

کلاس بندی: رگرسیون لجیستیکی

سید ناصر رضوی n.razavi@tabrizu.ac.ir

۱۳۹۵

کلاس بندی

۲

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

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

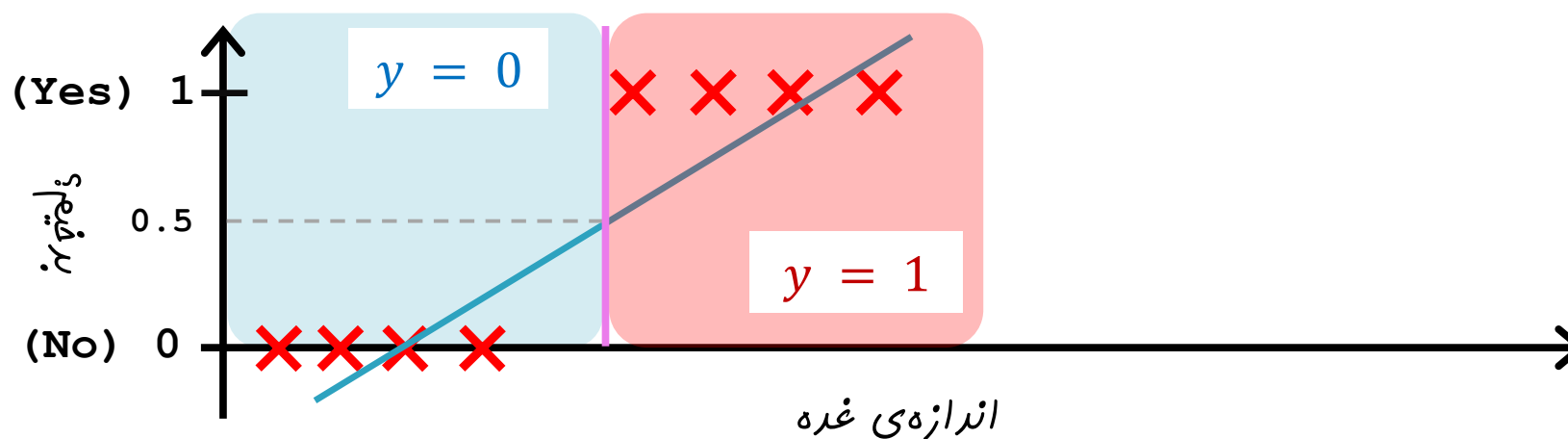
صفر: «کلاس منفی» (مانند غده‌ی فوش‌فیم)

یک: «کلاس مثبت» (مانند غده‌ی بدفیم)

- کلاس‌بندی. پیش‌بینی یک متغیر با مقادیر گسسته.

