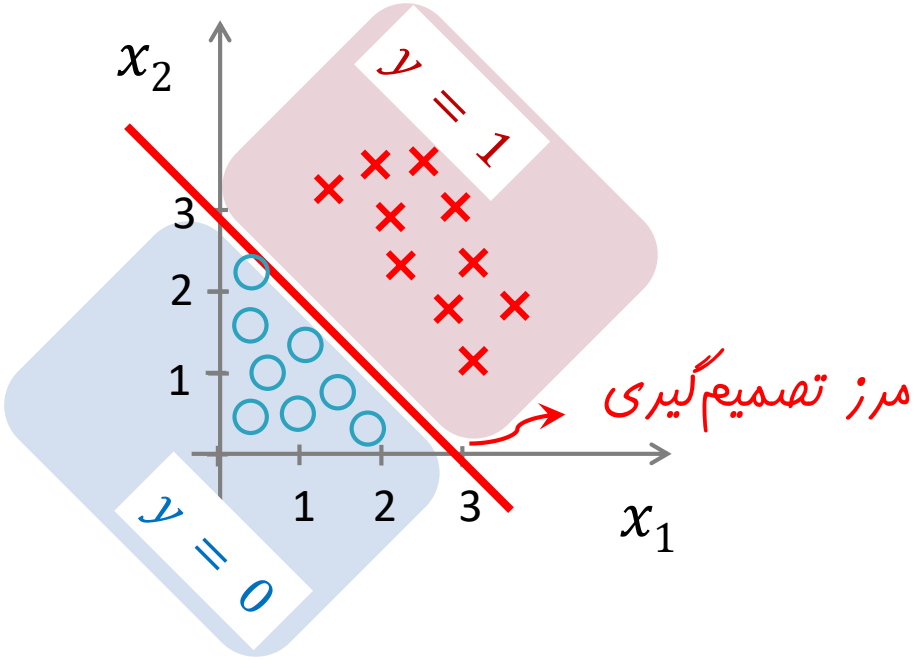
$$y = 0:\ h_\theta(x) < 0.5 \Rightarrow \theta^T x < 0$$

$$\theta^T x = 0$$

معادله مرز تصمیم‌گیری



مرز تصمیم‌گیری



- مرز تصمیم‌گیری.

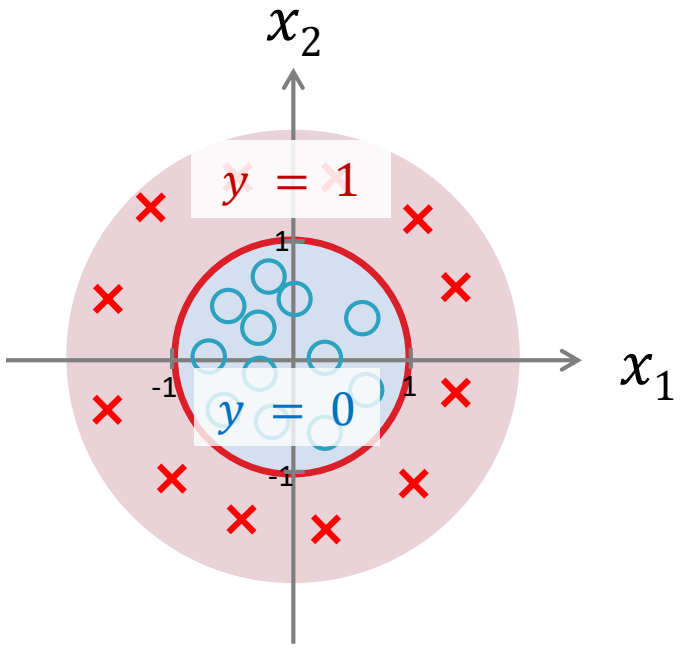
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

-3 1 1

- خروجی y برابر با ۱ است، اگر $-3 + x_1 + x_2 \geq 0$

- $x_1 + x_2 \geq 3 \Rightarrow y = 1$
- $x_1 + x_2 < 3 \Rightarrow y = 0$

مرز تصمیم‌گیری غیرخطی



- مرز تصمیم‌گیری:

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

$-1 \quad 0 \quad 0 \quad 1 \quad 1$

$$x_1^2 + x_2^2 \geq 1 \Rightarrow y = 1$$

$$x_1^2 + x_2^2 < 1 \Rightarrow y = 0$$

تابع هزینه

رگرسیون لجستیک

- مجموعه آموزشی.

