D1EAD – Análise Estatística para Ciência de Dados 2021.1

Regressão Logística

Aula baseada no curso *Machine Learning* do prof. Andrew Ng (Coursera)

Prof. Ricardo Sovat
sovat@ifsp.edu.br
Prof. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas

Classificação Binária

Vestibular: Aprovado / Reprovado

E-mail: Spam / Não-spam

Tumor: Maligno / Benigno

Assinatura: Real / Falsa

Tipo de app: Gratuito / Pago

....

Classificação Binária

Vestibular: Aprovado / Reprovado

E-mail: Spam / Não-spam

Tumor: Maligno / Benigno

Assinatura: Real / Falsa

Tipo de app: Gratuito / Pago

....

Variáveis categóricas com apenas 2 valores

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

0: Classe Negativa

1: Classe Positiva

Regressão Linear

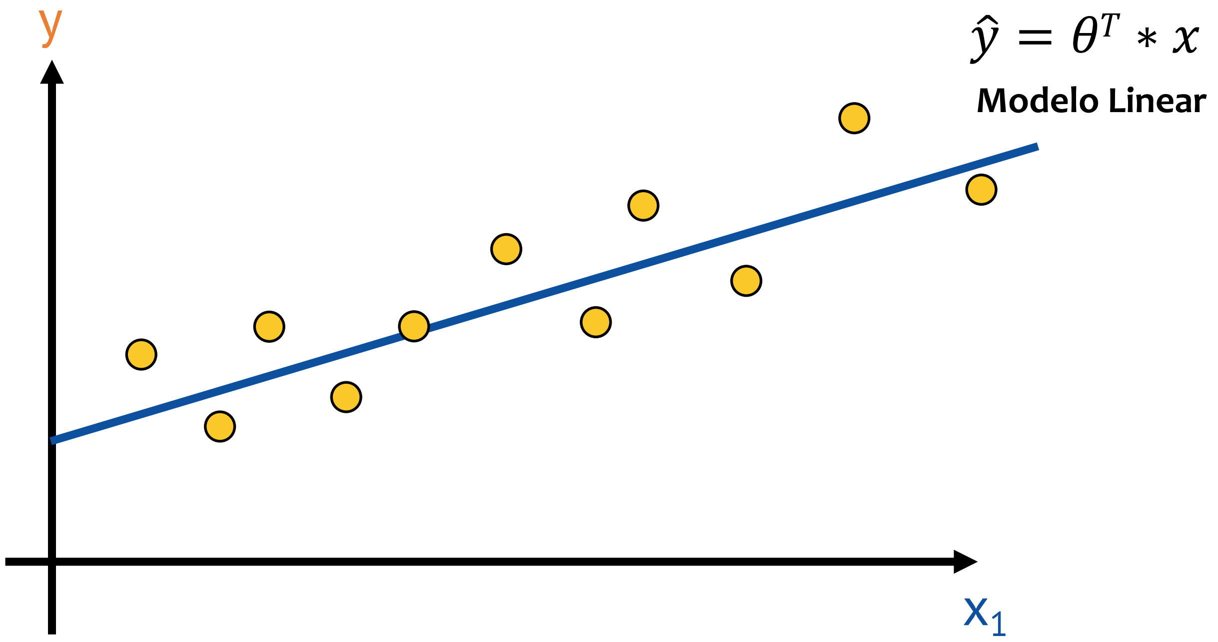
$$h_\theta(x) = \hat{y} = \theta_0 + \theta_1 * x_1 + \cdots + \theta_n * x_n$$

$$h_\theta(x) = \hat{y} = \theta^T * x$$

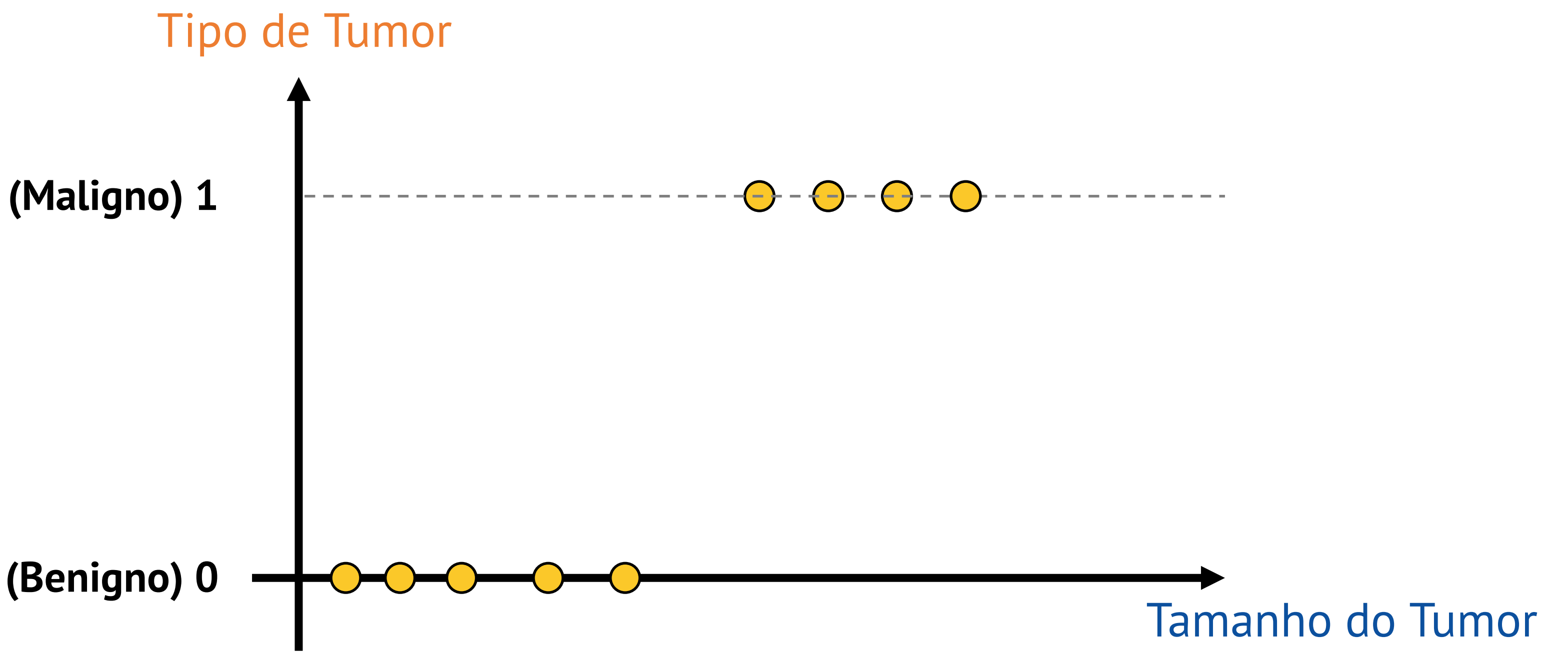
$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \cdots \\ \theta_n \end{bmatrix} \qquad x = \begin{bmatrix} \overset{x_0}{1} \\ x_1 \\ \cdots \\ x_n \end{bmatrix}$$

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

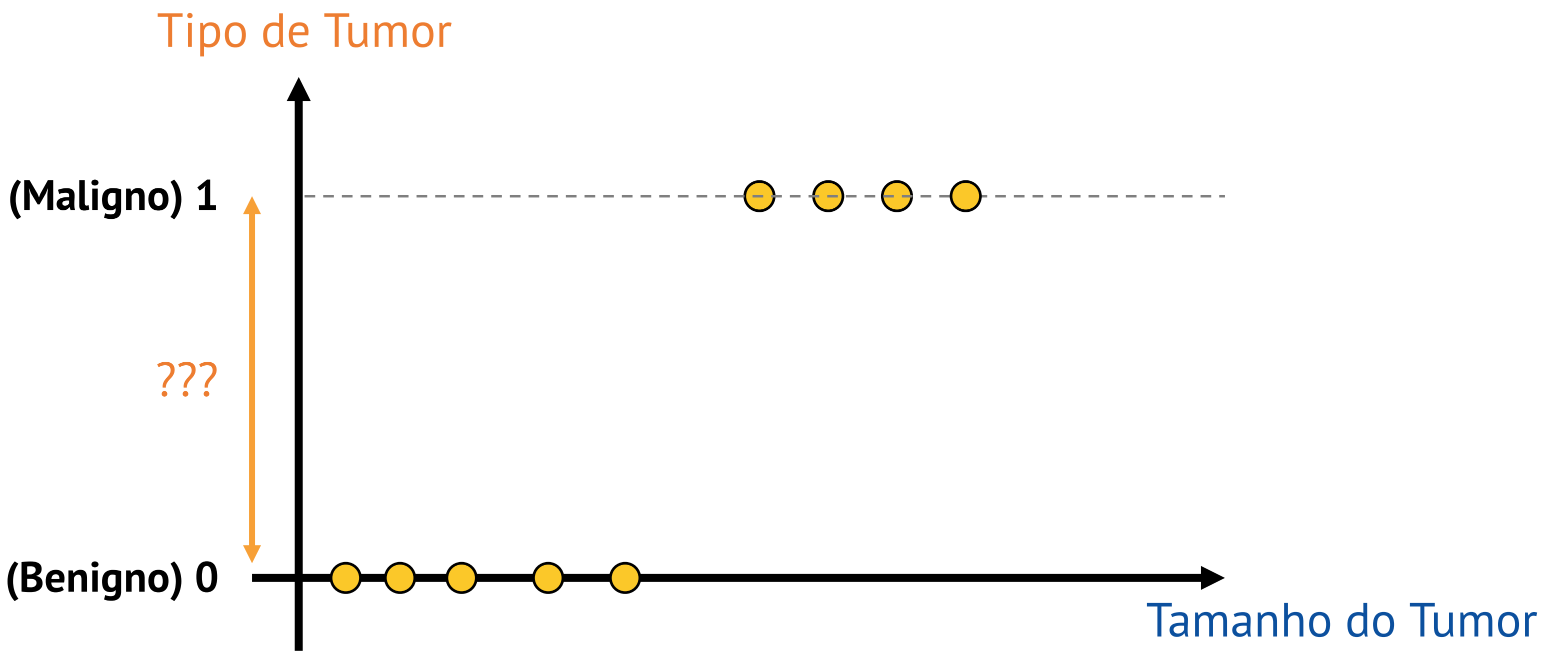
Em machine learning, **vetores** são comumente representados como **vetores coluna**.



