

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



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

Régressão Linear

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

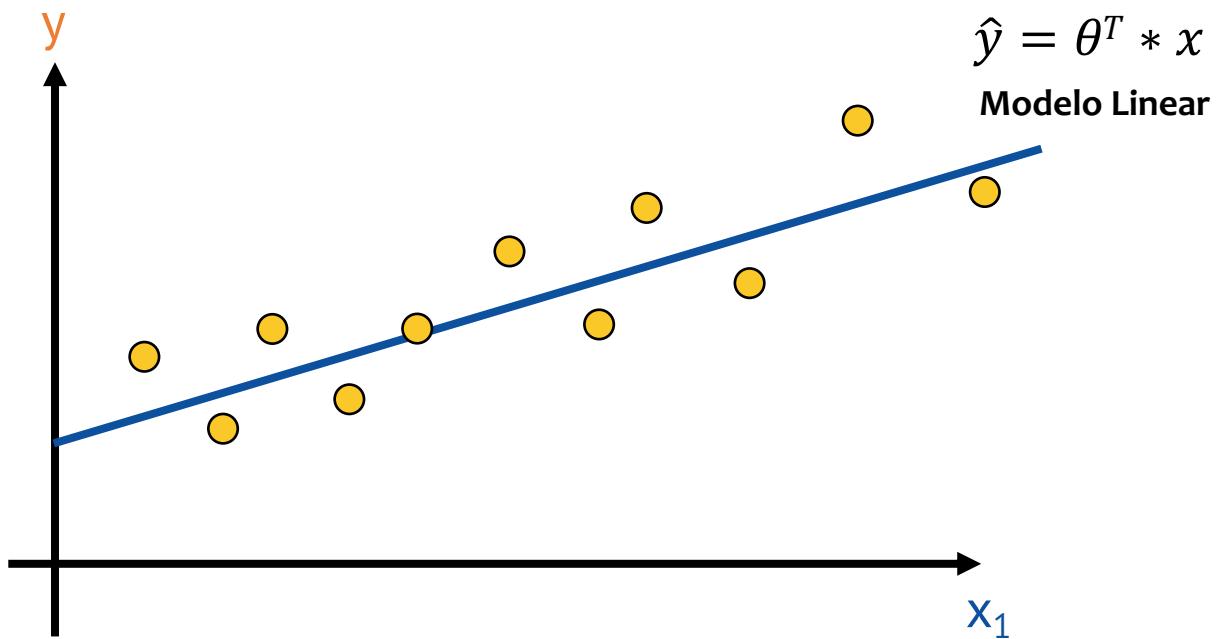
$$h_{\theta}(x) = \hat{y} = \theta^T * x$$

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

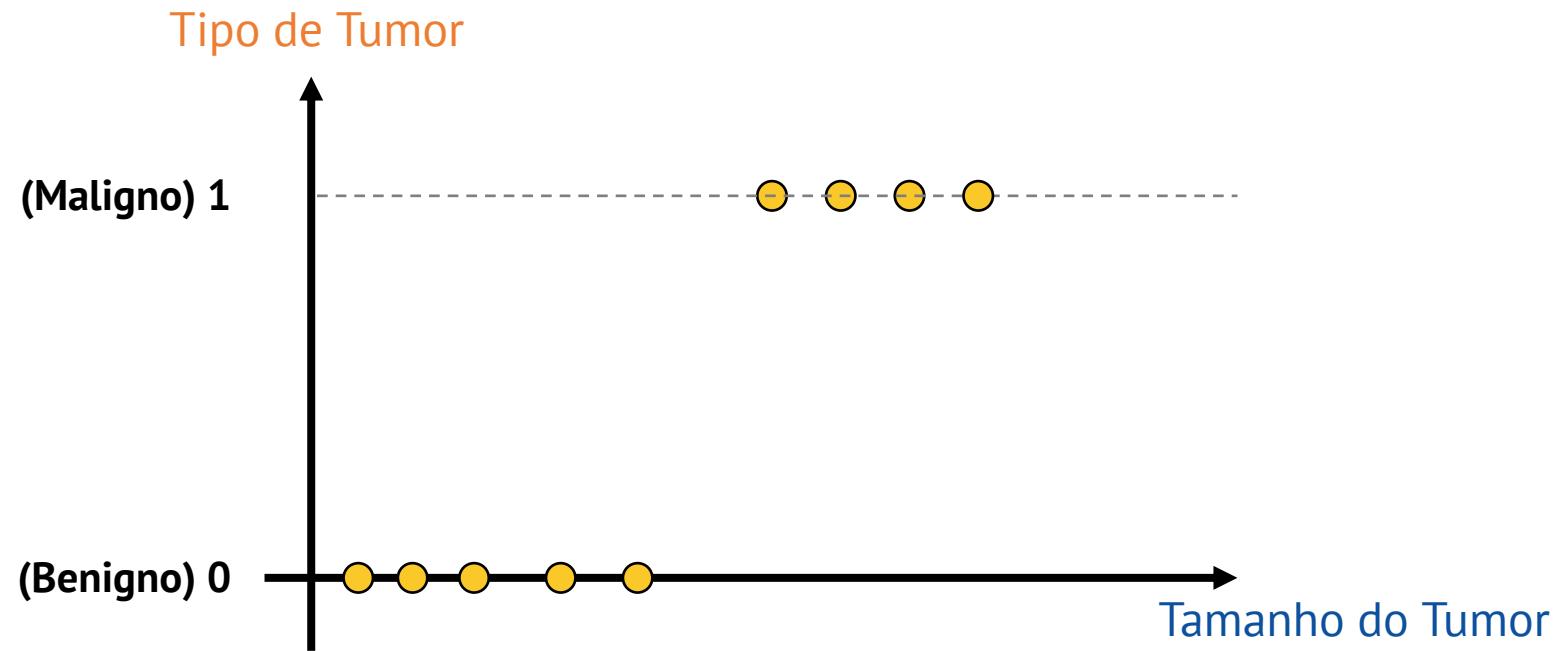
$$x = \begin{bmatrix} x_0 \\ 1 \\ x_1 \\ \dots \\ x_n \end{bmatrix}$$

$$\theta^T = [\theta_0 \ \theta_1 \dots \theta_n]$$

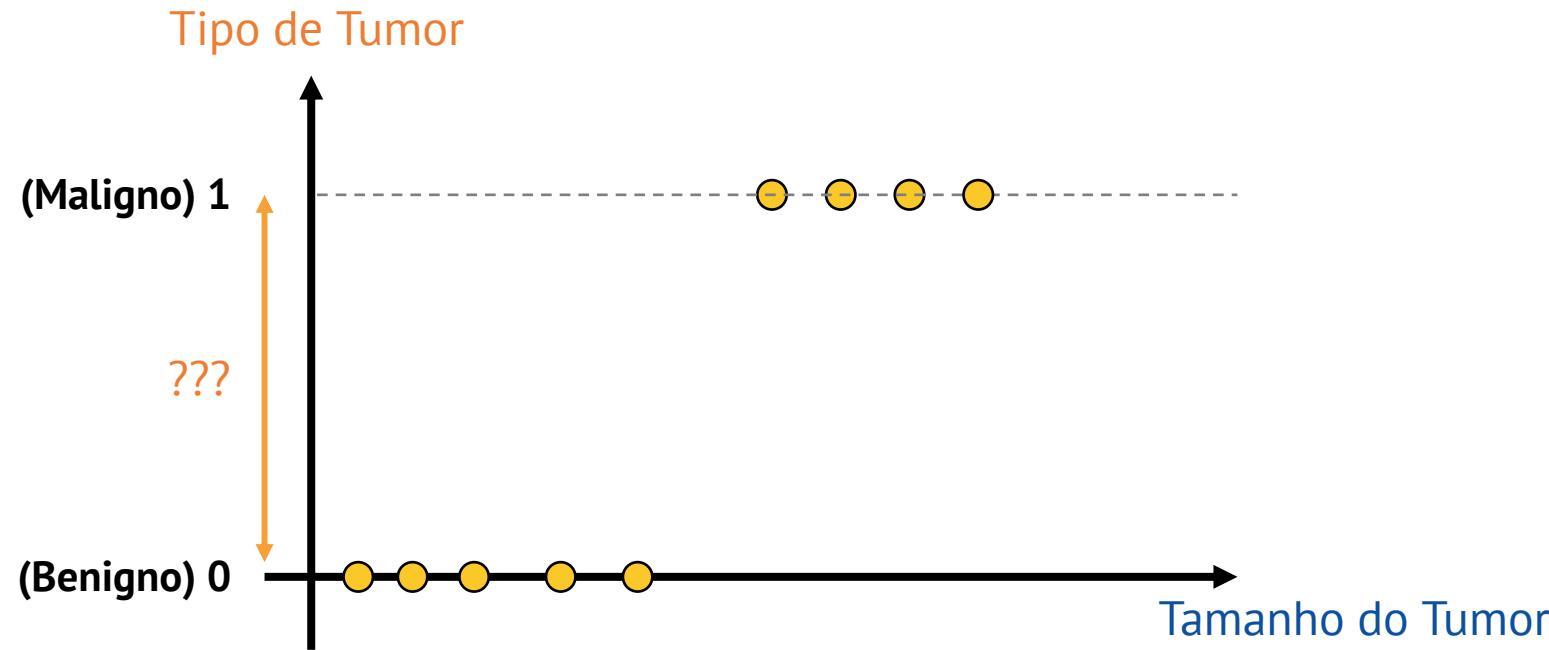
Em machine learning, **vetores** são comumente representados como **vetores coluna**.