□ کلاس‌بندی دودویی

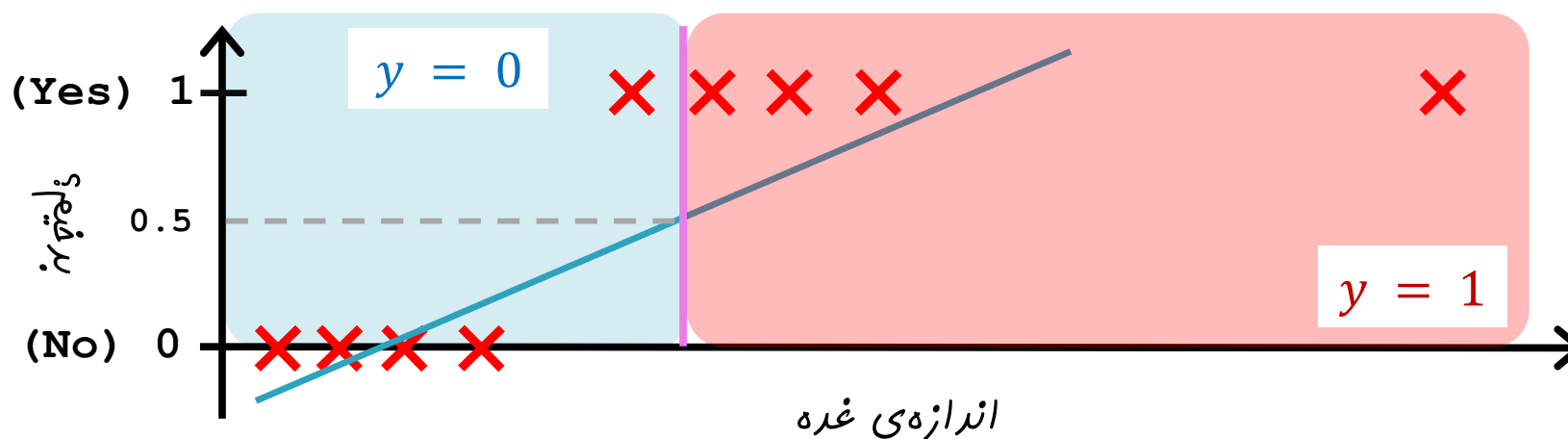
□ کلاس‌بندی چندکلاسی



□ 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 $y = 0$



□ 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 $y = 0$

□ در کلاس بندی دودویی داریم:

$$y = 0 \text{ یا } y = 1$$

□ اما در رگرسیون ممکن است:

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

□ رگرسیون لجستیکی (کلاس بندی).

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

بازنمایی فرضیه در رگرسیون لجستیک

بازنمایی فرضیه

۷

□ هدف.

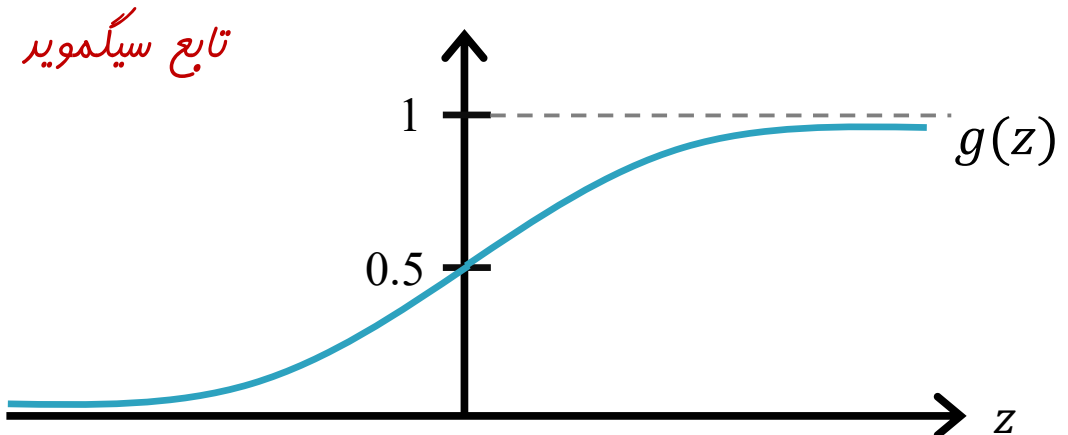
$$0 \leq h_{\theta}(x) \leq 1$$

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

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

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

□ فرضیه.



□ تفسیر خروجی فرضیه.

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

□ مثال. اگر داشته باشیم:

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

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

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

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

□ تفسیر احتمالاتی فرضیه.

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

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

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

□ تابع درست‌نمایی.

$$\begin{aligned} L(\theta) = p(Y|X; \theta) &= \prod_{i=1}^m p(y^{(i)}|x^{(i)}; \theta) \\ &= \prod_{i=1}^m h_{\theta}(x^{(i)})^{y^{(i)}} (1 - h_{\theta}(x^{(i)}))^{1-y^{(i)}} \end{aligned}$$

تخمین بیشترین درست‌نمایی

۱۰

□ لگاریتم تابع درست‌نمایی.

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

□ تابع هزینه.

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

مرز تصمیم‌گیری

مرز تصمیم‌گیری

۱۲

□ رگرسیون لجستیک.