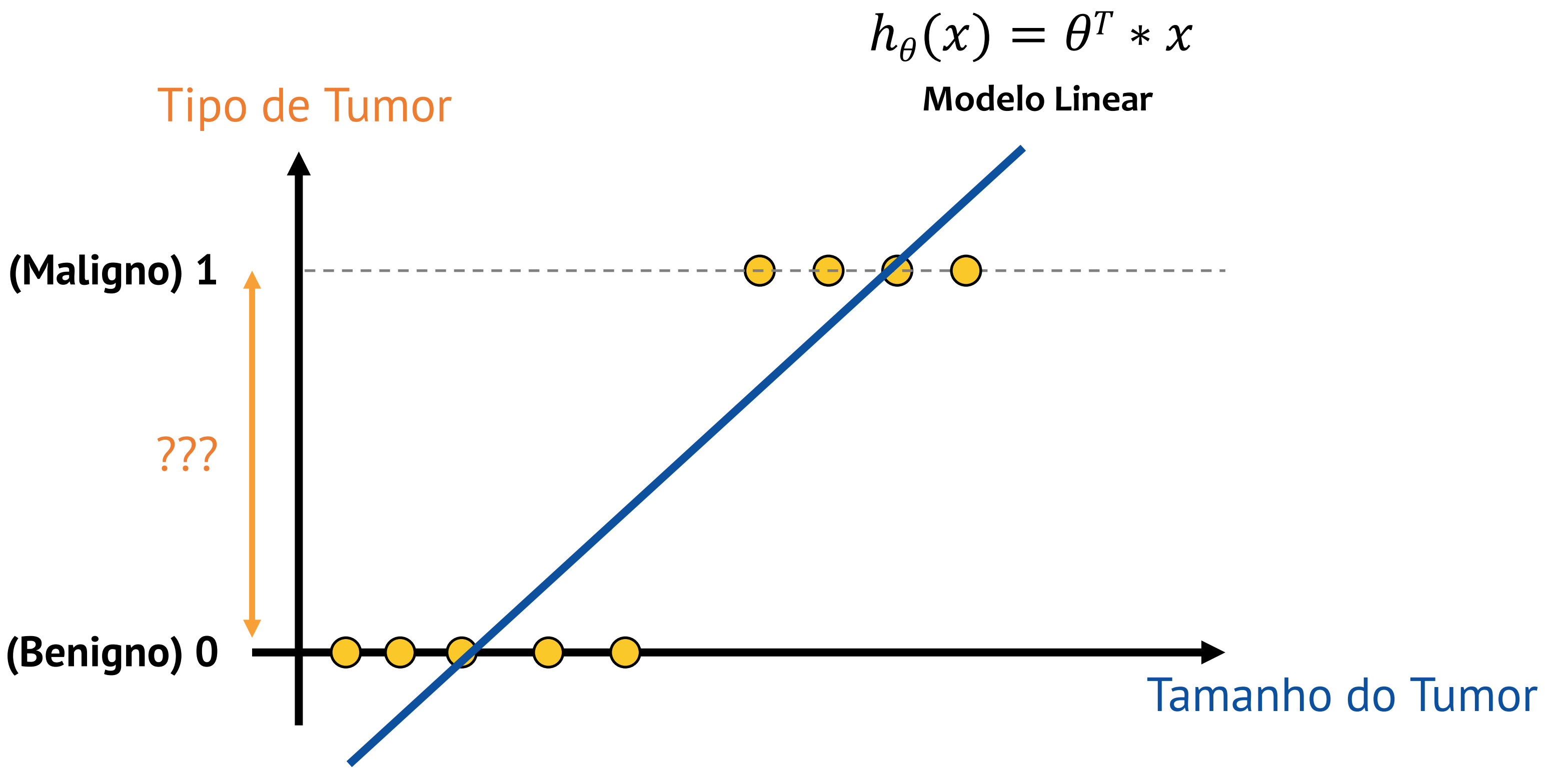
Problema de Classificação



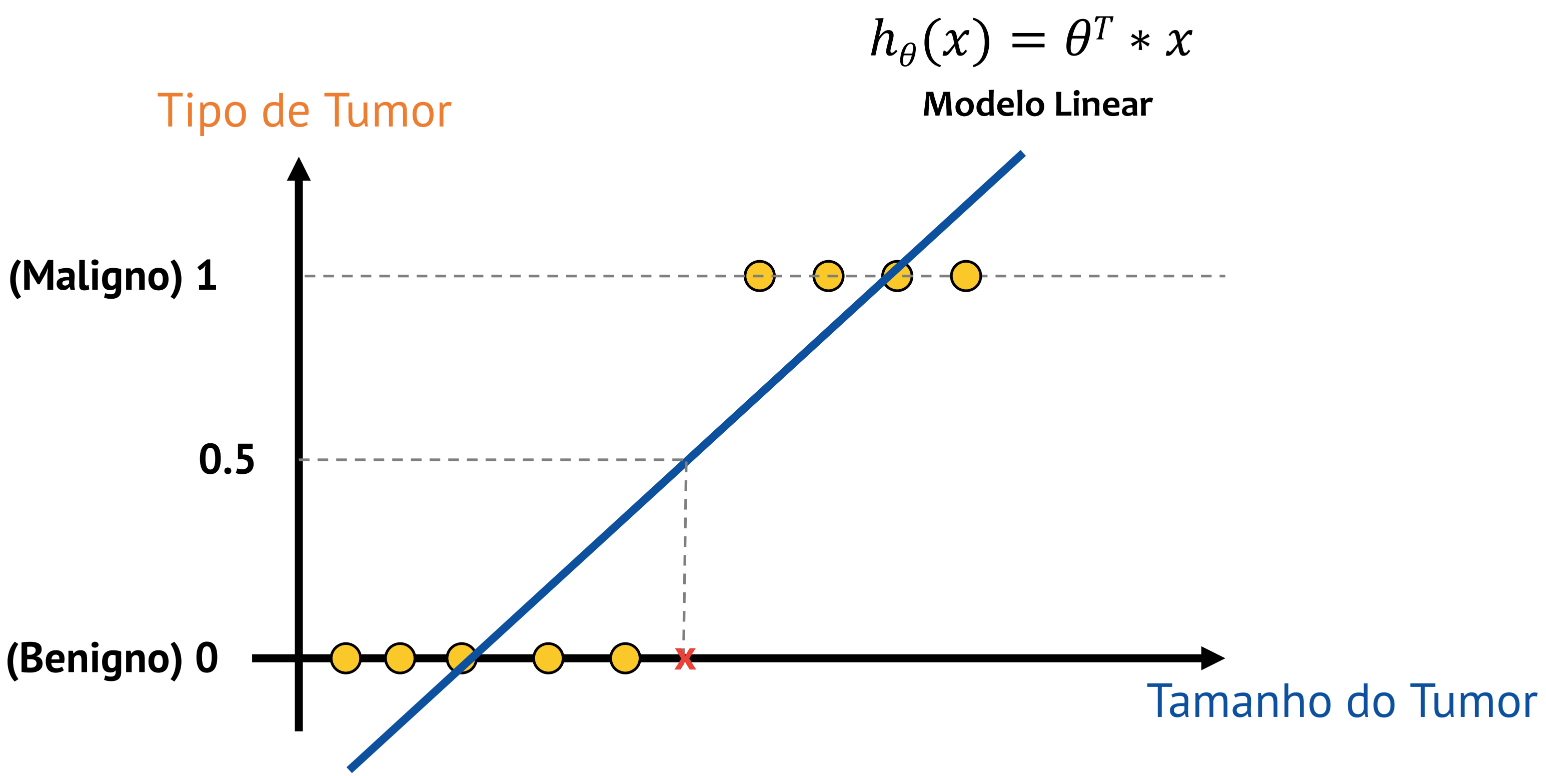
Problema de Classificação



Problema de Classificação



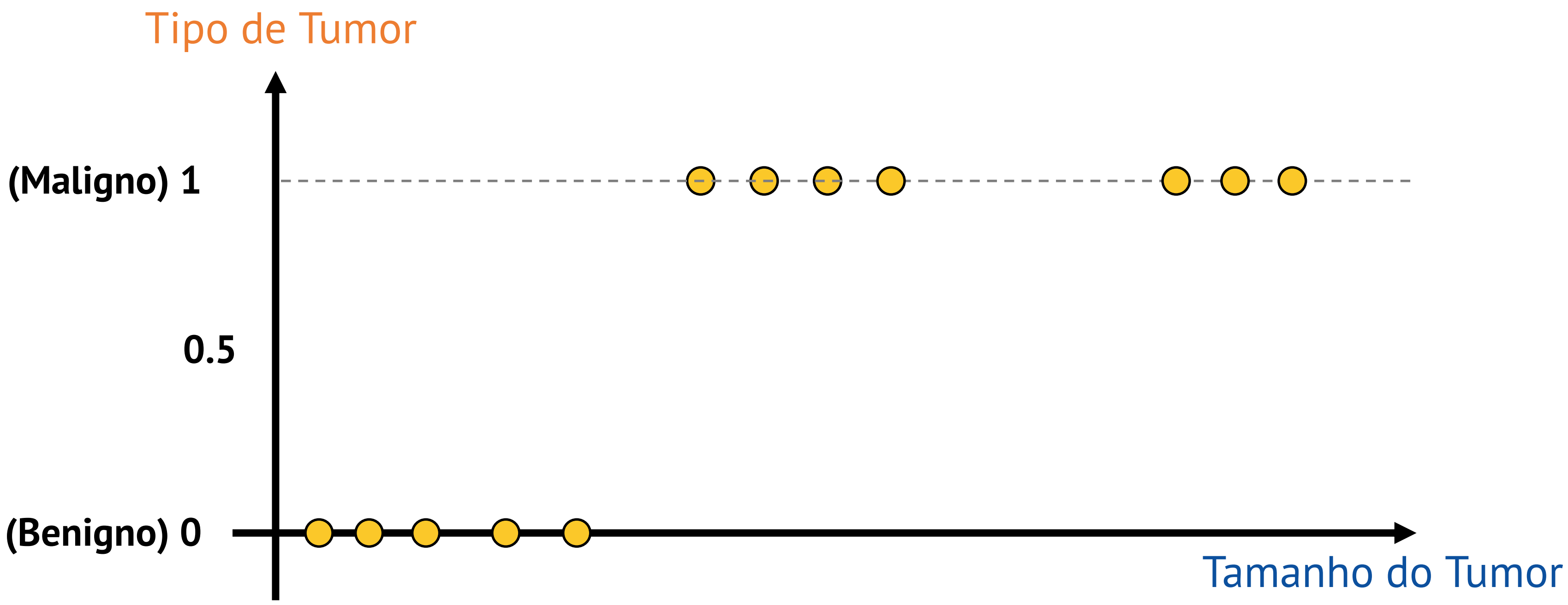
Problema de Classificação



Se $\boldsymbol{h_\theta(x)} \geq \mathbf{0.5}$, classifique $\hat{y} = 1$

Se $\boldsymbol{h_{\theta(x)}} < \mathbf{0}$, classifique $\hat{y} = 0$

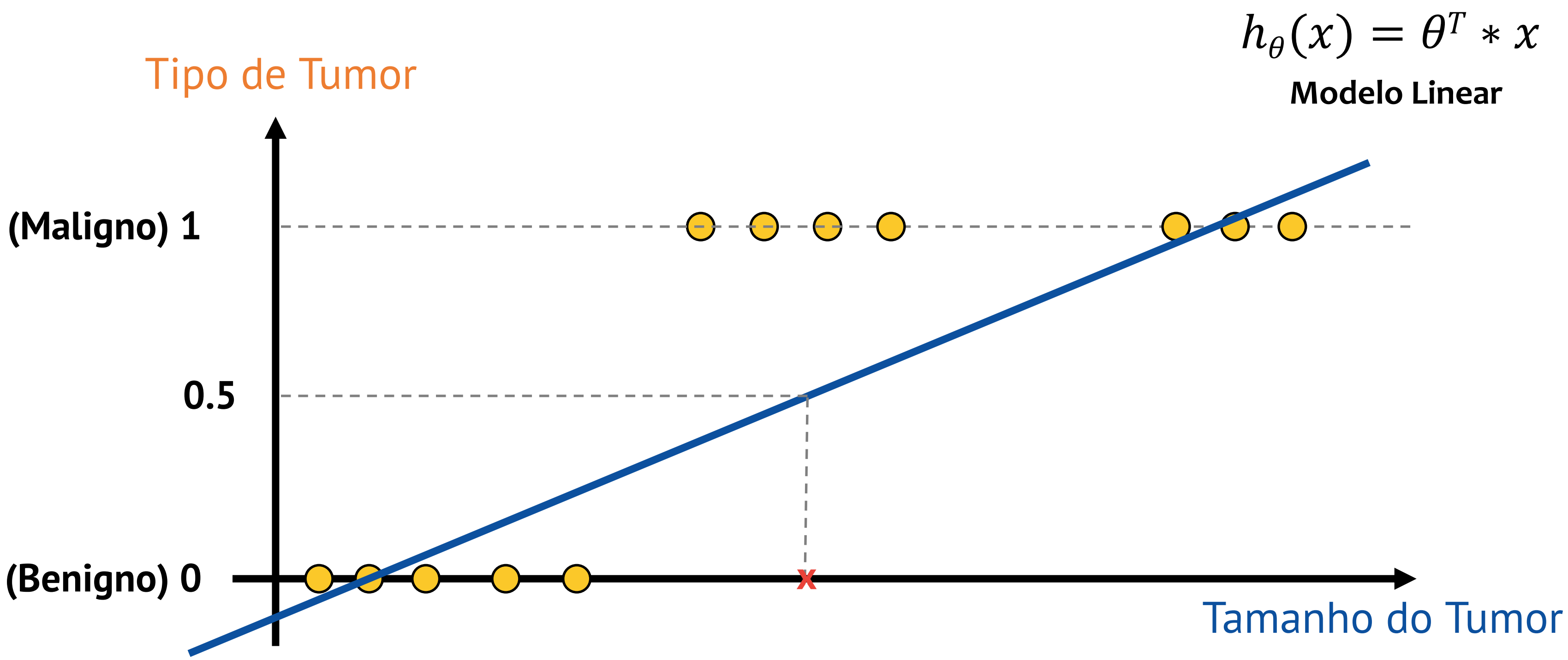
Problema de Classificação



Se $\boldsymbol{h_\theta}(\boldsymbol{x}) \geq \mathbf{0.5}$, classifique $\hat{y} = 1$

Se $\boldsymbol{h}_{\theta(\boldsymbol{x})} < \mathbf{0}$, classifique $\hat{y} = 0$

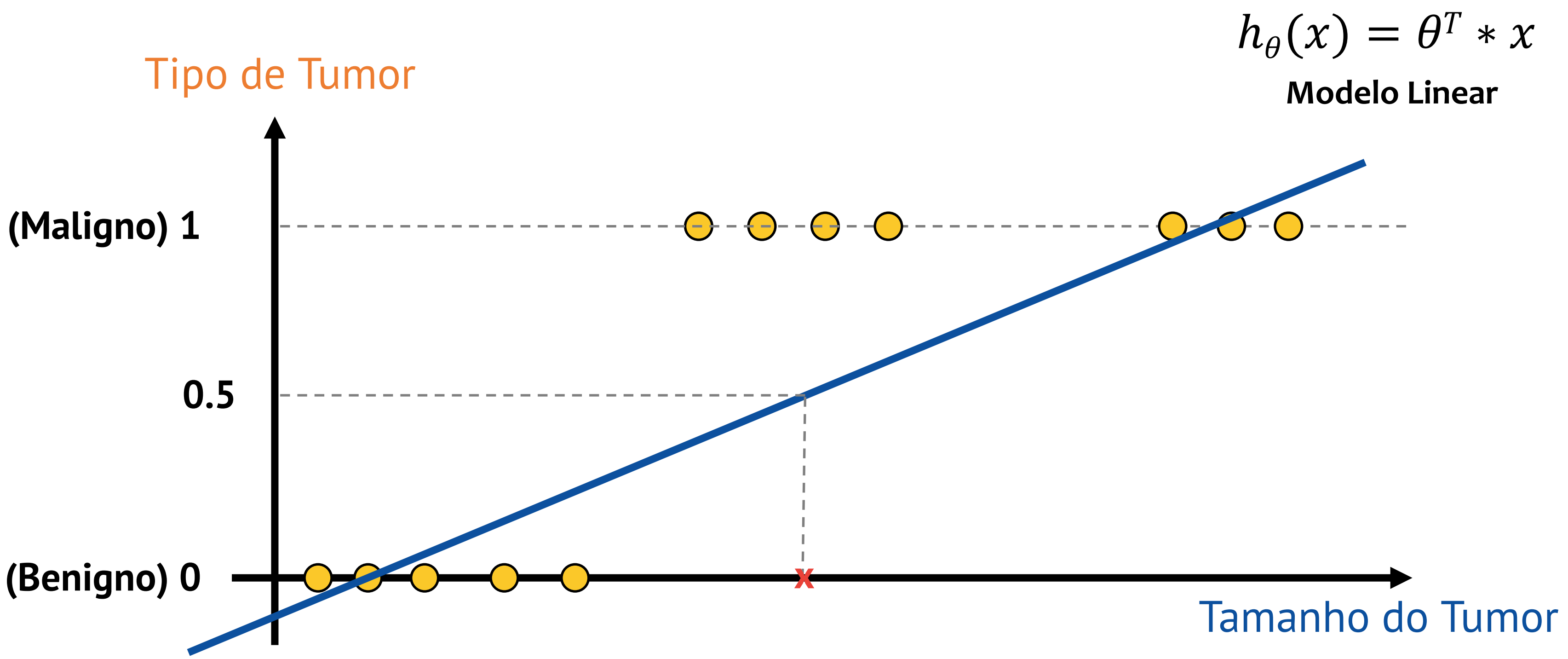
Problema de Classificação



Se $\boldsymbol{h_\theta(x)} \geq \mathbf{0.5}$, classifique $\hat{y} = 1$

Se $\boldsymbol{h_{\theta(x)}} < \mathbf{0}$, classifique $\hat{y} = 0$

Problema de Classificação

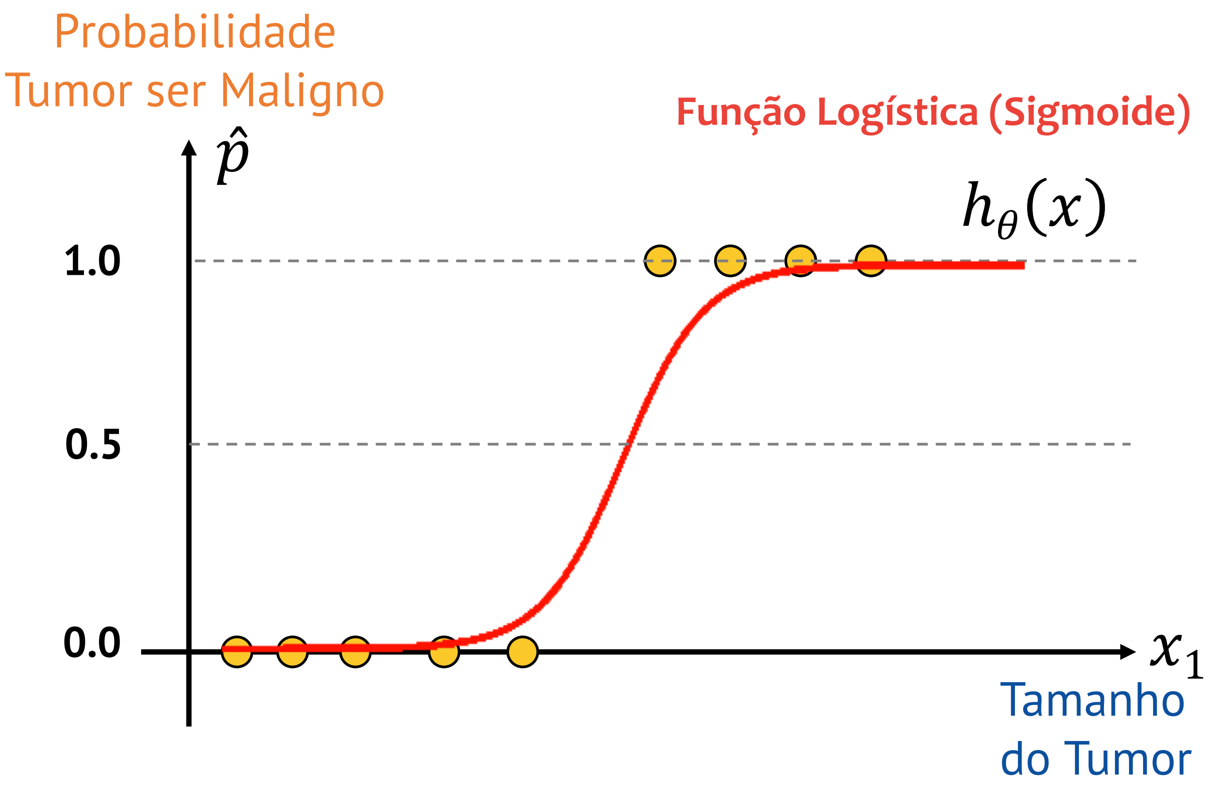


Aplicar uma **Regressão Linear** em problemas de **classificação** **não parece ser uma boa ideia**

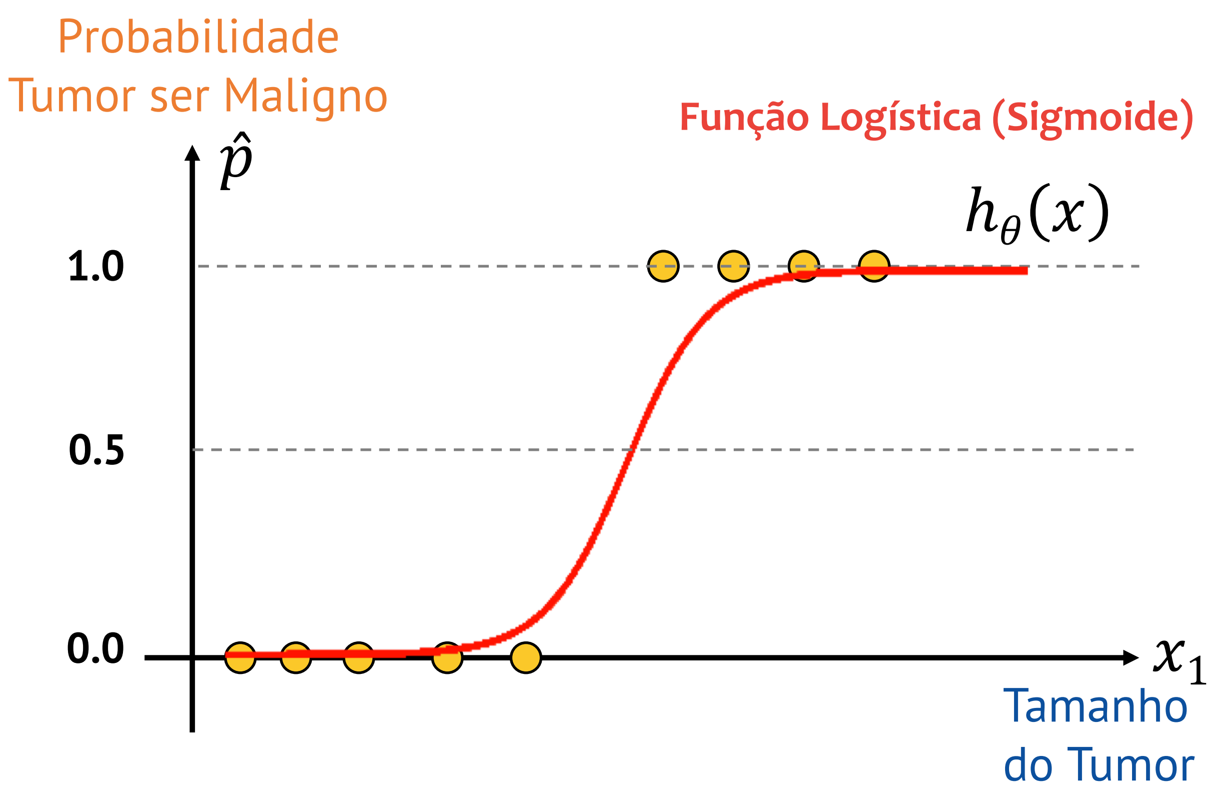
Se $\boldsymbol{h_\theta(x) \geq 0.5}$, classifique $\hat{y} = 1$

Se $\boldsymbol{h_\theta(x) < 0}$, classifique $\hat{y} = 0$

Regressão Logística



Regressão Logística



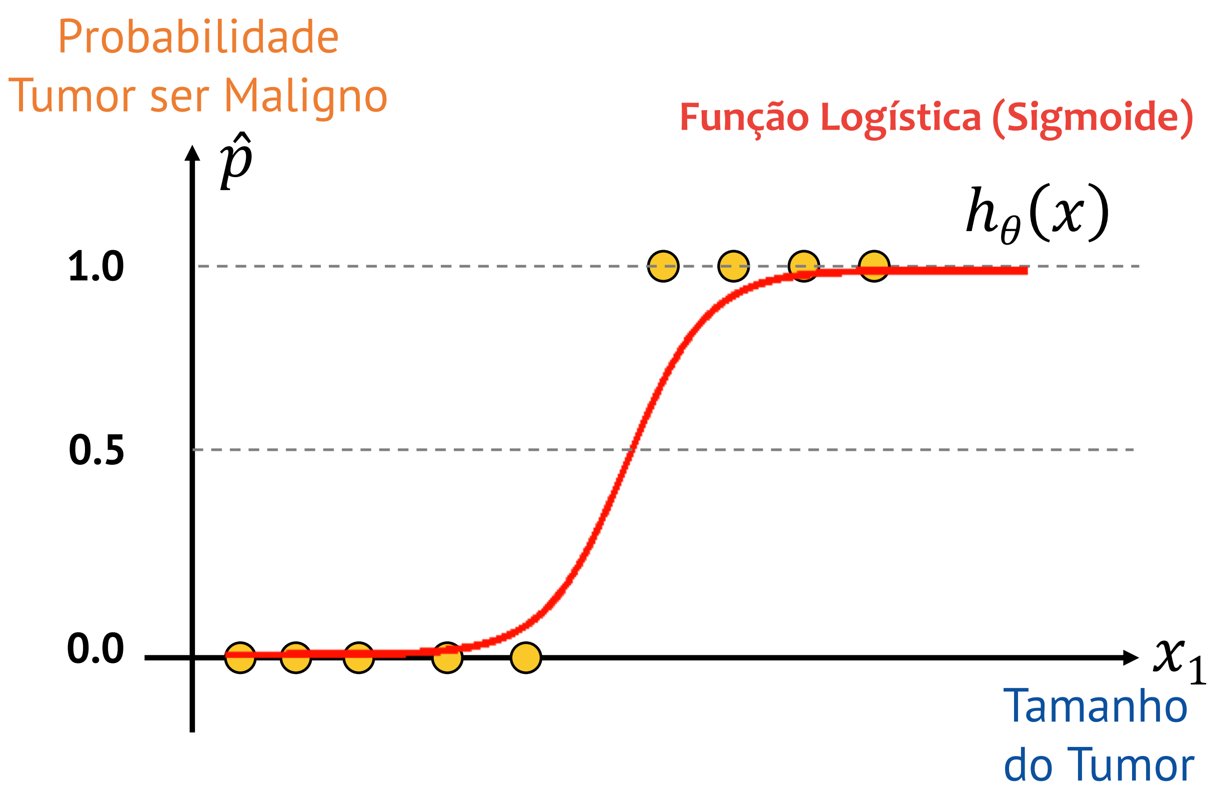
Modelo de Regressão Logística

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

$h_\theta(x)$ = Probabilidade estimada que a observação representada pelo vetor x, com os parâmetros θ, seja da **classe positiva ($\mathbf{y = 1}$)**