Problema de Classificação

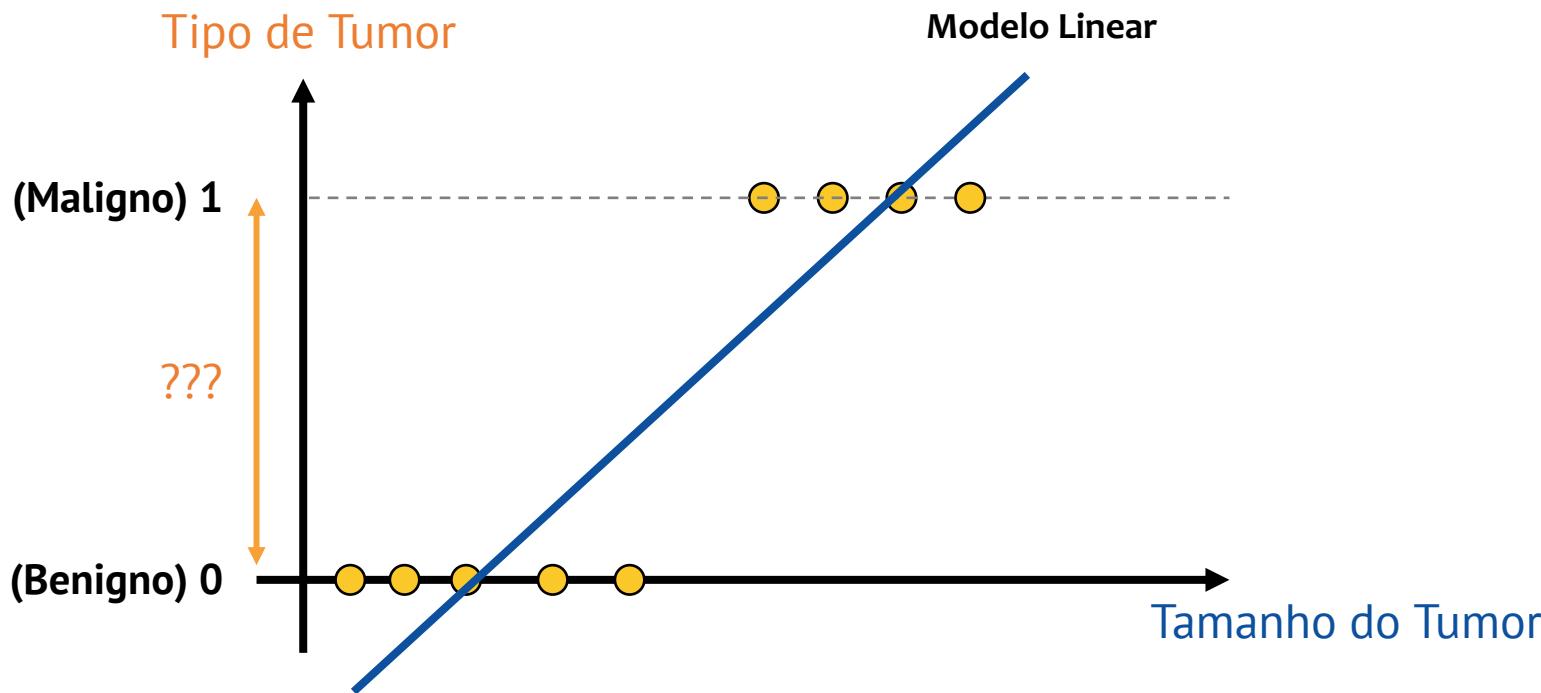


Problema de Classificação



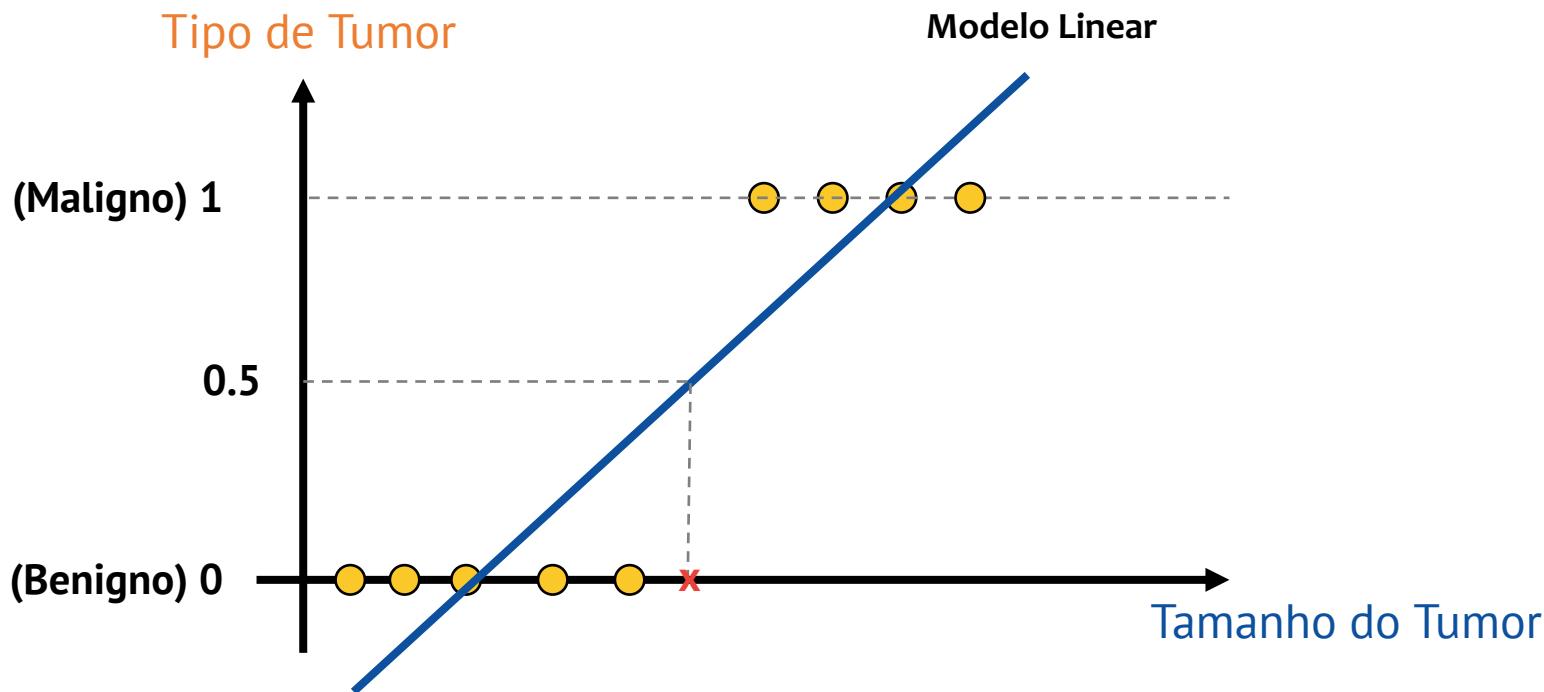
Problema de Classificação

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



Problema de Classificação

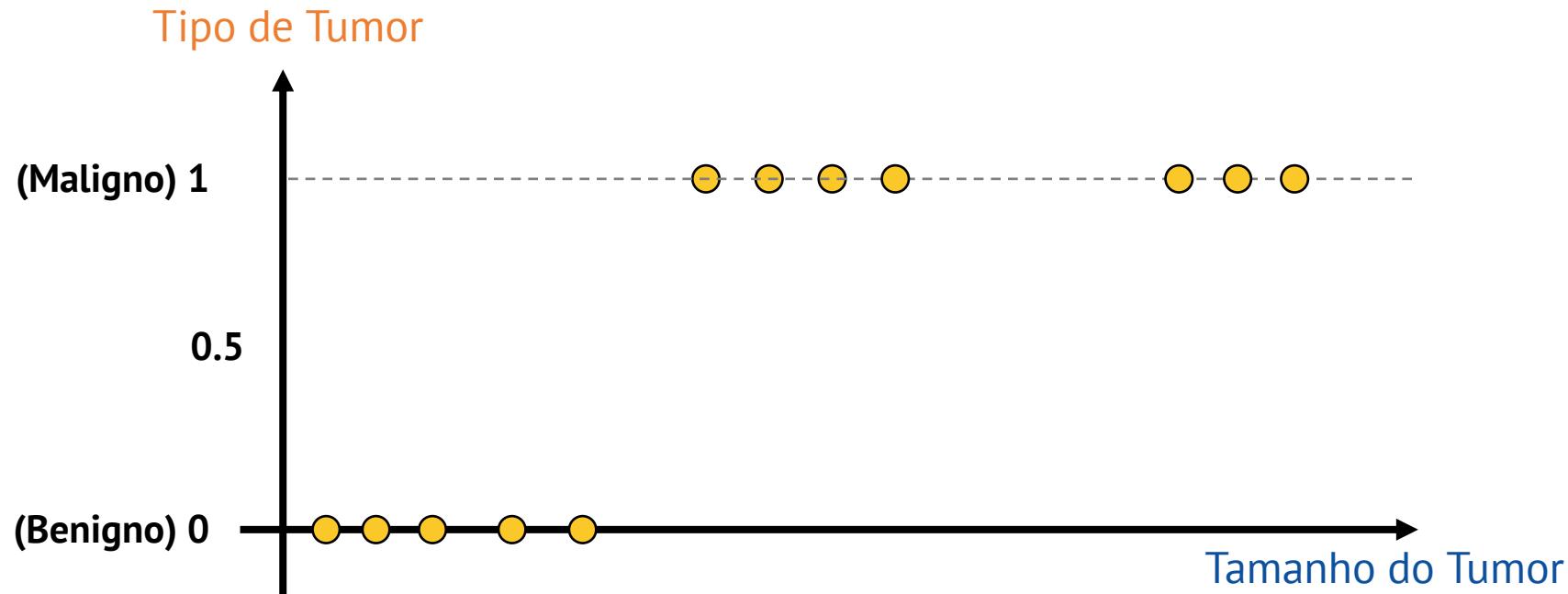
$$h_{\theta}(x) = \theta^T * x$$



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

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

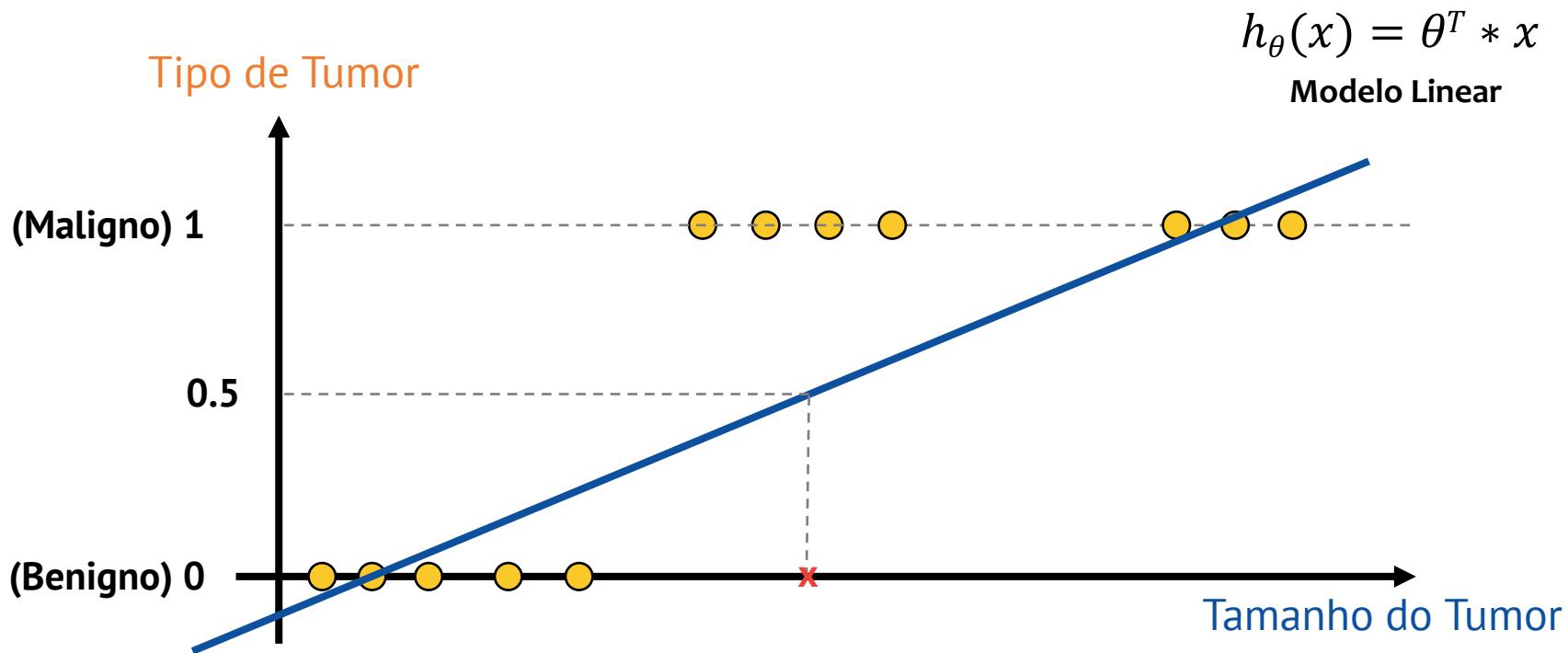
Problema de Classificação



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

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

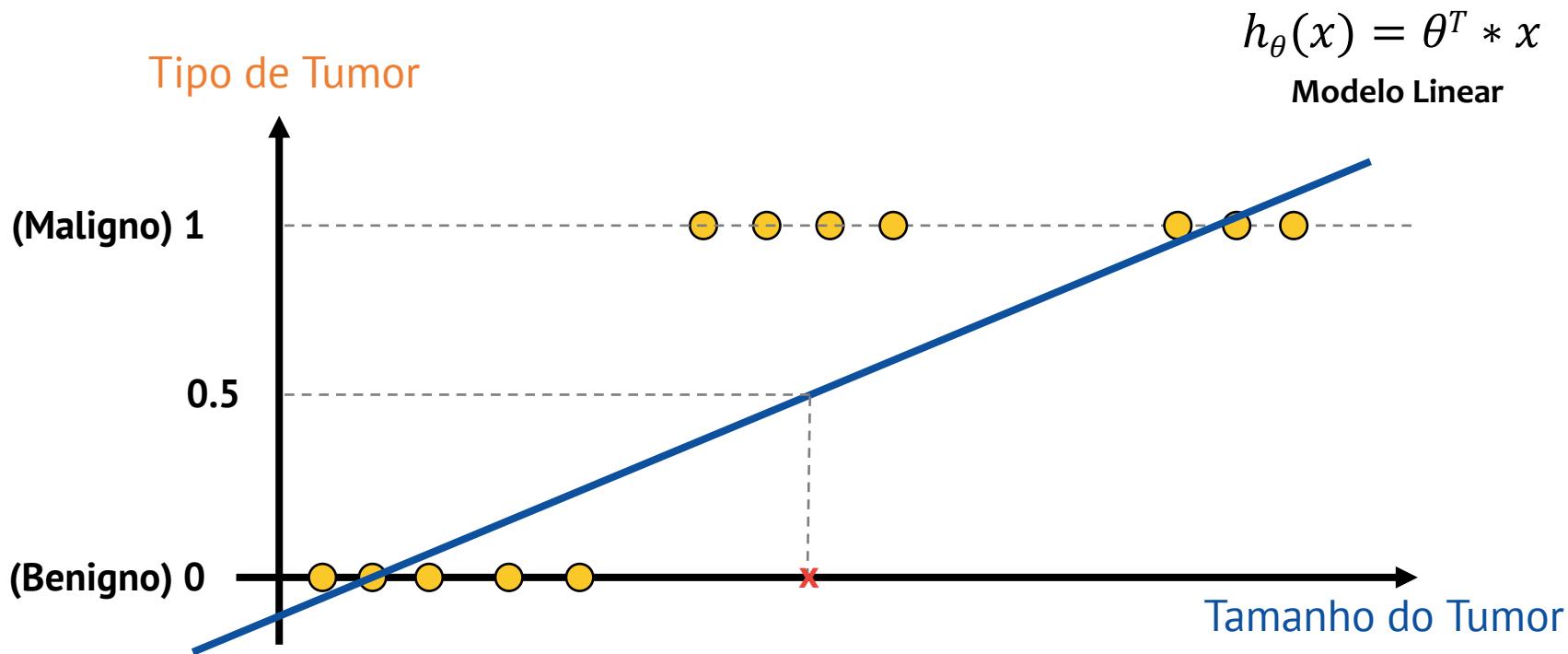
Problema de Classificação



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

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

Problema de Classificação

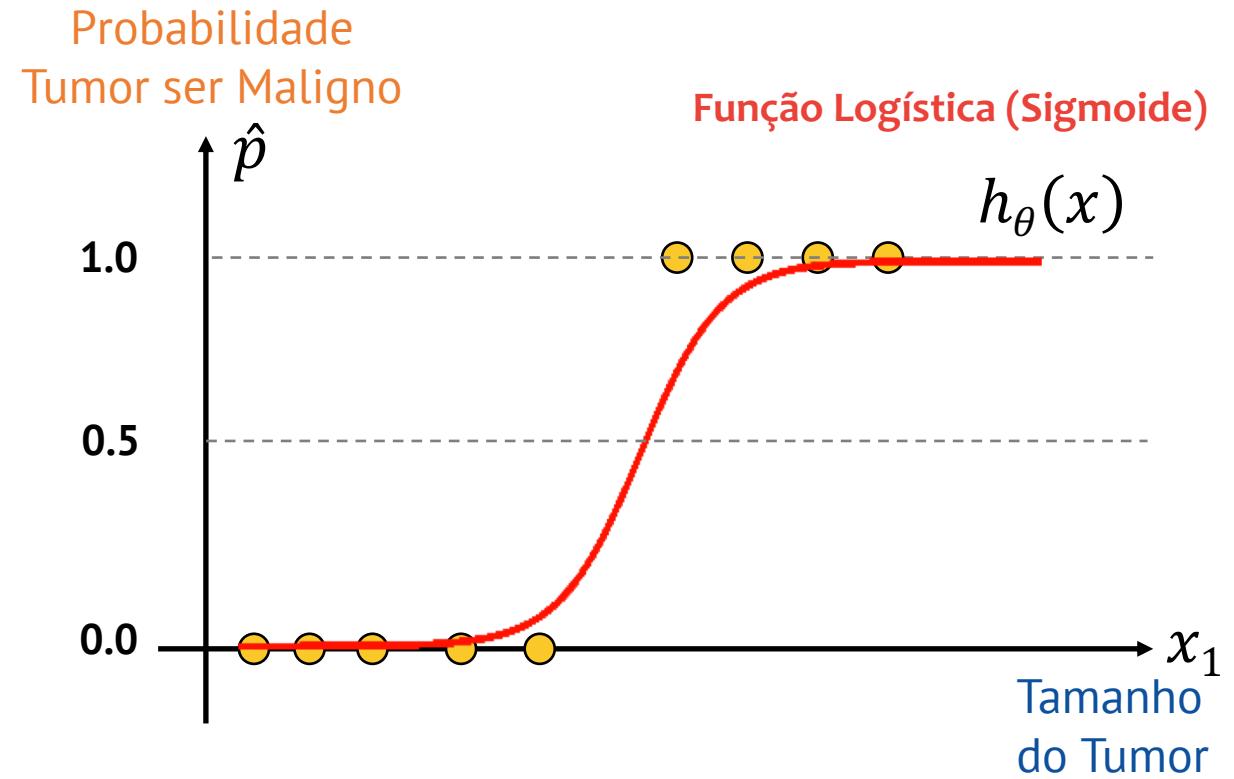


Aplicar uma **Regressão Linear**
em problemas de **classificação**
não parece ser uma boa ideia