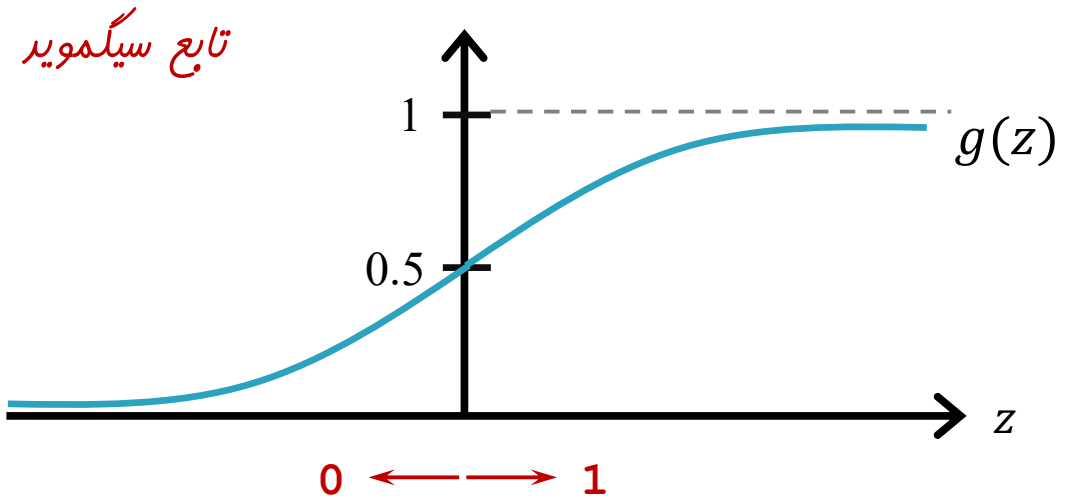
$$h_{\theta}(x) = g(\theta^T x)$$

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

□ قرار دادن یک آستانه بر روی خروجی کلاس‌بند:

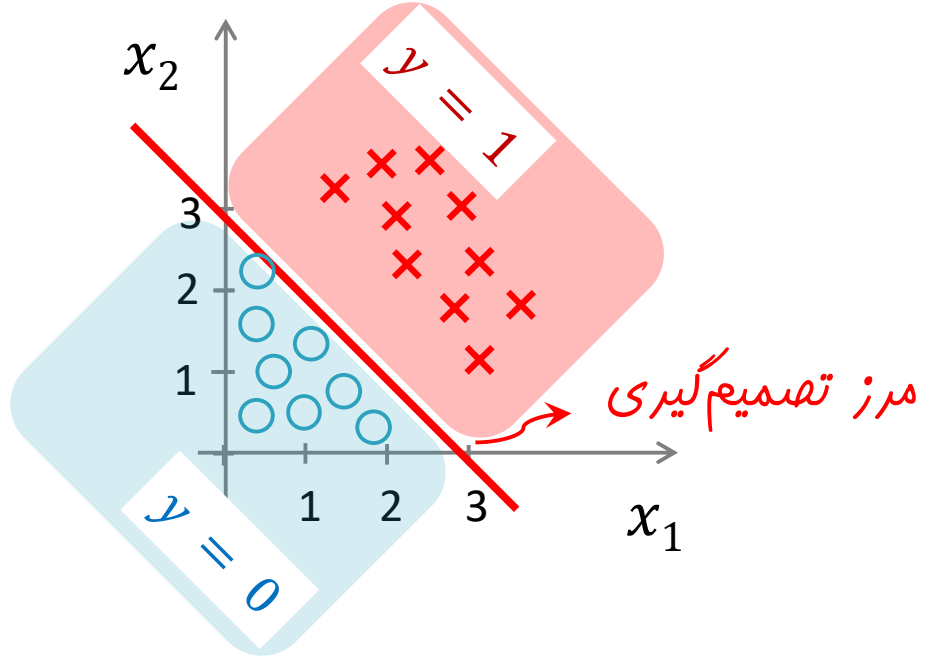
$$y = 1: h_{\theta}(x) \geq 0.5 \Rightarrow \theta^T x \geq 0$$