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

Regressão Logística



Modelo de Regressão Logística

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

$h_\theta(x)$ = Probabilidade estimada que a observação representada pelo vetor x, com os parâmetros θ, seja da **classe positiva ($\mathbf{y = 1}$)**

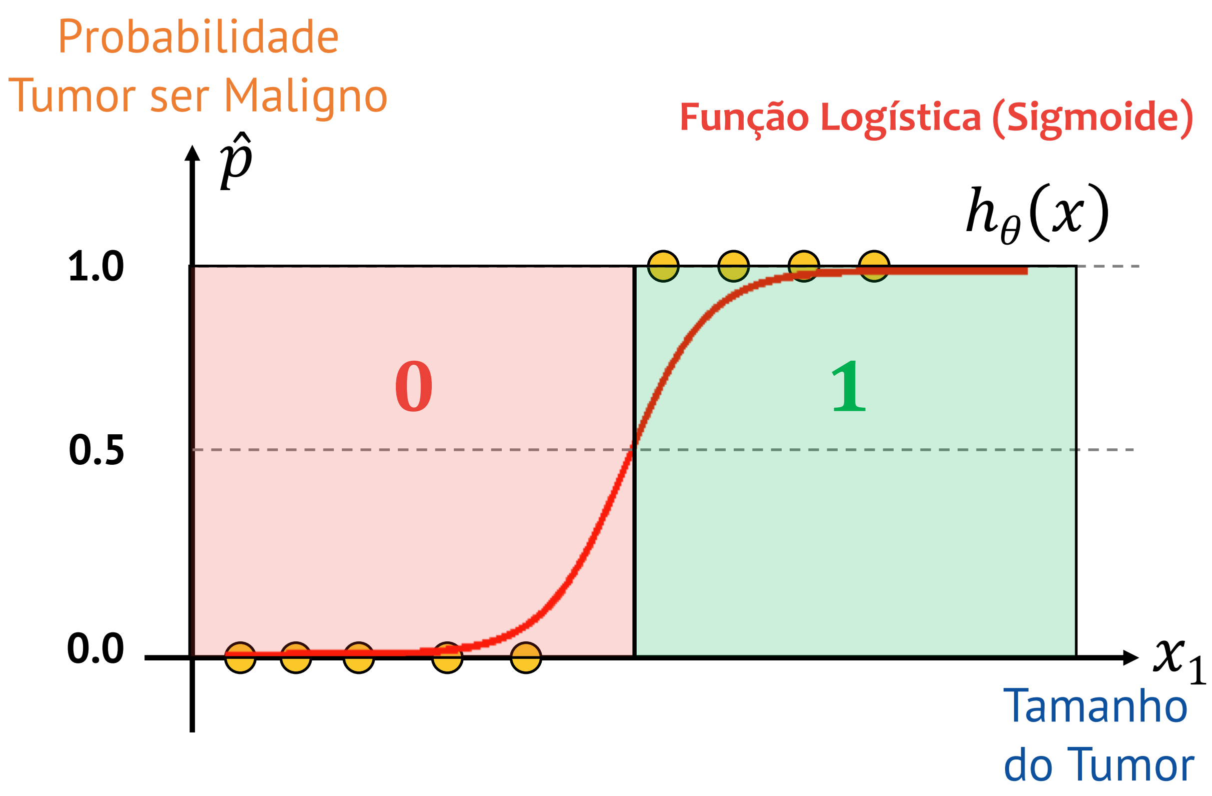
$h_\theta(x) = \mathrm{P}(\mathrm{y} = 1|\mathrm{x}; \theta)$

P.ex: Se

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ tamaho\ do \\ tumor \end{bmatrix} \qquad h_\theta(x) = \mathbf{0.7}$$

A **chance** do tumor ser **maligno** é de **70%**

Regressão Logística



Modelo de Regressão Logística

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

$h_\theta(x)$ = Probabilidade estimada que a observação representada pelo vetor x, com os parâmetros θ, seja da **classe positiva $(\mathbf{y = 1})$**

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

Classificação

$$\hat{y} = \begin{cases} 1 & se\ \hat{p} \geq 0.5 \\ 0 & se\ \hat{p} < 0.5 \end{cases}$$

P.ex: Se

$$x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ tamaho\ do \\ tumor \end{bmatrix} \qquad h_\theta(x) = \mathbf{0.7}$$

A **chance** do tumor ser **maligno** é de **70%**

Regressão Logística

Modelo de Regressão Logística

$$\hat{p} = h_\theta(x) = \sigma\big(f_\theta(x)\big) = \sigma(\theta^T * x) = \frac{1}{1 + e^{-(\theta^T * x)}}$$

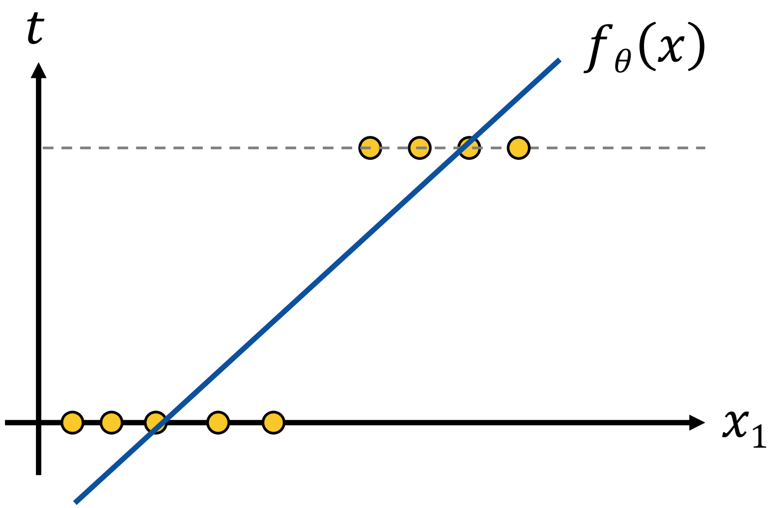
Regressão Logística

Modelo de Regressão Logística

$$\hat{p} = h_\theta(x) = \sigma\big(f_\theta(x)\big) = \sigma(\theta^T * x) = \frac{1}{1 + e^{-(\theta^T * x)}}$$

Modelo Linear

$$t = f_\theta(x) = \theta^T * x$$



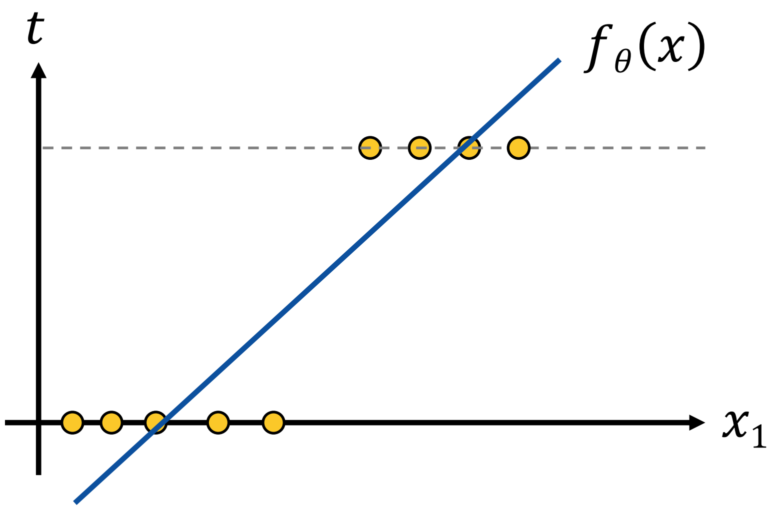
Regressão Logística

Modelo de Regressão Logística

$$\hat{p} = h_\theta(x) = \sigma\big(f_\theta(x)\big) = \sigma(\theta^T * x) = \frac{1}{1 + e^{-(\theta^T * x)}}$$

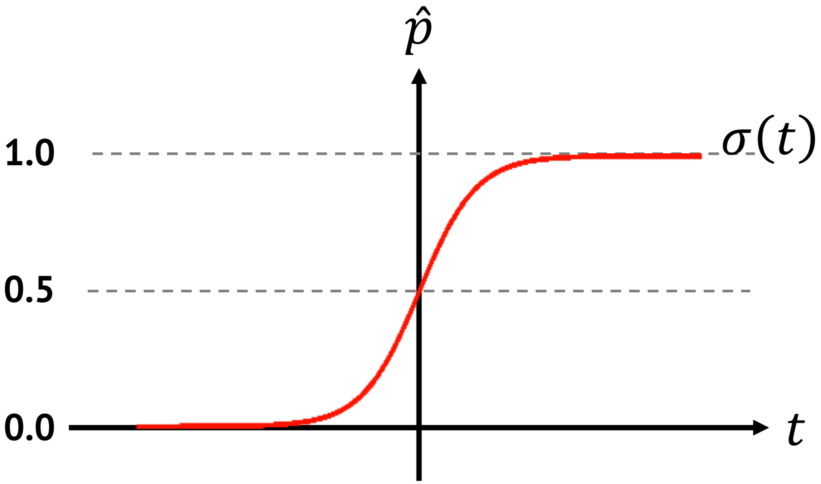
Modelo Linear

$$t = f_\theta(x) = \theta^T * x$$

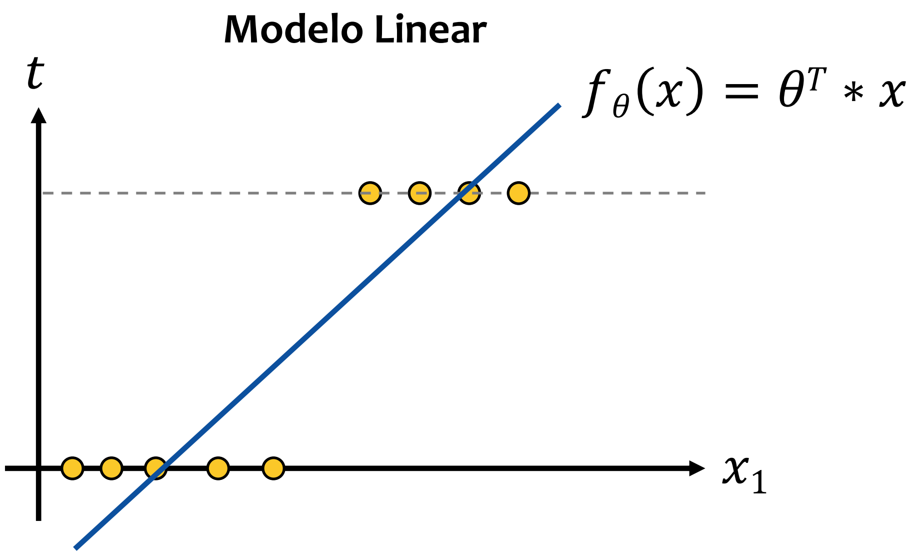


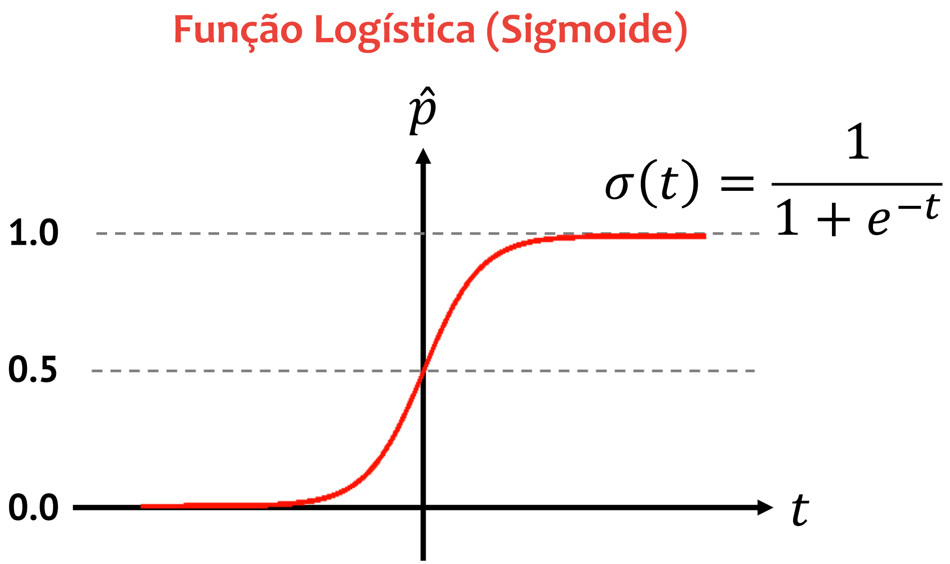
Função Logística (Sigmoide)

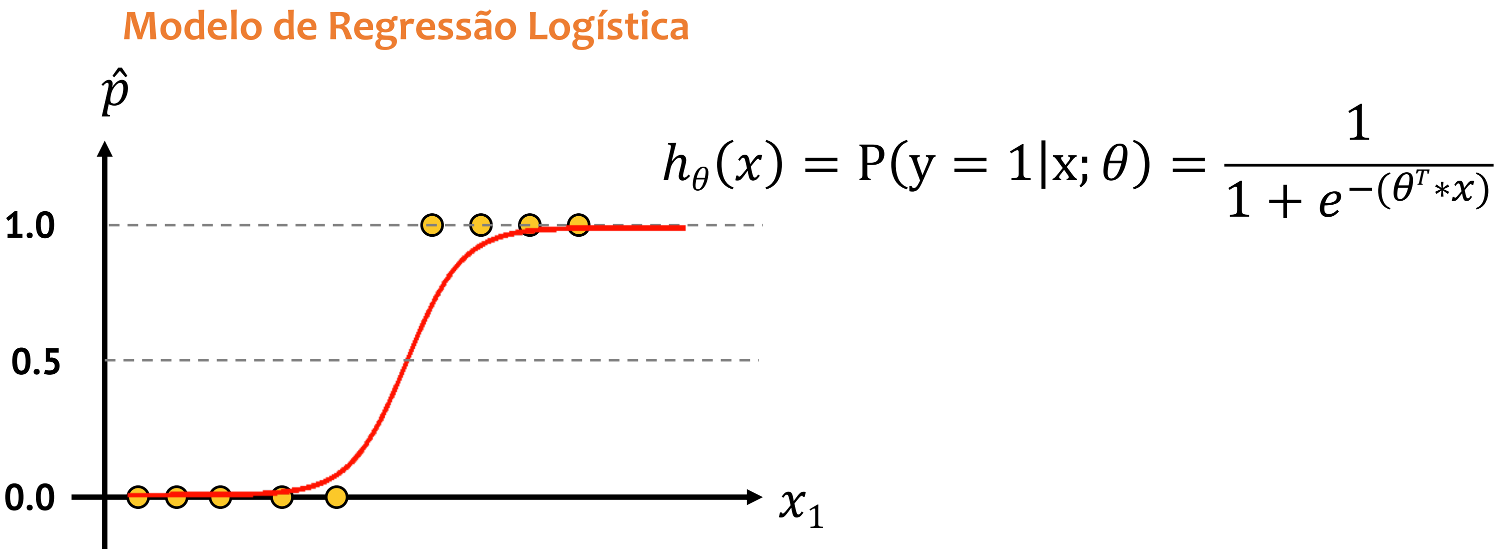
$$\hat{p} = \sigma(t) = \frac{1}{1 + e^{-t}}$$



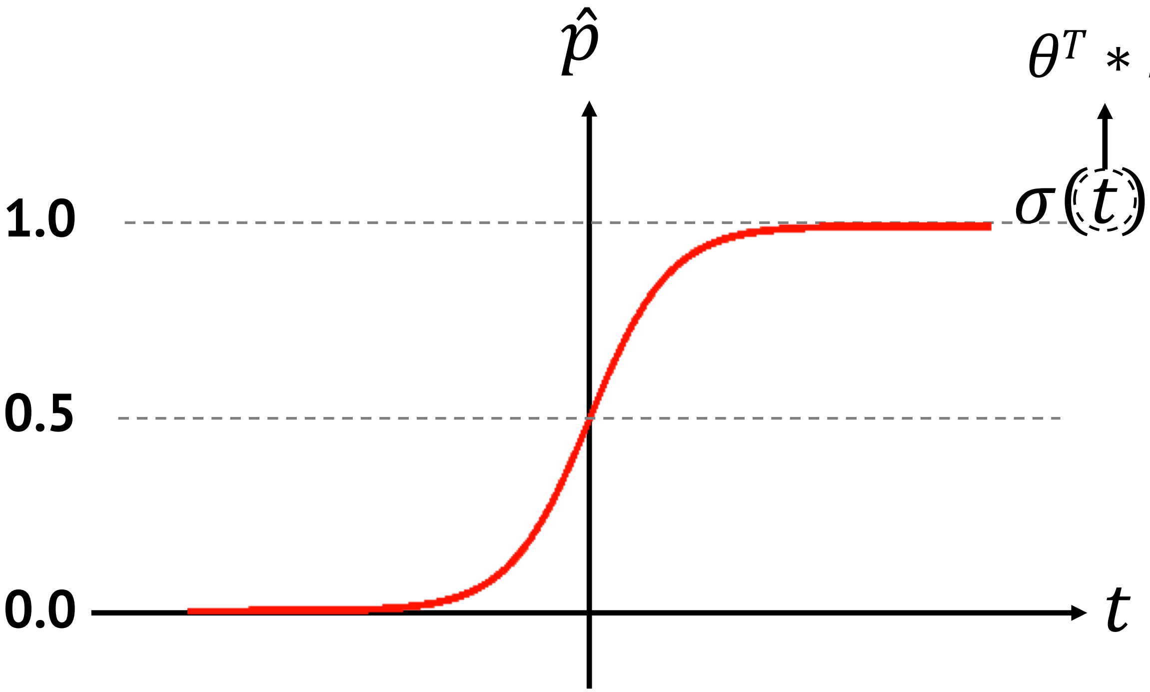
Regressão Logística







Função Logística (Sigmoide)



$\theta^T * x$

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

$\sigma(t)$ retorna um número entre **0** e **1**

$\sigma(t) < 0.5$ quando $t < 0$

$\sigma(t) \geq 0.5$ quando $t \geq 0$

O **modelo de Regressão Logística** prediz:

- **classe 1 (positiva)** se $(\theta^T * x) \geq 0$
- **classe 0 (negativa):** se $(\theta^T * x) < 0$

Treinamento e Função de Custo

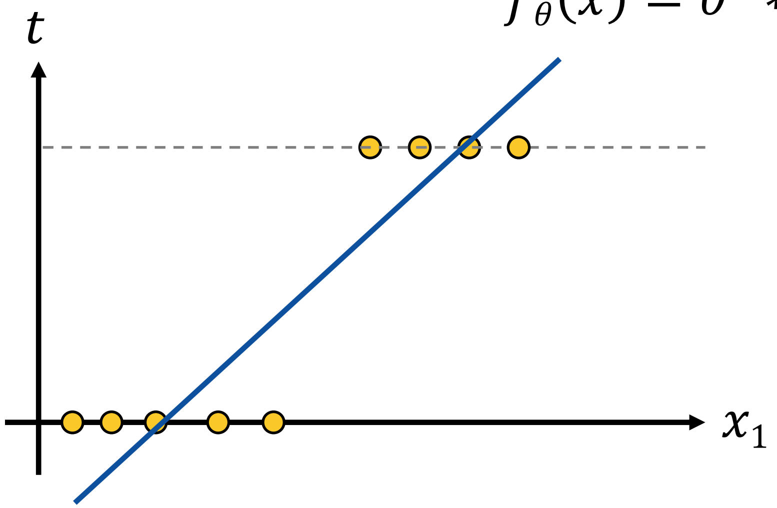
Achar o vetor de parâmetros $\boldsymbol{\theta}$ de maneira que o modelo estime:

- **probabilidades altas** para **instâncias positivas** $(\boldsymbol{y} = \mathbf{1})$;
- **probabilidades baixas** para **instâncias negativas** $(\boldsymbol{y} = \mathbf{0})$

Função de Custo

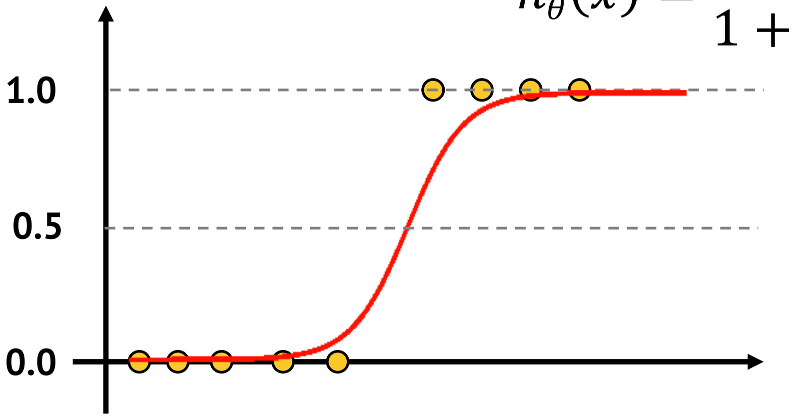
$$c(\theta) = \begin{cases} -\log(h_\theta(x)) & se\ y = 1 \\ -\log(1 - h_\theta(x)) & se\ y = 0 \end{cases}$$

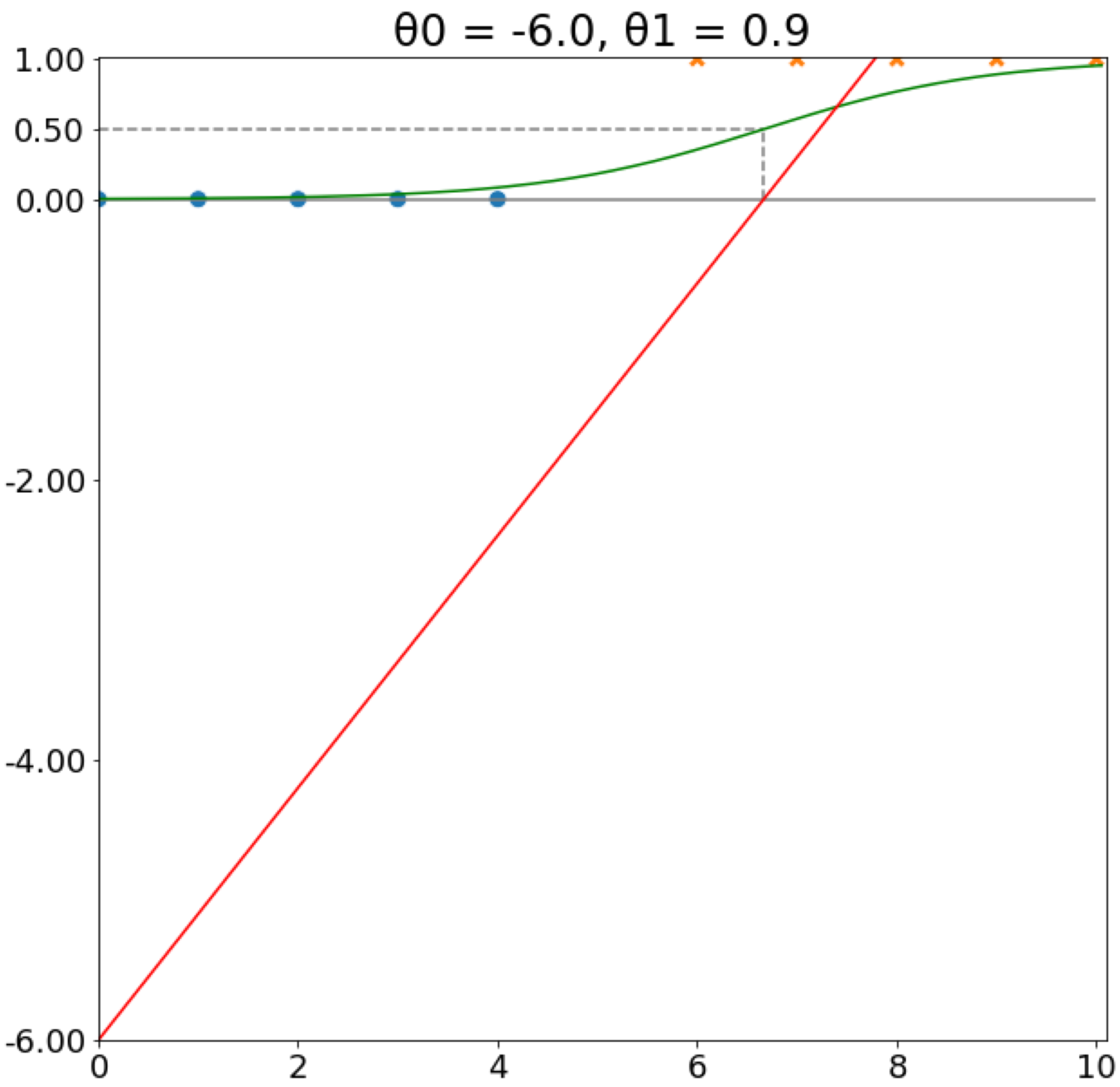
$$f_\theta(x) = \theta^T * x$$

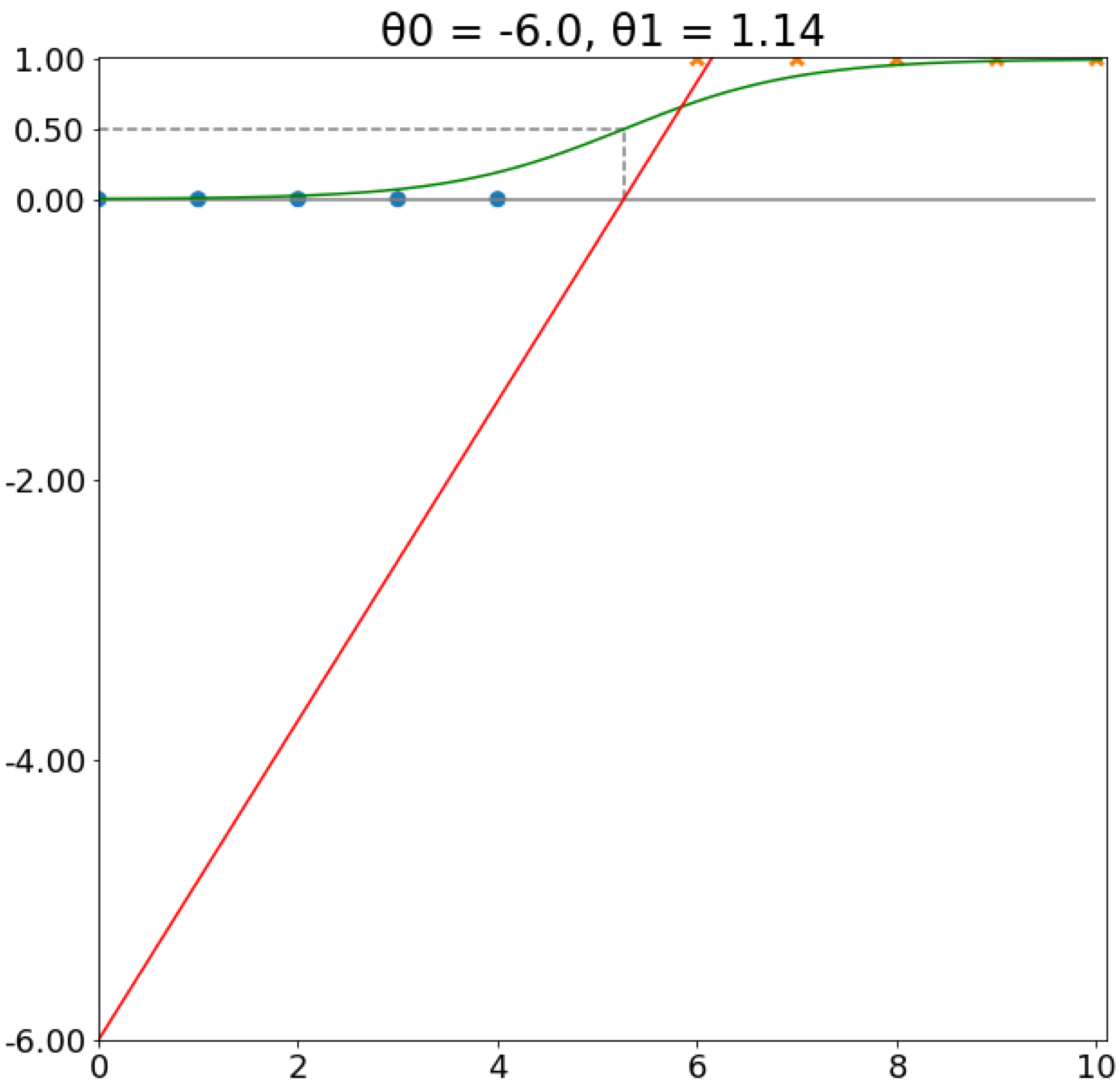


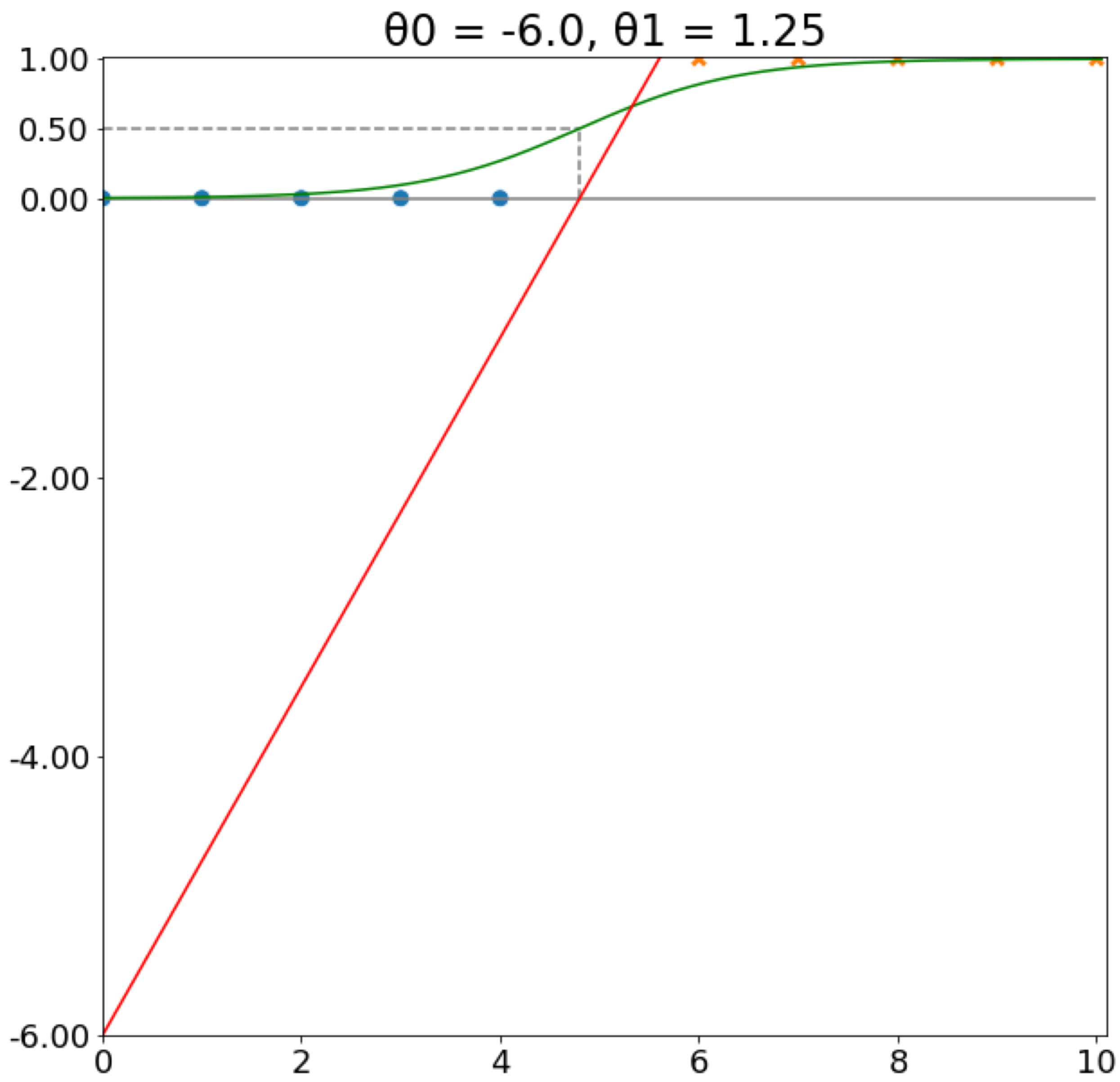
Modelo de Regressão Logística

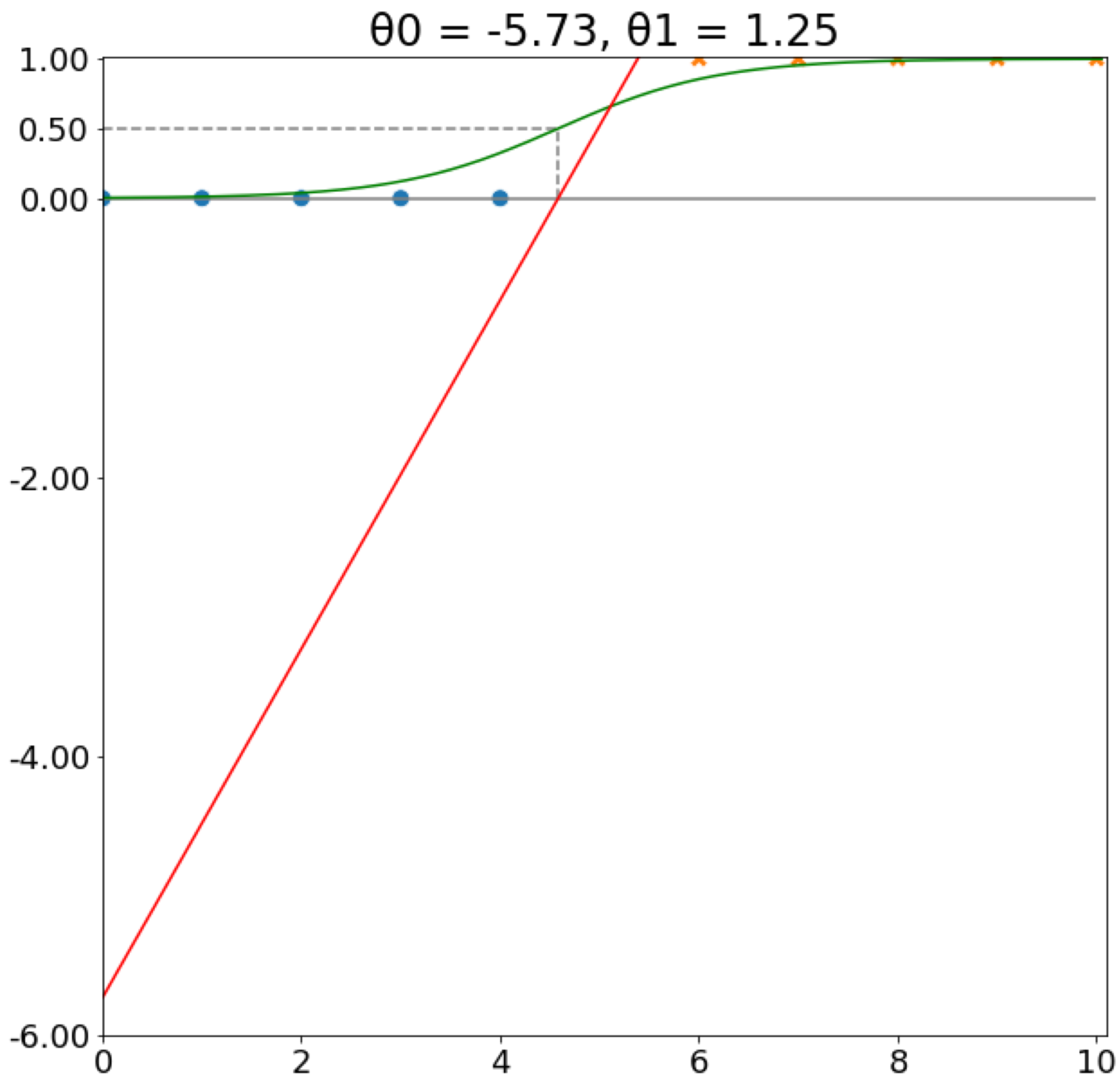
$$h_\theta(x) = \frac{1}{1 + e^{-(\theta^T * x)}}$$