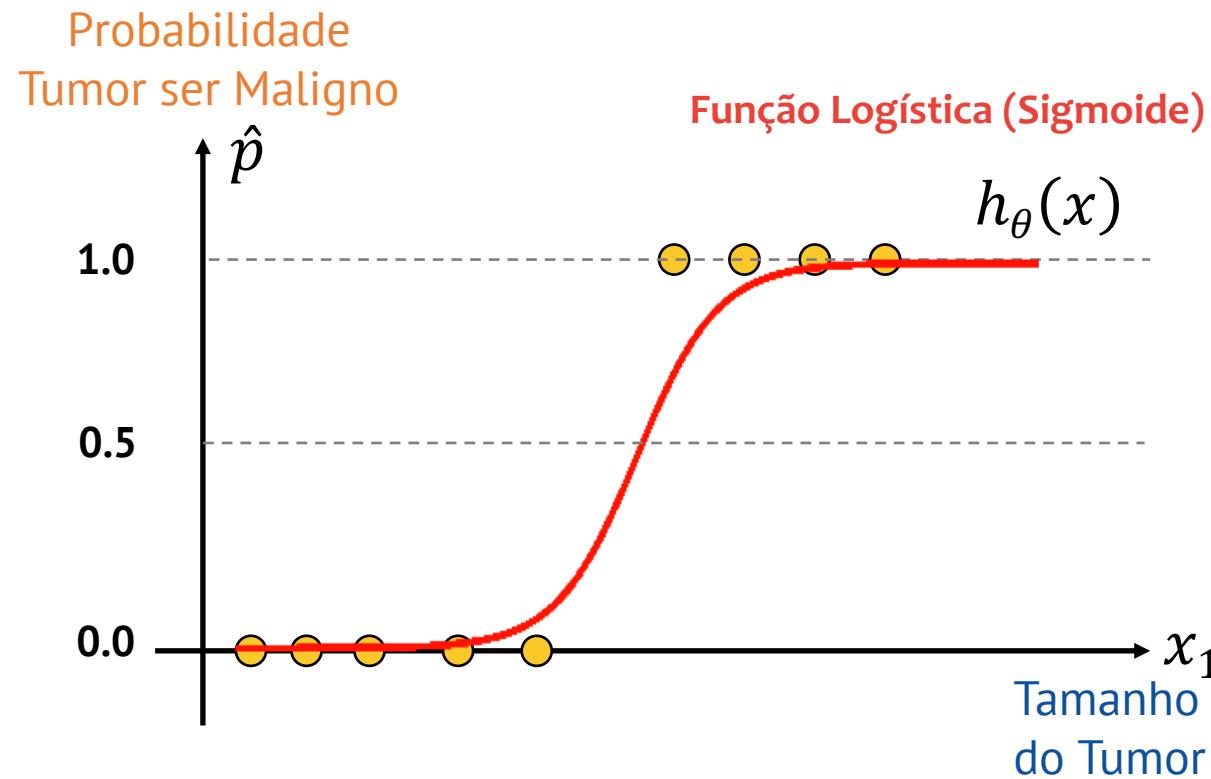
Se $h_{\theta}(x) \geq 0.5$, classifique $\hat{y} = 1$

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

Régressão Logística



Regressão Logística



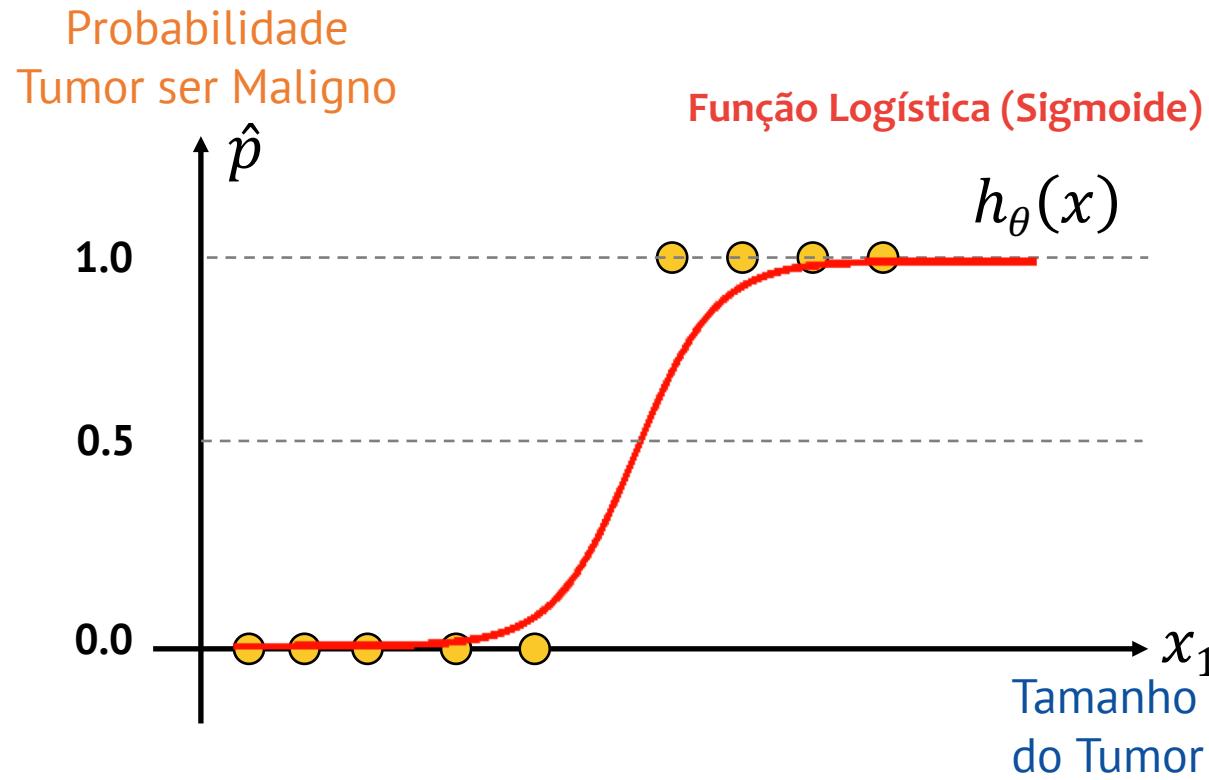
Modelo de Regressão Logística

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

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

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

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 ($y = 1$)**

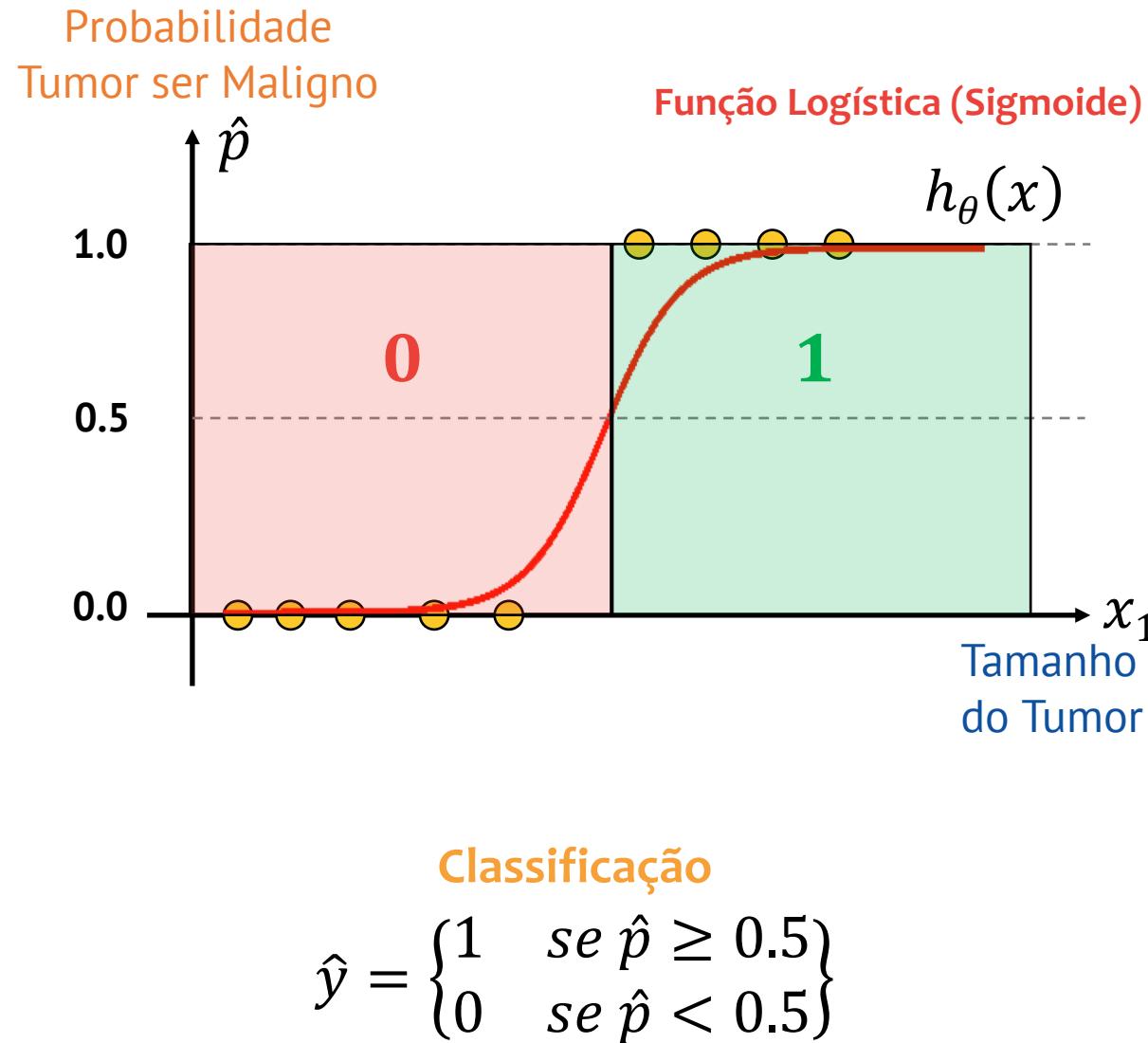
$$h_{\theta}(x) = P(y = 1|x; \theta)$$

P.ex: Se

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

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 ($y = 1$)**

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

P.ex: Se

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

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

Regressão Logística

Modelo de Regressão Logística

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

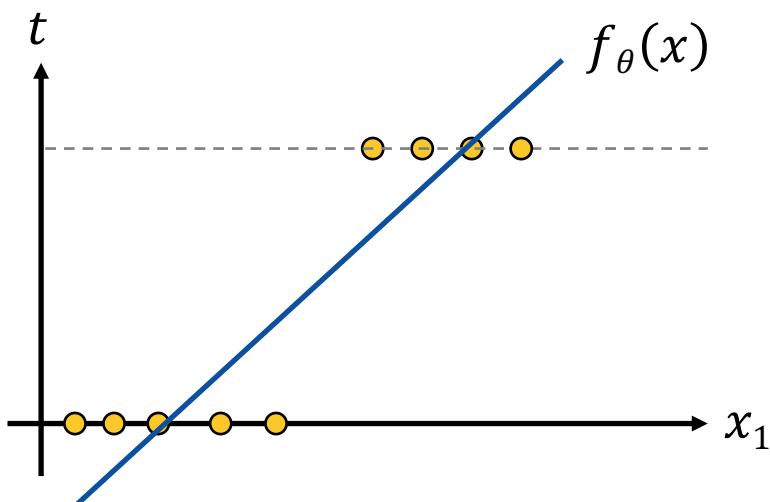
Regressão Logística

Modelo de Regressão Logística

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

Modelo Linear

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



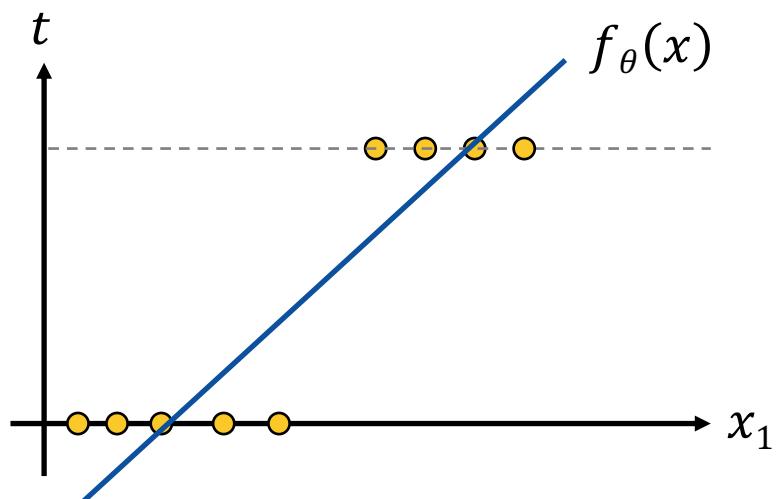
Regressão Logística

Modelo de Regressão Logística

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

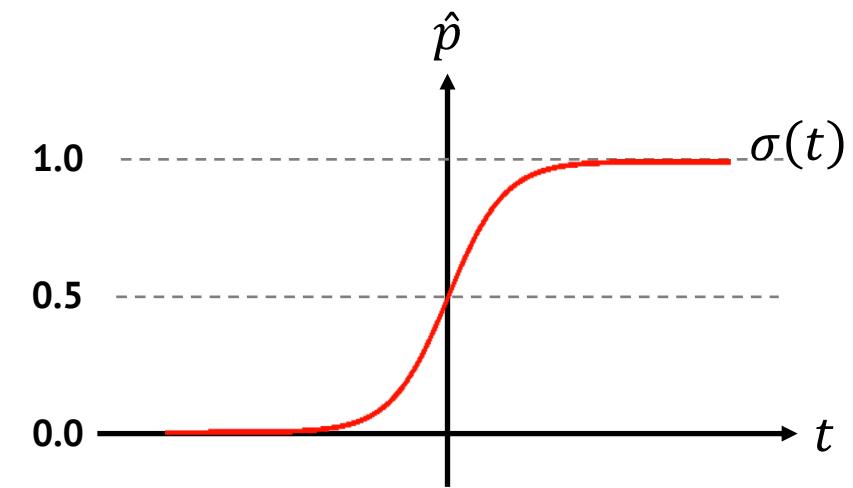
Modelo Linear

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



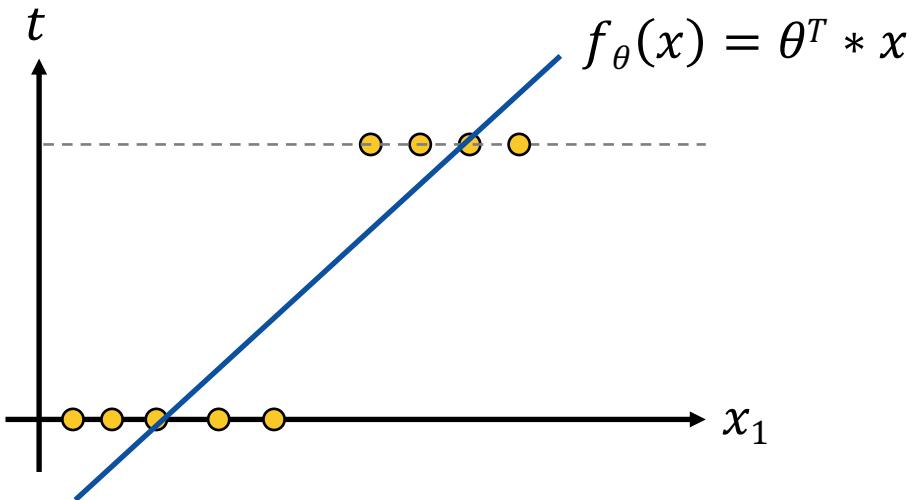
Função Logística (Sigmoidal)

