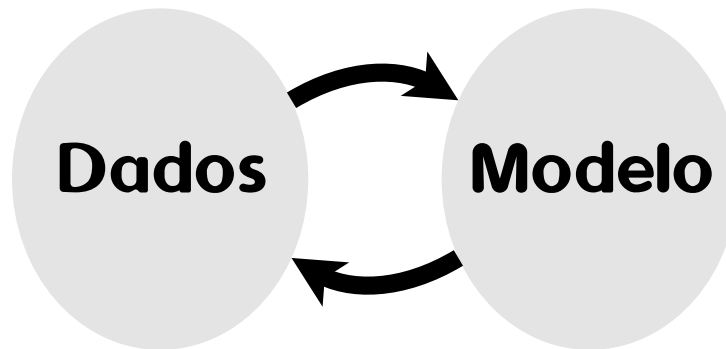


PRIMEIRA EDIÇÃO



---

# Métodos numéricos

---

PROBLEMAS NÃO LINEARES E INVERSOS

Fernando Pujaico Rivera

*email:*  
[fernando.pujaico.rivera@gmail.com](mailto:fernando.pujaico.rivera@gmail.com)

XXXXXXXXXXXXX 2020



Esta obra está liberada com uma Licença Creative Commons Atribuição - NãoComercial - SemDerivações 4.0 Internacional. Não é possível usar este arquivo excepto em conformidade com a Licença. Pode obter uma cópia da Licença em: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Limite de responsabilidade e exceção de garantia: O autor tem feito seu melhor esforço na preparação deste material. Esta edição deve ser proporcionada sem nenhuma modificação. Se distribui gratuitamente com a esperança de que seja útil, porém sem nenhuma garantia expressa ou implícita em relação à exatidão ou completitude do conteúdo.

Garanta o “download” gratuito da versão digital do livro em <https://trucomanx.github.io/metodos.numericos>

**Impresso no Brasil – ISBN: XXXXXXXXXXXX**

**Publicado:** Edição Independente

**Primeira impressão:** XXXXXXXXXXXX 2020

**Diagramação:** Fernando Pujaico Rivera

**Revisão de texto:** Fernando Pujaico Rivera

**Capa:** Fernando Pujaico Rivera

### Ficha catalográfica

Pujaico Rivera, Fernando, 1982.

Métodos numéricos: Problemas não lineares e inversos / Fernando Pujaico Rivera. – Lavras, Edição Independente, 2020.

30 p.: XXXxXXXcm.

Inclui Bibliografia

ISBN:XXXXXXXXXX

1. Métodos numéricos. 2. Problemas inversos. 3. Cálculo numérico. I. Título.

CDD: 515

CDU: 519.6

*Amicum lege feliciter, vivas, gaudeas, floreas in Deo.*

*Fernando*



# Agradecimentos

*Dou muitas graças a Deus*



# Patrocínio

Para investir nesta pesquisa e colaborar com o desenvolvimento e crescimento deste projeto, você pode comprar um exemplar do livro. Para ver uma lista com indicações sobre onde comprar:

- Uma versão impressa do livro, aceder a <https://trucomanx.github.io/metodos.numericos/comprar-impresso.html>
- Uma versão digital do livro, aceder a <https://trucomanx.github.io/metodos.numericos/comprar-digital.html>

Também pode colaborar com dinheiro em efetivo, desde 5 reais, pelo seguinte método:

- <https://apoia.se/metodosnumericos>

Para verificar a integridade do arquivo da versão digital deste livro, pode seguir as indicações publicadas no sitio oficial do projeto:

- <https://trucomanx.github.io/metodos.numericos/verificar.html>

Se já colaborou com a pesquisa, e se assim o deseja, sintase livre de me mandar um e-mail a [fernando.pujaico.rivera@gmail.com](mailto:fernando.pujaico.rivera@gmail.com), sugerindo abordar um novo assunto ou aprofundar em outro. Se seu pedido está dentro das minhas capacidades este será agregado sem falta na seguinte edição do livro.

Fernando Pujaico Rivera







# Sumário

## I

## Teoria geral

<b>1</b>	<b>Classificação de dados</b>	<b>13</b>
1.1	<b>Regressão logística com MSE de <math>f_c(x) : \mathbb{R} \rightarrow \mathbb{R}</math></b>	<b>14</b>
1.1.1	Exemplos de classificação com uma função $f_c(x) : \mathbb{R} \rightarrow \mathbb{R}$	14
1.2	<b>Regressão logística polinomial com MSE de <math>f_c(x) : \mathbb{R} \rightarrow \mathbb{R}</math></b>	<b>16</b>
1.2.1	Exemplos de classificação com uma função $f_c(x) : \mathbb{R} \rightarrow \mathbb{R}$	16
1.3	<b>Regressão logística com MSE de <math>f_c(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}</math></b>	<b>18</b>
1.3.1	Exemplos de classificação com uma função $f_c(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$	18
1.4	<b>Regressão logística polinomial com MSE de <math>f_c(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}</math></b>	<b>20</b>
1.4.1	Exemplos de classificação com uma função $f_c(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$	21

## II

## Referências

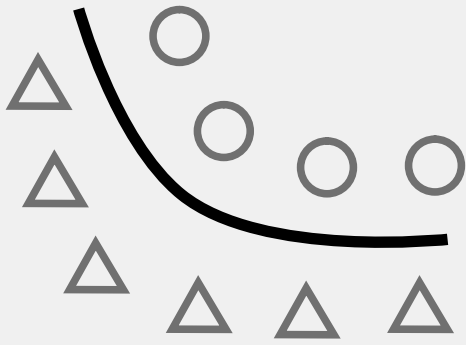
<b>Bibliografia</b>	<b>25</b>
Livros	25
Artigos	25
Outras fontes	25
<b>Índice</b>	<b>27</b>



# Teoria geral

<b>1</b>	<b>Classificação de dados .....</b>	<b>13</b>
1.1	Regressão logística com MSE de $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$	
1.2	Regressão logística polinomial com MSE de $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$	
1.3	Regressão logística com MSE de $f_{\mathbf{c}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$	
1.4	Regressão logística polinomial com MSE de $f_{\mathbf{c}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$	