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



مرز تصمیم‌گیری

۱۳



□ 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 $-3 + x_1 + x_2 \geq 0$

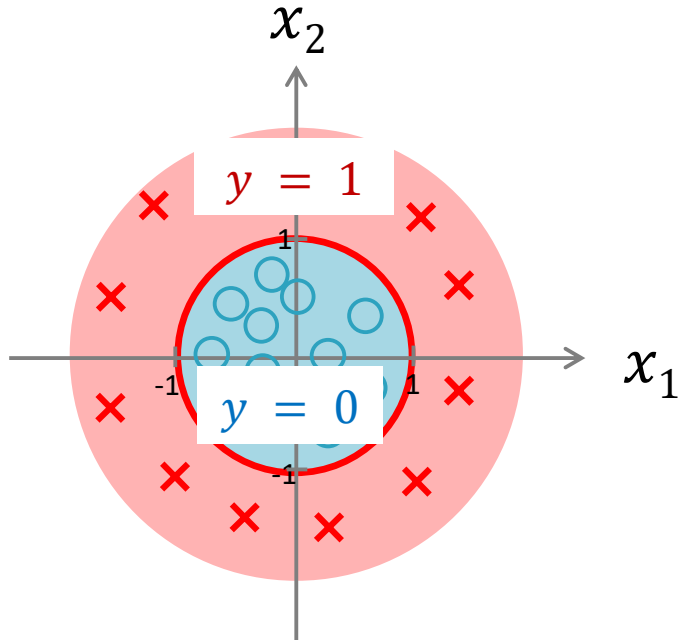
$$\square x_1 + x_2 \geq 3 \Rightarrow y = 1$$

$$\square x_1 + x_2 < 3 \Rightarrow y = 0$$

مرز تصمیم‌گیری غیر خطی

۱۴

□ مرز تصمیم‌گیری.



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

$\downarrow$ $\downarrow$ $\downarrow$ $\downarrow$ $\downarrow$
-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

□ مجموعه‌ی آموزشی.

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

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

□ نمونه‌ی آموزشی.

□ فرضیه.

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

□ س. مقادیر پارامترهای θ را چگونه انتخاب کنیم؟

□ تابع هزینه.

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

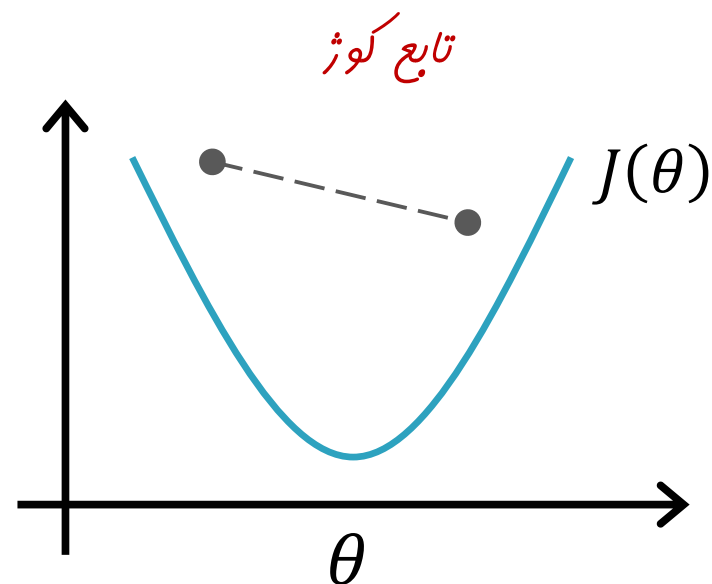
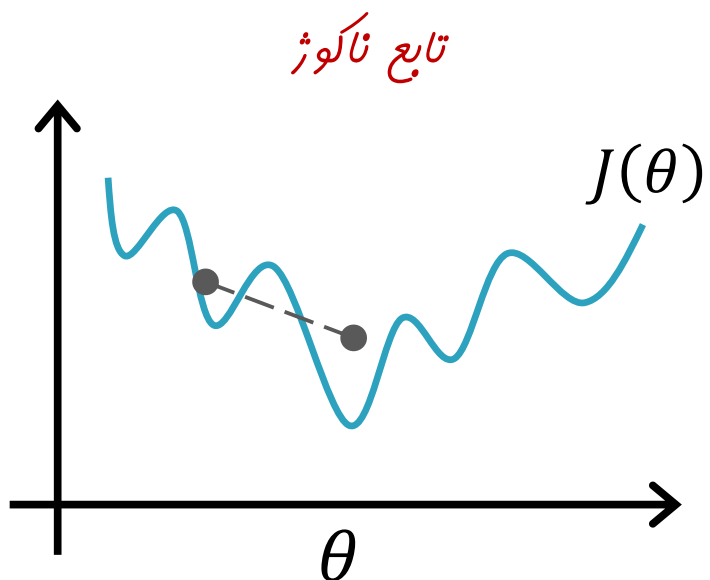
$$\text{cost}(h_{\theta}(x^{(i)}), y^{(i)}) = \frac{1}{2} (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

