REGRESSÃO LOGÍSTICA
CLASSIFICAÇÃO

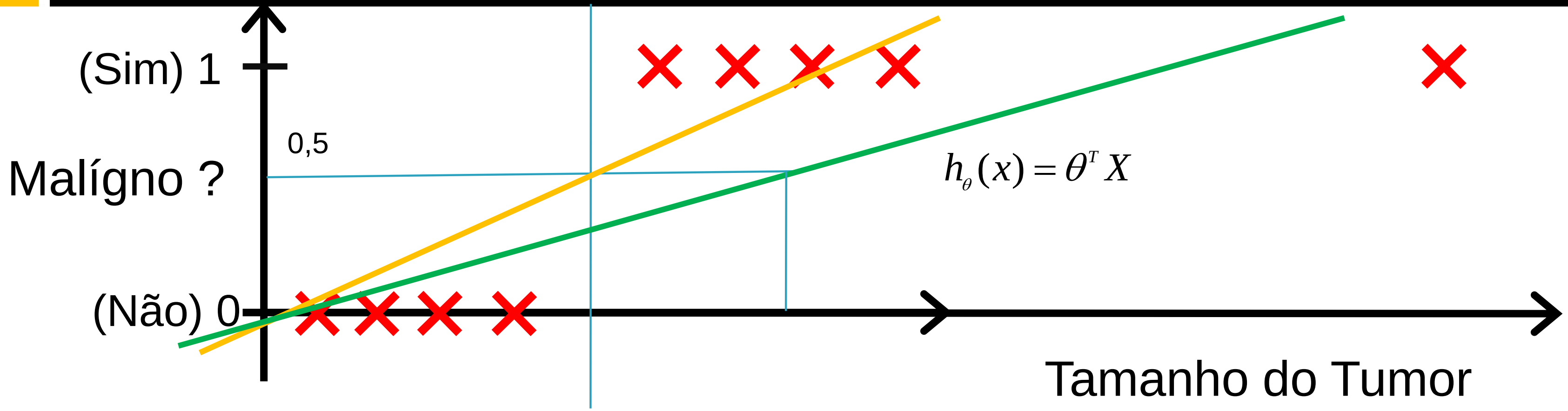
Rogério Gomes e Paulo Almeida

Classificação

Email: Spam / Não Spam?
Transações Online: Fraudulenta (Sim / Não)?
Tumor: Maligno / Benigno ?

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

0: “Classe Negativa” (e.g., tumor benígno)
1: “Classe Positiva” (e.g., tumor malígno)



Limiar de saída do Classificador $h_\theta(x)$ em 0.5:

Se $h_\theta(x) \geq 0.5$, predizer "y = 1"

Se $h_\theta(x) < 0.5$, predizer "y = 0"

Classificação: y = 0 ou 1

Mas na regressão linear:

$h_\theta(x)$ pode ser > 1 ou < 0

Então usar:

Regressão Logística: $0 \leq h_\theta(x) \leq 1$

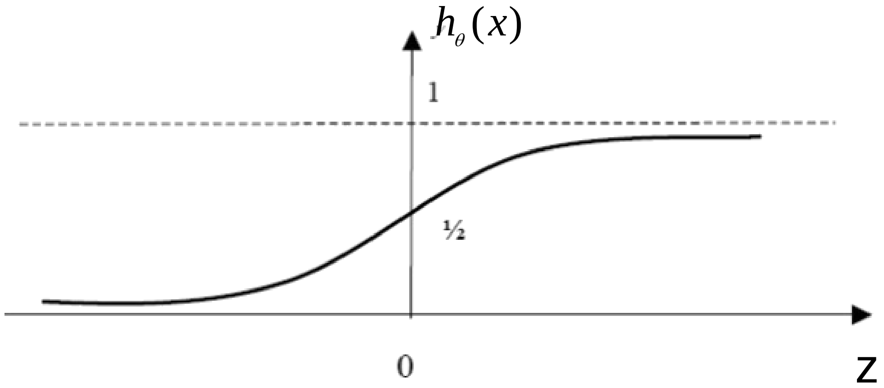
Regressão Logística
Representação

Modelo : Regressão Logística

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

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

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



Função Sigmoide ou
Função Logística

Saída da Hipótese - Interpretação

$h_\theta(x)$ = probabilidade da saída ser y = 1 para uma entrada x

Exemplo: se $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$

$$h_\theta(x) = 0.7$$

O paciente tem 70% de chance de ter um tumor malígno

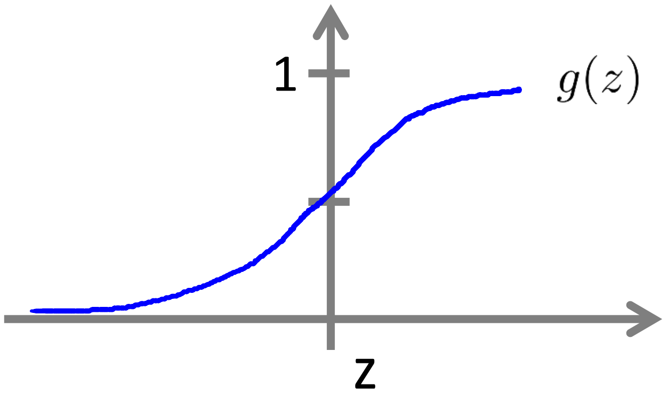
$h_\theta(x) = P(y = 1 | x; \theta)$ “probabilidade de y = 1, dado x, parametrizado por θ ”

$$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$$
$$P(y = 0|x; \theta) = 1 - P(y = 1|x; \theta)$$

Regressão Logística
Fronteira de Decisão

Regressão Logística

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



Suponha: predizer "$y = 1$" se $h_\theta(x) \geq 0.5$

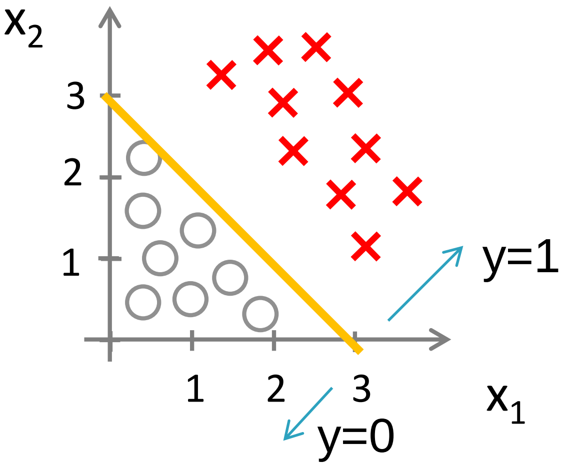
$$\theta^T X \geq 0 \rightarrow g(z) \geq 0{,}5$$

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

predizer "$y = 0$" se $h_\theta(x) < 0.5$

$$\theta^T X < 0 \rightarrow g(z) < 0{,}5$$

Fronteira de Decisão



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

Predizer "$y = 1$" se $-3 + x_1 + x_2 \geq 0$

$$x_1 + x_2 \geq 3$$

Para $y = 0$ $\quad x_1 + x_2 < 3$

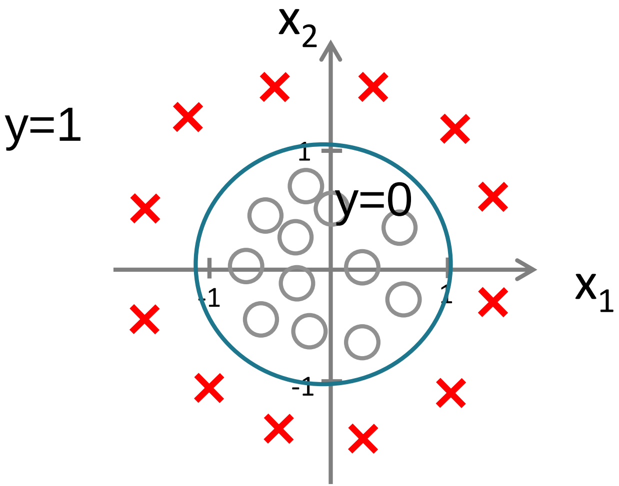
Fazendo

$$h_\vartheta(x) = 0{,}5$$

Temos :