# 1. Classificação de dados

**Definição 1.0.1 — Função sigmoid ou Função logística**  $f : \mathbb{R} \rightarrow \mathbb{R}$ : Esta é uma função  $f(u)$ , definida como na Eq. (1.1), com domínio que atinge todos os números reais e contradomínio entre 0 e 1. [Kur+01, pp. 27],

$$f(u) = \frac{1}{1 + e^{-u}}. \quad (1.1)$$

A derivada da função logística cumpre a seguinte propriedade,

$$\frac{df(u)}{du} = f(u)(1 - f(u)). \quad (1.2)$$

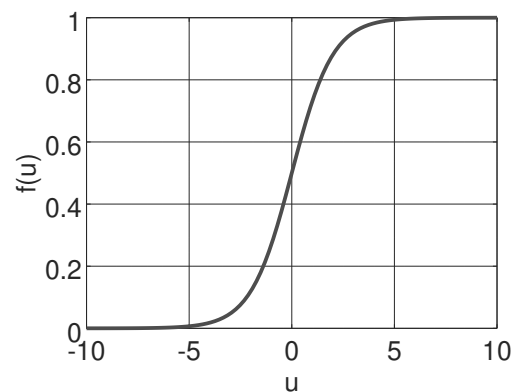


Figura 1.1: Função sigmoide.

**Definição 1.0.2 — Função logit**  $\mathbb{R} \rightarrow \mathbb{R}$ : A função  $\text{logit}(f)$  é definida como na Eq. (1.3), com domínio entre 0 e 1 e contradomínio que atinge todos os números reais. [KK10, pp. 17],

$$\begin{aligned} u &= \text{logit}(f), \\ &= \ln\left(\frac{f}{1-f}\right). \end{aligned} \quad (1.3)$$

A função logit, pode entende-se como a função inversa da função logística.

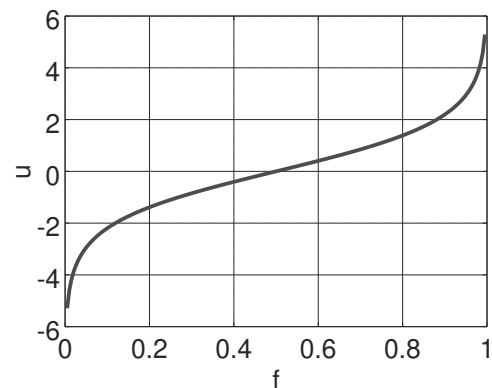
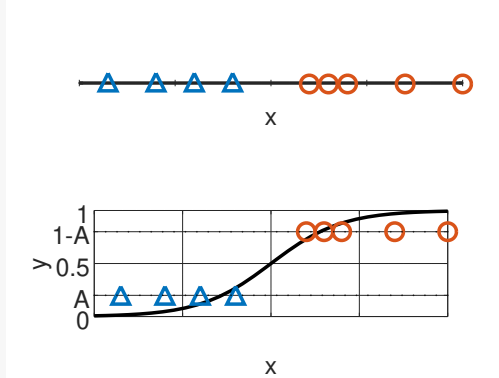


Figura 1.2: Função logit.

## 1.1 Regressão logística com MSE de $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$

### Teorema 1.1 — Classificação de dados em $\mathbb{R}$ :



Dados, um conjunto de  $L$  dados  $x_l \in \mathbb{R}, 1 \leq l \leq L$ , repartidos em dois grupos etiquetados com os símbolos  $\triangle$  e  $\circ$ , e separáveis por um hiperplano. Se desejamos criar um classificador mediante a função  $f_{\mathbf{c}} : \mathbb{R} \rightarrow \mathbb{R}$ , com domínio  $x \in \mathbb{R}$ , contra-domínio  $y \in \mathbb{R}$  e parâmetros agrupados no vetor  $\mathbf{c} = [c_1 \ c_2]^T \in \mathbb{R}^2$ , como definido na Eq. (1.4),

$$y \equiv f_{\mathbf{c}}(x) = \frac{1}{1 + e^{-h_{\mathbf{c}}(x)}}, \quad h_{\mathbf{c}}(x) = c_1 + c_2x, \quad (1.4)$$

ou seu equivalente:  $\text{logit}(y) = h_{\mathbf{c}}(x)$ .

Podemos atribuir a cada valor  $x_l$  uma etiqueta  $y_l \in \{A, 1 - A\}$ , onde  $0 < A \ll 0.5$  é escolhido por nós, e afirmar que o vetor  $\mathbf{c} = \hat{\mathbf{c}}$ , que minimiza o erro quadrático médio<sup>a</sup>  $e(\mathbf{c})$ ,

$$e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L w_l \|h_{\mathbf{c}}(x_l) - \text{logit}(y_l)\|^2, \quad (1.5)$$

ponderado com os pesos  $w_l \in \mathbb{R}_+$ , pode ser achado com

$$\hat{\mathbf{c}} = [\mathbf{A}^T \mathbf{W} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{W} \mathbf{z}, \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_l \\ \vdots & \vdots \\ 1 & x_L \end{bmatrix}, \quad \mathbf{W} = \text{diag} \left( \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_l \\ \vdots \\ w_L \end{bmatrix} \right), \quad \mathbf{z} = \begin{bmatrix} \text{logit}(y_1) \\ \text{logit}(y_2) \\ \vdots \\ \text{logit}(y_l) \\ \vdots \\ \text{logit}(y_L) \end{bmatrix}. \quad (1.6)$$

<sup>a</sup>Do inglês “mean square error” com siglas *MSE*.



- Dado que a função de classificação  $f_{\mathbf{c}}(x)$  vai entre 0 e 1, podemos reinterpretar este valor como se fosse uma probabilidade; neste caso,  $f_{\mathbf{c}}(x)$  representa a probabilidade de que um valor  $x$  pertença ao grupo  $\circ$ .
- O limiar da classificação de  $f_{\mathbf{c}}(x)$  está no hiperplano  $c_1 + c_2x = 0$ , provocando neste ponto um  $f_{\mathbf{c}}(x) = 0.5$ .