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

تابع هزینه

۱۸

□ توابع کوژ و ناکوژ.

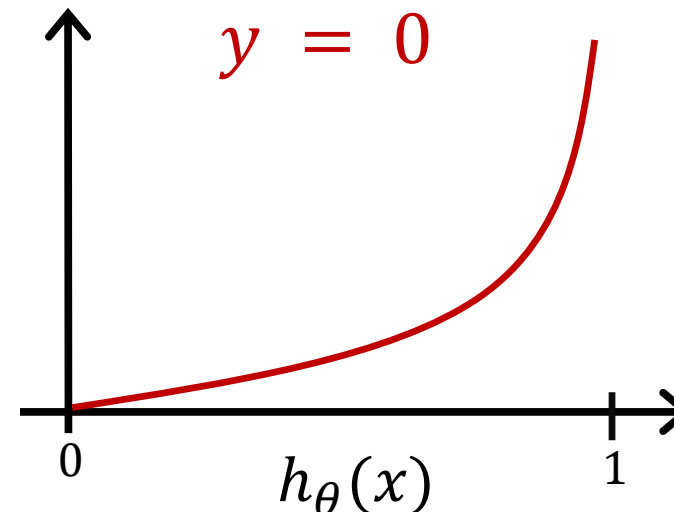
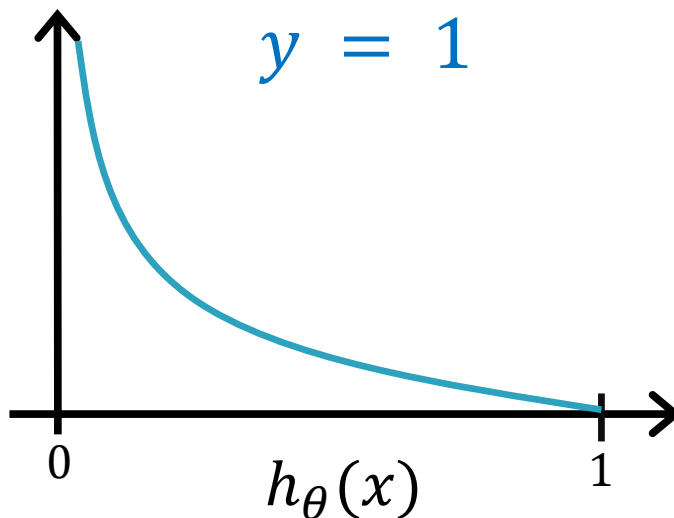


تابع هزینه در رگرسیون لجستیکی

۱۹

□ تابع هزینه.

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



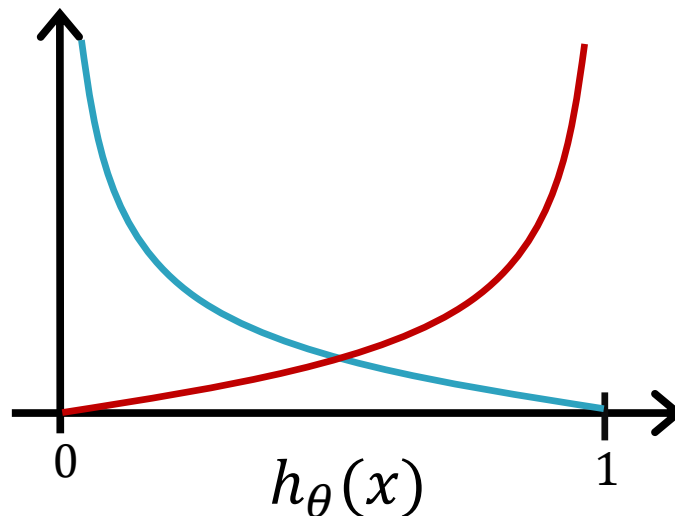
تابع هزینه در رگرسیون لجستیکی

۲۰

□ ساده‌سازی تابع هزینه.