$$x_1 + x_2 = 3$$

Fronteiras de decisão Não-lineares

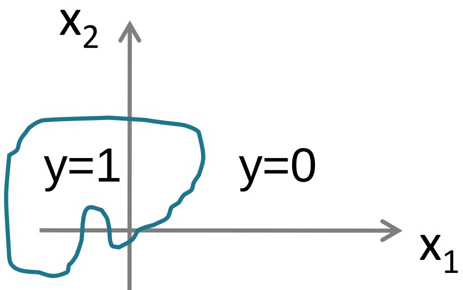


$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

$$\theta = \begin{bmatrix} -1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Predizer "$y = 1$" se $-1 + x_1^2 + x_2^2 \geq 0$

$$x_1^2 + x_2^2 = 1 \qquad x_1^2 + x_2^2 \geq 1$$



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2$$
$$+\theta_4 x_1^2 x_2 + \theta_5 x_1^2 x_2^2 + \theta_6 x_1^3 x_2 + \ldots)$$

Regressão Logística
Função Custo

Conjunto de Treinamento: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$

m exemplos

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

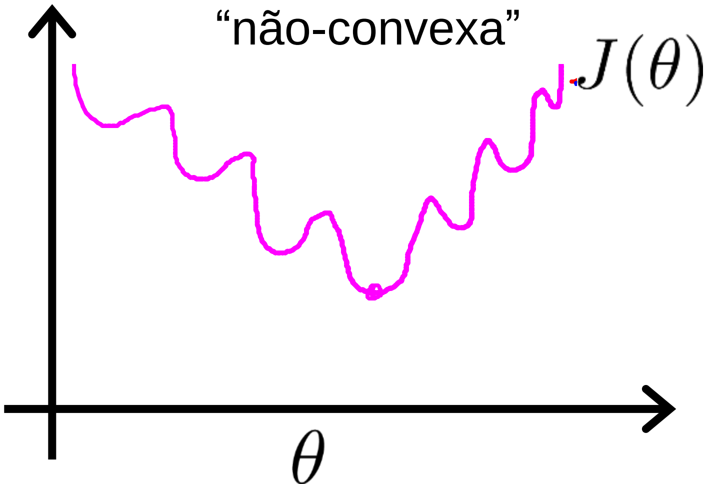
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

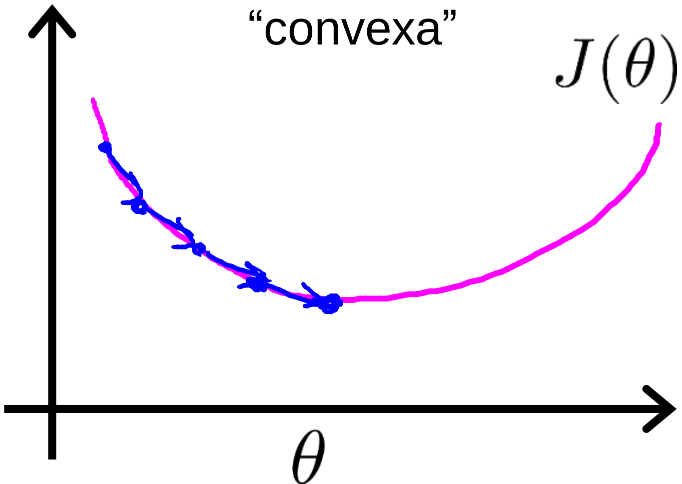
Como escolher os parâmetros θ ?