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

- نمونه آموزشی.

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad x_0 = 1, \qquad y \in \{0,1\}$$

- فرضیه.

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

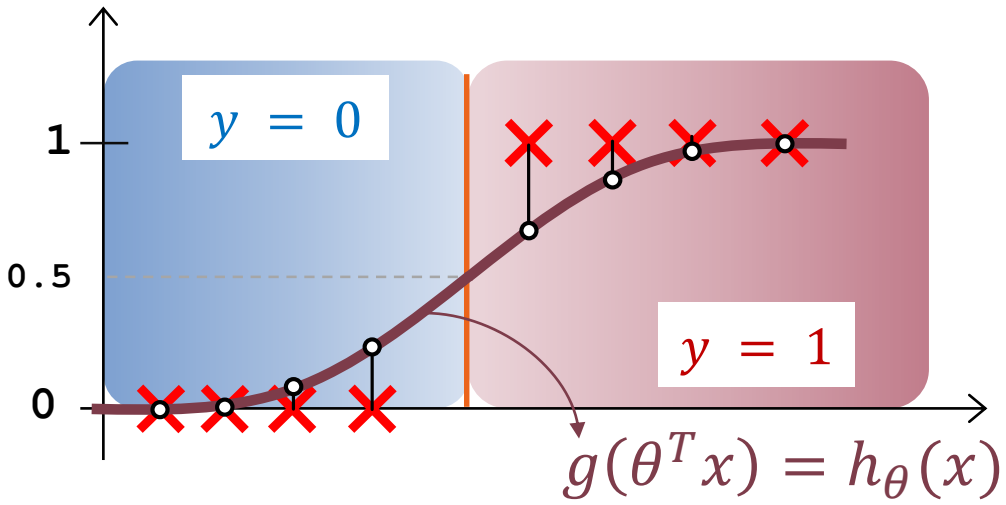
- س. مقادیر پارامترهای θ را چگونه انتخاب کنیم؟

رگرسیون لجستیک

- تابع هزینه.

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

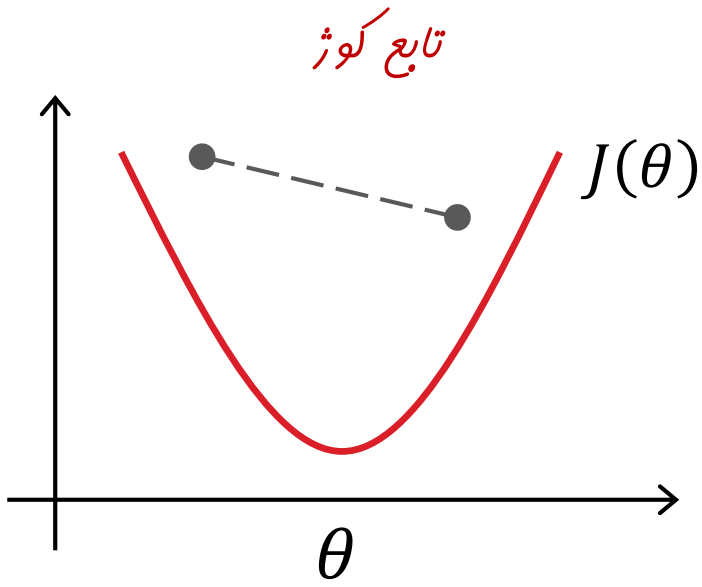
$$cost\left(h_\theta \left(x^{(i)} \right), y^{(i)} \right) = \frac{1}{2} \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right)^2$$