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

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



تابع هزینه در رگرسیون لجستیکی

۲۱

□ تابع هزینه.

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

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

□ تعیین مقدار پارامترها.

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

□ پیش‌بینی برای ورودی جدید x .

تابع هزینه در رگرسیون لجستیکی

۲۲

□ تابع هزینه.

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

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

$$\nabla J(\theta) \in \mathbb{R}^{n+1}$$

$$H = X^T \text{diag}(h_{\theta}(X)(1 - h_{\theta}(X))) X$$

$$H \in \mathbb{R}^{(n+1) \times (n+1)}$$

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

الگوریتم گرادیان کاهشی

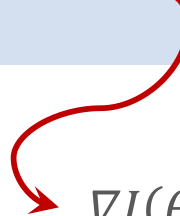
۲۳

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

□ الگوریتم گرادیان کاهشی. [شکل برداری]

```
repeat until convergence {
```

```
     $\theta := \theta - \alpha \underbrace{\nabla J(\theta)}$   
}
```


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

الگوریتم گرادیان کاهشی

۲۴

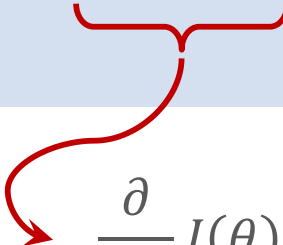
$$J(\theta) = \sum_{i=1}^m \left[-y^{(i)} \log(h_{\theta}(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right]$$

□ الگوریتم گرادیان کاهشی.

repeat until convergence {

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

}


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

الگوریتم گرادیان کاهشی

۲۵

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

□ الگوریتم گرادیان کاهشی.

repeat until convergence {

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

}

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

روش‌های بهینه‌سازی پیشرفته

روش‌های بهینه‌سازی پیشرفته

۲۷

□ 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:

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

repeat until convergence {

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

}

گرادیان کاهشی

روش‌های بهینه‌سازی پیشرفته

۲۸

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

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

□ 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

۲۹

□ مثال.

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

```
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);
```

روش‌های بهینه‌سازی پیشرفته

۳۰

□ مثال.

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

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

تابع هزینه

۳۱

□ 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  $J(\theta)$  ];
    gradient = zeros(n+1, 1);