Função Custo

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

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



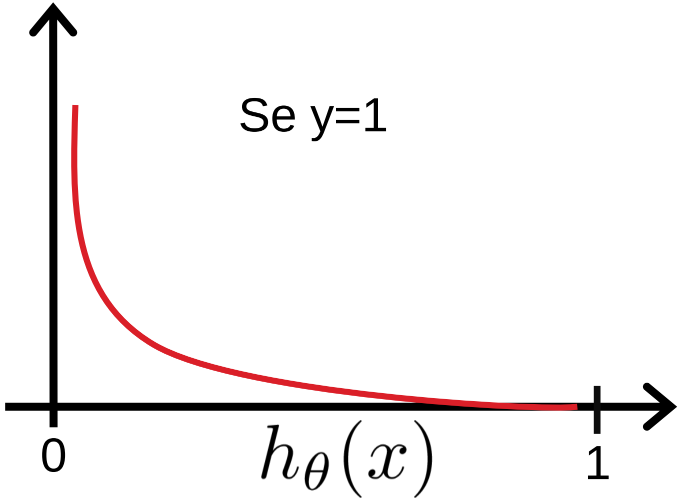


Regressão Logística – Função Custo

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

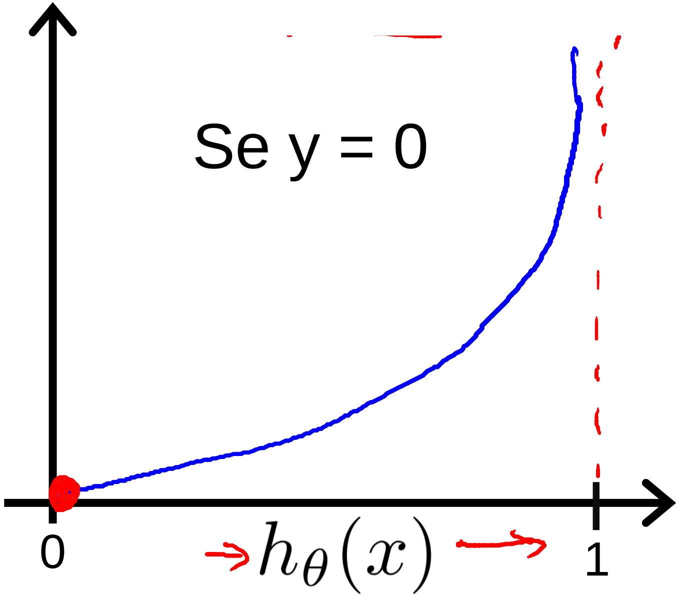
Custo=0 se $y = 1, h_\theta(x) = 1$

Mas quando $h_\theta(x) \to 0, Cost \to \infty$



Regressão Logística – Função Custo

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



Regressão Logística - Função Custo Simplificada e Gradiente Descendente

Regressão Logística – Função Custo

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

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

$$Observe: y = 0 \quad ou \quad y = 1 \quad Sempre$$

$$Cost(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

Regressão Logística – Função Custo

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

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

Para ajustar os parâmetros θ, faça:

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

Para predizer um novo valor de x:

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

Gradiente Descendente

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

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

Repeat $\{$

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

$\}$ (simultaneamente atualize todos os θ_j)

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

Gradiente Descendente

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

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

Repeat $\{$

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

$\}$ (simultaneamente atualize todos os θ_j)

O algoritmo parece idêntico à regressão linear!

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

Regressão Logística - Otimização Avançada

Algoritmo de otimização

Função Custo $J(\theta)$. Queremos $\min_\theta J(\theta)$.

Dado θ , Nós temos que computar

- $J(\theta)$
- $\frac{\partial}{\partial \theta_j} J(\theta)$ (for $j = 0, 1, \ldots, n$)

Gradiente descendente:

Repeat $\{$

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

$\}$

Algoritmo de otimização

Dado θ, temos que computar

- $J(\theta)$
- $\frac{\partial}{\partial \theta_j} J(\theta)$ (for $j = 0, 1, \ldots, n$)

Algoritmos de Otimização:

- Gradiente descendente
- Gradiente Conjugado
- BFGS
- L-BFGS

Vantagens:

- Não precisa definir α manualmente
- Usualmente mais rápido que o gradiente descendente.