### 1.1.1 Exemplos de classificação com uma função $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$

■ **Exemplo 1.1** Conhecida as  $L = 10$  amostras  $x_l$  e seus respectivos grupos indicados pelos símbolos  $\triangle$  e  $\circ$ , mostrados na Tabela 1.1, achar o classificador  $f_{\mathbf{c}}(x)$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|c_1 + c_2x_l - \text{logit}(y_l)\|^2$ . ■

**Solução 1.1 — Relativa ao Exemplo 1.1:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(x)$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|c_1 + c_2x_l - \text{logit}(y_l)\|^2$  com os  $L = 10$  dados  $x_l$  da Tabela 1.1, usamos a Eq. (1.6) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.1, 0.9\}$ , 0.1 para  $\triangle$  e 0.9 para  $\circ$ , obtendo um vetor  $\hat{\mathbf{c}} = [-5.5043 \quad 2.4248]^T$ . Assim, podemos representar a função

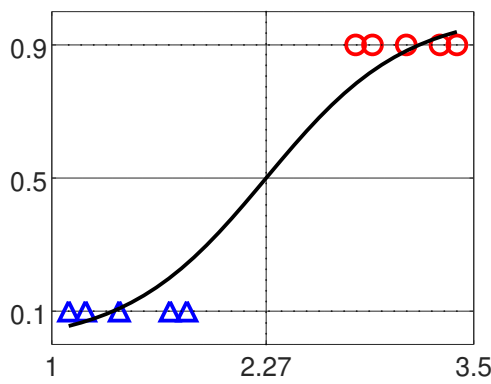
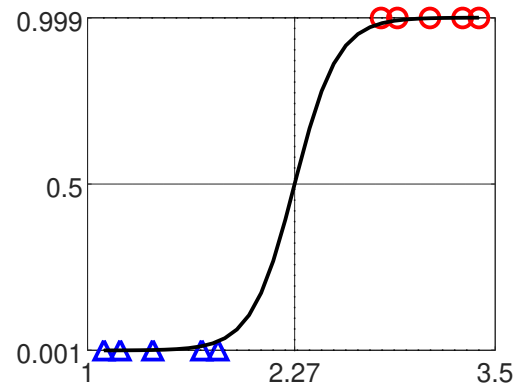
$l$	1	2	3	4	5	6	7	8	9	10
$x_l$	1.1	1.2	1.4	1.7	1.8	2.8	2.9	3.1	3.3	3.4
$y_l$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$

Tabela 1.1: Valores  $x_l$ .

$f_{\mathbf{c}}(x)$  que classifica os dados  $x_l$ , como é mostrado na Figura 1.3a e na Eq. (1.7),

$$f_{\mathbf{c}}(x) = \frac{1}{1 + e^{5.5043 - 2.4248x}}. \quad (1.7)$$

É interessante ressaltar que para um valor  $A = 0.1$  a pendente é pequena e a classificação é pouco definida, com limiar de classificação em 2.27.

(a) Gráfico da classificação usando  $y_l \in \{0.1, 0.9\}$ .(b) Gráfico da classificação usando  $y_l \in \{0.001, 0.999\}$ .Figura 1.3: Classificação usando a função  $f_{\mathbf{c}}(x)$ .

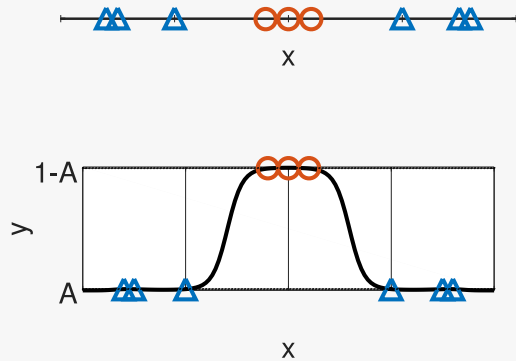
**Solução 1.2 — Relativa ao Exemplo 1.1:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(x)$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|c_1 + c_2 x_l - \text{logit}(y_l)\|^2$  com os  $L = 10$  dados  $x_l$  da Tabela 1.1, usamos a Eq. (1.6) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.001, 0.999\}$ , 0.001 para  $\triangle$  e 0.999 para  $\circ$ , obtendo um vetor  $\hat{\mathbf{c}} = [-17.3022 \quad 7.6221]^T$ . Assim, podemos representar a função  $f_{\mathbf{c}}(x)$  que classifica os dados  $x_l$ , como é mostrado na Figura 1.3b e na Eq. (1.8),

$$f_{\mathbf{c}}(x) = \frac{1}{1 + e^{17.3022 - 7.6221x}}. \quad (1.8)$$

É interessante ressaltar que para um valor  $A = 0.001$  a pendente é abrupta para cada grupo e a classificação é bem definida, com limiar de classificação em 2.27.

## 1.2 Regressão logística polinomial com MSE de $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$

### Teorema 1.2 — Classificação de dados em $\mathbb{R}$ :



Dados, um conjunto de  $L$  dados  $x_l \in \mathbb{R}$ ,  $1 \leq l \leq L$ , repartidos em dois grupos etiquetados com os símbolos  $\triangle$  e  $\circ$ , não separáveis por um hiperplano. Se desejamos criar um classificador mediante a função  $f_{\mathbf{c}} : \mathbb{R} \rightarrow \mathbb{R}$ , com domínio  $x \in \mathbb{R}$ , contradomínio  $y \in \mathbb{R}$  e parâmetros agrupados no vetor  $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_{M+1}]^T \in \mathbb{R}^{M+1}$ , como definido na Eq. (1.9),

$$y \equiv f_{\mathbf{c}}(x) = \frac{1}{1 + e^{-h_{\mathbf{c}}(x)}}, \quad h_{\mathbf{c}}(x) = \sum_{m=0}^M c_{m+1} x^m, \quad (1.9)$$

ou seu equivalente:  $\text{logit}(y) = h_{\mathbf{c}}(x)$ .

Podemos atribuir a cada valor  $x_l$  uma etiqueta  $y_l \in \{A, 1-A\}$ , onde  $0 < A \ll 0.5$  é escolhido por nós, e afirmar que o vetor  $\mathbf{c} = \hat{\mathbf{c}}$ , que minimiza o erro quadrático médio<sup>a</sup>  $e(\mathbf{c})$ ,

$$e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L w_l \|h_{\mathbf{c}}(x_l) - \text{logit}(y_l)\|^2, \quad (1.10)$$

ponderado com os pesos  $w_l \in \mathbb{R}_+$  pode ser calculado com a Eq. (1.11); onde a matriz  $\mathbf{W} = \text{diag}([w_1, w_2, \dots, w_L]^T)$ ;

$$\hat{\mathbf{c}} = [\mathbf{A}^T \mathbf{W} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{W} \mathbf{z}, \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^M \\ 1 & x_2 & x_2^2 & \dots & x_2^M \\ 1 & x_3 & x_3^2 & \dots & x_3^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_l & x_l^2 & \dots & x_l^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_L & x_L^2 & \dots & x_L^M \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \text{logit}(y_1) \\ \text{logit}(y_2) \\ \text{logit}(y_3) \\ \vdots \\ \text{logit}(y_l) \\ \vdots \\ \text{logit}(y_L) \end{bmatrix}. \quad (1.11)$$

<sup>a</sup>Do inglês “mean square error” com siglas *MSE*.



- Dado que a função de classificação  $f_{\mathbf{c}}(x)$  vai entre 0 e 1, podemos reinterpretar este valor como se fosse uma probabilidade; neste caso,  $f_{\mathbf{c}}(x)$  representa a probabilidade de que um valor  $x$  pertença ao grupo  $\circ$ .
- Os limiares de classificação na função  $f_{\mathbf{c}}(x)$  estão nas raízes  $h_{\mathbf{c}}(x) = 0$ , provocando nestes pontos um  $f_{\mathbf{c}}(x) = 0.5$ .

### 1.2.1 Exemplos de classificação com uma função $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$

■ **Exemplo 1.2** Conhecidas as  $L = 12$  amostras  $x_l$  e seus respectivos grupos indicados pelos símbolos  $\triangle$  e  $\circ$ , mostrados na Tabela 1.2, achar o classificador  $f_{\mathbf{c}}(x)$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_{\mathbf{c}}(x) - \text{logit}(y_l)\|^2$ . ■



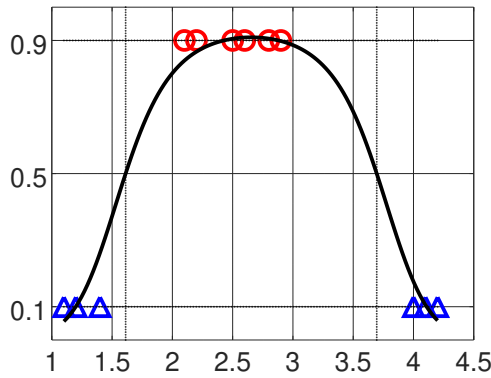
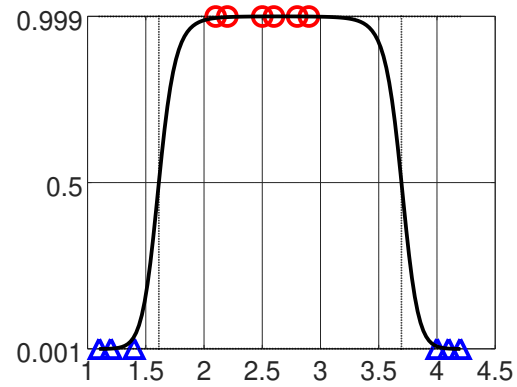
$l$	1	2	3	4	5	6	7	8	9	10	11	12
$x_l$	1.1	1.2	1.4	4.0	4.1	4.2	2.1	2.2	2.5	2.6	2.8	2.9
$y_l$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$

Tabela 1.2: Valores  $x_l$ .

**Solução 1.3 — Relativa ao Exemplo 1.2:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(x)$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_{\mathbf{c}}(x_l) - \text{logit}(y_l)\|^2$  com os  $L = 12$  dados  $x_l$  da Tabela 1.2, usamos a Eq. (1.11) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.1, 0.9\}$ , 0.1 para  $\triangle$  e 0.9 para  $\circ$ , obtendo um vetor  $\hat{\mathbf{c}} = [-12.6667 \quad 11.2848 \quad -2.1263]^T$ . Assim, podemos representar a função  $f_{\mathbf{c}}(x)$  que classifica os dados  $x_l$ , como é mostrado na Figura 1.4a e na Eq. (1.12),

$$f_{\mathbf{c}}(x) = \frac{1}{1 + e^{-(-12.6667 + 11.2848x - 2.1263x^2)}}. \quad (1.12)$$

É interessante ressaltar que para um valor  $A = 0.1$  a pendente é pequena e a classificação é pouco definida, com limiares de classificação em 1.6122 e 3.6951.

(a) Gráfico da classificação usando  $y_l \in \{0.1, 0.9\}$ .(b) Gráfico da classificação usando  $y_l \in \{0.001, 0.999\}$ .Figura 1.4: Classificação usando a função  $f_{\mathbf{c}}(x)$ .

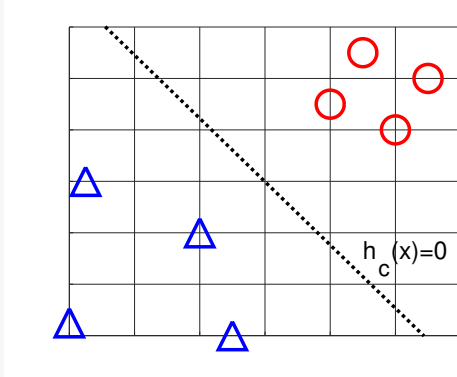
**Solução 1.4 — Relativa ao Exemplo 1.2:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(x)$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_{\mathbf{c}}(x_l) - \text{logit}(y_l)\|^2$  com os  $L = 12$  dados  $x_l$  da Tabela 1.2, usamos a Eq. (1.11) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.001, 0.999\}$ , 0.001 para  $\triangle$  e 0.999 para  $\circ$ , obtendo um vetor  $\hat{\mathbf{c}} = [-39.8164 \quad 35.4727 \quad -6.6838]^T$ . Assim, podemos representar a função  $f_{\mathbf{c}}(x)$  que classifica os dados  $x_l$ , como é mostrado na Figura 1.4b e na Eq. (1.13),

$$f_{\mathbf{c}}(x) = \frac{1}{1 + e^{-(-39.8164 + 35.4727x - 6.6838x^2)}}. \quad (1.13)$$

É interessante ressaltar que para um valor  $A = 0.001$  a pendente é abrupta para cada grupo com uma classificação bem definida, e limiares de classificação em 1.6122 e 3.6951.

## 1.3 Regressão logística com MSE de $f_c(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$

### Teorema 1.3 — Classificação de dados em $\mathbb{R}^N$ :



Dados, um conjunto de  $L$  pontos  $\mathbf{x}_l \in \mathbb{R}^N$ ,  $1 \leq l \leq L$ , repartidos em dois grupos etiquetados com os símbolos  $\Delta$  e  $\bigcirc$ , separáveis por um hiperplano em  $\mathbb{R}^N$ . Se desejamos criar um classificador mediante a função  $f_c : \mathbb{R}^N \rightarrow \mathbb{R}$ , com domínio  $\mathbf{x} \in \mathbb{R}^N$ , contradomínio  $y \in \mathbb{R}$  e parâmetros agrupados no vetor  $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_{N+1}]^T \in \mathbb{R}^{N+1}$ , como definido na Eq. (1.14),

$$y \equiv f_c(\mathbf{x}) = \frac{1}{1 + e^{-h_c(\mathbf{x})}}, \quad h_c(\mathbf{x}) = [1 \ \mathbf{x}^T] \mathbf{c}, \quad (1.14)$$

ou seu equivalente:  $\text{logit}(y) = h_c(\mathbf{x})$ .

Podemos atribuir a cada valor  $\mathbf{x}_l$  uma etiqueta  $y_l \in \{A, 1-A\}$ ,  $A$  para  $\Delta$  e  $1-A$  para  $\bigcirc$ , onde  $0 < A \ll 0.5$  é escolhido por nós, e afirmar que o vetor  $\mathbf{c} = \hat{\mathbf{c}}$ , que minimiza o erro quadrático médio<sup>a</sup>  $e(\mathbf{c})$ ,

$$e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L w_l \|h_c(\mathbf{x}_l) - \text{logit}(y_l)\|^2, \quad (1.15)$$

pode ser achado mediante a Eq. (1.16). Os erros de cada amostra são ponderados com os pesos  $w_l \in \mathbb{R}_+$  e agrupados na matriz  $\mathbf{W} = \text{diag}([w_1, w_2, \dots, w_L]^T)$ ;

$$\hat{\mathbf{c}} = [\mathbf{A}^T \mathbf{W} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{W} \mathbf{z}, \quad \mathbf{A} = \begin{bmatrix} 1 & \mathbf{x}_1^T \\ 1 & \mathbf{x}_2^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_l^T \\ \vdots & \vdots \\ 1 & \mathbf{x}_L^T \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \text{logit}(y_1) \\ \text{logit}(y_2) \\ \vdots \\ \text{logit}(y_l) \\ \vdots \\ \text{logit}(y_L) \end{bmatrix}. \quad (1.16)$$

<sup>a</sup>Do inglês “mean square error” com siglas *MSE*.



- Dado que a função de classificação  $f_c(\mathbf{x})$  vai entre 0 e 1, podemos reinterpretar este valor como se fosse uma probabilidade; neste caso,  $f_c(\mathbf{x})$  representa a probabilidade de que um ponto  $\mathbf{x}$  pertença ao grupo  $\bigcirc$ .
- O limiar de classificação na função  $f_c(\mathbf{x})$  está no hiperplano  $h_c(\mathbf{x}) = 0$ , provocando nestes pontos um  $f_c(\mathbf{x}) = 0.5$ .

### 1.3.1 Exemplos de classificação com uma função $f_c(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$

■ **Exemplo 1.3** Conhecidas as  $L = 10$  amostras  $\mathbf{x}_l$  e seus respectivos grupos indicados pelos símbolos  $\Delta$  e  $\bigcirc$ , mostrados na Tabela 1.3, achar o classificador  $f_c(\mathbf{x})$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_c(\mathbf{x}_l) - \text{logit}(y_l)\|^2$ . ■

**Solução 1.5 — Relativa ao Exemplo 1.3:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_c(\mathbf{x})$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_c(\mathbf{x}_l) - \text{logit}(y_l)\|^2$  com os  $L = 10$  dados  $\mathbf{x}_l$  da

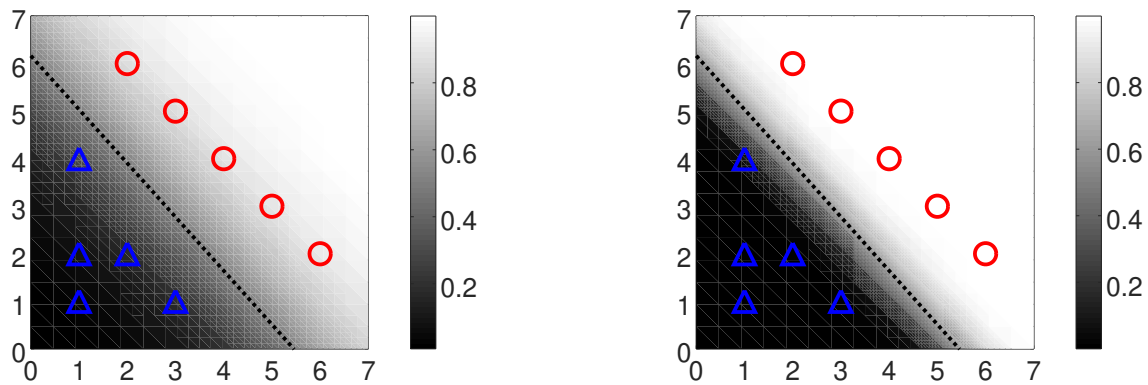
$l$	1	2	3	4	5	6	7	8	9	10
$\mathbf{x}_l$	1	3	1	2	1	4	5	6	3	2
	1	1	4	2	2	4	3	2	5	6
$y_l$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$

Tabela 1.3: Valores  $\mathbf{x}_l$ .

Tabela 1.3, usamos a Eq. (1.16) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.1, 0.9\}$ , 0.1 para  $\triangle$  e 0.9 para  $\circ$ , obtendo um vetor  $\hat{\mathbf{c}} = [-5.21014 \quad 0.95532 \quad 0.84509]^T$ . Assim, podemos representar a função  $f_{\mathbf{c}}(\mathbf{x})$  que classifica os dados  $\mathbf{x}_l$ , como é mostrado na Figura 1.5a,

$$f_{\mathbf{c}}(\mathbf{x}) = \frac{1}{1 + e^{-(5.21014 + 0.95532x_1 + 0.84509x_2)}}. \quad (1.17)$$

É interessante ressaltar que a pendente é pequena e a classificação é pouco definida, com limiares de classificação no hiperplano  $-5.21014 + 0.95532x_1 + 0.84509x_2 = 0$ .

(a) Gráfico da classificação usando  $y_l \in \{0.1, 0.9\}$ .(b) Gráfico da classificação usando  $y_l \in \{0.001, 0.999\}$ .Figura 1.5: Classificação usando a função  $f_{\mathbf{c}}(\mathbf{x})$ .

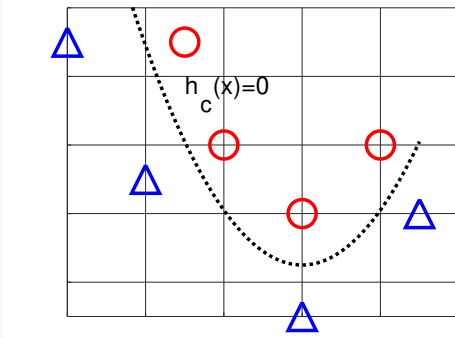
**Solução 1.6 — Relativa ao Exemplo 1.3:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(\mathbf{x})$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_{\mathbf{c}}(\mathbf{x}_l) - \text{logit}(y_l)\|^2$  com os  $L = 10$  dados  $\mathbf{x}_l$  da Tabela 1.3, usamos a Eq. (1.16) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.001, 0.999\}$ , 0.001 para  $\triangle$  e 0.999 para  $\circ$ , obtendo um vetor  $\hat{\mathbf{c}} = [-16.3776 \quad 3.0029 \quad 2.6564]^T$ . Assim, podemos representar a função  $f_{\mathbf{c}}(\mathbf{x})$  que classifica os dados  $\mathbf{x}_l$ , como é mostrado na Figura 1.5b e na Eq. (1.18),

$$f_{\mathbf{c}}(\mathbf{x}) = \frac{1}{1 + e^{-(-16.3776 + 3.0029x_1 + 2.6564x_2)}}. \quad (1.18)$$

É interessante ressaltar que a pendente é abrupta para cada grupo com uma classificação bem definida, e limiares de classificação no hiperplano  $-16.3776 + 3.0029x_1 + 2.6564x_2 = 0$ .

## 1.4 Regressão logística polinomial com MSE de $f_{\mathbf{c}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$

### Teorema 1.4 — Classificação de dados em $\mathbb{R}^N$ :



Dados, um conjunto de  $L$  pontos  $\mathbf{x}_l \in \mathbb{R}^N$ ,  $1 \leq l \leq L$ , repartidos em dois grupos etiquetados com os símbolos  $\Delta$  e  $\bigcirc$ , não separáveis por um hiperplano em  $\mathbb{R}^N$ . Se desejamos criar um classificador mediante a função  $f_{\mathbf{c}} : \mathbb{R}^N \rightarrow \mathbb{R}$ , com domínio  $\mathbf{x} \in \mathbb{R}^N$ , contra-domínio  $y \in \mathbb{R}$  e parâmetros agrupados no vetor  $\mathbf{c}$ , como definido na Eq. (1.19),

$$y \equiv f_{\mathbf{c}}(\mathbf{x}) = \frac{1}{1 + e^{-h_{\mathbf{c}}(\mathbf{x})}}, \quad h_{\mathbf{c}}(\mathbf{x}) = \mathbf{a}_M(\mathbf{x})\mathbf{c}, \quad (1.19)$$

ou seu equivalente:  $\text{logit}(y) = h_{\mathbf{c}}(\mathbf{x})$ , sendo  $h_{\mathbf{c}}(\mathbf{x})$  um polinômio multivariante de grau total  $M$  [GCL07, pp. 47] [Zip12, pp. 108] com coeficientes agrupados no vetor  $\mathbf{c}$ . Podemos atribuir a cada valor  $\mathbf{x}_l$  uma etiqueta  $y_l \in \{A, 1-A\}$ ,  $A$  para  $\Delta$  e  $1-A$  para  $\bigcirc$ , onde  $0 < A \ll 0.5$  é escolhido por nós, e afirmar que o vetor  $\mathbf{c} = \hat{\mathbf{c}}$ , que minimiza o erro quadrático médio<sup>a</sup>  $e(\mathbf{c})$ ,

$$e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L w_l \|h_{\mathbf{c}}(\mathbf{x}_l) - \text{logit}(y_l)\|^2, \quad (1.20)$$

pode ser achado mediante a Eq. (1.21). Os erros de cada amostra são ponderados com os pesos  $w_l \in \mathbb{R}_+$  e agrupados na matriz  $\mathbf{W} = \text{diag}([w_1, w_2, \dots, w_L]^T)$ ;

$$\hat{\mathbf{c}} = [\mathbf{A}^T \mathbf{W} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{W} \mathbf{z}, \quad \mathbf{A} = \begin{bmatrix} a_M(\mathbf{x}_1) \\ a_M(\mathbf{x}_2) \\ \vdots \\ a_M(\mathbf{x}_L) \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} \text{logit}(y_1) \\ \text{logit}(y_2) \\ \vdots \\ \text{logit}(y_L) \end{bmatrix}, \quad (1.21)$$

$$\mathbf{a}(\mathbf{x}) = [\mathbf{b}_0(\mathbf{x}) \quad \mathbf{b}_1(\mathbf{x}) \quad \dots \quad \mathbf{b}_m(\mathbf{x}) \quad \dots \quad \mathbf{b}_M(\mathbf{x})], \quad (1.22)$$

$$\mathbf{b}_0(\mathbf{x}) = [1], \quad \mathbf{b}_1(\mathbf{x}) = [x_1 \quad x_2 \quad \dots \quad x_N], \quad \mathbf{b}_m(\mathbf{x}) = \bigcup_{\alpha_1 + \alpha_2 + \dots + \alpha_N = m}^{\rightarrow} x_1^{\alpha_1} x_2^{\alpha_2} \dots x_N^{\alpha_N}. \quad (1.23)$$

$\mathbf{b}_m(\mathbf{x})$  representa um vetor linha com elementos  $x_1^{\alpha_1} x_2^{\alpha_2} \dots x_N^{\alpha_N}$  de modo que em cada elemento se cumpre que  $\alpha_1 + \alpha_2 + \dots + \alpha_N = m$ , sendo que  $\alpha_n \in \mathbb{N}$ .

<sup>a</sup>Do inglês “mean square error” com siglas *MSE*.



- Dado que a função de classificação  $f_{\mathbf{c}}(\mathbf{x})$  vai entre 0 e 1, podemos reinterpretar este valor como se fosse uma probabilidade; neste caso,  $f_{\mathbf{c}}(\mathbf{x})$  representa a probabilidade de que um ponto  $\mathbf{x}$  pertença ao grupo  $\bigcirc$ .
- O limiar de classificação na função  $f_{\mathbf{c}}(\mathbf{x})$  está no hiperplano  $h_{\mathbf{c}}(\mathbf{x}) = 0$ .
- A ordem dos elementos do vetor  $\mathbf{a}_M(\mathbf{x})$  podem ser alterados, isto só modificará a posição dos elementos no vetor  $\mathbf{c}$ .