    gradient(1) = [code to compute  $\frac{\partial}{\partial \theta_0} J(\theta)$  ];
    gradient(2) = [code to compute  $\frac{\partial}{\partial \theta_1} J(\theta)$  ];
    ...
    gradient(n+1) = [code to compute  $\frac{\partial}{\partial \theta_n} J(\theta)$  ];
```

کلاس بندی با چند کلاس

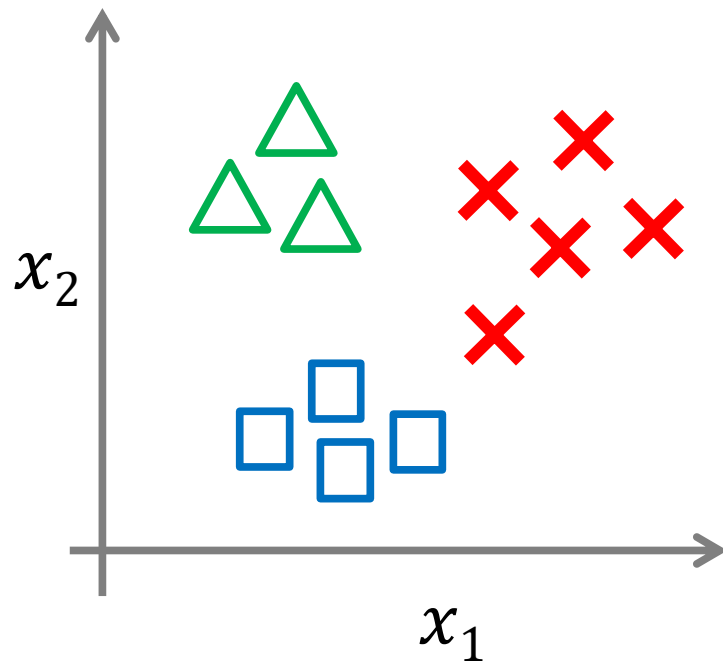
کلاس‌بندی با چند کلاس

۳۳

□ ایمیل: کاری، خانوادگی، سرگرمی

□ نمودارهای پزشکی: سالم، سرما خوردگی، آنفلوآنزا

□ هوا: آفتابی، ابری، بارانی، برفی

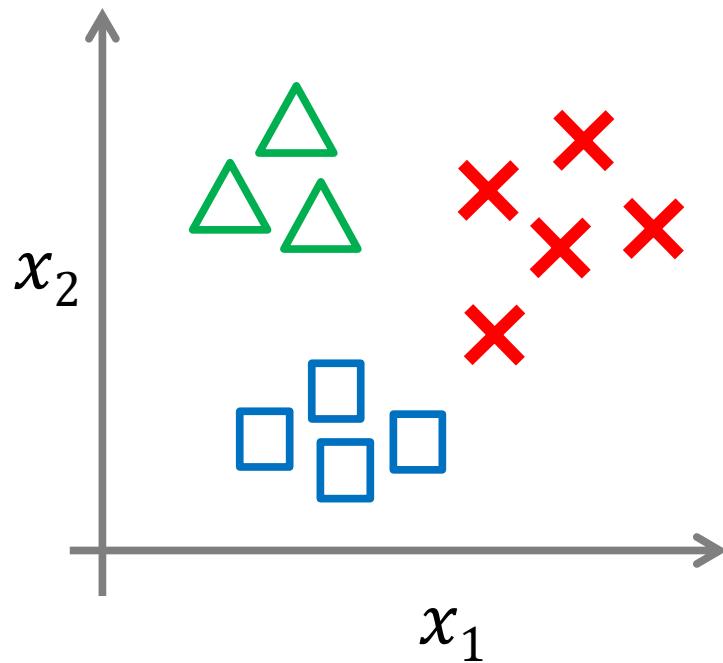


$$y \in \{1, 2, 3, \dots, k\}$$

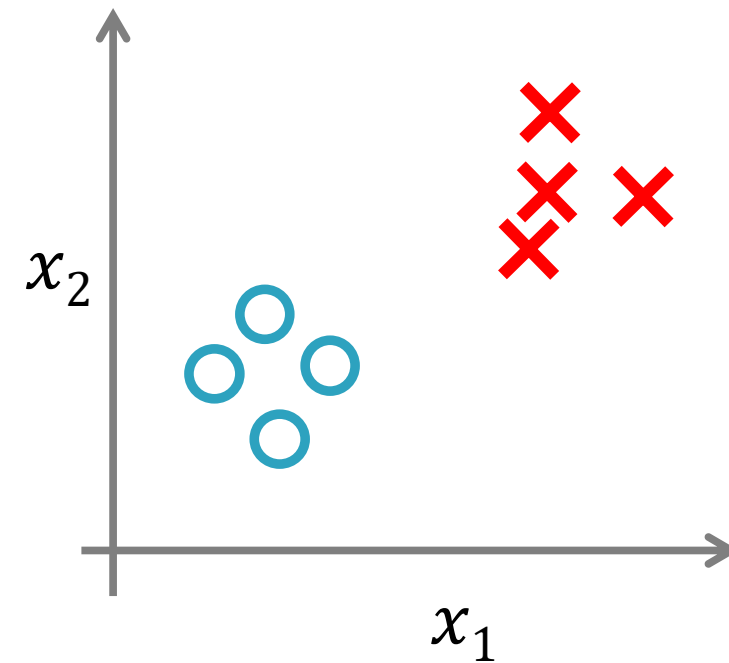
کلاس‌بندی با چند کلاس

۳۴

کلاس‌بندی چند کلاسی



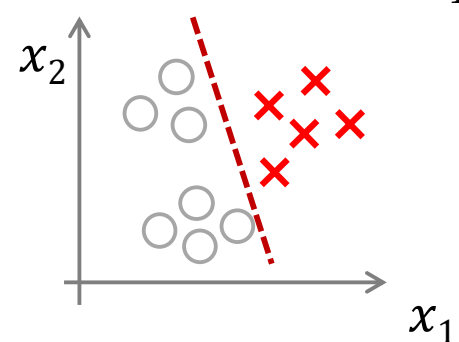