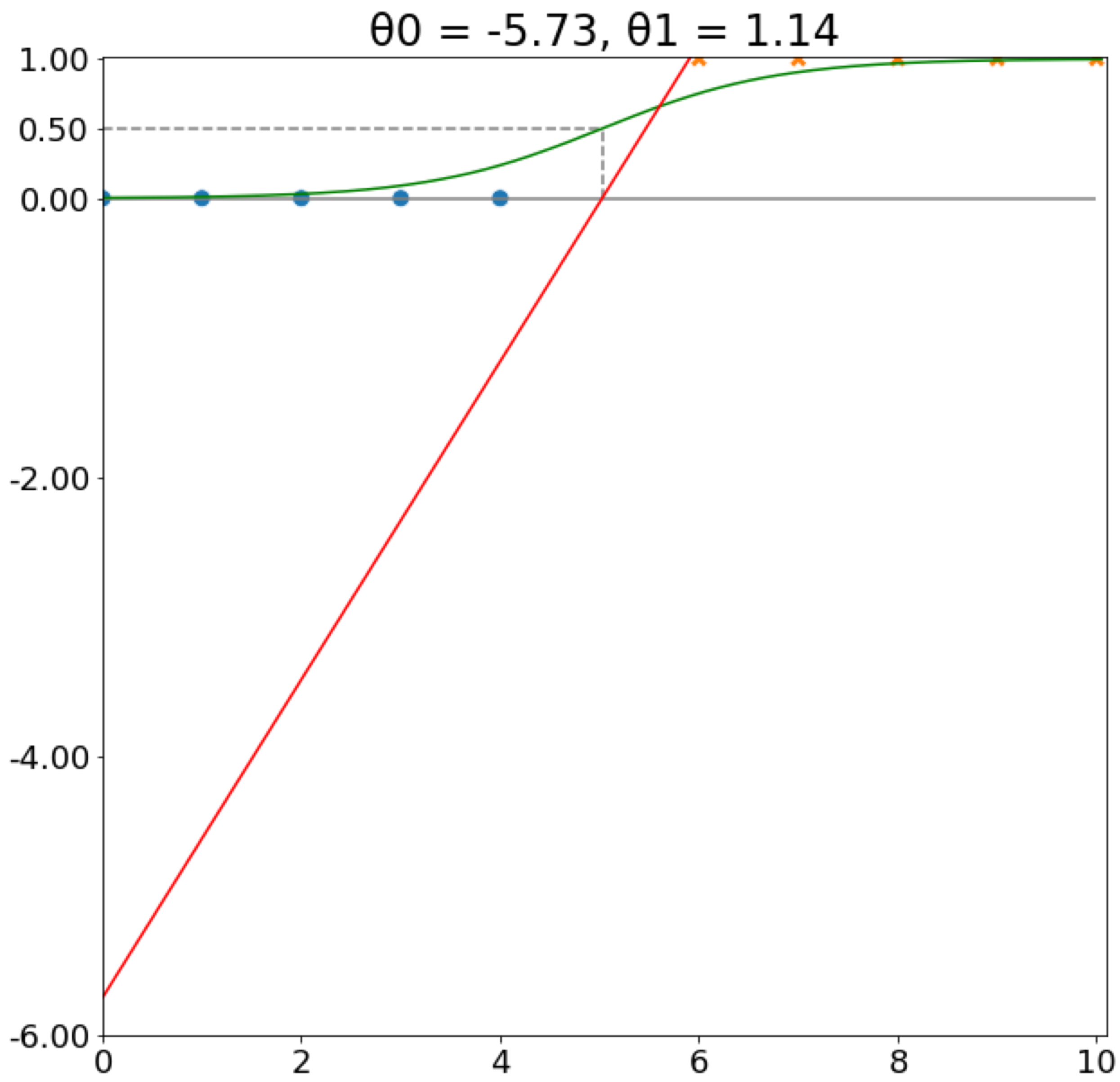


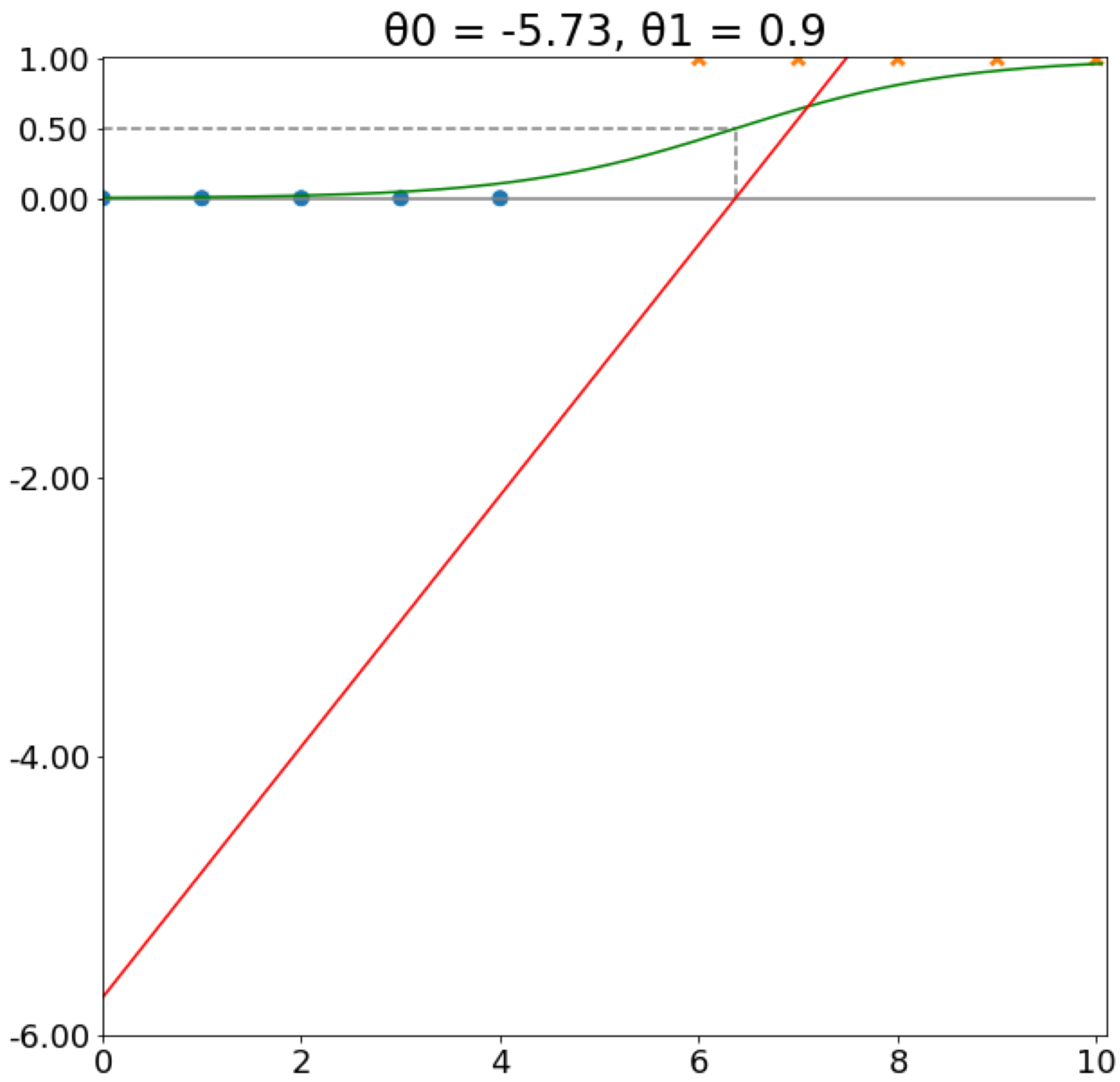

θ0 = -6.0, θ1 = 0.9
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

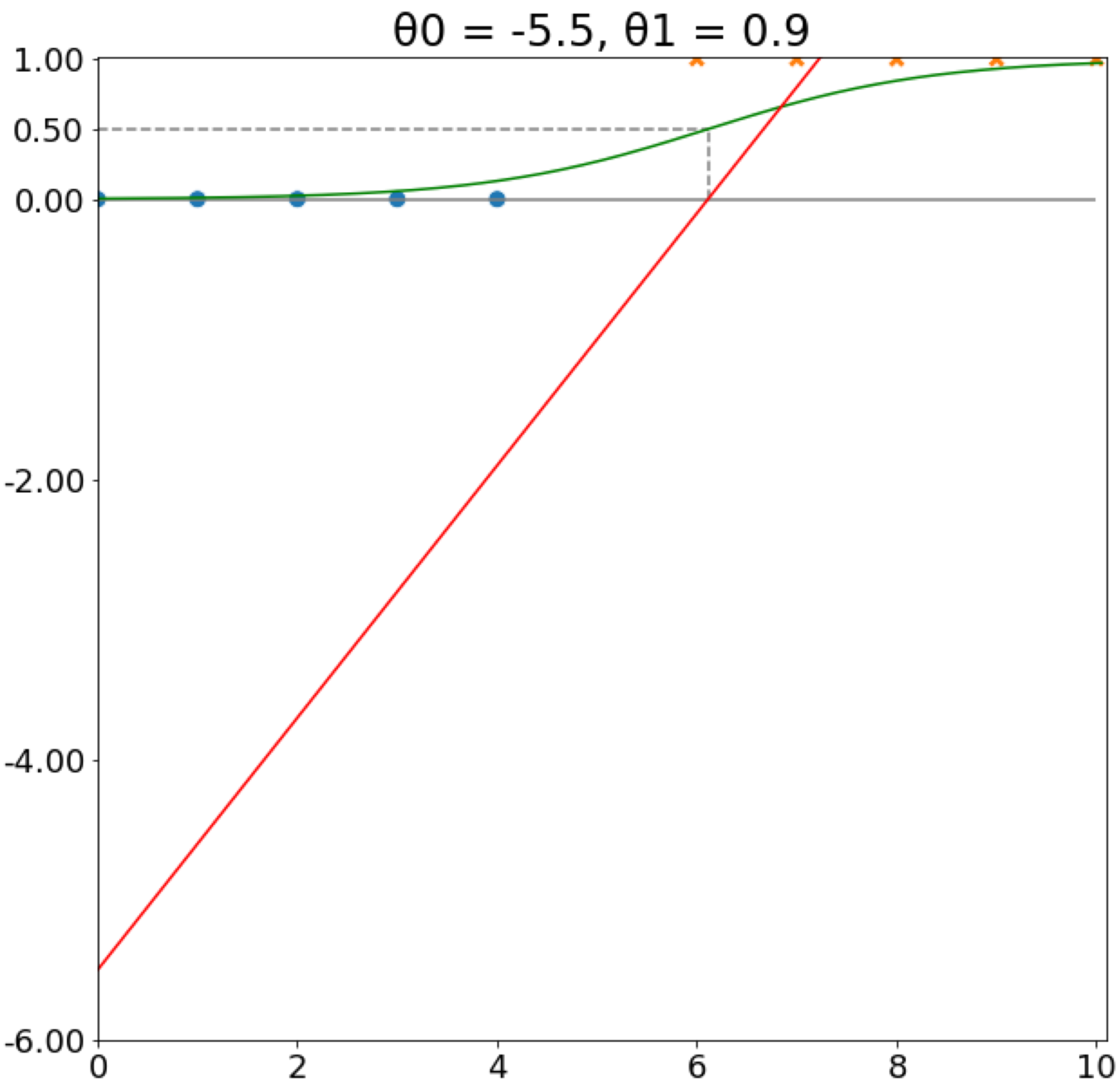

θ0 = -6.0, θ1 = 1.14
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10


θ0 = -6.0, θ1 = 1.25
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

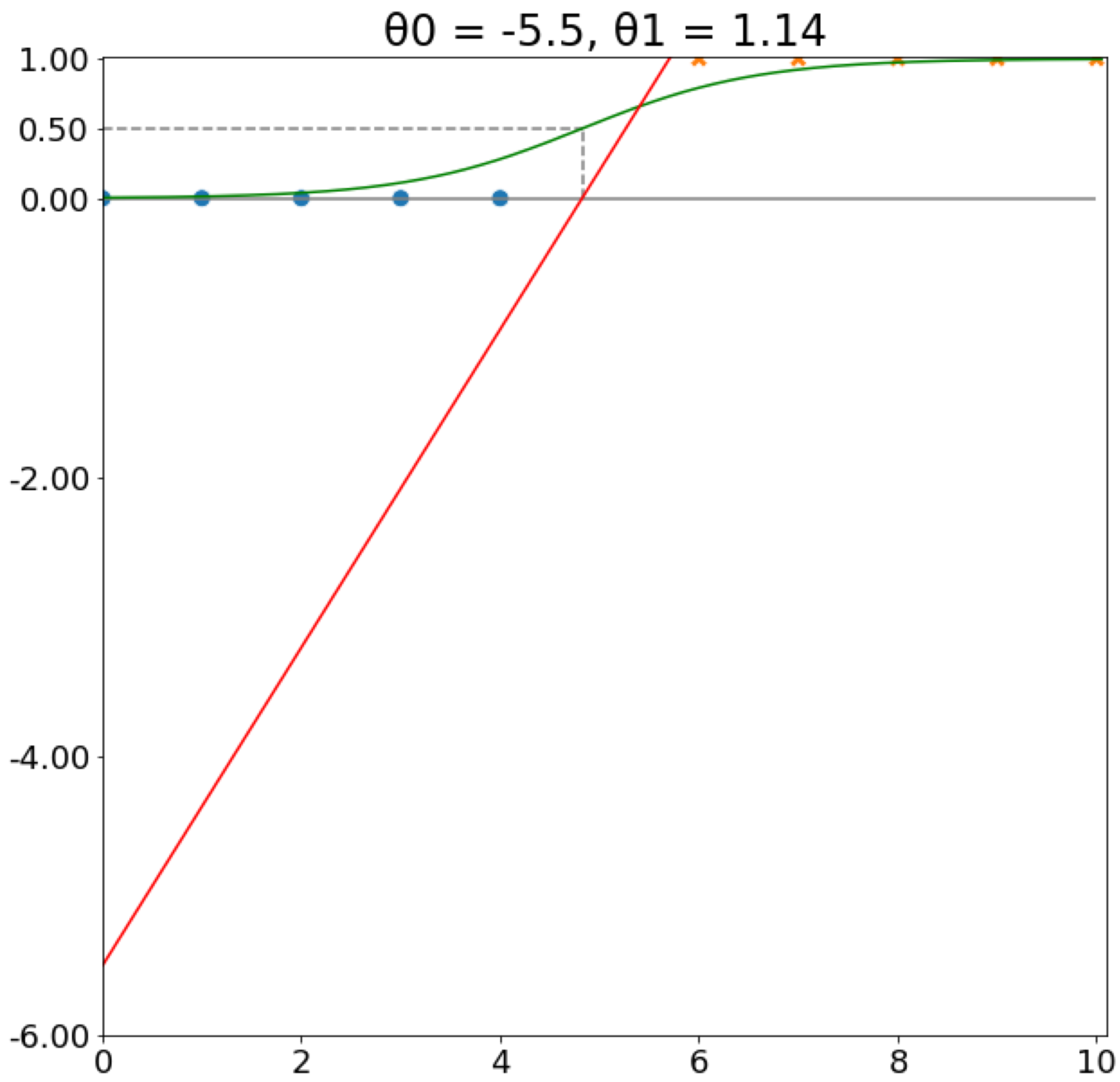

θ0 = -5.73, θ1 = 1.25
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

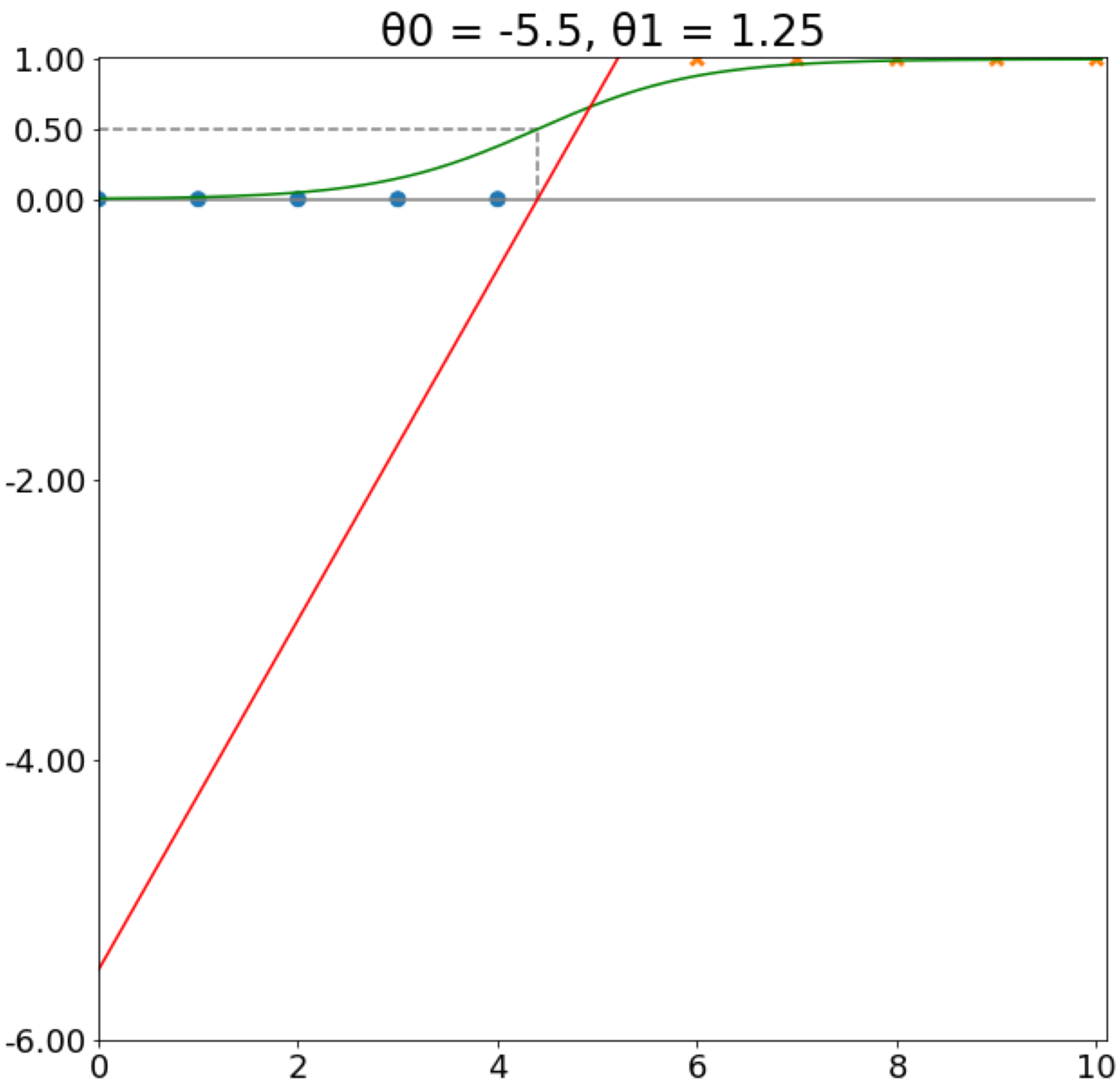

θ0 = -5.73, θ1 = 1.14
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

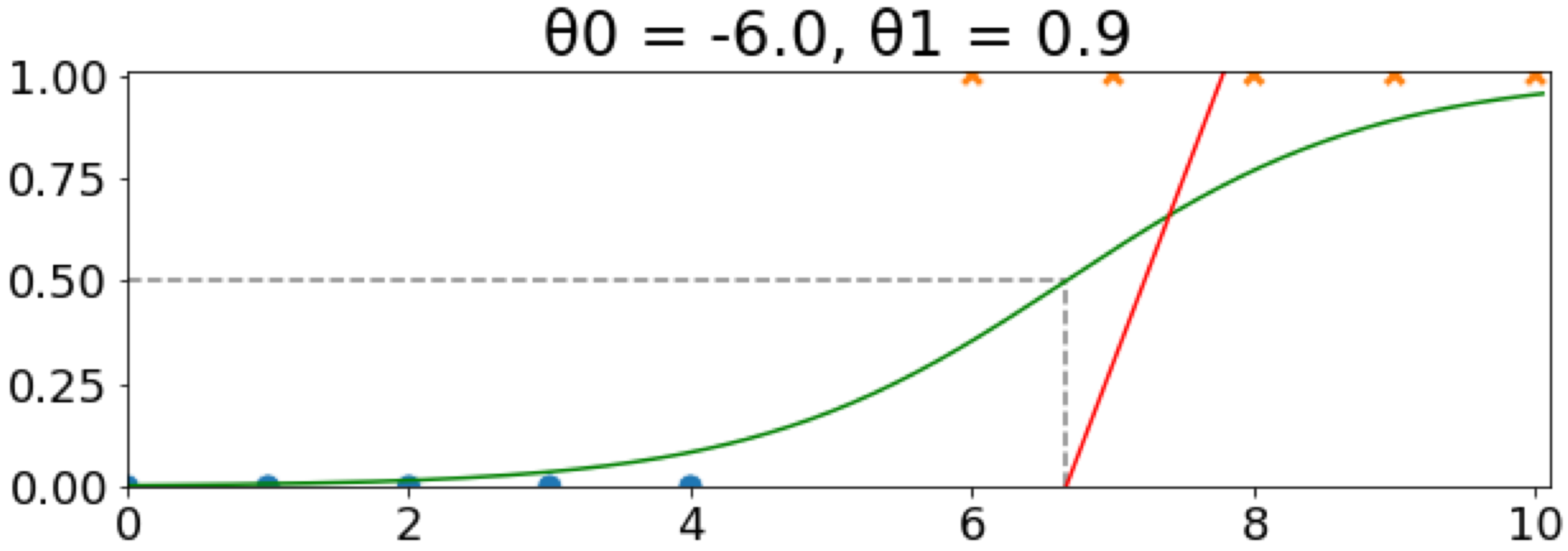

θ0 = -5.73, θ1 = 0.9
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

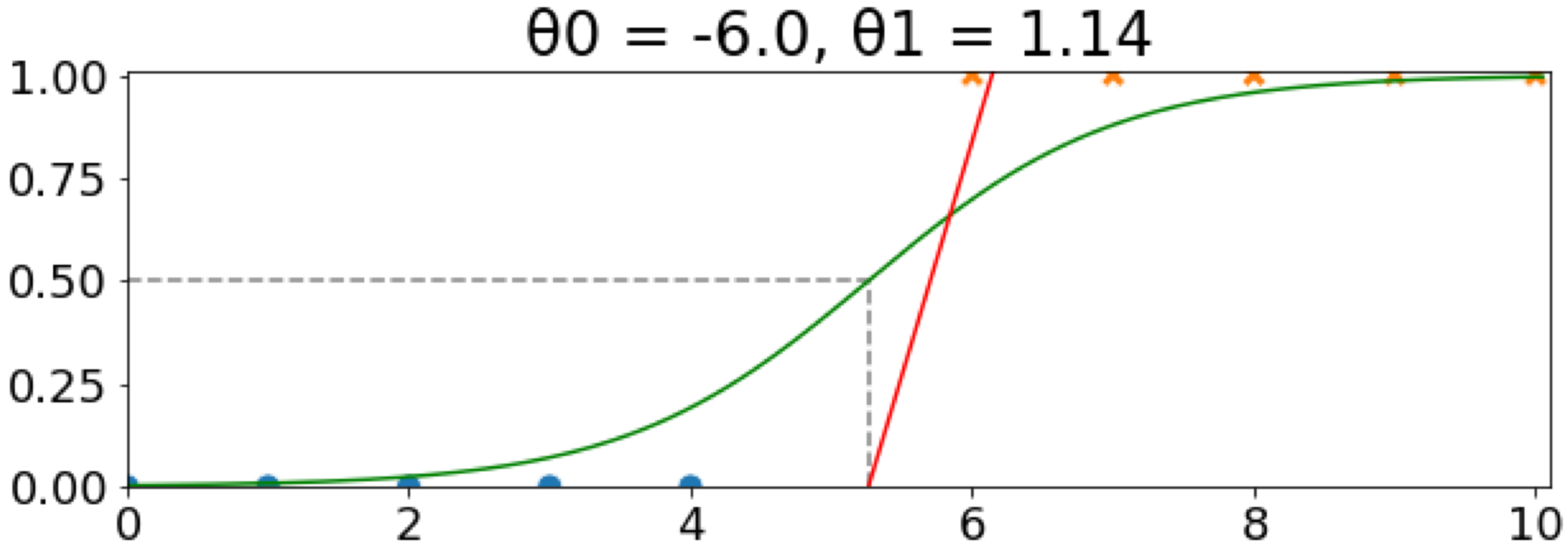


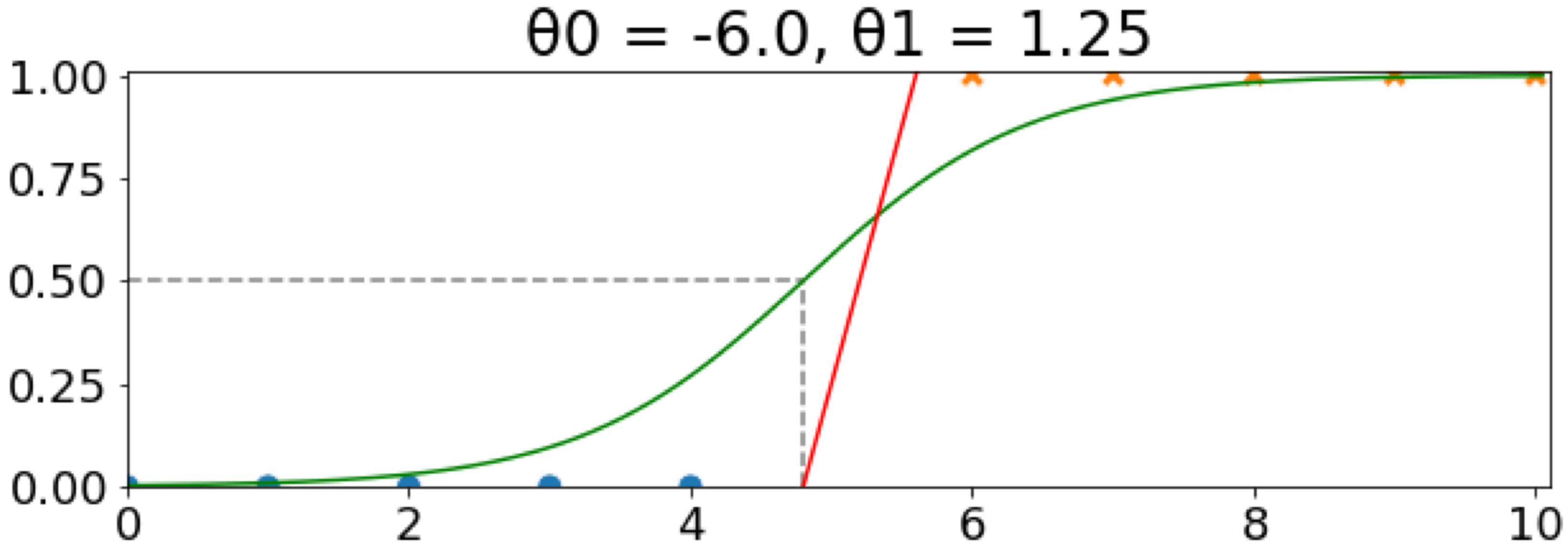
θ0 = -5.5, θ1 = 0.9
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