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

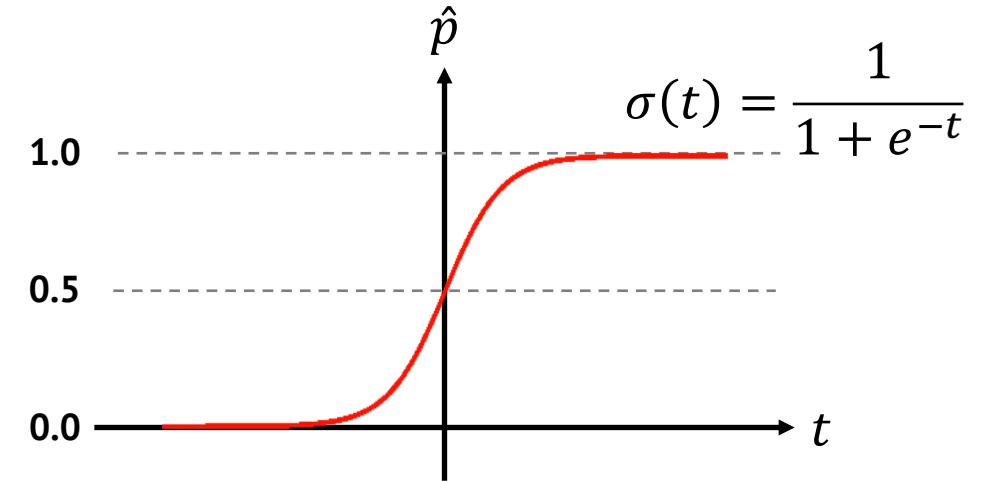


Regressão Logística

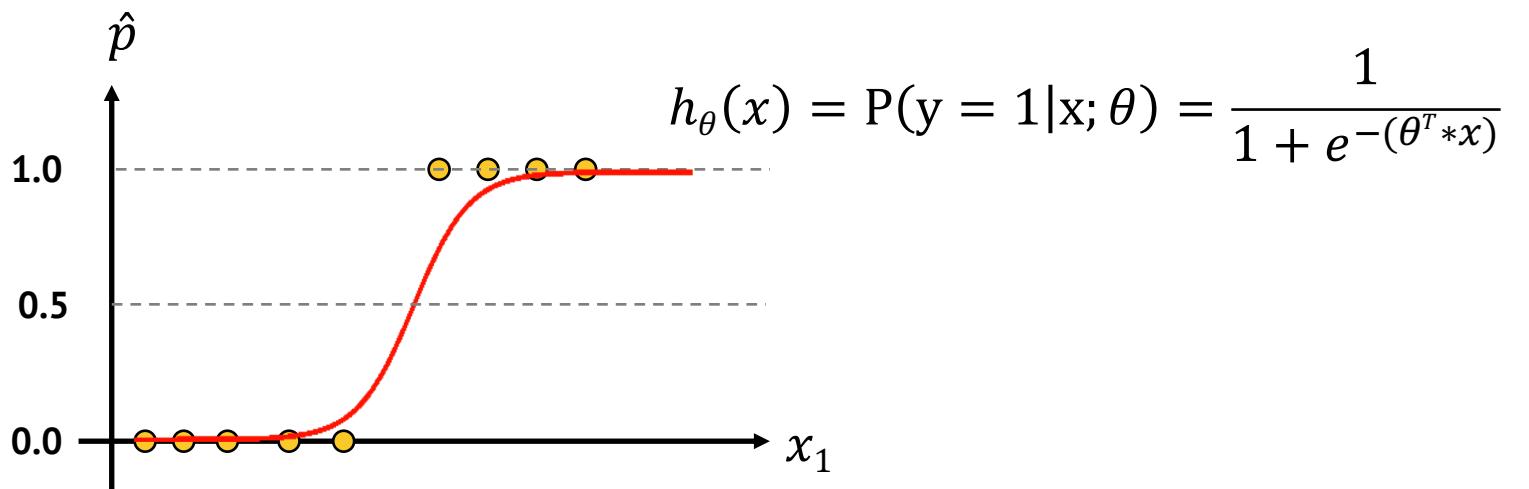
Modelo Linear



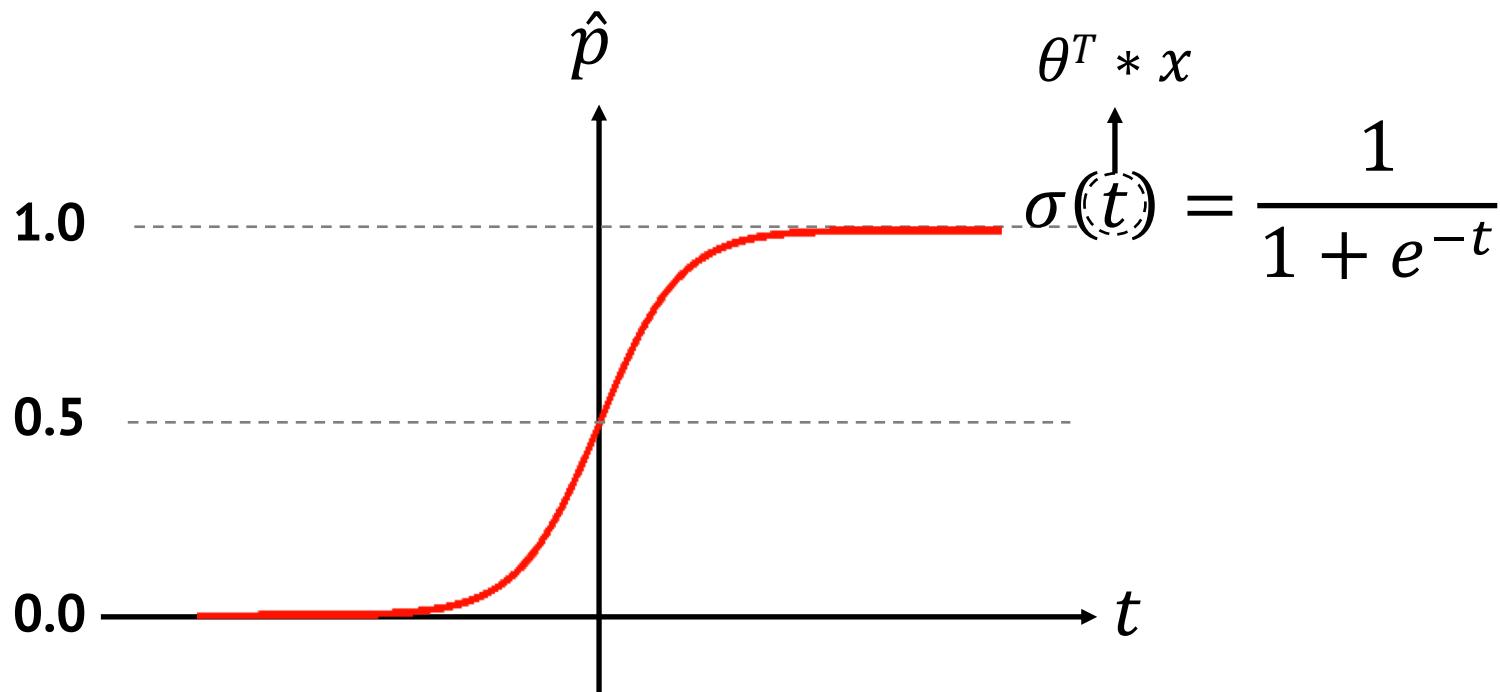
Função Logística (Sísmoide)



Modelo de Regressão Logística



Função Logística (Sísmoide)



$\sigma(t)$ 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 $(\theta^T * x) < 0$

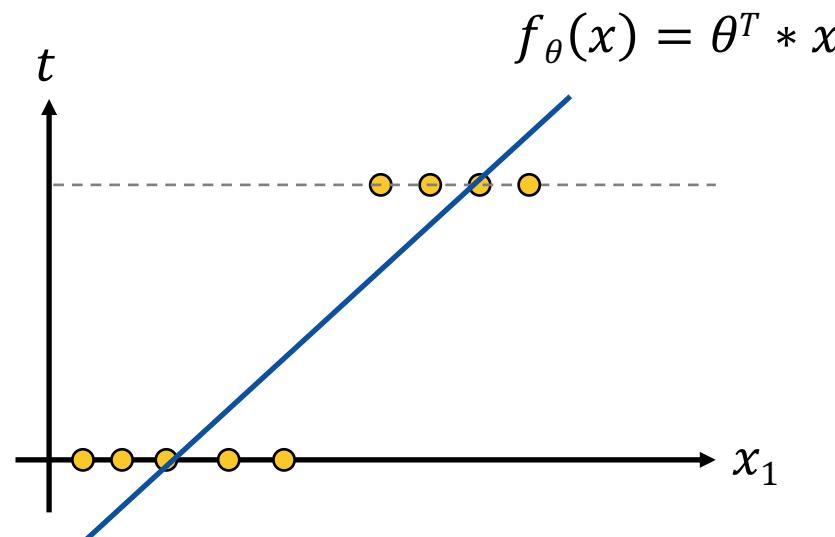
Treinamento e Função de Custo

Achar o vetor de parâmetros θ de maneira que o modelo estime:

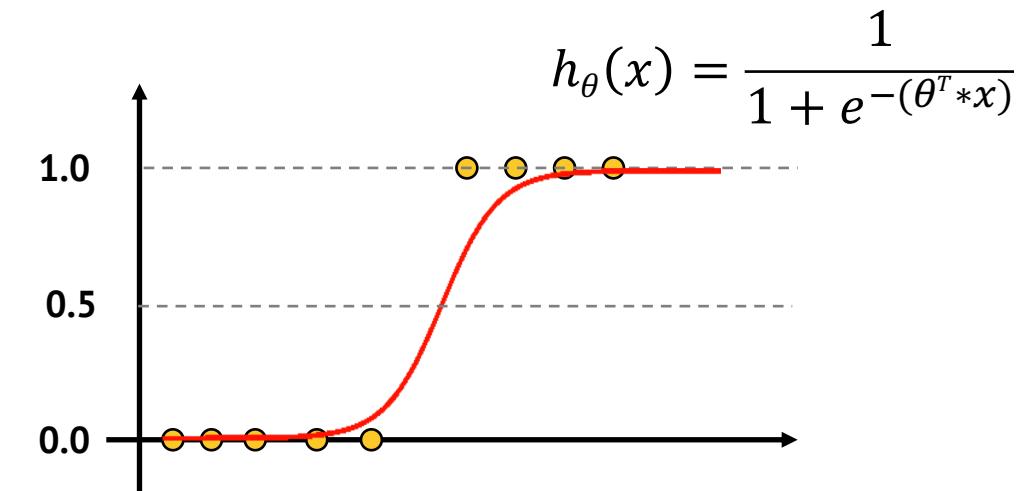
- **probabilidades altas** para **instâncias positivas ($y = 1$)**;
- **probabilidades baixas** para **instâncias negativas ($y = 0$)**

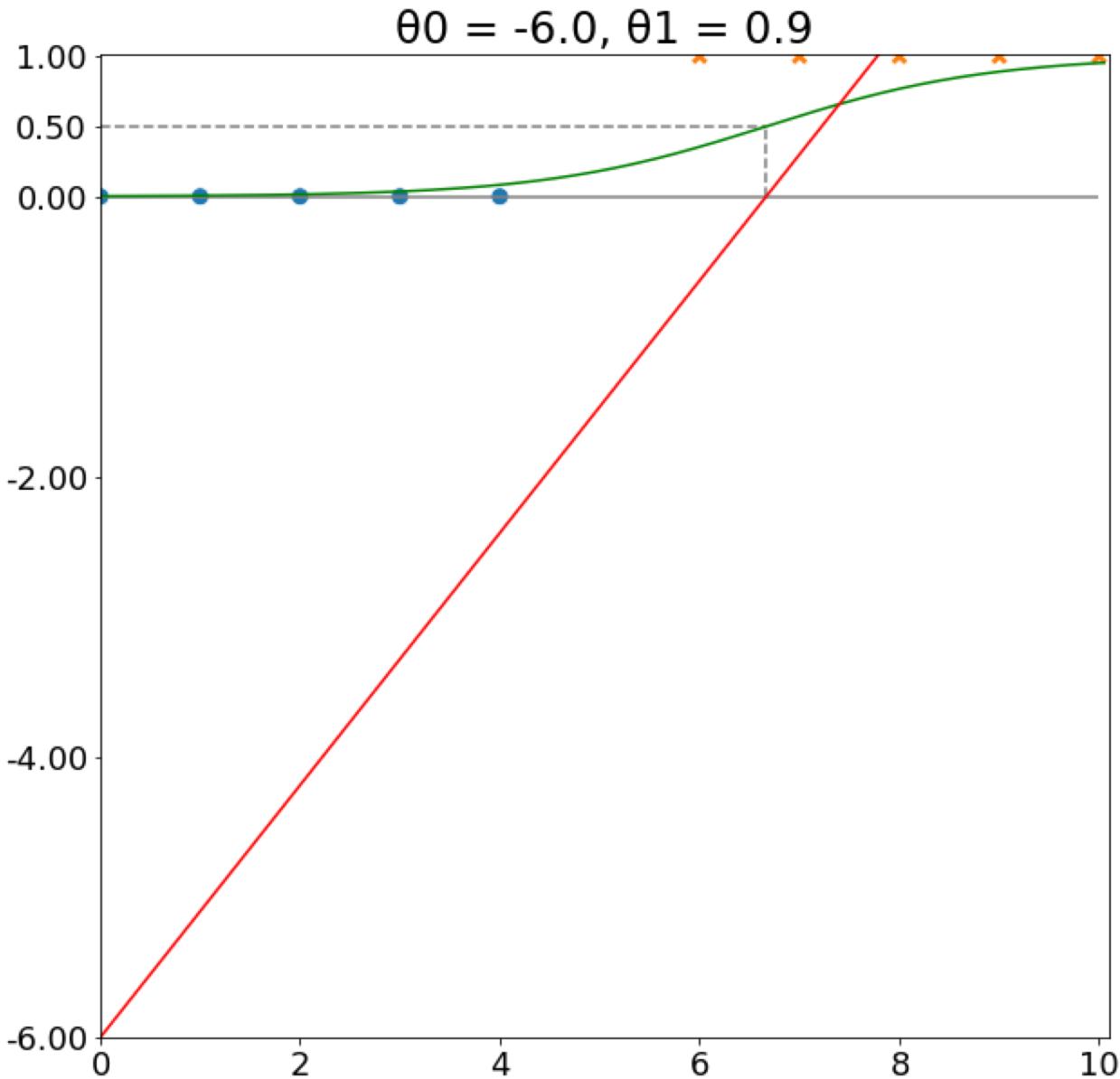
Função de Custo

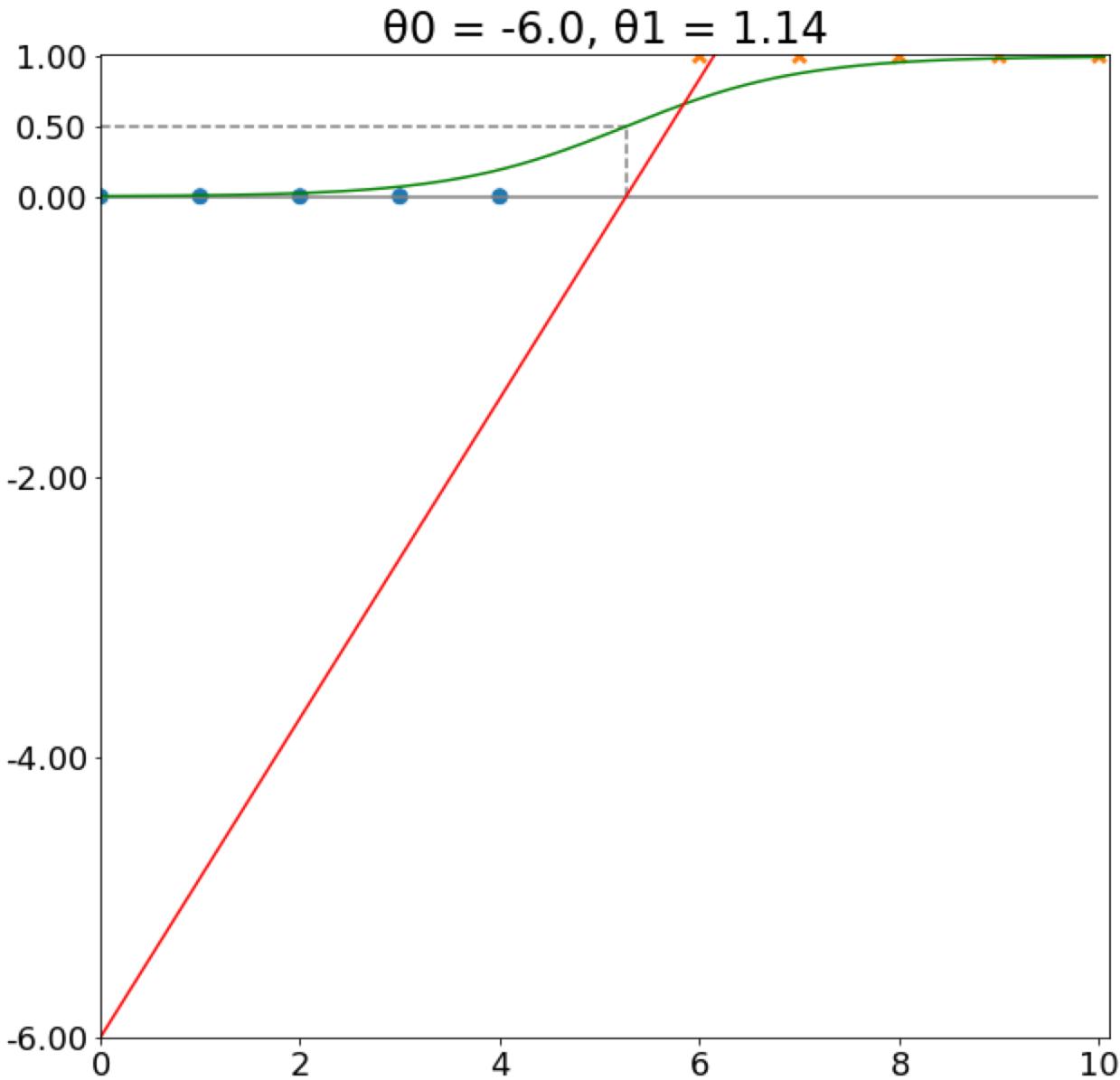
$$c(\theta) = \begin{cases} -\log(h_\theta(x)) & \text{se } y = 1 \\ -\log(1 - h_\theta(x)) & \text{se } y = 0 \end{cases}$$