■ **Exemplo 1.4 — Polinômios multivariante:**

- Polinômio univariado de grau total 2:  $P_{\mathbf{c}}(x) = c_1 + c_2x + c_3x^2$ .

$$\mathbf{b}_0(x) = [1], \quad \mathbf{b}_1(x) = [x], \quad \mathbf{b}_2(x) = [x^2]. \quad (1.24)$$

- Polinômio bivariado de grau total 2:  $P_{\mathbf{c}}(\mathbf{x}) = c_1 + c_2x_1 + c_3x_2 + c_4x_1^2 + c_5x_1x_2 + c_6x_2^2$ .

$$\mathbf{b}_0(\mathbf{x}) = [1], \quad \mathbf{b}_1(\mathbf{x}) = [x_1 \ x_2], \quad \mathbf{b}_2(\mathbf{x}) = [x_1^2 \ x_2^2 \ x_1x_2]. \quad (1.25)$$

- Polinômio trivariado de grau total 2:  $P_{\mathbf{c}}(\mathbf{x}) = c_1 + c_2x_1 + c_3x_2 + c_4x_3 + c_5x_1^2 + c_6x_2^2 + c_7x_3^2 + c_8x_1x_2 + c_9x_1x_3 + c_{10}x_2x_3$ .

$$\mathbf{b}_0(\mathbf{x}) = [1], \quad \mathbf{b}_1(\mathbf{x}) = [x_1 \ x_2 \ x_3], \quad \mathbf{b}_2(\mathbf{x}) = [x_1^2 \ x_2^2 \ x_3^2 \ x_1x_2 \ x_1x_3 \ x_2x_3]. \quad (1.26)$$

**Teorema 1.5 — Número de elementos de  $\mathbf{b}_m(\mathbf{x})$ :** Dada uma função vetorial  $\mathbf{b}_m(\mathbf{x})$  de parâmetro  $\mathbf{x} \in \mathbb{R}^N$ , a quantidade de elementos  $L_{\mathbf{b}}(N, m)$  do vetor  $\mathbf{b}_m(\mathbf{x})$  pode ser calculado com

$$L_{\mathbf{b}}(N, m) = \frac{N(N+1)(N+2)\dots(N+m-1)}{m!} = \binom{N+m-1}{m}. \quad (1.27)$$

**Prova:**  $L_{\mathbf{b}}(N, m)$  representa o número de combinações (grupos ou multiconjuntos) com repetição, onde cada grupo tem  $m$  elementos, escolhidos de um total de  $N$  elementos, quando a ordem de escolha não importa, e os elementos podem ser escolhidos mais de uma vez em cada grupo.

	$\mathbf{b}_0(\mathbf{x})$	$\mathbf{b}_1(\mathbf{x})$	$\mathbf{b}_2(\mathbf{x})$	$\mathbf{b}_3(\mathbf{x})$
$L_{\mathbf{b}}(N, m)$	1	$N$	$\frac{N(N+1)}{2}$	$\frac{N(N+1)(N+2)}{3!}$

Tabela 1.4: Número de elementos de  $\mathbf{b}_m(\mathbf{x})$ .

### 1.4.1 Exemplos de classificação com uma função $f_{\mathbf{c}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$

- **Exemplo 1.5** Conhecidas as  $L = 10$  amostras  $\mathbf{x}_l$  e seus respectivos grupos indicados pelos símbolos  $\triangle$  e  $\bigcirc$ , mostrados na Tabela 1.5, achar o classificador  $f_{\mathbf{c}}(\mathbf{x})$  que usa um polinômio multivariante  $h_{\mathbf{c}}(\mathbf{x})$  de grau total  $M = 2$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L ||h_{\mathbf{c}}(\mathbf{x}_l) - \logit(y_l)||^2$ .

$l$	1	2	3	4	5	6	7	8	9	10
$\mathbf{x}_l$	1	1	2	4	5	2	2	3	5	6
	5	4	2	1	1	6	5	3	2	2
$\mathbf{y}_l$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\triangle$	$\bigcirc$	$\bigcirc$	$\bigcirc$	$\bigcirc$	$\bigcirc$

Tabela 1.5: Valores  $\mathbf{x}_l$ .

**Solução 1.7 — Relativa ao Exemplo 1.5:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(\mathbf{x})$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_{\mathbf{c}}(\mathbf{x}_l) - \text{logit}(y_l)\|^2$  com os  $L = 10$  dados  $\mathbf{x}_l$  da Tabela 1.5, usamos as Eqs. (1.26) e (1.21) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.1, 0.9\}$ , 0.1 para  $\triangle$  e 0.9 para  $\circ$ , obtendo um vetor

$$\hat{\mathbf{c}} = [-11.71919 \quad 2.49938 \quad 2.49938 \quad -0.30282 \quad -0.30282 \quad 0.50579]^T. \quad (1.28)$$

Assim, podemos representar a função  $f_{\mathbf{c}}(\mathbf{x})$  que classifica os dados  $\mathbf{x}_l$ , como é mostrado na Figura 1.6a,

$$f_{\mathbf{c}}(\mathbf{x}) = \frac{1}{1 + e^{-h_{\mathbf{c}}(\mathbf{x})}}, \quad h_{\mathbf{c}}(\mathbf{x}) = \begin{aligned} &-11.71919 \\ &+ 2.49938x_1 + 2.49938x_2 \\ &- 0.30282x_1^2 - 0.30282x_2^2 + 0.50579x_1x_2. \end{aligned} \quad (1.29)$$

É interessante ressaltar que para um valor  $A = 0.1$  a pendente na mudança de classificação é pouco definida, com limiares de classificação na superfície  $h_{\mathbf{c}}(\mathbf{x}) = 0$ .