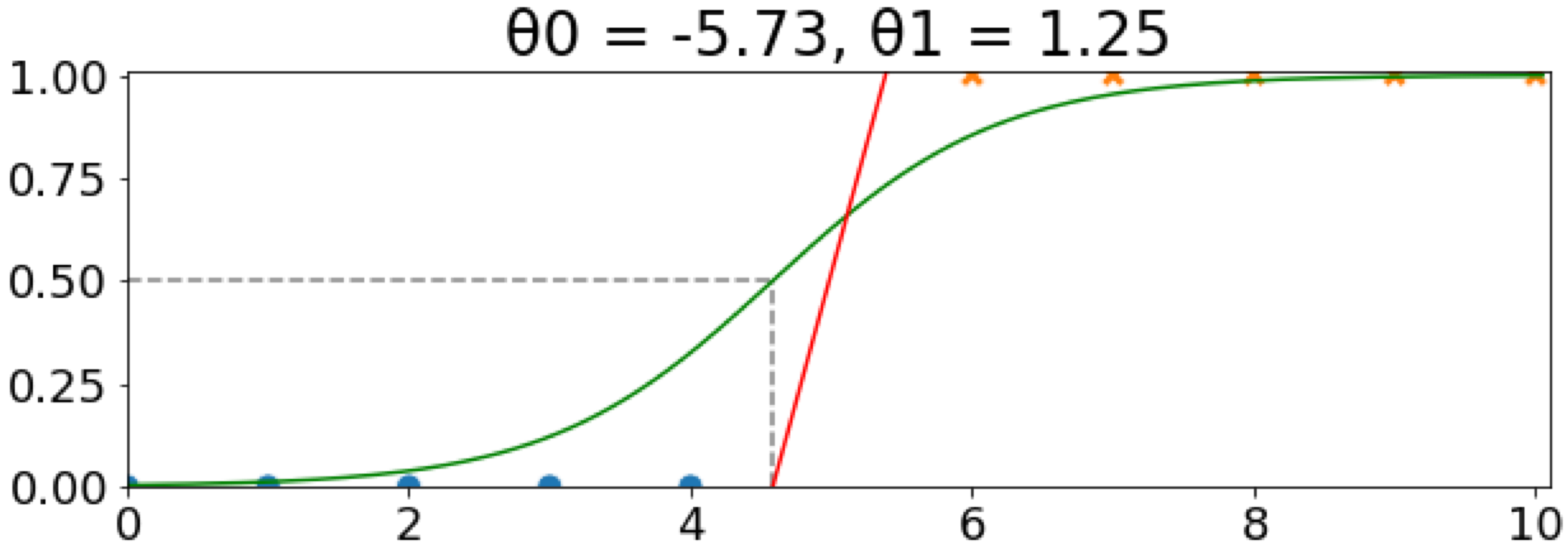

θ0 = -5.5, θ1 = 1.14
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

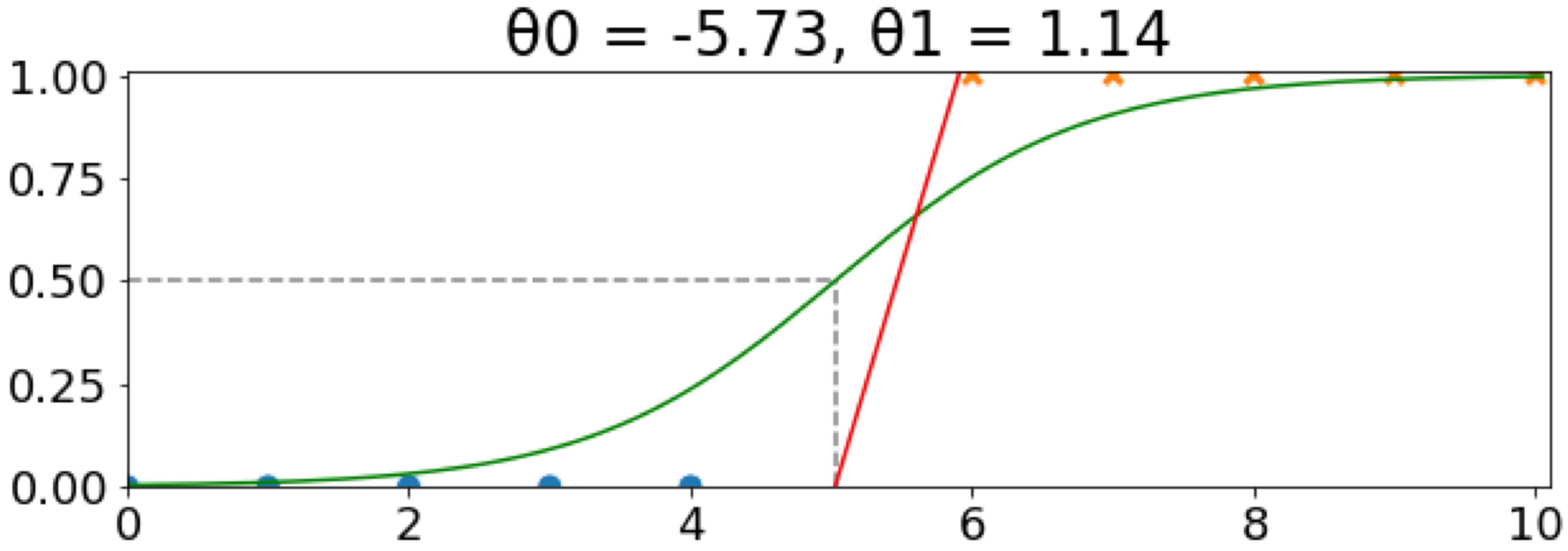

θ0 = -5.5, θ1 = 1.25
1.00
0.50
0.00
-2.00
-4.00
-6.00
0
2
4
6
8
10

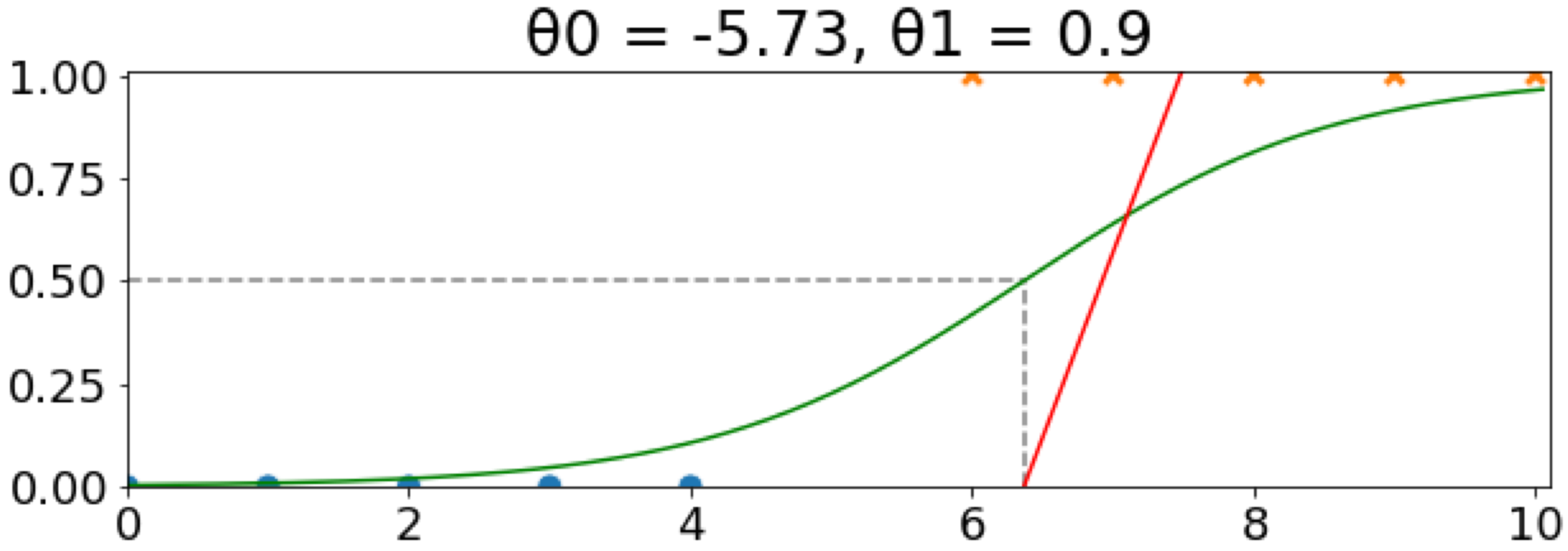

θ0 = -6.0, θ1 = 0.9
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10

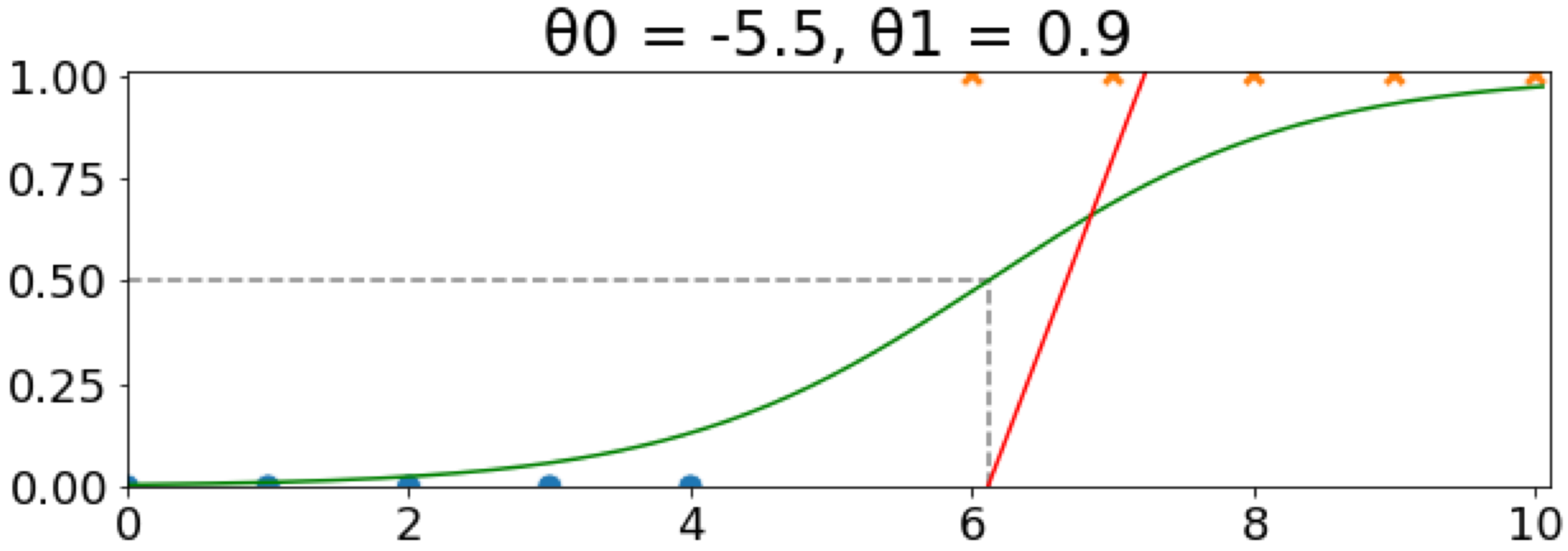

θ0 = -6.0, θ1 = 1.14
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10

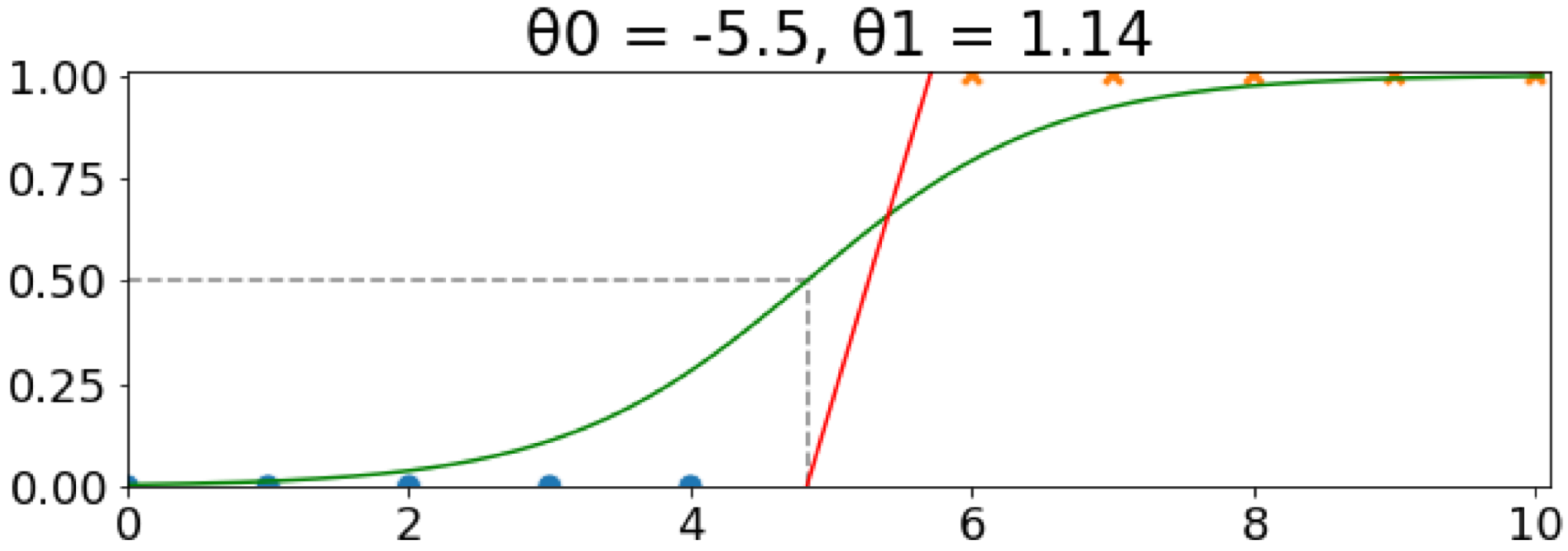

θ0 = -6.0, θ1 = 1.25
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10

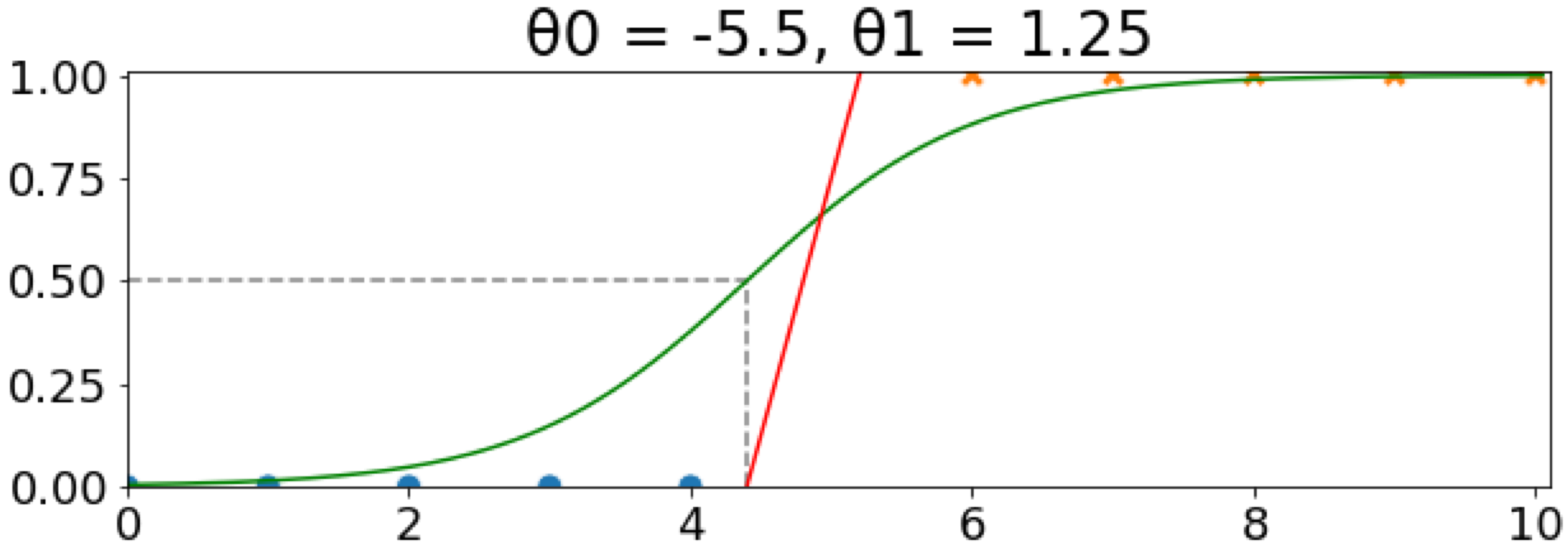

θ0 = -5.73, θ1 = 1.25
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10


θ0 = -5.73, θ1 = 1.14
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10


θ0 = -5.73, θ1 = 0.9
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10