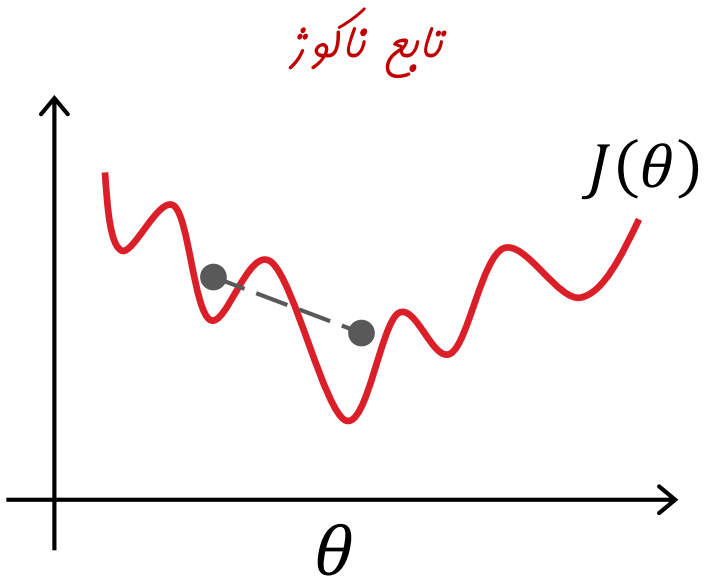


توجه. از آنجا که $h_\theta(x^{(i)})$ یک تابع غیرخطی از پارامترها است، تابع هزینه دیگر یک تابع کوژ نخواهد بود.

تابع هزینه

- توابع کوژ و ناکوژ.

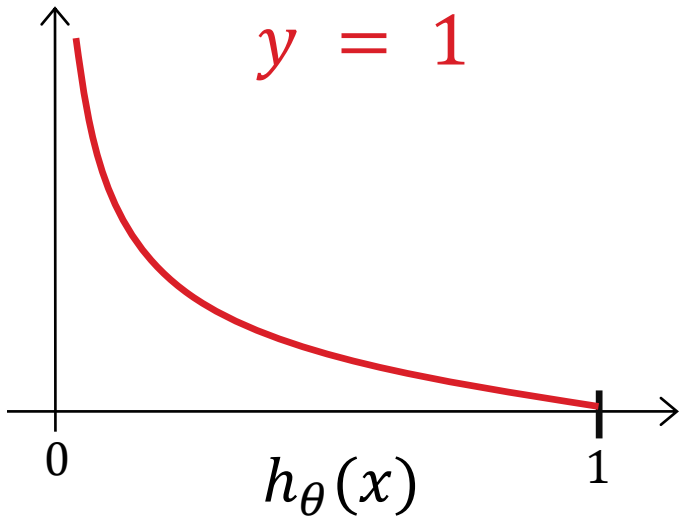


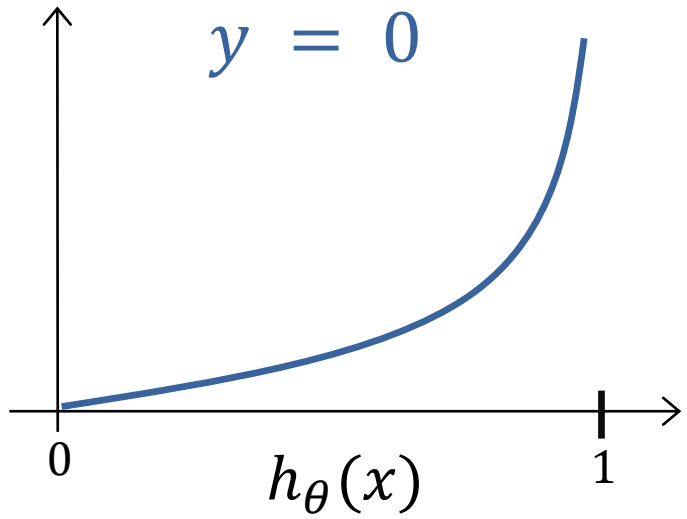


تابع هزینه در رگرسیون لجستیک

- تابع هزینه.

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)), & y = 1 \\ -\log(1 - h_\theta(x)), & y = 0 \end{cases}$$