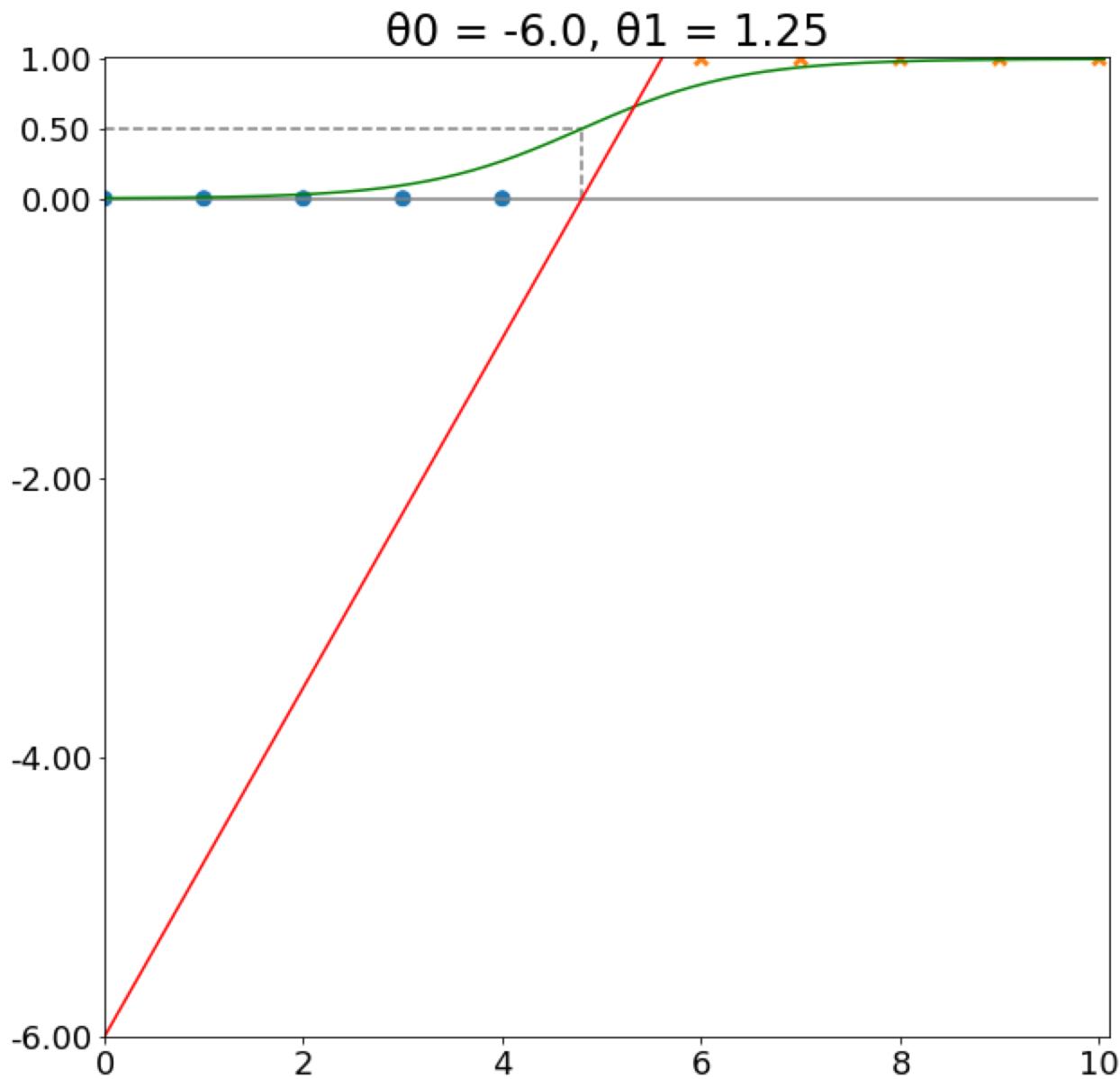


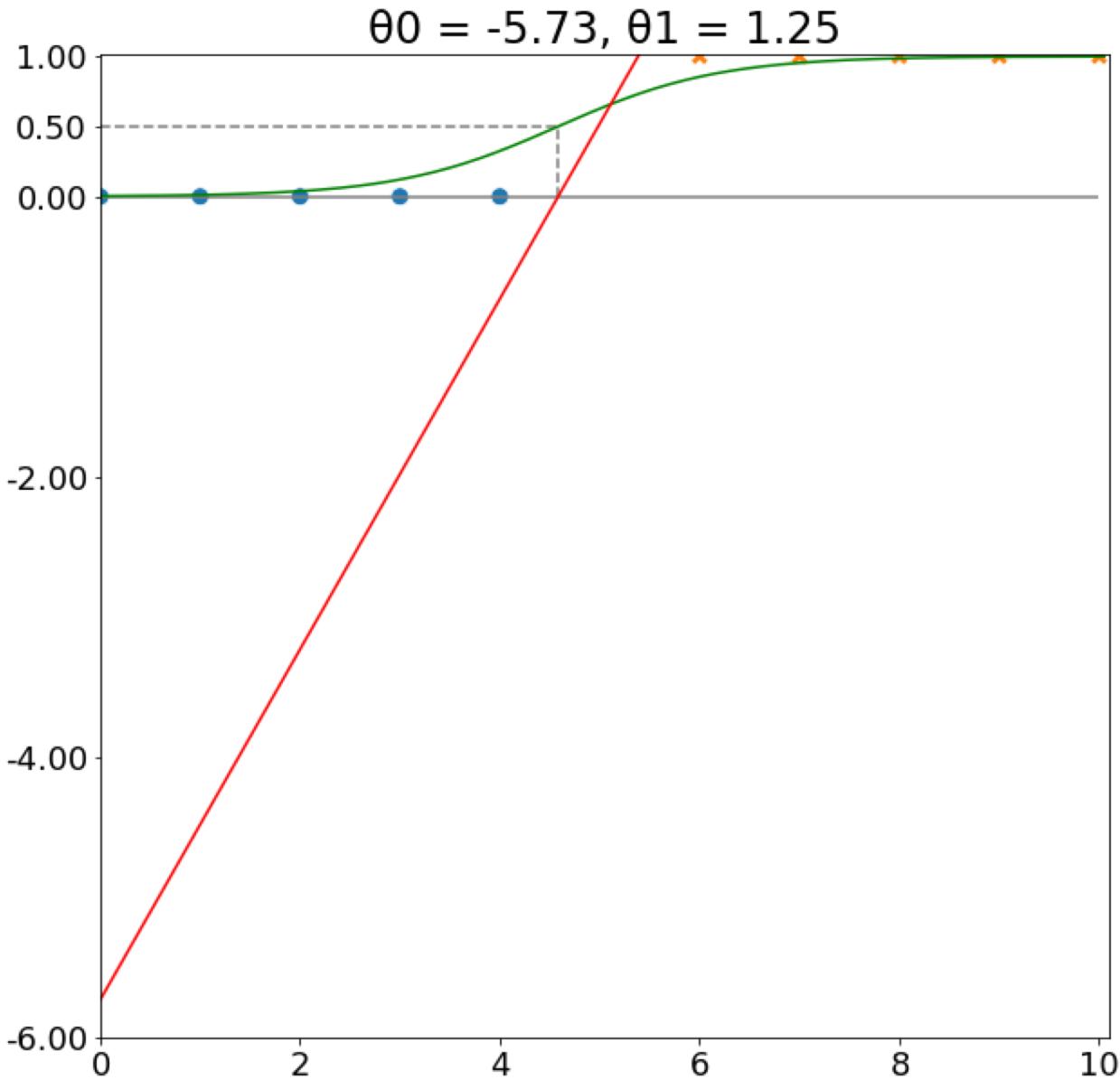
Modelo de Regressão Logística

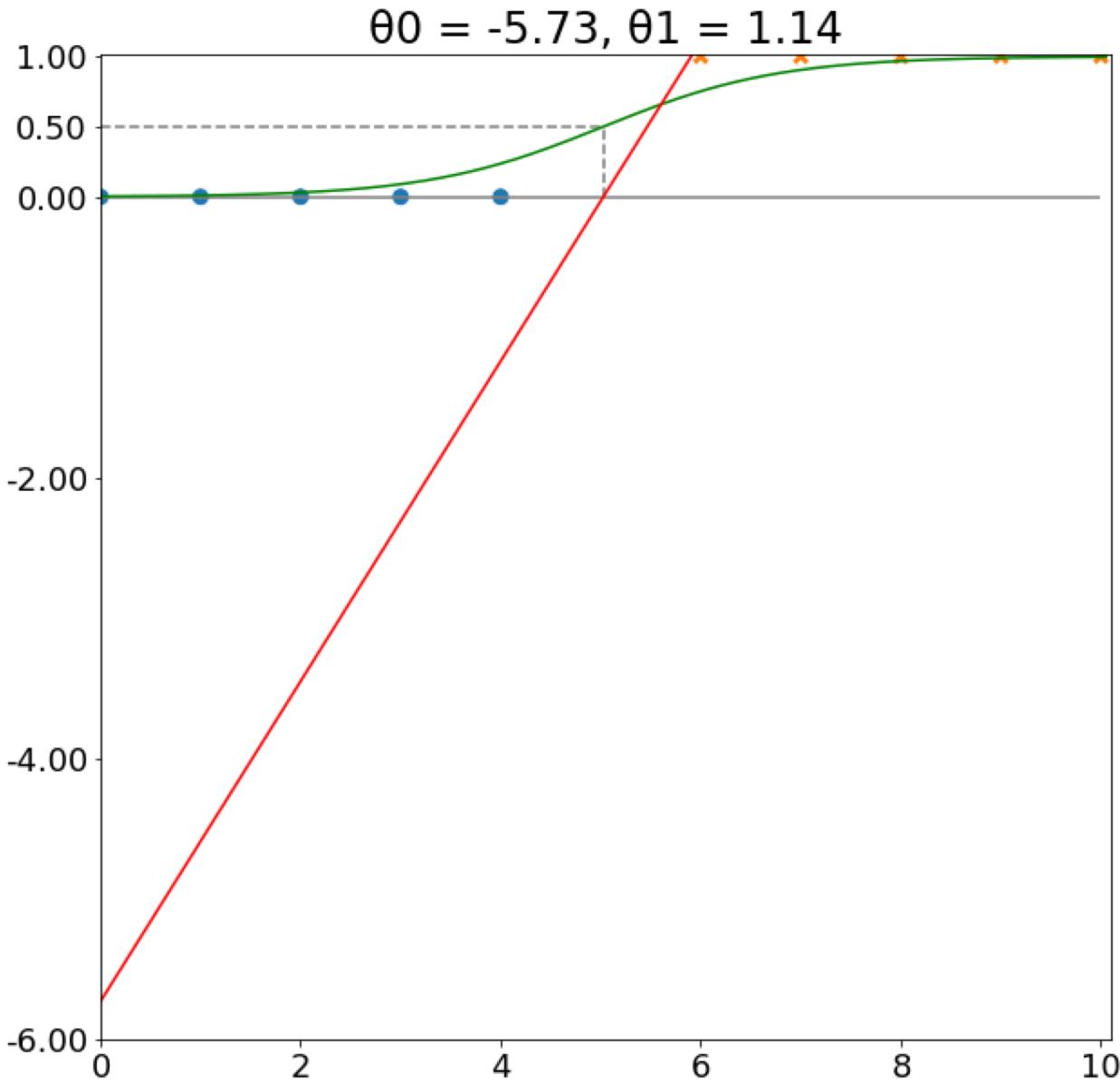


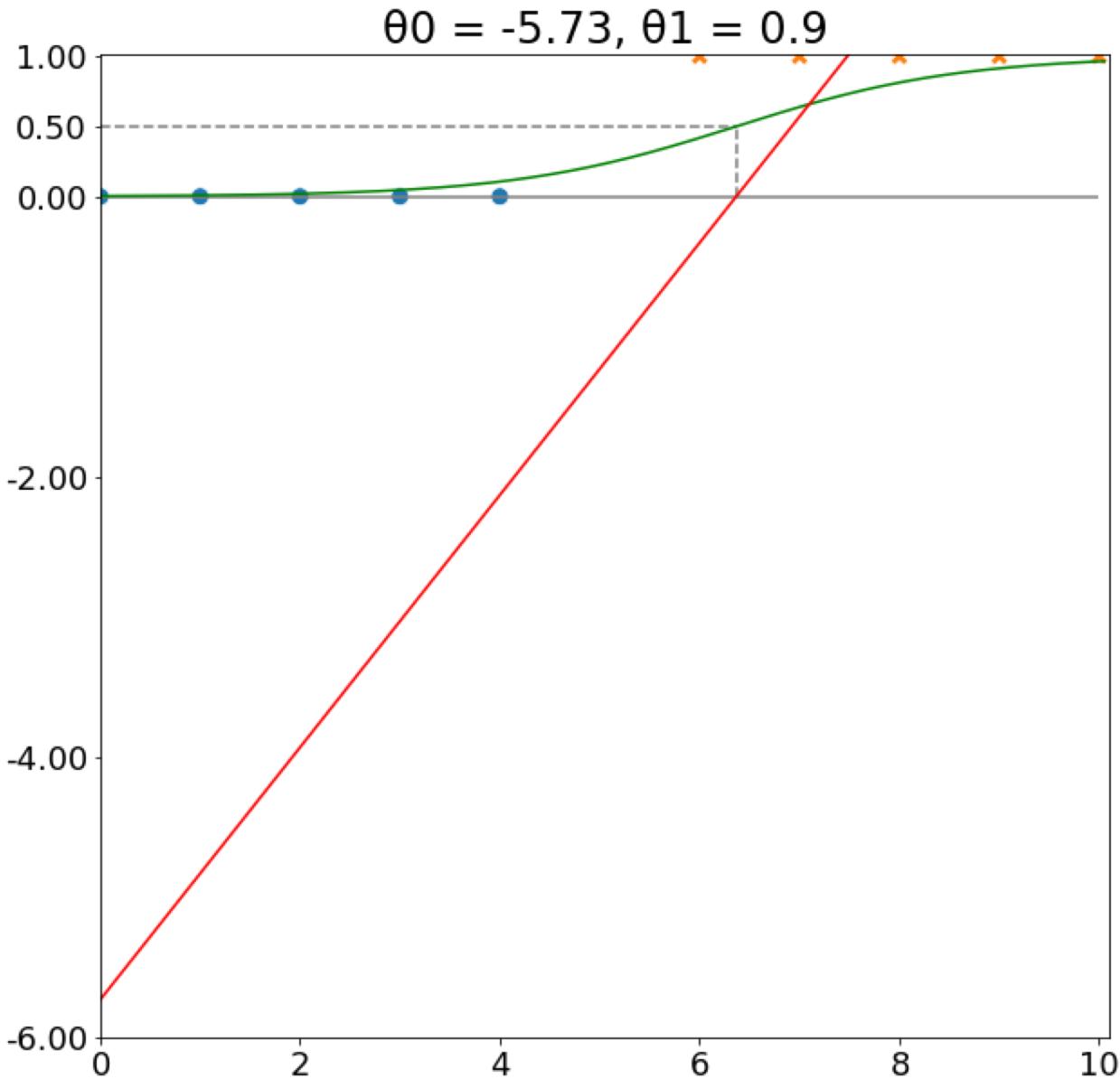


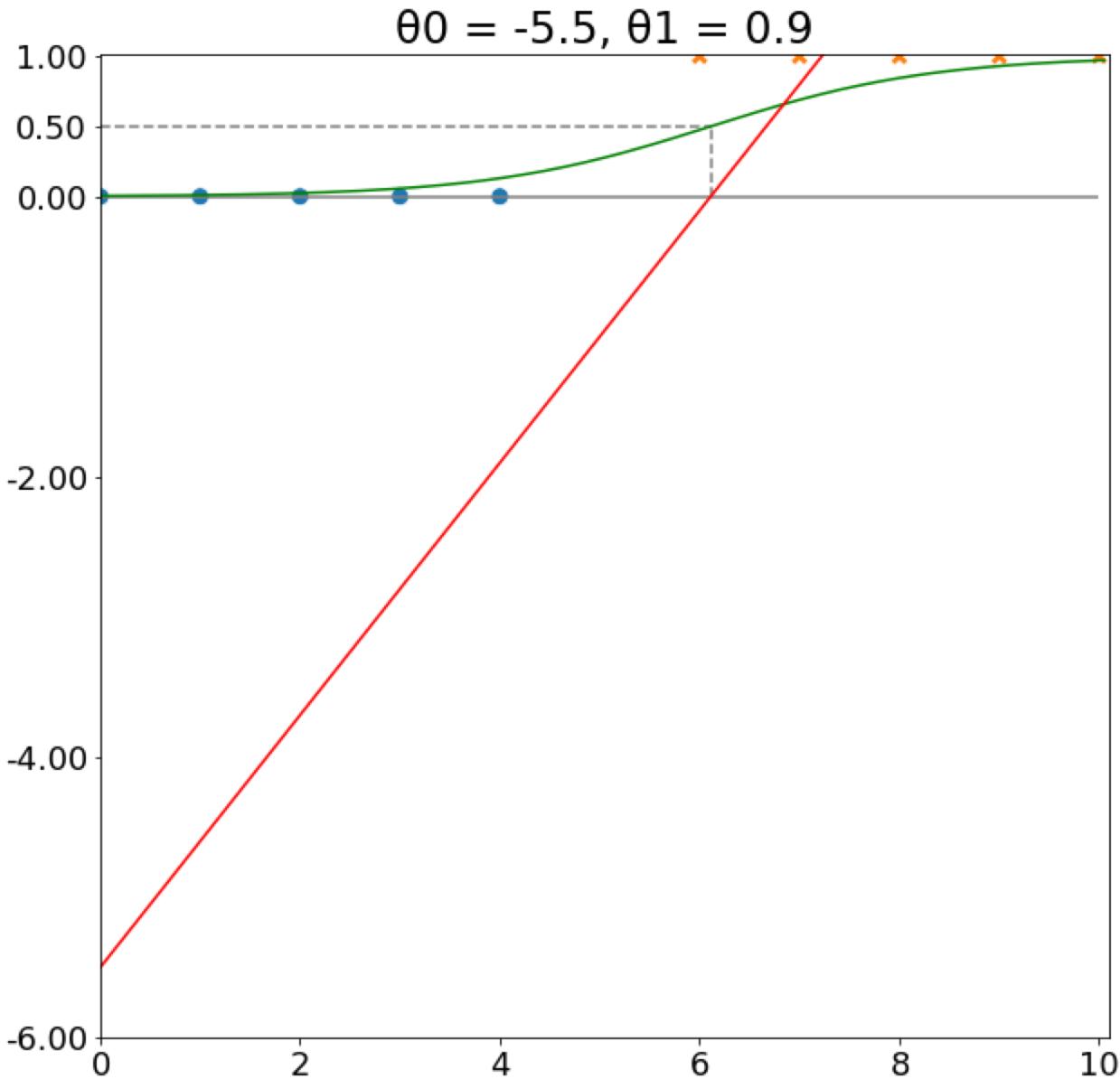


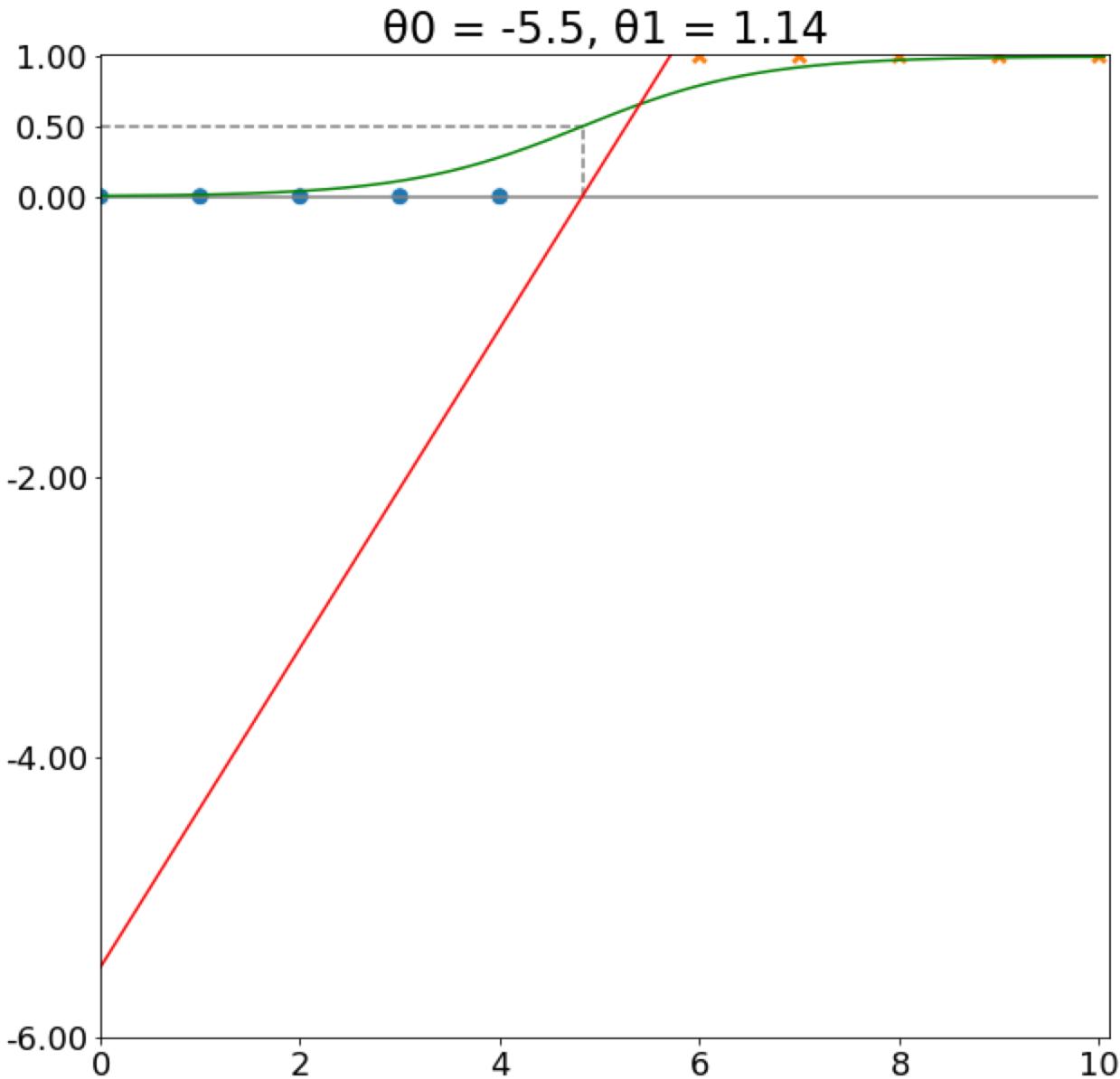


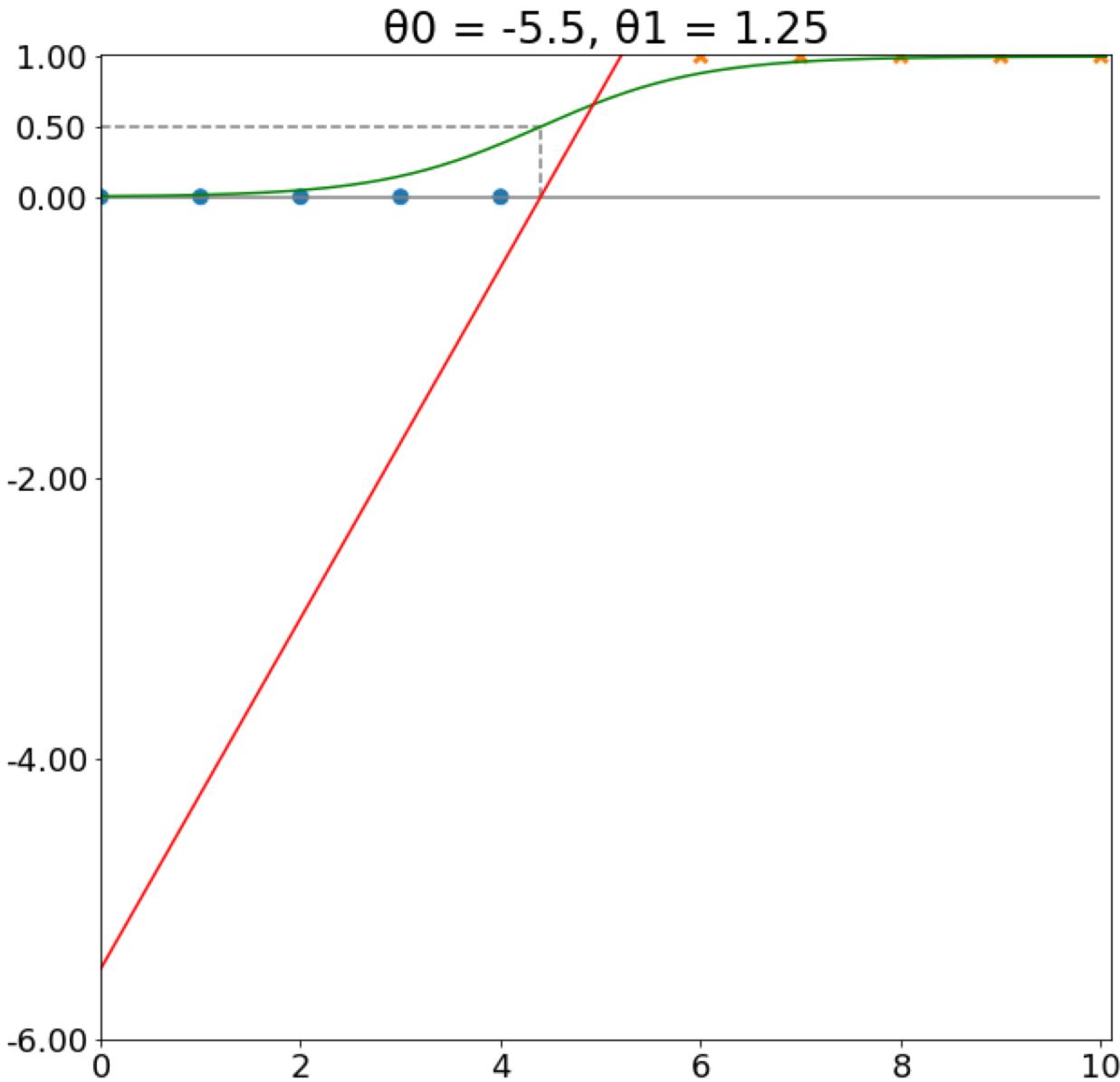




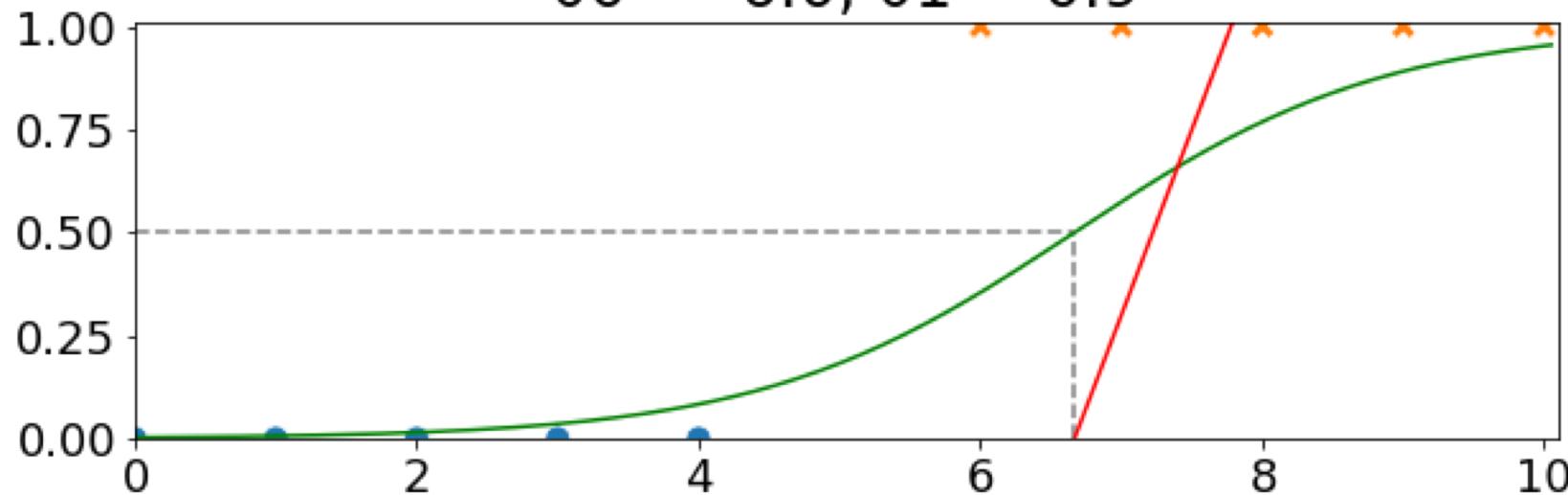




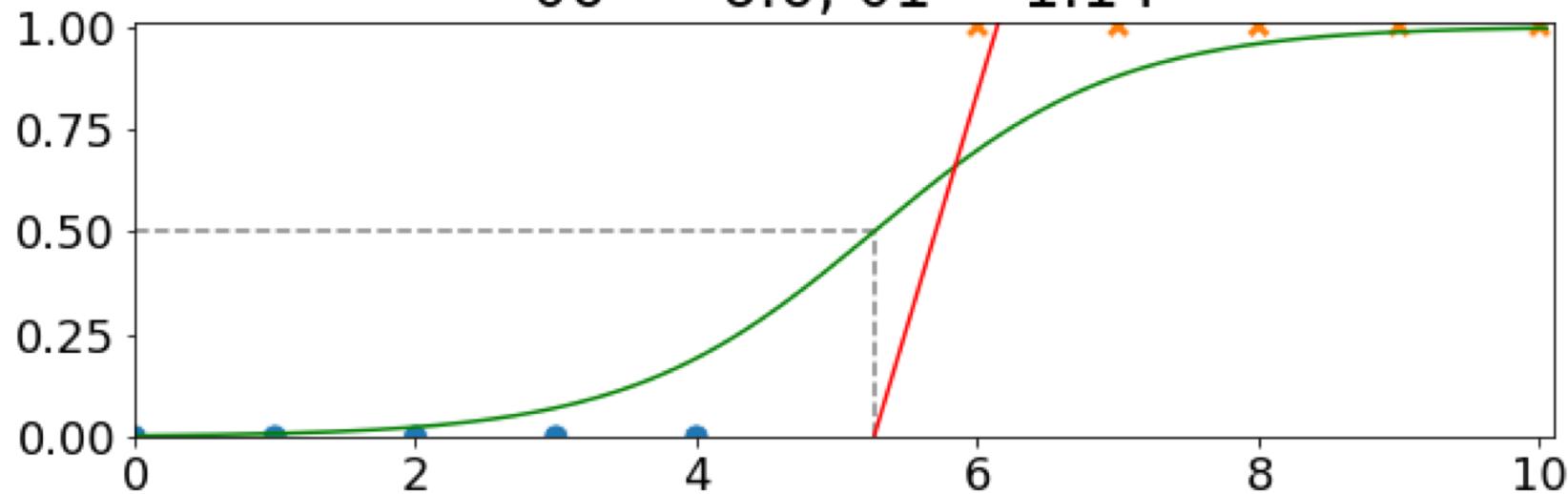


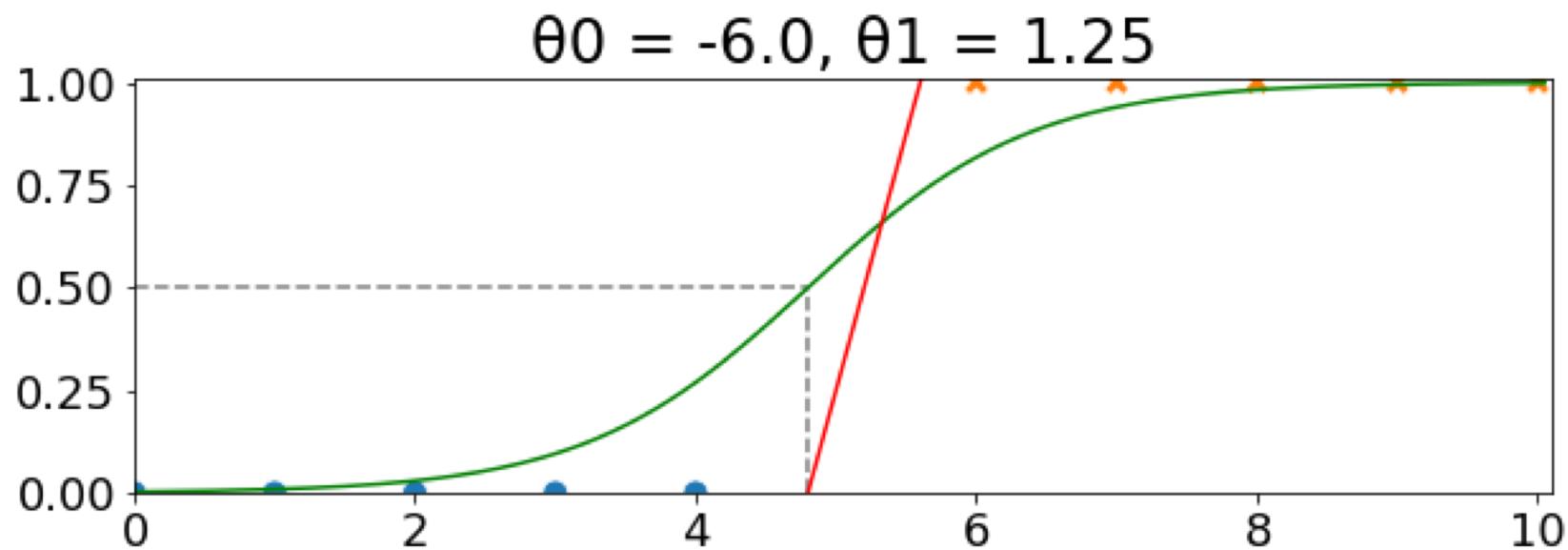


$$\theta_0 = -6.0, \theta_1 = 0.9$$

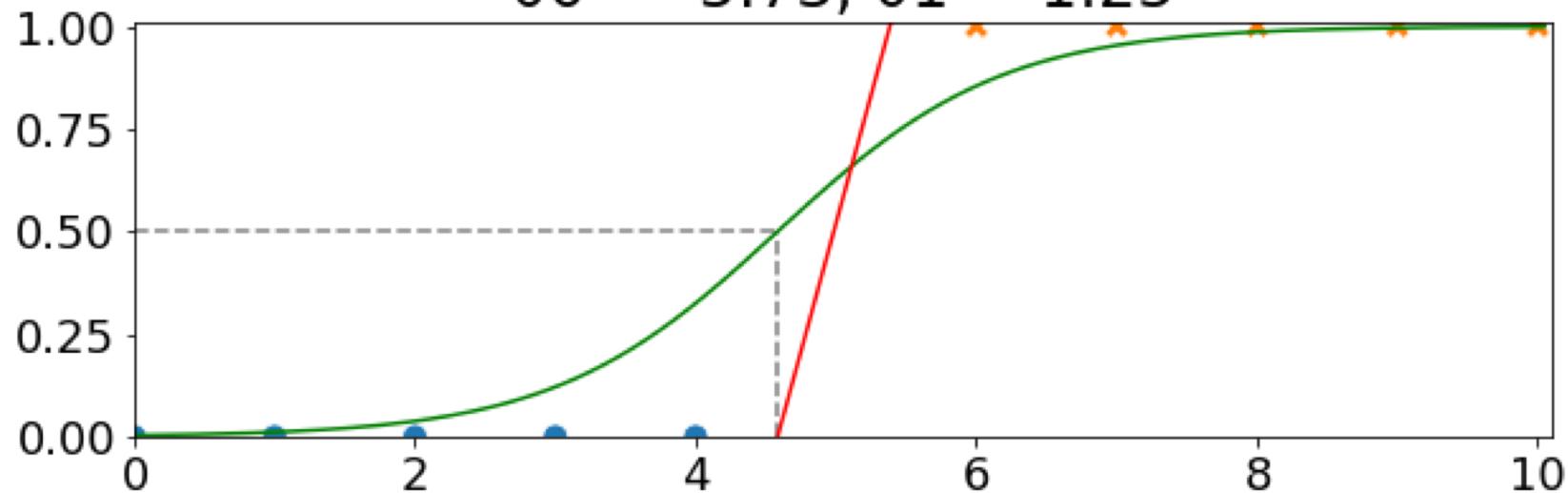


$$\theta_0 = -6.0, \theta_1 = 1.14$$

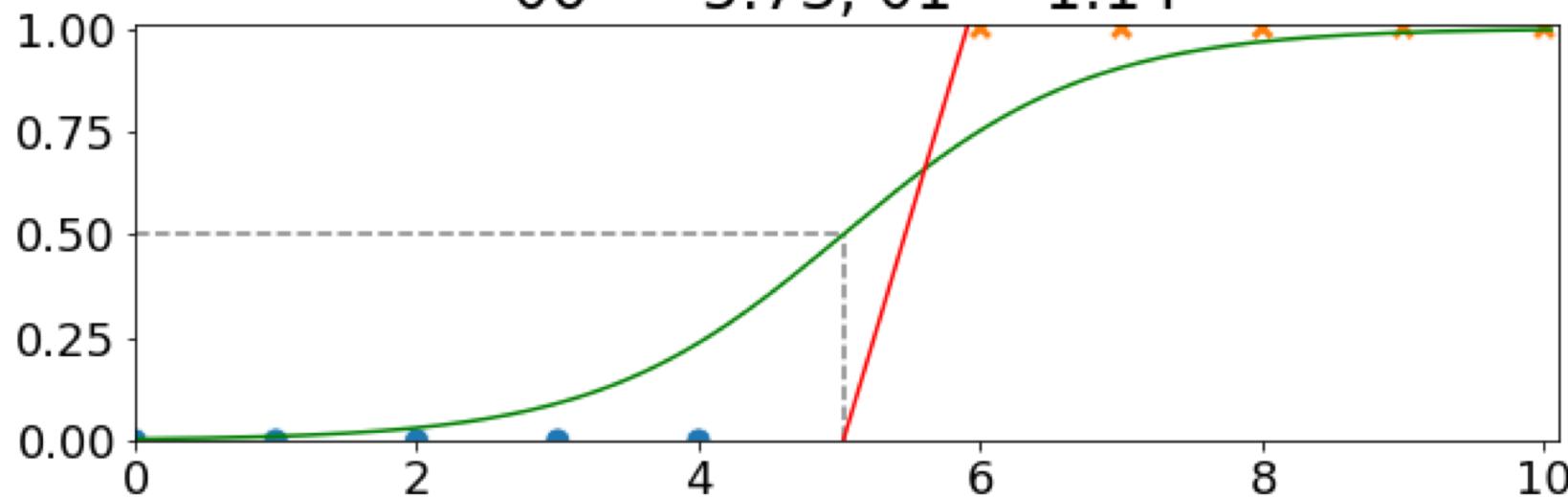




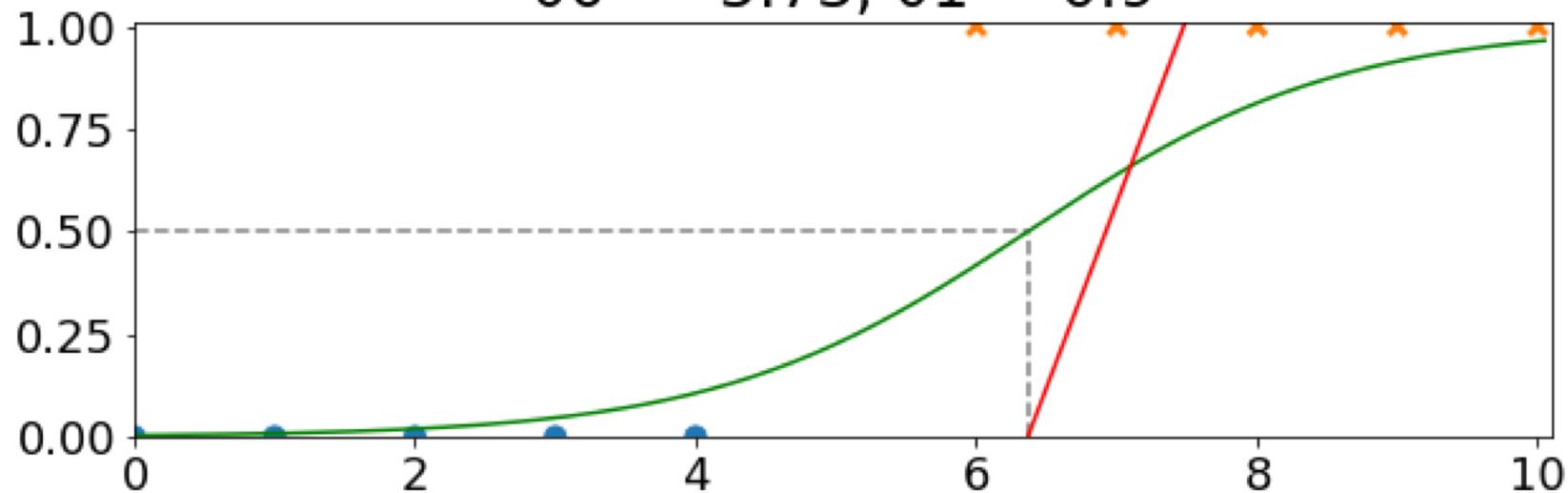
$$\theta_0 = -5.73, \theta_1 = 1.25$$