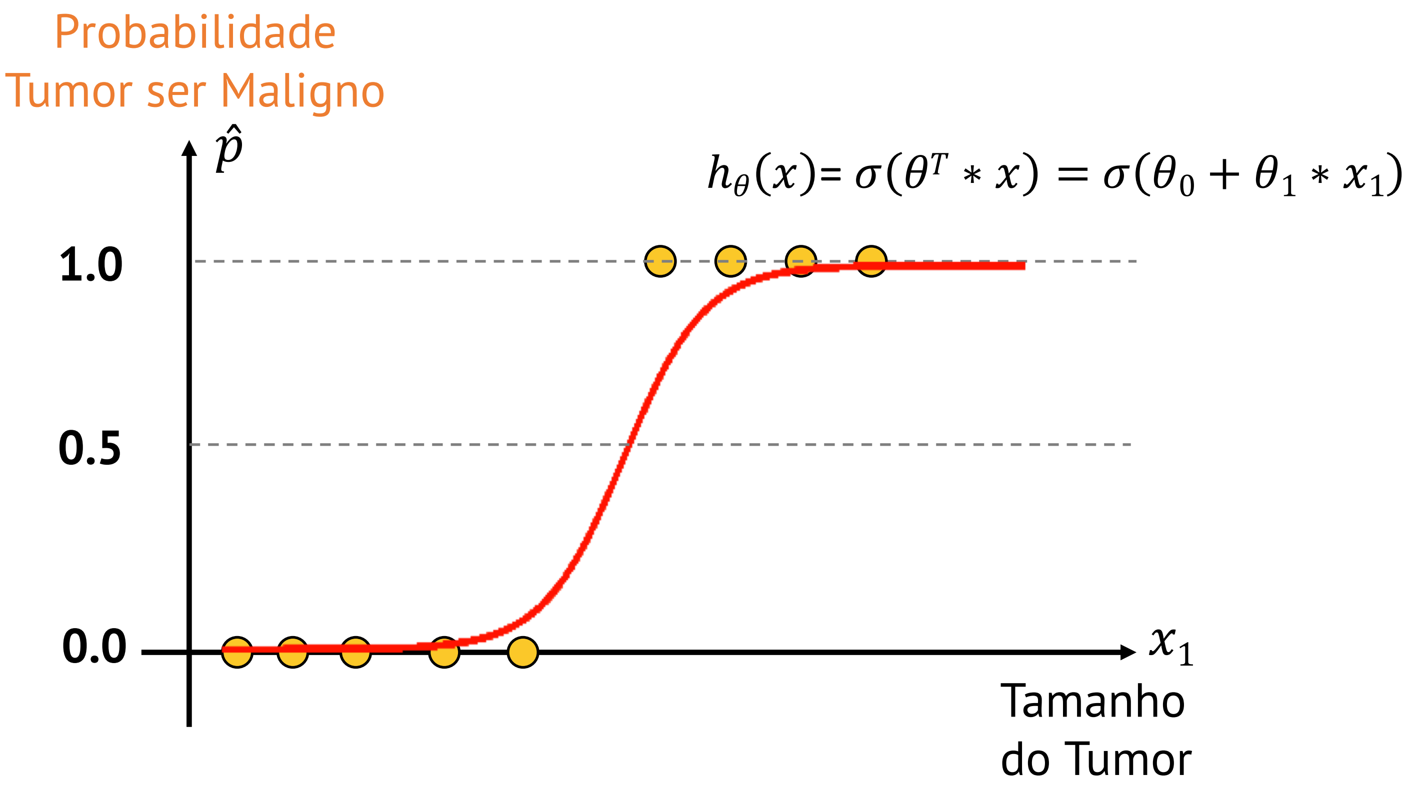

θ0 = -5.5, θ1 = 0.9
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10


θ0 = -5.5, θ1 = 1.14
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10

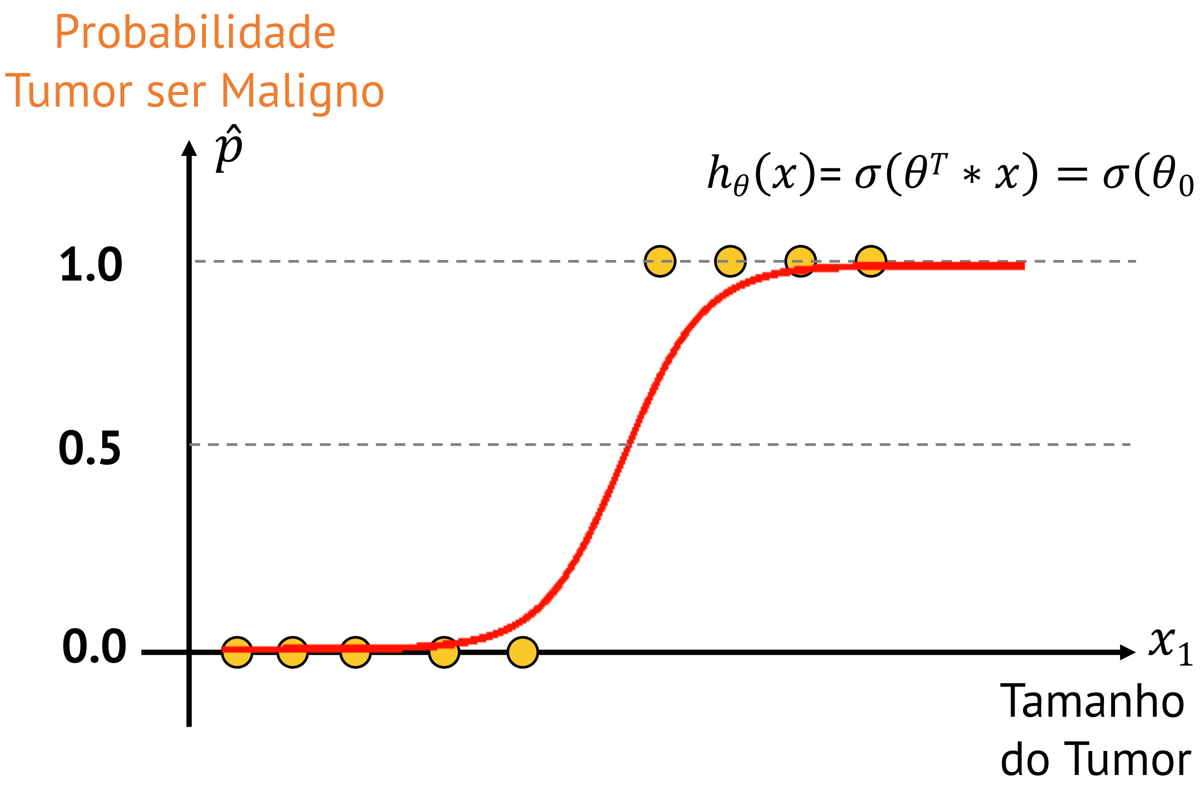


θ0 = -5.5, θ1 = 1.25
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10

Decision Boundary



Decision Boundary



$h_\theta(x) = \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1)$

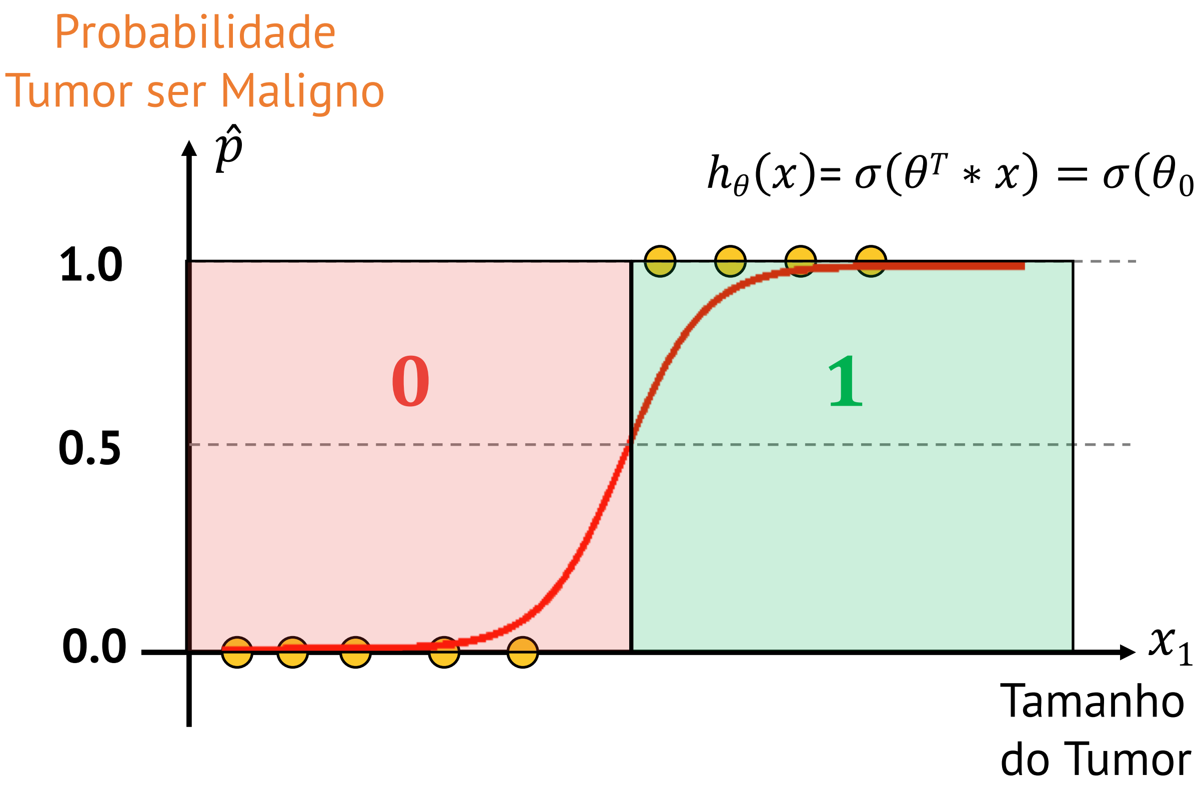
$\sigma(t)$ retorna um número entre **0** e **1**

$\sigma(t) < 0.5$ quando $t < 0$ **(classe negativa)**

$\sigma(t) \geq 0.5$ quando $t \geq 0$ **(classe positiva)**

$\sigma(t) = 0.5$ quando $t = 0$

Decision Boundary



$h_\theta(x) = \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1)$

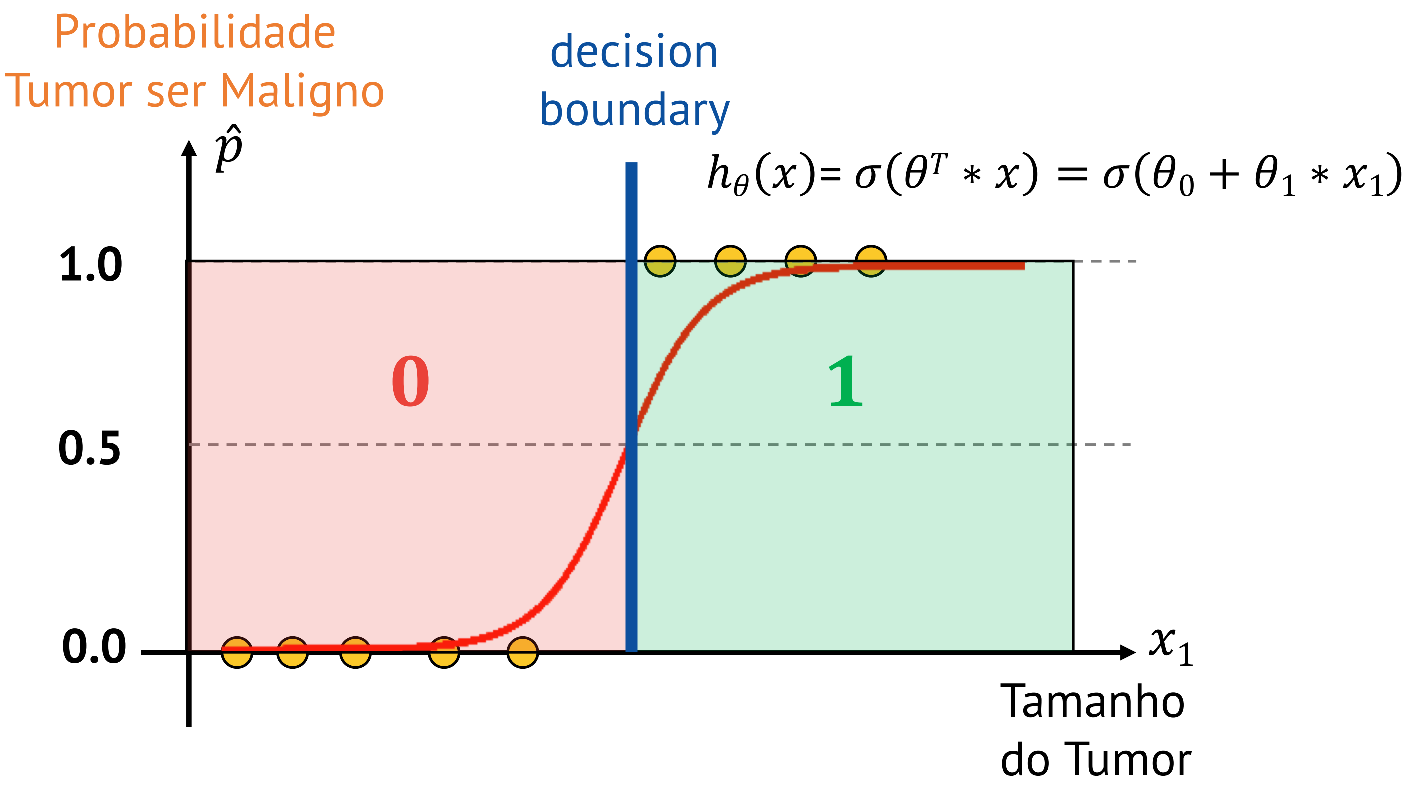
$\sigma(t)$ retorna um número entre **0** e **1**