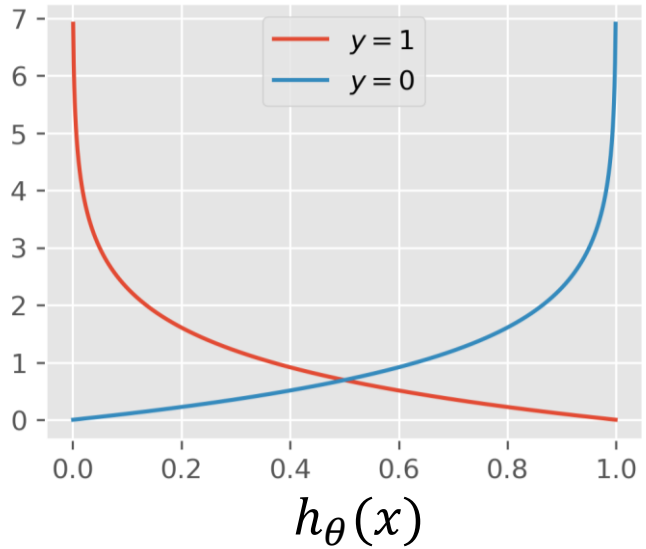


تابع هزینه در رگرسیون لجستیک

- ساده‌سازی تابع هزینه.

$$J(\theta) = \sum_{i=1}^{m} cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$$

$$cost\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = -y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$



تابع هزینه در رگرسیون لجستیک

- تابع هزینه.

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

- تعیین مقدار پارامترها.

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

- پیش‌بینی برای ورودی جدید x.

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

تابع هزینه در رگرسیون لجستیک

- تابع هزینه.

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

$$\nabla J(\theta) = X^T \left(h_\theta(X) - y\right) \qquad \nabla J(\theta) \in \mathbb{R}^{n+1}$$

$$H = X^T \operatorname{diag}\left(h_\theta(X)\left(1 - h_\theta(X)\right)\right) X \qquad H \in \mathbb{R}^{(n+1)\times(n+1)}$$

توجه. ماتریس هسین یک ماتریس مثبت معین است، بنابراین تابع هزینه یک تابع کوژ است.

الگوریتم گرادیان کاهشی

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

- الگوریتم گرادیان کاهشی. [شکل برداری]

```
repeat until convergence {
    θ := θ − α ∇J(θ)
}
```

$$\nabla J(\theta) = X^T (h_\theta(X) - y)$$

الگوریتم گرادیان کاهشی

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

- الگوریتم گرادیان کاهشی.

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \qquad (j = 0,1, \dots, n)$$

}

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

الگوریتم گرادیان کاهشی

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

- الگوریتم گرادیان کاهشی.

repeat until convergence {

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

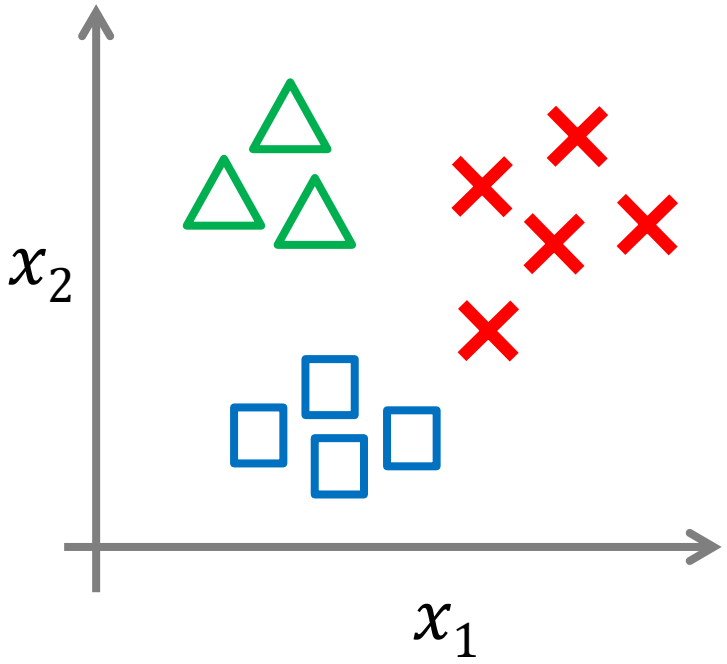
}

توجه. این الگوریتم درست مانند الگوریتم رگرسیون خطی است و تنها تفاوت در تابع فرضیه است.

دسته‌بندی با چند دسته

دسته‌بندی با چند دسته

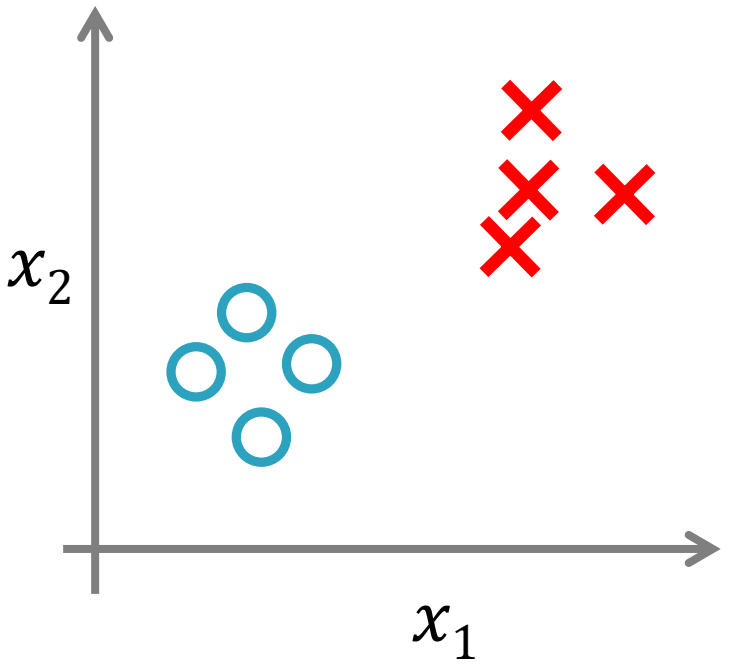
- ایمیل: کاری، خانوادگی، سرگرمی
- نمودارهای پزشکی: سالم، سرما خوردگی، آنفلوآنزا
- هوا: آفتابی، ابری، بارانی، برفی



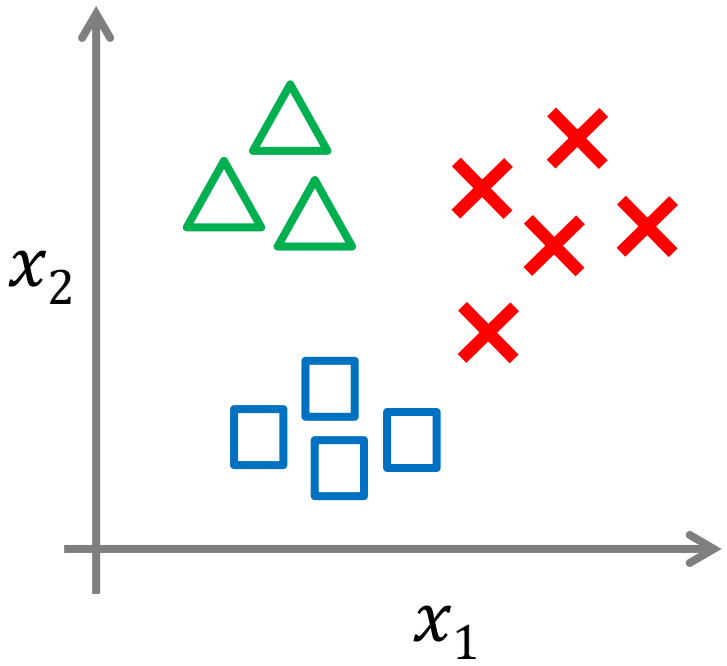
$$y \in \{1, 2, 3, \dots, k\}$$

دسته‌بندی با چند دسته

دسته‌بندی دودویی

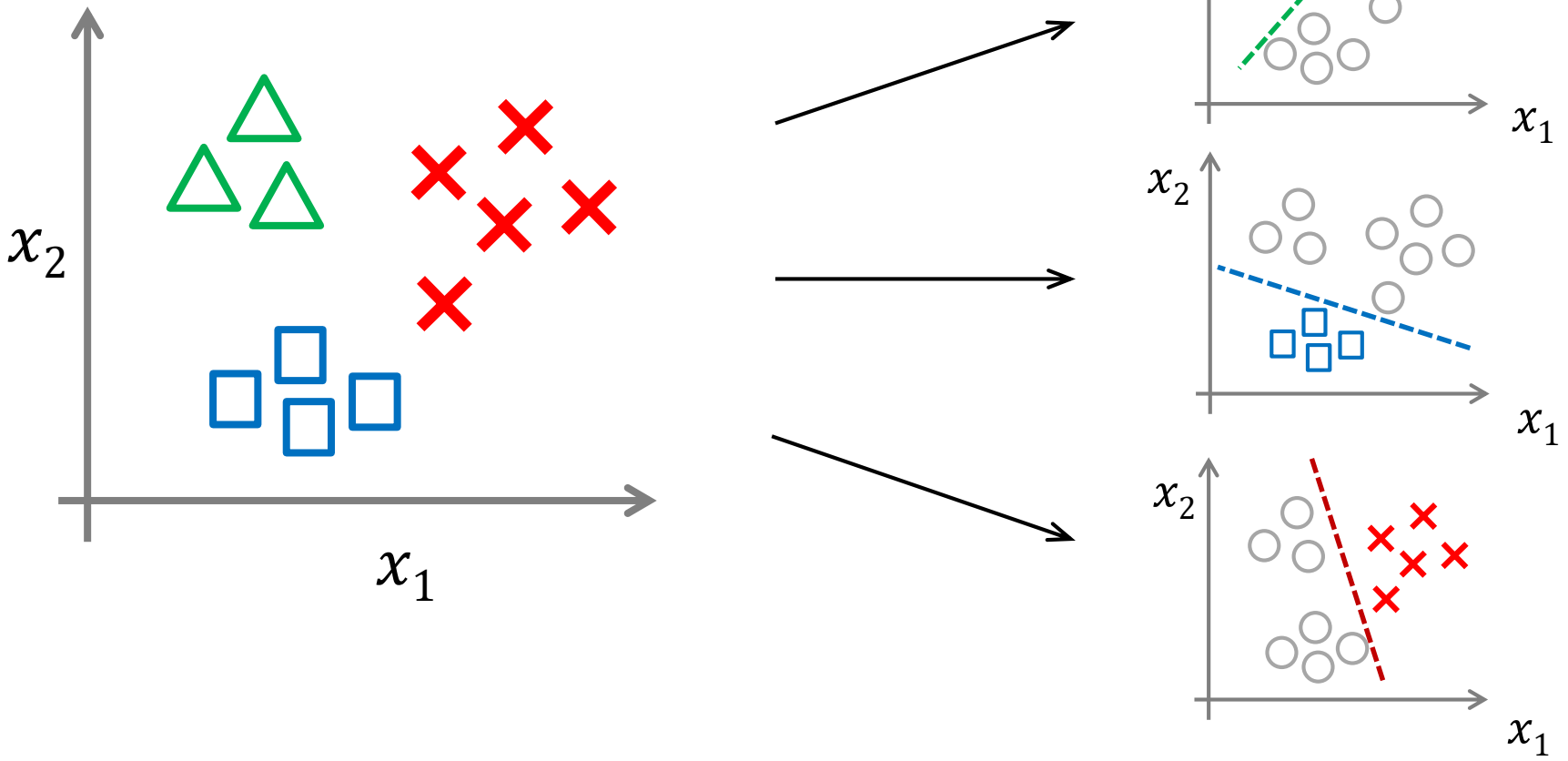


دسته‌بندی چنددسته‌ای



دسته‌بندی با چند دسته: یکی در برابر همه

$$h_\theta^{(i)}(x) = P(y = i|x;\theta) \qquad (i = 1,2,3)$$



دسته‌بندی با چند دسته: یکی در برابر همه

- یکی در برابر همه. به ازای هر دسته i، دسته‌بند رگرسیون لجیستیک $h_\theta^{(i)}(x)$ را به منظور تخمین احتمال تعلق ورودی x به دسته i آموزش بده.

- پیش‌بینی. به منظور دسته‌بندی ورودی جدید x، دسته i را انتخاب کن به گونه‌ای که:

$$y = \arg\max_i h_\theta^{(i)}(x)$$

$$h_\theta^{(1)}(x) = 0.25$$

$$h_\theta^{(2)}(x) = 0.70 \quad \Rightarrow \quad y = 2$$

$$h_\theta^{(3)}(x) = 0.45$$

روش‌های بهینه‌سازی پیشرفته

روش‌های بهینه‌سازی پیشرفته

- هدف. یافتن مقدار θ به منظور کمینه‌سازی تابع هزینه.

$$\min_{\theta} J(\theta)$$

- فرض. برنامه‌ای داریم که با داشتن مقادیر θ، می‌تواند مقادیر زیر را محاسبه کند:

$$J(\theta) \qquad \frac{\partial}{\partial \theta_j} J(\theta)$$

```
repeat until convergence {
```

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \qquad (j = 0,1, \dots, n)$$

```
}
```

گرادیان کاهشی

روش‌های بهینه‌سازی پیشرفته

- فرض. برنامه‌ای داریم که با داشتن مقادیر θ، می‌تواند مقادیر زیر را محاسبه کند:

$$J(\theta) \qquad \frac{\partial}{\partial \theta_j} J(\theta)$$

- الگوریتم‌های بهینه‌سازی پیشرفته.
  - گرادیان مزدوج
  - BFGS
  - L-BFGS

- مزایا. این روش‌ها نیاز به انتخاب نرخ یادگیری ندارند و معمولاً نسبت به الگوریتم گرادیان کاهشی زودتر همگرا می‌شوند.

روش‌های بهینه‌سازی پیشرفته

- مثال.

$$J(\theta) = (\theta_0 - 5)^2 + (\theta_1 - 5)^2$$

$$\frac{\partial}{\partial \theta_0} J(\theta) = 2(\theta_0 - 5)$$

$$\frac{\partial}{\partial \theta_1} J(\theta) = 2(\theta_1 - 5)$$

```
def J(theta):
    return (theta[0] - 5) ** 2 + (theta[1] - 5) ** 2
```

```
def grads(theta):
    return np.array([2 * (theta[0] - 5), 2 * (theta[1] - 5)])
```

روش‌های بهینه‌سازی پیشرفته

- مثال.

$$J(\theta) = (\theta_0 - 5)^2 + (\theta_1 - 5)^2$$

$$\frac{\partial}{\partial \theta_0} J(\theta) = 2(\theta_0 - 5)$$

$$\frac{\partial}{\partial \theta_1} J(\theta) = 2(\theta_1 - 5)$$

```
from scipy.optimize import minimize

minimize(J, x0=[0, 0], method='CG', jac=gra
```

```
     fun: 2.477476329894505e-18
     jac: array([1.71271335e-08, 1.71271335e-08])
 message: 'Optimization terminated successfully.'
    nfev: 20
     nit: 2
    njev: 5
  status: 0
 success: True
       x: array([5., 5.])
```

روش‌های بهینه‌سازی پیشرفته

- مثال.

$$J(\theta) = (\theta_0 - 5)^2 + (\theta_1 - 5)^2$$

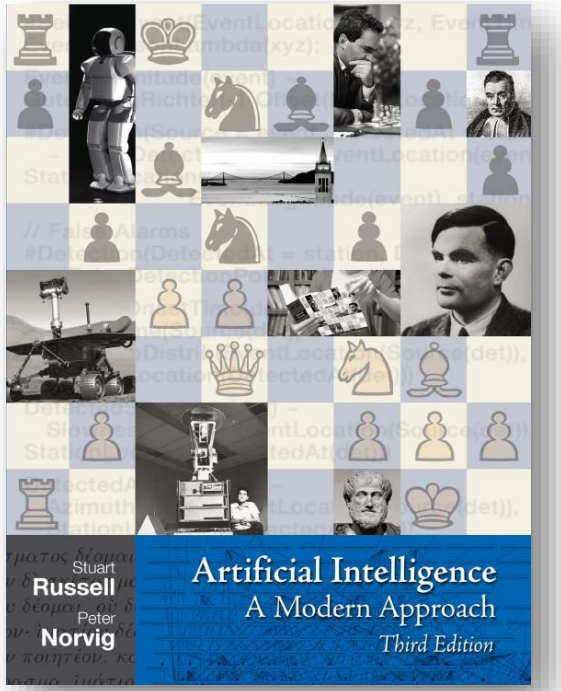
$$\frac{\partial}{\partial \theta_0} J(\theta) = 2(\theta_0 - 5)$$

$$\frac{\partial}{\partial \theta_1} J(\theta) = 2(\theta_1 - 5)$$

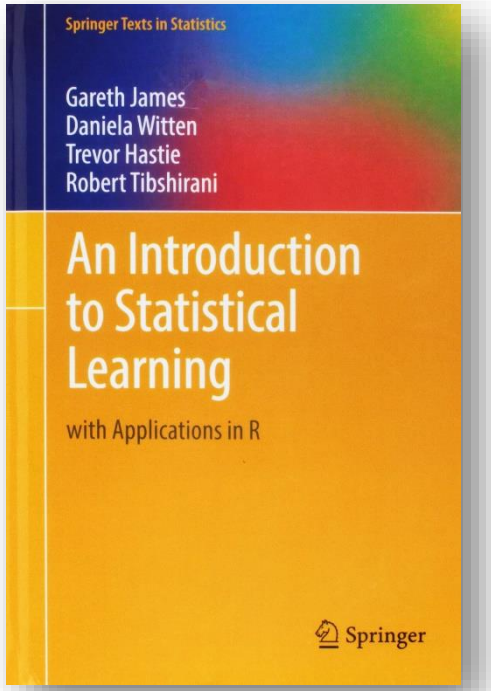
```
from scipy.optimize import minimize

minimize(J, x0=[0, 0], method='BFGS', jac=
```

```
     fun: 3.5538794606501983e-16
hess_inv: array([[ 0.75, -0.25],
       [-0.25,  0.75]])
     jac: array([-1.17592194e-08, -1.17592194e-08])
 message: 'Optimization terminated successfully.'
    nfev: 16
     nit: 3
    njev: 4
  status: 0
 success: True
       x: array([4.99999999, 4.99999999])
```

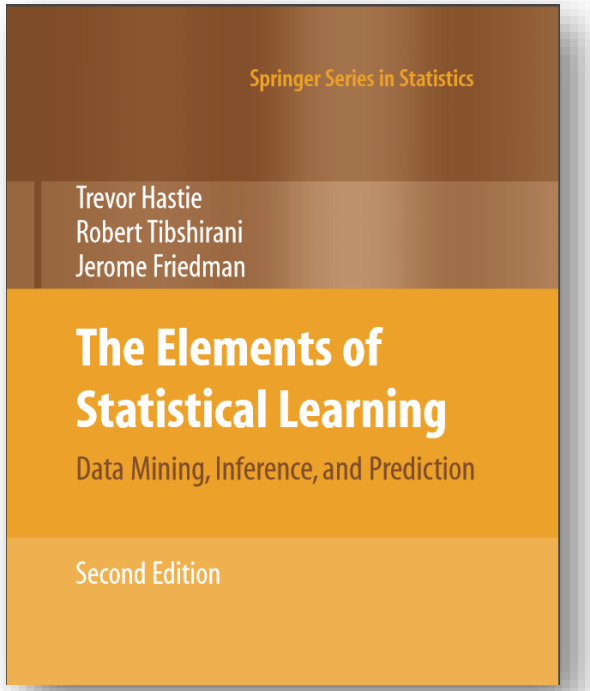
مطالعه بیشتر



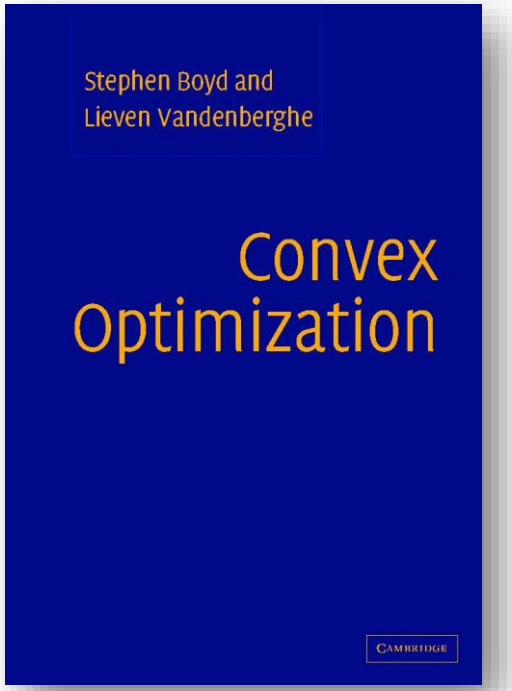
[صفحات ۷۲۵ تا ۷۲۷]



[صفحات ۱۳۰ تا ۱۳۷]



[صفحات ۱۱۹ تا ۱۲۸]



[بهینه‌سازی محدب]