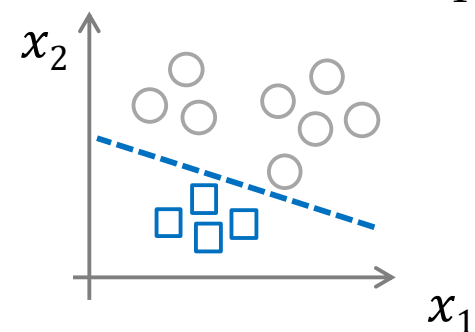
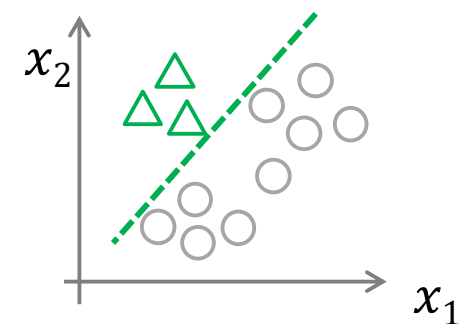
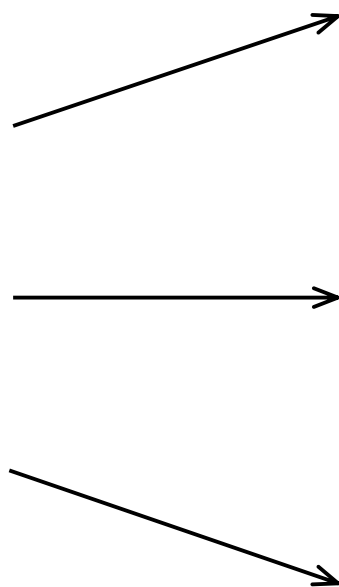
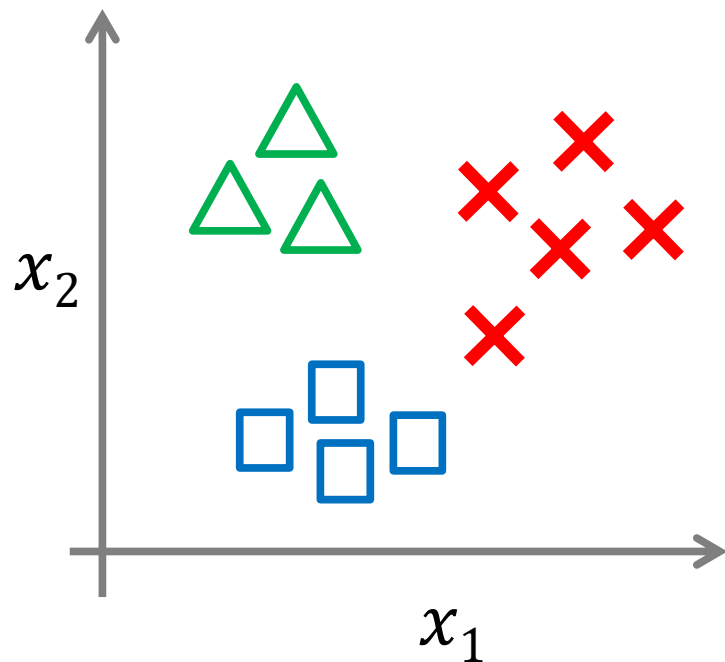
کلاس‌بندی دودویی



کلاس‌بندی با چند کلاس: یکی در برابر همه

۳۵

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



Class 1: 
Class 2: 
Class 3: 

کلاس‌بندی با چند کلاس: یکی در برابر همه

۳۶

□ 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:

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

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

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

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

$$\Rightarrow y = 2$$