$\sigma(t) < 0.5$ quando $t < 0$ **(classe negativa)**

$\sigma(t) \geq 0.5$ quando $t \geq 0$ **(classe positiva)**

$\sigma(t) = 0.5$ quando $t = 0$

Decision Boundary



$\sigma(t)$ retorna um número entre **0** e **1**

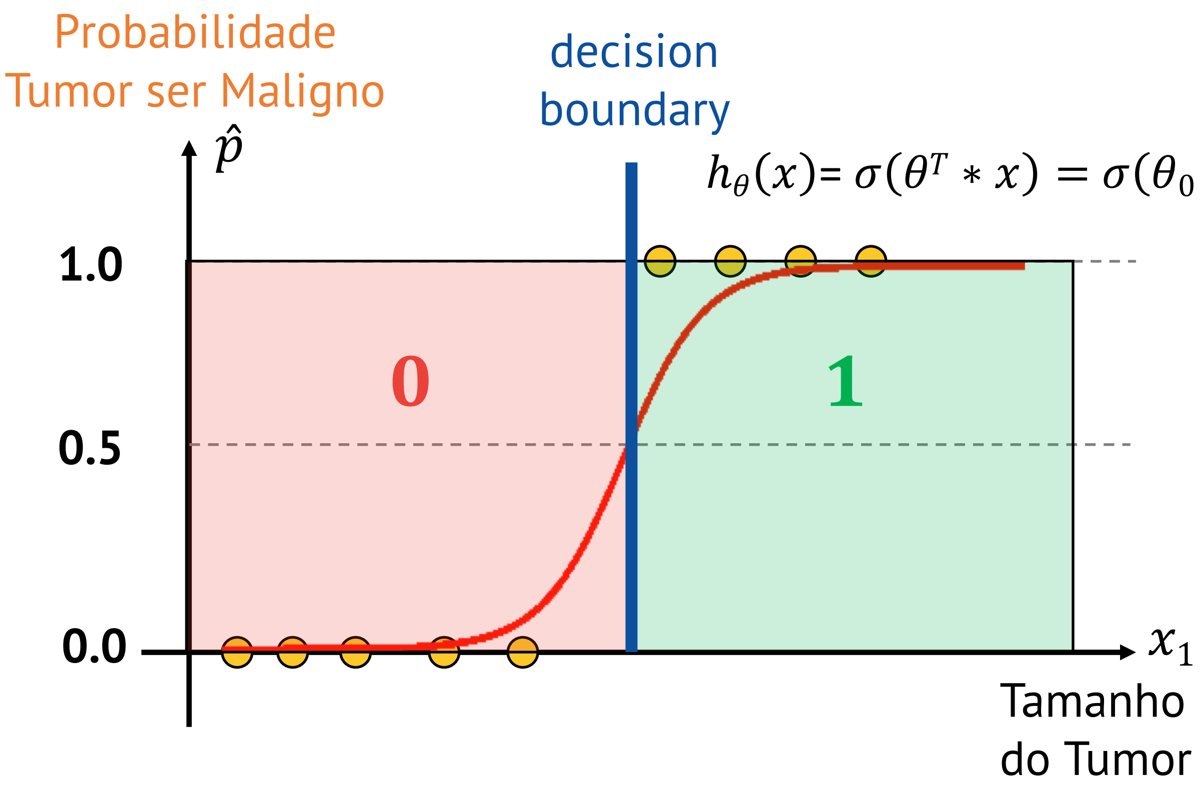
$\sigma(t) < 0.5$ quando $t < 0$ **(classe negativa)**

$\sigma(t) \geq 0.5$ quando $t \geq 0$ **(classe positiva)**

$\sigma(t) = 0.5$ quando $t = 0$

Logo, o hiperplano $(\theta^T * x) = 0$
forma a ***decision boundary*** do **classificador.**

Decision Boundary



$h_\theta(x) = \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1)$

$\sigma(t)$ retorna um número entre **0** e **1**

$\sigma(t) < 0.5$ quando $t < 0$ **(classe negativa)**

$\sigma(t) \geq 0.5$ quando $t \geq 0$ **(classe positiva)**

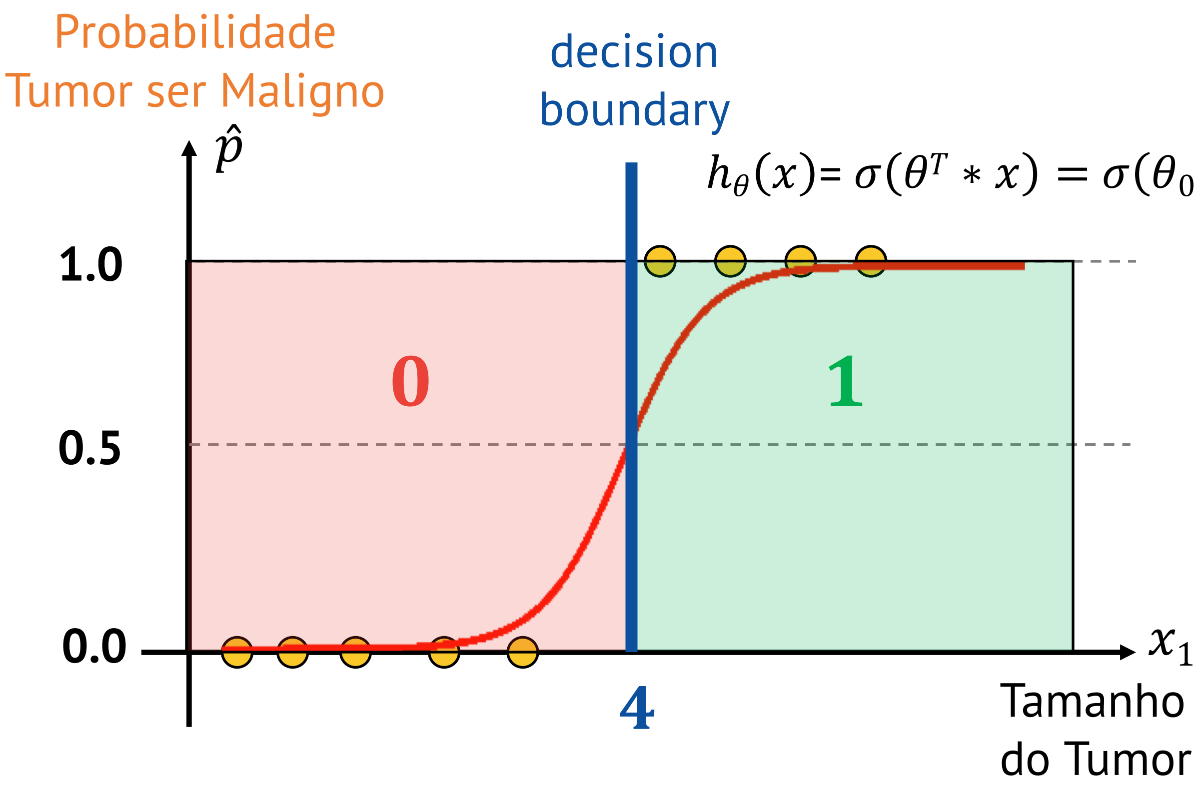
$\sigma(t) = 0.5$ quando $t = 0$

Logo, o hiperplano $(\theta^T * x) = 0$
forma a ***decision boundary*** do **classificador.**

Suponha que o melhor vetor θ encontrado foi:

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -4 \\ 1 \end{bmatrix} \qquad x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \end{bmatrix}$$

Decision Boundary



$h_\theta(x) = \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1)$

$\sigma(t)$ retorna um número entre **0** e **1**

$\sigma(t) < 0.5$ quando $t < 0$ (classe negativa)

$\sigma(t) \geq 0.5$ quando $t \geq 0$ (classe positiva)

$\sigma(t) = 0.5$ quando $t = 0$

Logo, o hiperplano $(\theta^T * x) = 0$
forma a ***decision boundary*** do **classificador.**

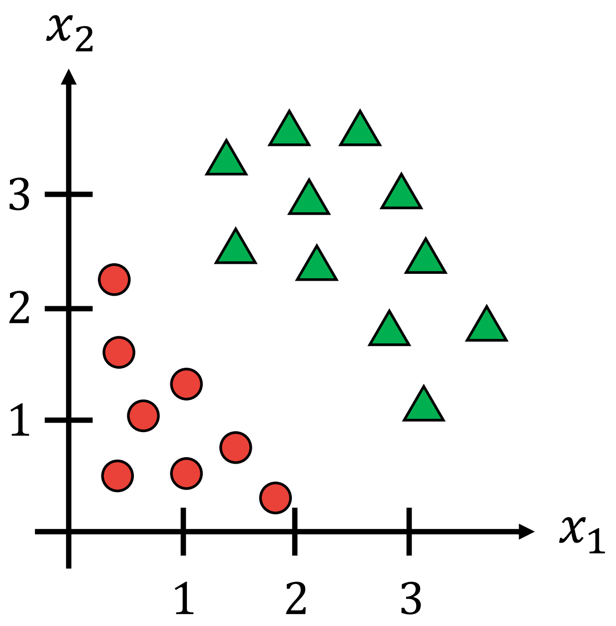
$$\begin{aligned} \theta^T * x &= 0 \\ -4 + 1 * x_1 &= 0 \\ \boxed{x_1 = 4} \end{aligned}$$

decision boundary

Suponha que o melhor vetor θ encontrado foi:

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -4 \\ 1 \end{bmatrix} \qquad x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \end{bmatrix}$$

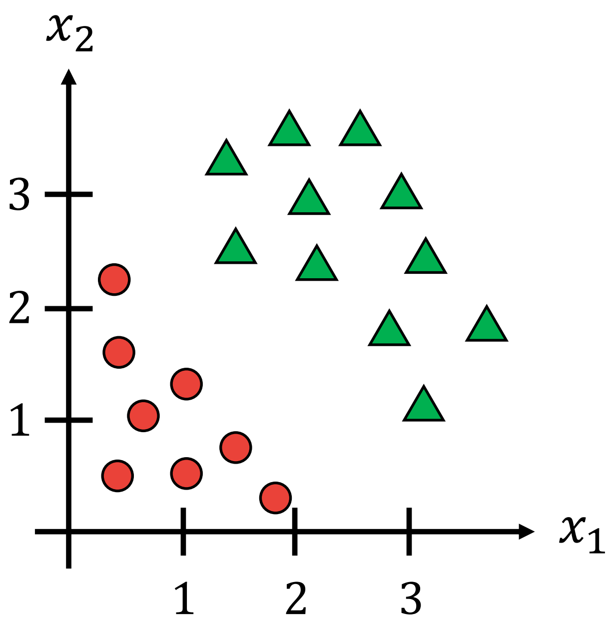
Decision Boundary: outro exemplo



$h_\theta(x)= \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1 + \theta_2 * x_2)$

O hiperplano $(\theta^T * x) = 0$
forma a ***decision boundary*** do **classificador.**

Decision Boundary: outro exemplo



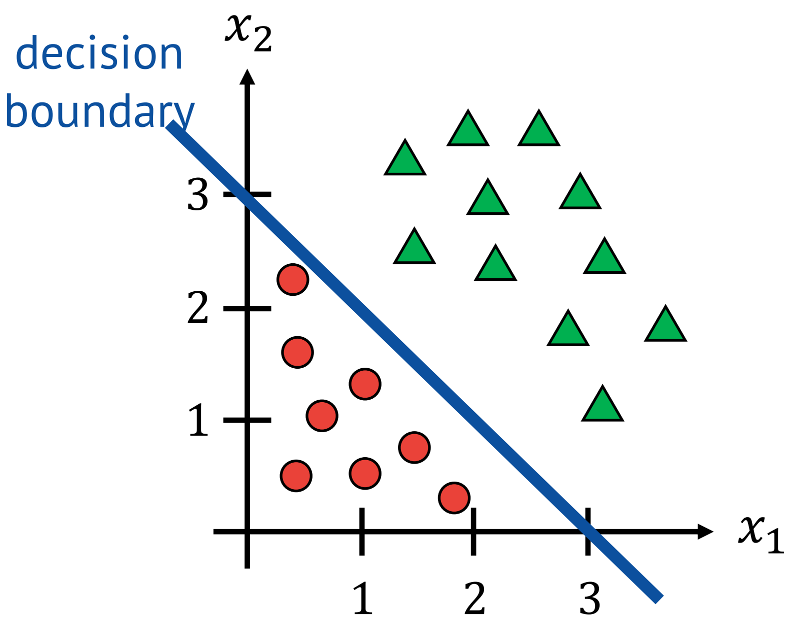
$h_\theta(x)= \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1 + \theta_2 * x_2)$

Suponha que o melhor vetor θ encontrado foi:

$$\theta = \begin{bmatrix}\theta_0\\ \theta_1\\ \theta_2\end{bmatrix} = \begin{bmatrix}-3\\ 1\\ 1\end{bmatrix} \quad x = \begin{bmatrix}x_0\\ x_1\\ x_2\end{bmatrix} = \begin{bmatrix}1\\ x_1\\ x_2\end{bmatrix}$$

O hiperplano $(\theta^T * x) = 0$
forma a ***decision boundary*** do **classificador.**

Decision Boundary: outro exemplo



$h_\theta(x) = \sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1 + \theta_2 * x_2)$

Suponha que o melhor vetor θ encontrado foi:

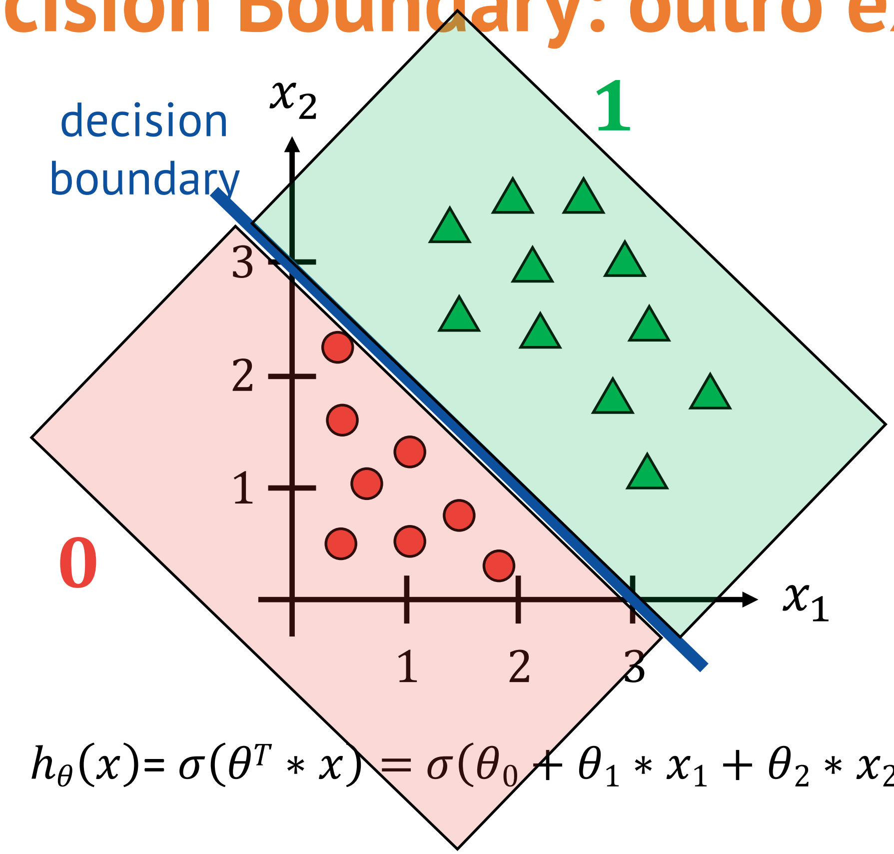
$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \quad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

O hiperplano $(\theta^T * x) = 0$ forma a ***decision boundary*** do **classificador.**

$$\begin{aligned} \theta^T * x &= 0 \\ -3 + 1 * x_1 + 1 * x_2 &= 0 \\ \boxed{x_1 + x_2 = 3} \end{aligned}$$

decision boundary

Decision Boundary: outro exemplo



$h_\theta(x)$= $\sigma(\theta^T * x) = \sigma(\theta_0 + \theta_1 * x_1 + \theta_2 * x_2)$

Suponha que o melhor vetor θ encontrado foi:

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \qquad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

O hiperplano $(\theta^T * x) = 0$
forma a ***decision boundary*** do **classificador.**

$$\begin{aligned} \theta^T * x &= 0 \\ -3 + 1 * x_1 + 1 * x_2 &= 0 \\ \boxed{x_1 + x_2 = 3} \end{aligned}$$

decision boundary

D1EAD – Análise Estatística para Ciência de Dados 2021.1

Regressão Logística

Aula baseada no curso *Machine Learning* do prof. Andrew Ng (Coursera)

Prof. Ricardo Sovat
sovat@ifsp.edu.br
Prof. Samuel Martins (Samuka)
samuel.martins@ifsp.edu.br

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas