فصل چهارم : کلاس بندی

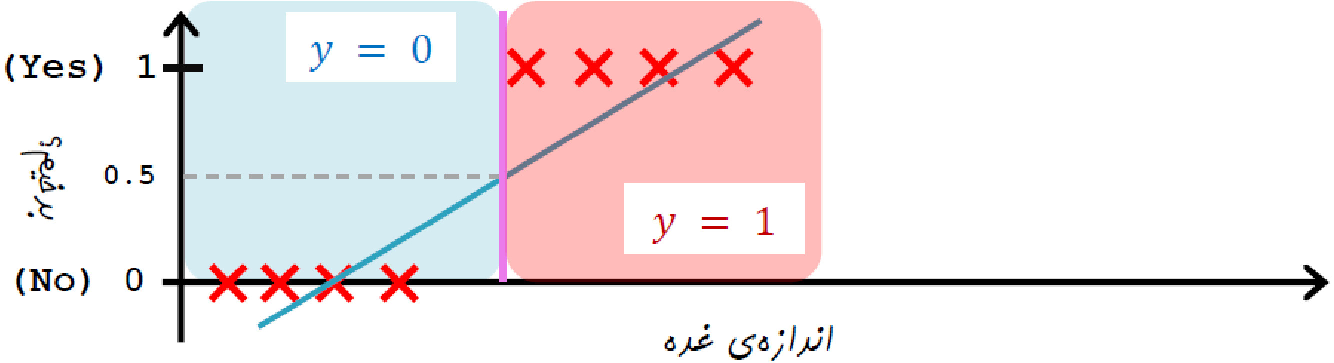
- ایمیل: هرزنامه (بله / خیر؟)
- تراکنش برخط: کلاهبرداری (بله / خیر؟)
- غده سرطانی: خوش‌خیم / بدخیم ؟
- در این مثال‌ها، متغیری که می‌خواهیم مقدارش را پیش‌بینی کنیم دارای دو مقدار است:

صفر: «کلاس منفی» (مانند غده‌ی خوش‌خیم)
یک: «کلاس مثبت» (مانند غده‌ی بدخیم)

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

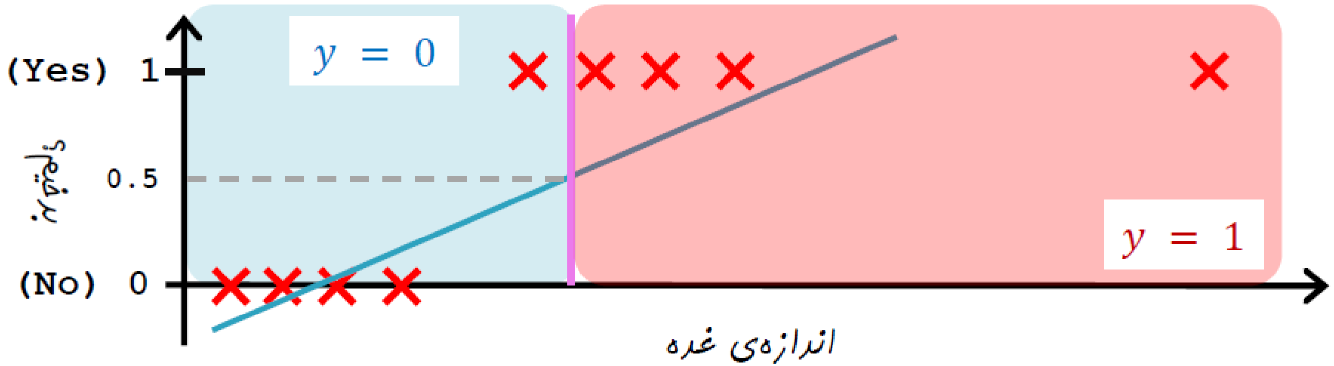
- کلاس‌بندی. پیش‌بینی یک متغیر با مقادیر گسسته.
  - کلاس‌بندی دودویی
  - کلاس‌بندی چندکلاسی

فصل چهارم : کلاس بندی



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

فصل چهارم : کلاس بندی



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

فصل چهارم : کلاس بندی

- در کلاس‌بندی دودویی داریم:

$y = 0$ یا $y = 1$

- اما در رگرسیون ممکن است:

$h_\theta(x) < 0$ یا $h_\theta(x) > 1$

- رگرسیون لجستیکی (کلاس‌بندی).

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

بازنمایی فرضیه در رگرسیون لجستیکی

فصل چهارم : بازنمایی فرضیه

- هدف.

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

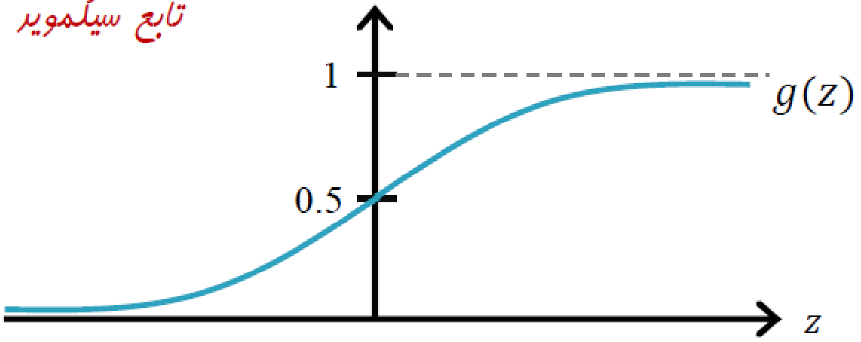
- فرضیه.

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

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

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

تابع سیگموید



فصل چهارم : فرضیه

- تفسیر خروجی فرضیه.

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

- مثال. اگر داشته باشیم:

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ tumorSize \end{bmatrix}, \qquad h_\theta(x) = 0.7$$

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

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

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

فصل چهارم : 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.

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

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

فصل چهارم : تخمین بیشترین درست نمایی

- لگاریتم تابع درست‌نمایی.

$$
\begin{aligned}
l(\theta) = \log L(\theta) &= \log \prod_{i=1}^{m} h_\theta(x^{(i)})^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}} \\
&= \sum_{i=1}^{m} \log\left(h_\theta(x^{(i)})^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}\right) \\
&= \sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right)
\end{aligned}
$$

- تابع هزینه.

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

مرز تصمیم‌گیری

فصل چهارم : مرز تصمیم گیری

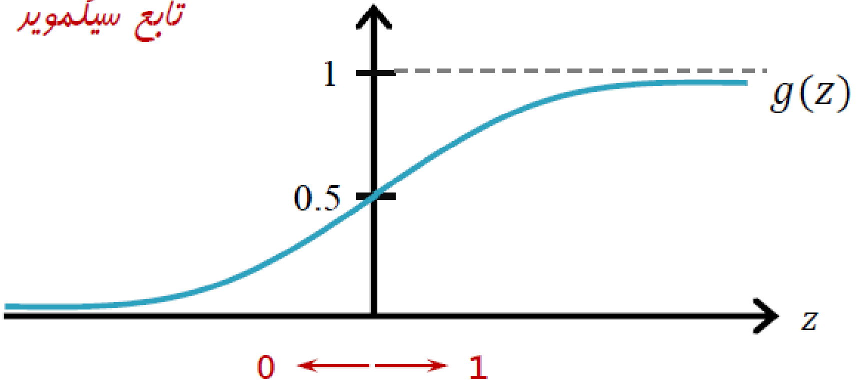
- رگرسیون لجستیک.

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

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

- قرار دادن یک آستانه بر روی خروجی کلاس‌بند:

تابع سیگموید

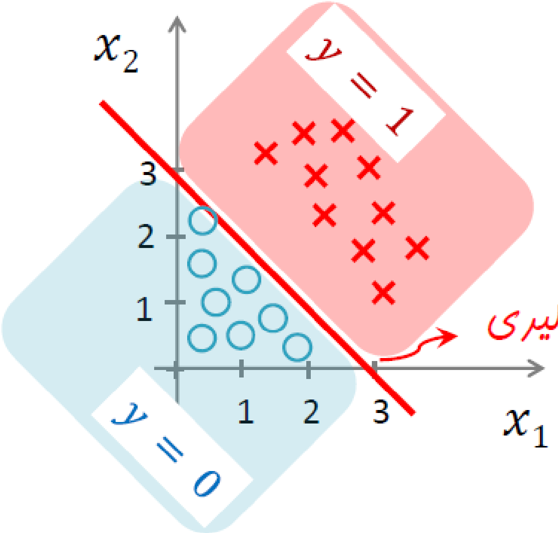


$$y = 1\text{:}\ \ h_\theta(x) \geq 0.5 \Rightarrow \theta^T x \geq 0$$

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

فصل چهارم : مرز تصمیم گیری

- مرز تصمیم‌گیری.



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

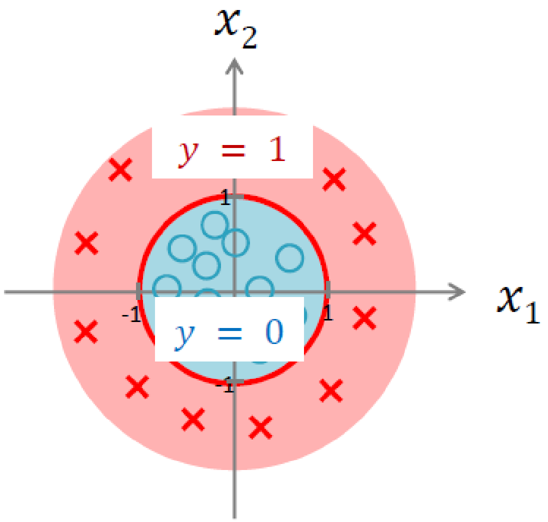
θ_0: -3, θ_1: 1, θ_2: 1

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

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

فصل چهارم : مرز تصمیم گیری غیر خطی

- 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 0 0 1 1

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

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

تابع هزینه

فصل چهارم : رگرسیون لجستیکی

- مجموعه‌ی آموزشی.

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

- نمونه‌ی آموزشی.

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

- فرضیه.

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

- س. مقادیر پارامترهای θ را چگونه انتخاب کنیم؟

- تابع هزینه.

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

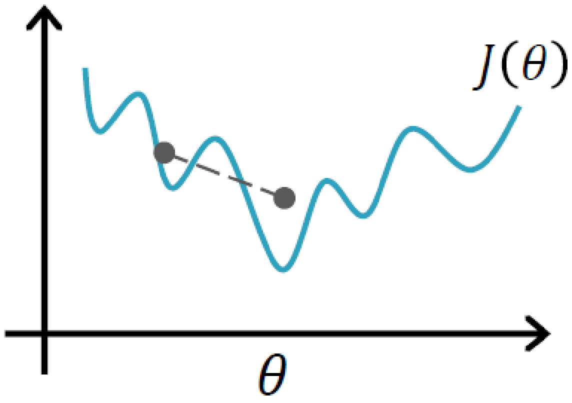
$$cost\left(h_\theta(x^{(i)}), y^{(i)}\right) = \frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

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

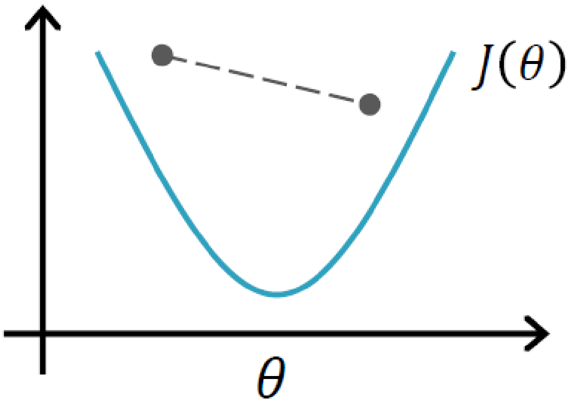
فصل چهارم : تابع هزینه

- توابع کوژ و ناکوژ.

تابع ناکوژ



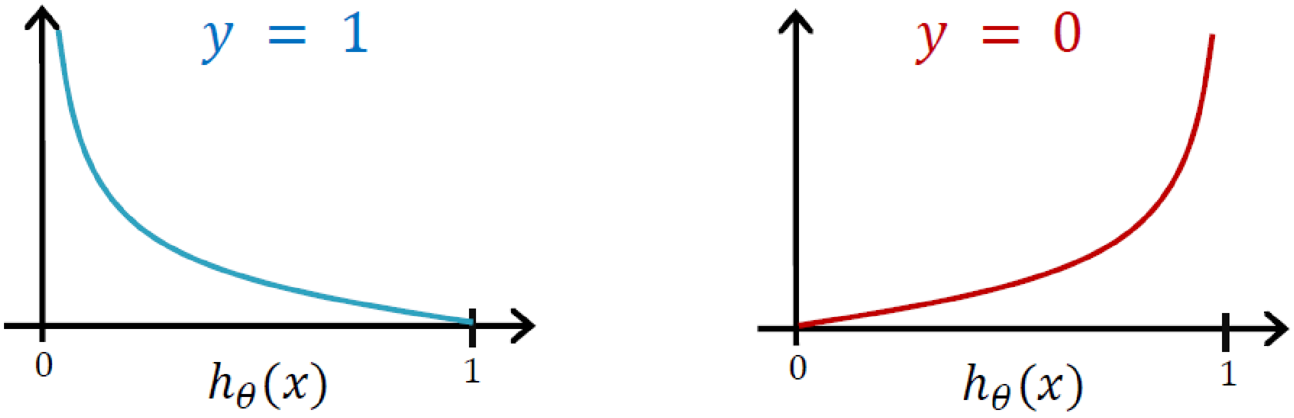
تابع کوژ



فصل چهارم : تابع هزینه در رگرسیون لجستیکی

- تابع هزینه.

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

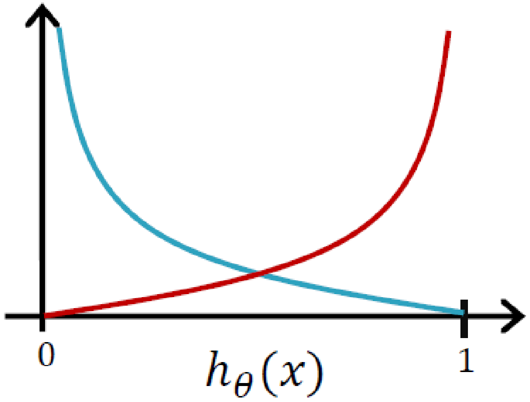


فصل چهارم : تابع هزینه در رگرسیون لجستیکی

- سادهسازی تابع هزینه.

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

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



فصل چهارم : تابع هزینه در رگرسیون لجستیکی

- 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.

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

فصل چهارم : تابع هزینه در رگرسیون لجستیکی

- تابع هزینه.

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

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

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

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

فصل چهارم : الگوریتم گرادیان کاهشی

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

- الگوریتم گرادیان کاهشی. [شکل برداری]

```
repeat until convergence {
```

$$\theta := \theta - \alpha \nabla J(\theta)$$

```
}
```

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

فصل چهارم : الگوریتم گرادیان کاهشی

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

- الگوریتم گرادیان کاهشی.

```
repeat until convergence {
```

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

```
}
```

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

فصل چهارم : الگوریتم گرادیان کاهشی

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

- الگوریتم گرادیان کاهشی.

repeat until convergence {

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

}

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

روش‌های بهینه‌سازی پیشرفته

فصل چهارم : روش های بهین هسازی 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:

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

```
repeat until convergence {
```

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

```
}
```

گرادیان کاهشی

فصل چهارم : روش های بهین هسازی پیشرفته

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

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

- 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.

- مثال.

$$J(\theta) = (\theta_1 - 5)^2 + (\theta_2 - 5)^2 \Rightarrow \begin{cases} \dfrac{\partial}{\partial \theta_1} J(\theta) = 2(\theta_1 - 5) \\ \dfrac{\partial}{\partial \theta_2} J(\theta) = 2(\theta_2 - 5) \end{cases}$$

```
function [jVal, gradient] = costFunction(theta)
      jVal = (theta(1) - 5) ^ 2 + (theta(2) - 5) ^ 2;
      gradient = zeros(2, 1);
      gradient(1) = 2 * (theta(1) - 5);
      gradient(2) = 2 * (theta(2) - 5);
```

فصل چهارم : روش های بهینه سازی پیشرفته

- مثال.

$$J(\theta) = (\theta_1 - 5)^2 + (\theta_2 - 5)^2 \quad \Rightarrow \quad \begin{aligned} \frac{\partial}{\partial \theta_1} J(\theta) &= 2(\theta_1 - 5) \\ \frac{\partial}{\partial \theta_2} J(\theta) &= 2(\theta_2 - 5) \end{aligned}$$

```
options = optimset('GradObj', 'on', 'MaxIter', 100);
initialTheta = zeros(2,1);
[optTheta, functionVal, exitFlag] ...
     = fminunc(@costFunction, initialTheta, options);
```

فصل چهارم : تابع هزینه

- تابع هزینه با n ویژگی.

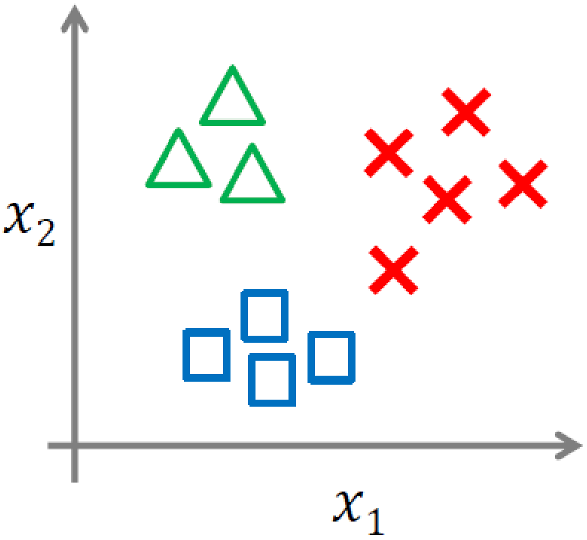
$$\theta^T = [\theta_0 \quad \theta_1 \quad \cdots \quad \theta_n]^T$$

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

    jVal = [code to compute J(θ) ];

    gradient = zeros(n+1, 1);

    gradient(1)   = [code to compute ∂/∂θ_0 J(θ) ];

    gradient(2)   = [code to compute ∂/∂θ_1 J(θ) ];

    ...

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

کلاس‌بندی با چند کلاس

فصل چهارم : کلاس بندی با چند کلاس

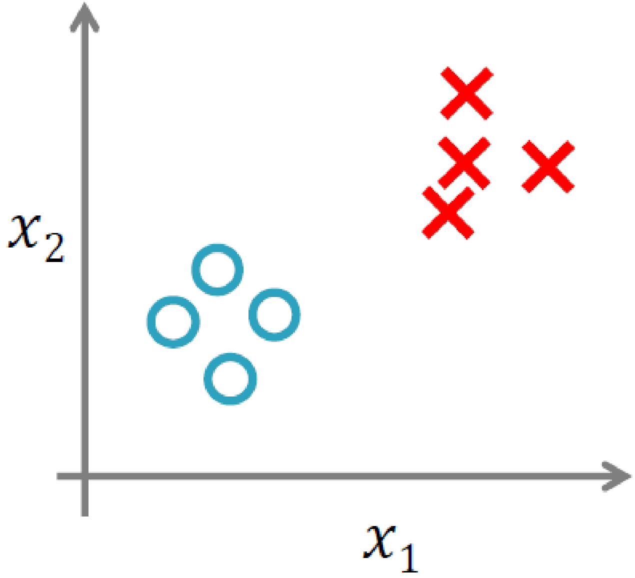
- ایمیل: کاری، خانوادگی، سرگرمی
- نمودارهای پزشکی: سالم، سرما خوردگی، آنفلوآنزا
- هوا: آفتابی، ابری، بارانی، برفی



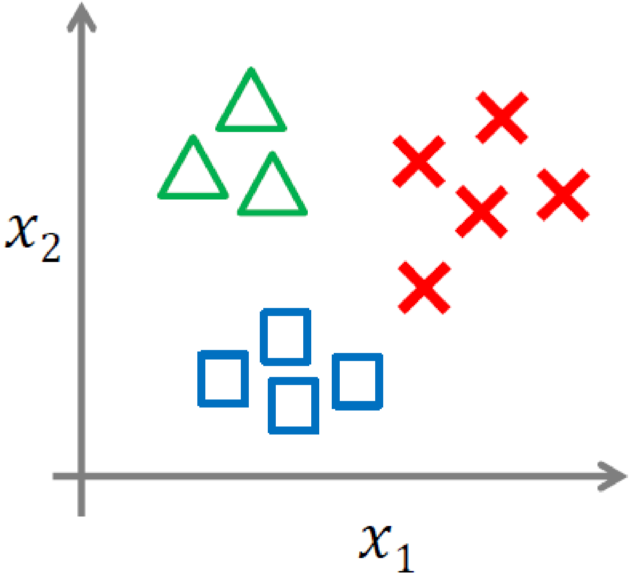
$$y \in \{1, 2, 3, \ldots, k\}$$

فصل چهارم : کلاس بندی با چند کلاس

کلاس‌بندی دودویی

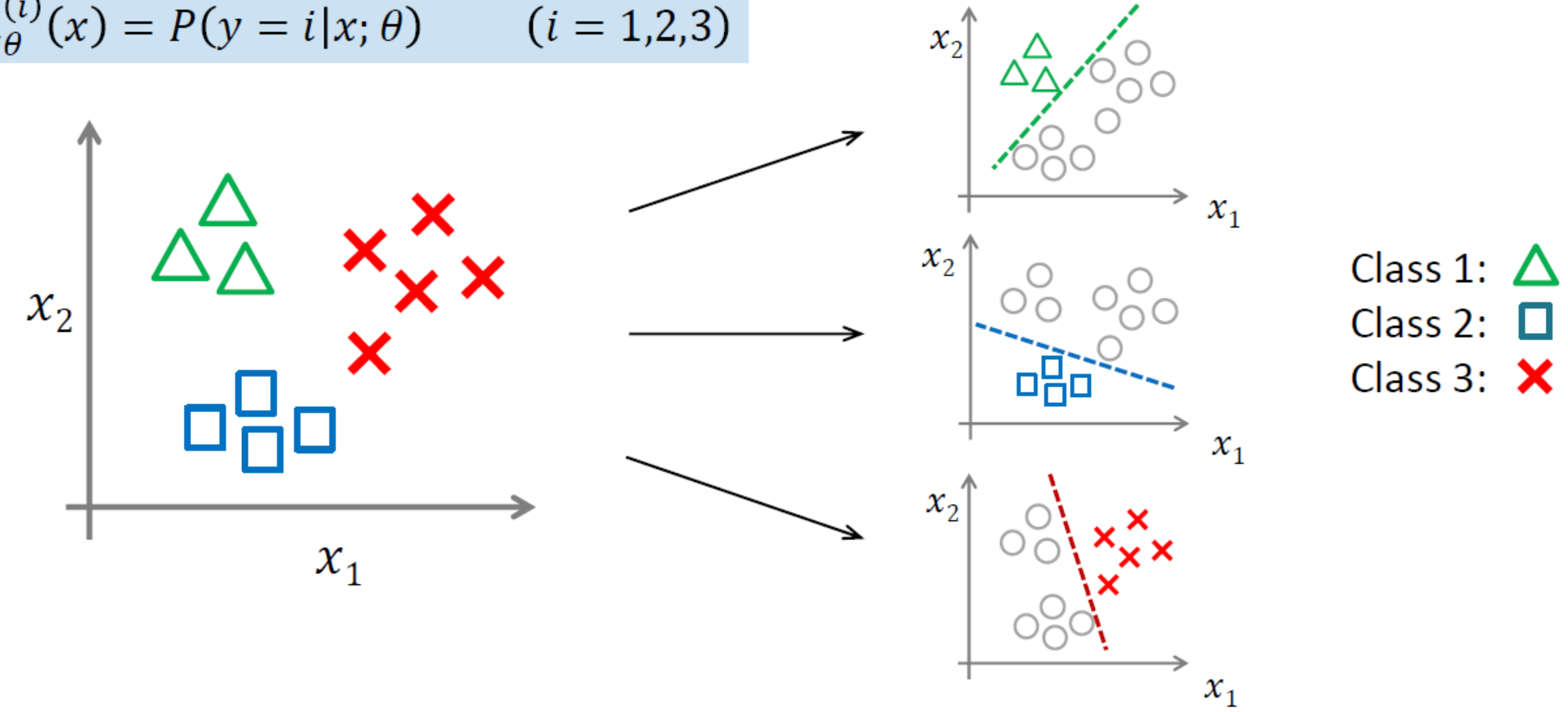


کلاس‌بندی چند کلاسی



فصل چهارم : کلاس بندی با چند کلاس(یکی در برابر همه)

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



فصل چهارم : کلاس بندی با چند کلاس(یکی در برابر همه)

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

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

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

$$h_\theta^{(1)}(x) = 0.25$$
$$h_\theta^{(2)}(x) = 0.70 \rightarrow y = 2$$
$$h_\theta^{(3)}(x) = 0.45$$

با تشکر از توجه شما