Desvantagens:

- Mais complexo

Exemplo:

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$J(\theta) = (\theta_1 - 5)^2 + (\theta_2 - 5)^2$$

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

$$\frac{\partial}{\partial \theta_2} J(\theta) = 2(\theta_2 - 5)$$

```
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);
```

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

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

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

jVal = [code to compute $J(\theta)$];

gradient(1) = [code to compute $\frac{\partial}{\partial\theta_0} J(\theta)$];

gradient(2) = [code to compute $\frac{\partial}{\partial\theta_1} J(\theta)$];

$\vdots$

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

Regressão Logística
Classificação Multi-classes:
Um-vs-todos

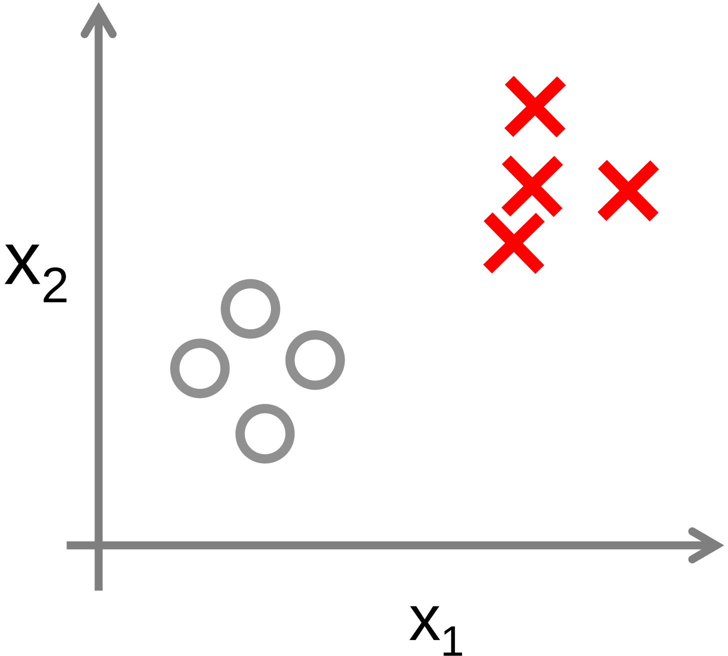
Classificação Multiclasses

Email foldering/tagging: Work, Friends, Family, Hobby

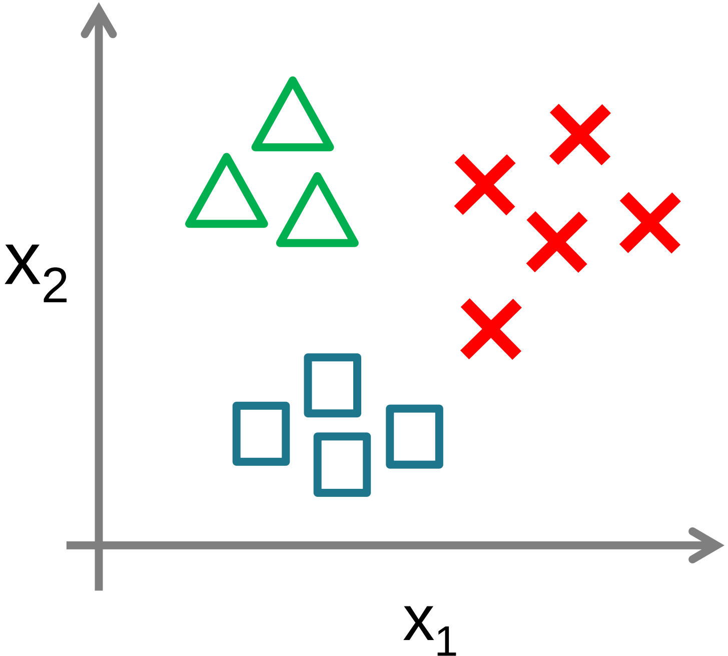
Medical diagrams: Not ill, Cold, Flu

Weather: Sunny, Cloudy, Rain, Snow

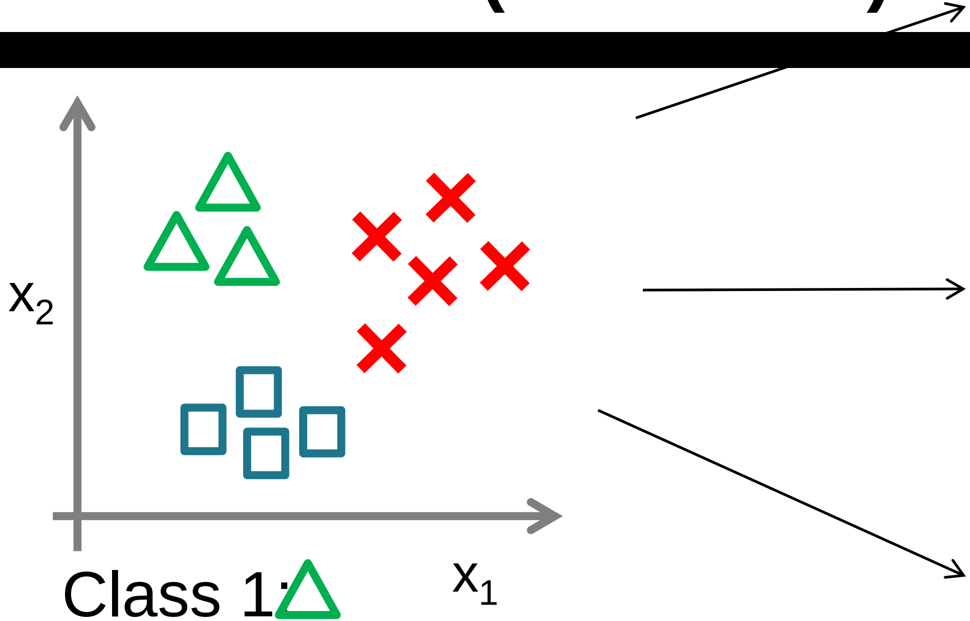
Classificação Binária:

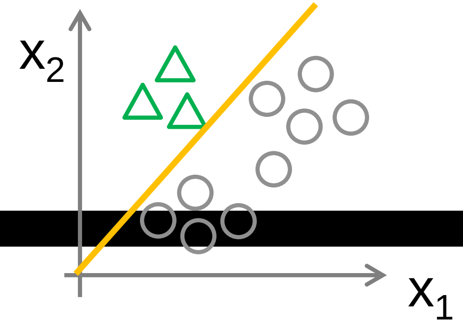


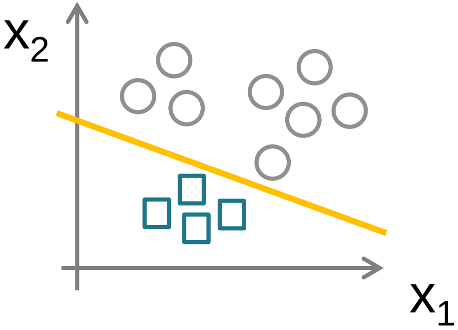
Classificação Multi-classe:

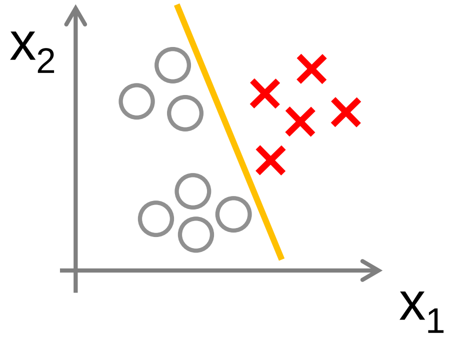


Um-vs-todos (One-vs-all)









Class 1:

Class 2:

Class 3:

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

Um-vs-todos (One-vs-all)

Treine o classificador regressão logística $h_\theta^{(i)}(x)$ para cada classe i para predizer a probablidade de $y = i$.

Para cada nova entrada x , faça a predição e pegue a classe i que maximiza

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