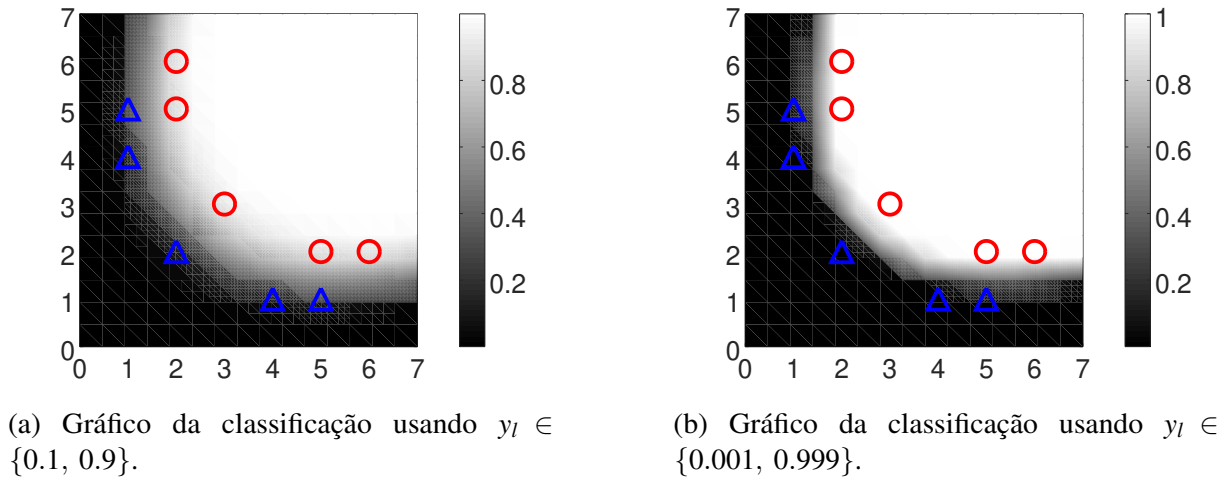


Figura 1.6: Classificação usando a função  $f_{\mathbf{c}}(\mathbf{x})$ .

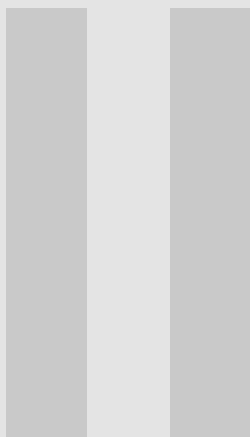
**Solução 1.8 — Relativa ao Exemplo 1.5:** Para obter o vetor de parâmetros  $\mathbf{c} = \hat{\mathbf{c}}$  da função  $f_{\mathbf{c}}(\mathbf{x})$ , que gere o menor erro  $e(\mathbf{c}) = \frac{1}{L} \sum_{l=1}^L \|h_{\mathbf{c}}(\mathbf{x}_l) - \text{logit}(y_l)\|^2$  com os  $L = 10$  dados  $\mathbf{x}_l$  da Tabela 1.5, usamos as Eqs. (1.26) e (1.21) onde escolhemos  $w_l = 1$  e valores  $y_l \in \{0.001, 0.999\}$ , 0.001 para  $\triangle$  e 0.999 para  $\circ$ , obtendo um vetor

$$\hat{\mathbf{c}} = [-36.83809 \quad 7.85656 \quad 7.85656 \quad -0.95187 \quad -0.95187 \quad 1.58990]^T. \quad (1.30)$$

Assim, podemos representar a função  $f_{\mathbf{c}}(\mathbf{x})$  que classifica os dados  $\mathbf{x}_l$ , como é mostrado na Figura 1.6b,

$$f_{\mathbf{c}}(\mathbf{x}) = \frac{1}{1 + e^{-h_{\mathbf{c}}(\mathbf{x})}}, \quad h_{\mathbf{c}}(\mathbf{x}) = \begin{aligned} &-36.83809 \\ &+ 7.85656x_1 + 7.85656x_2 \\ &- 0.95187x_1^2 - 0.95187x_2^2 + 1.58990x_1x_2. \end{aligned} \quad (1.31)$$

É interessante ressaltar que para um valor  $A = 0.001$  a pendente na mudança de classificação está bem definida, com limiares de classificação na superfície  $h_{\mathbf{c}}(\mathbf{x}) = 0$ .



# Referências

**Bibliografia** ..... 25

Livros

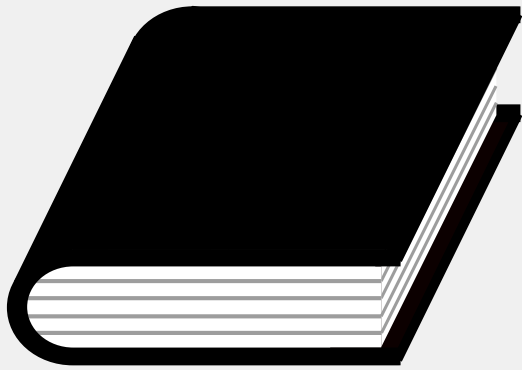
Artigos

Outras fontes

**Índice** ..... 27







## Bibliografia

### Livros

---

- [GCL07] K.O. Geddes, S.R. Czapor e G. Labahn. *Algorithms for Computer Algebra*. Springer US, 2007. ISBN: 9780585332475. URL: <https://books.google.com.br/books?id=9f0UwkkRxT4C> (ver página 20).
- [KK10] D.G. Kleinbaum e M. Klein. *Logistic Regression: A Self-Learning Text*. Statistics for Biology and Health. Springer New York, 2010. ISBN: 9781441917423. URL: <https://books.google.com.br/books?id=FTVDAAAQBAJ> (ver página 13).
- [Kur+01] V. Kurkova et al. *Artificial Neural Nets and Genetic Algorithms: Proceedings of the International Conference in Prague, Czech Republic, 2001*. Springer computer science. Springer Vienna, 2001. ISBN: 9783211836514. URL: <https://books.google.com.br/books?id=qXguv62cn1YC> (ver página 13).
- [Zip12] R. Zippel. *Effective Polynomial Computation*. The Springer International Series in Engineering and Computer Science. Springer US, 2012. ISBN: 9781461531883. URL: <https://books.google.com.br/books?id=mmYECAAAQBAJ> (ver página 20).

### Artigos

---

### Outras fontes

---





## Índice Remissivo

### F

#### Função

Logística.....	13
Logit .....	13
Sigmoide.....	13

### M

Multiconjunto .....	21
Multiset.....	21

### P

Polinômio multivariante .....	20
-------------------------------	----

### R

#### Regressão

Logística $f_{\mathbf{c}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$ .....	18
Logística $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$ .....	14
Logística polinomial $f_{\mathbf{c}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^{20}$	
Logística polinomial $f_{\mathbf{c}}(x) : \mathbb{R} \rightarrow \mathbb{R}$	16





*Este livro foi produzido por Fernando Pujaico Rivera, editado e diagramado usando  $\LaTeX$ , um tipo de fonte com codificação: T1, família: ptm, série: m, e tamanho: 12 pt , para ser impresso num papel tamanho **XXXxXXXcm**, **XXXXXXXXXXXX** 2020.*