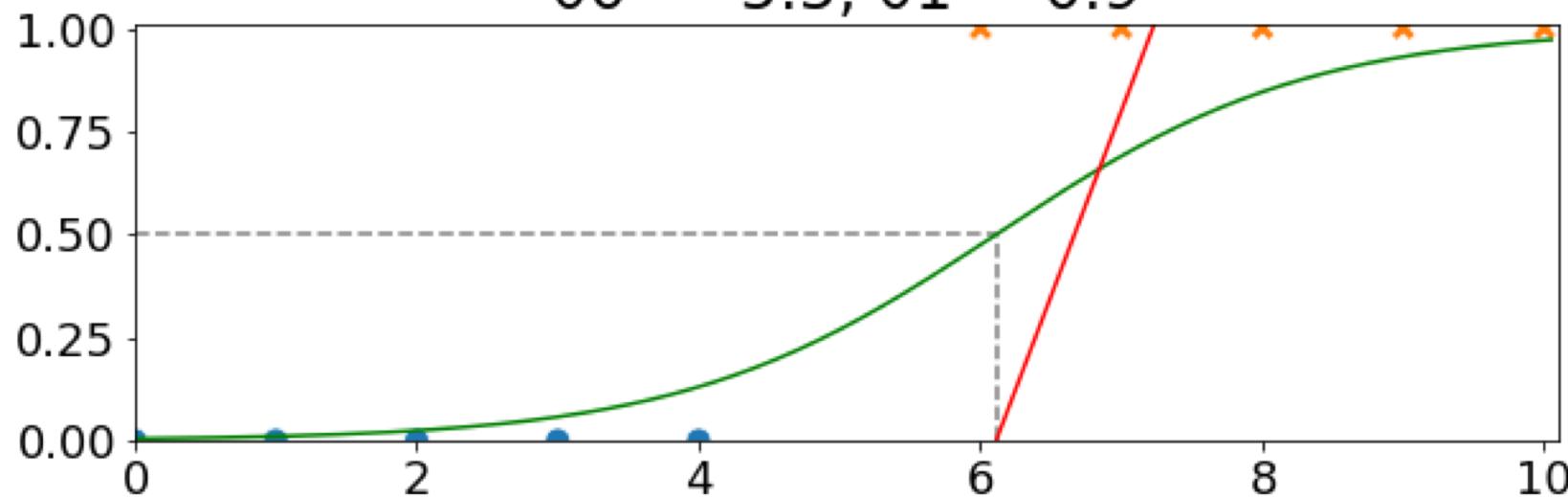
$$\theta_0 = -5.73, \theta_1 = 1.14$$



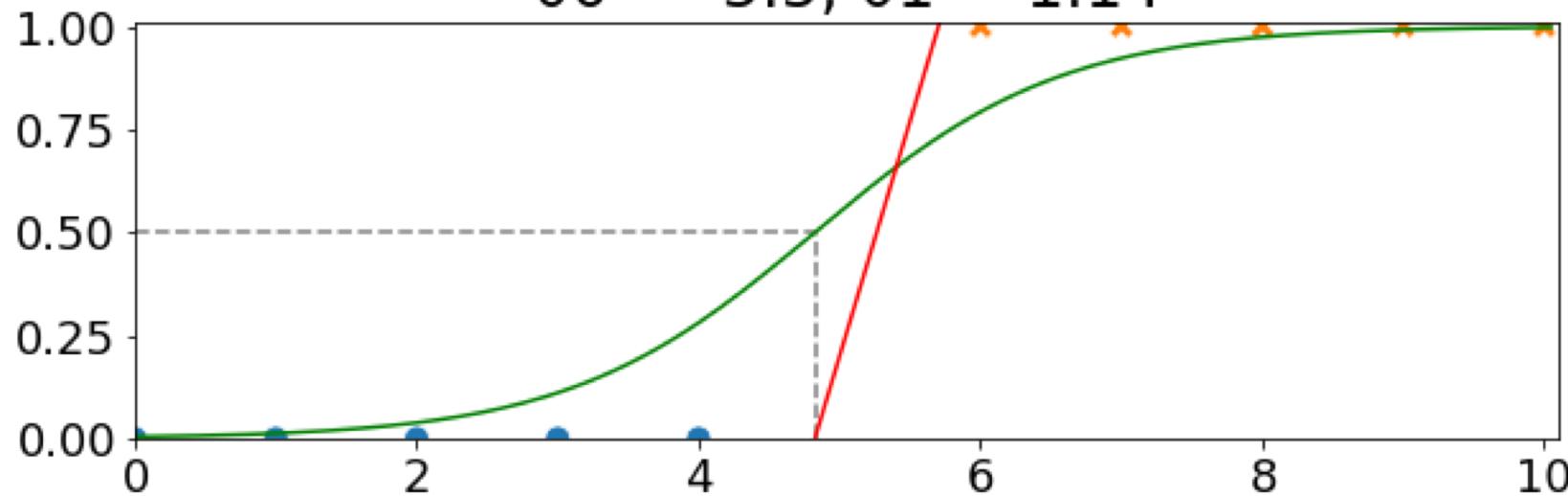
$$\theta_0 = -5.73, \theta_1 = 0.9$$

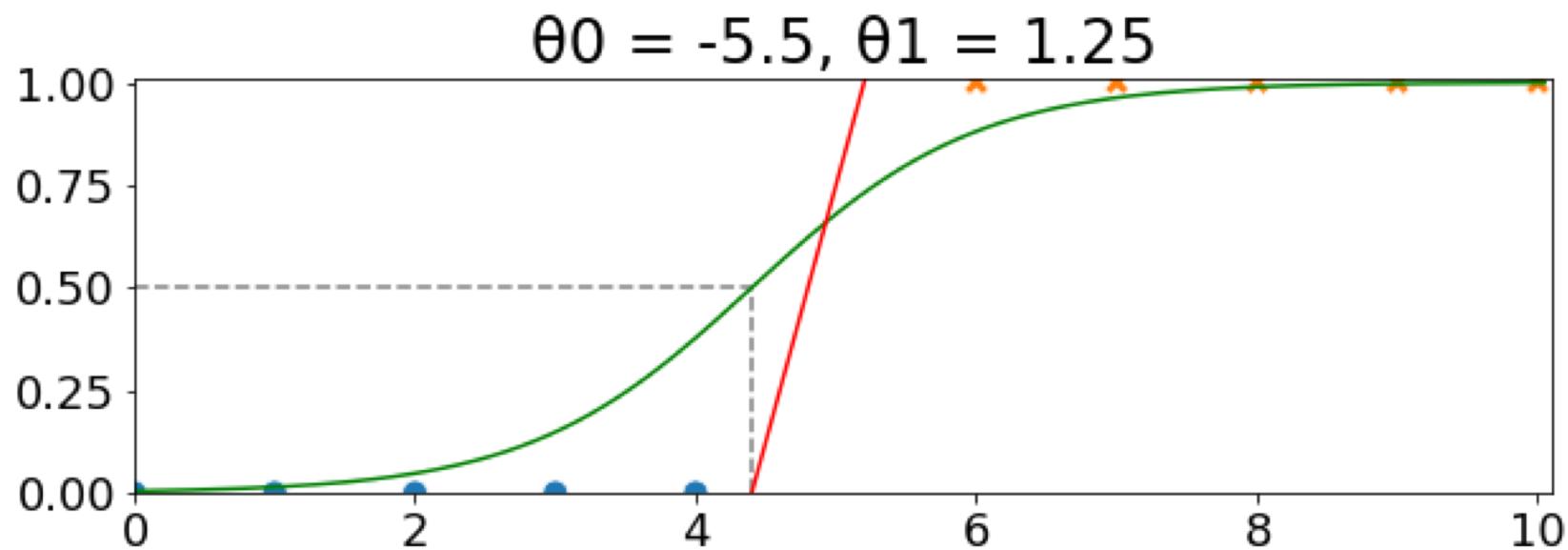


$$\theta_0 = -5.5, \theta_1 = 0.9$$



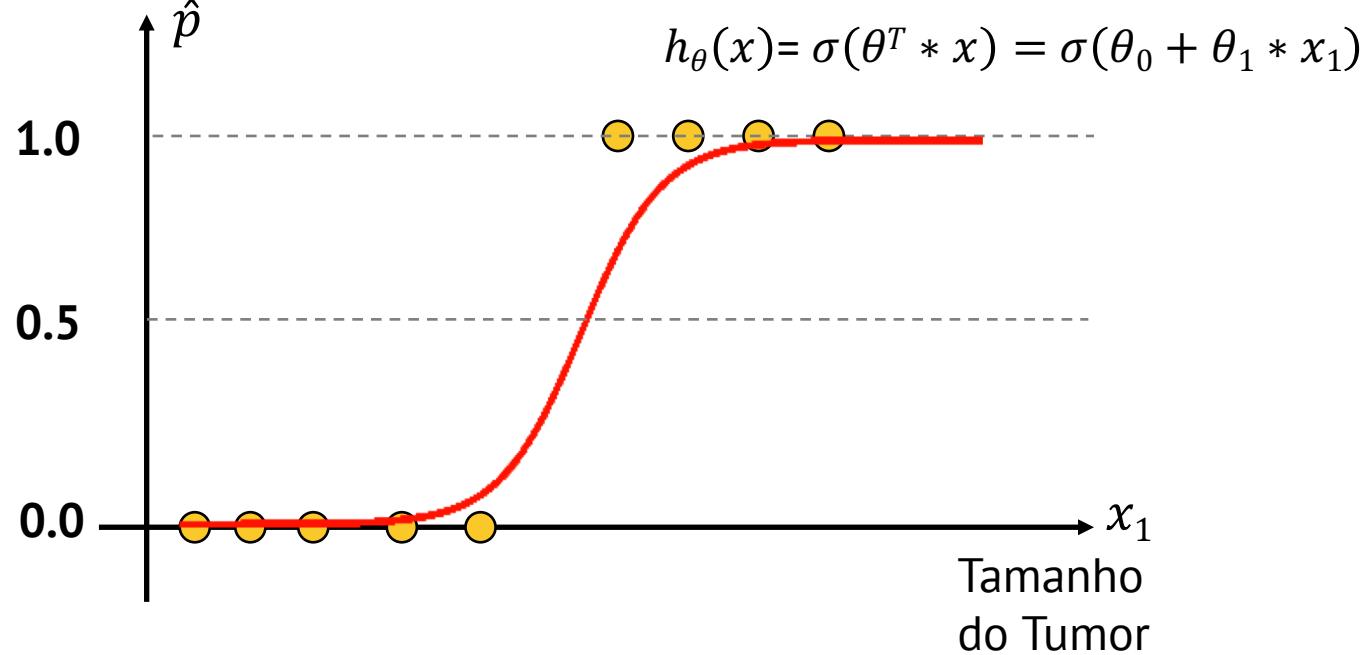
$$\theta_0 = -5.5, \theta_1 = 1.14$$





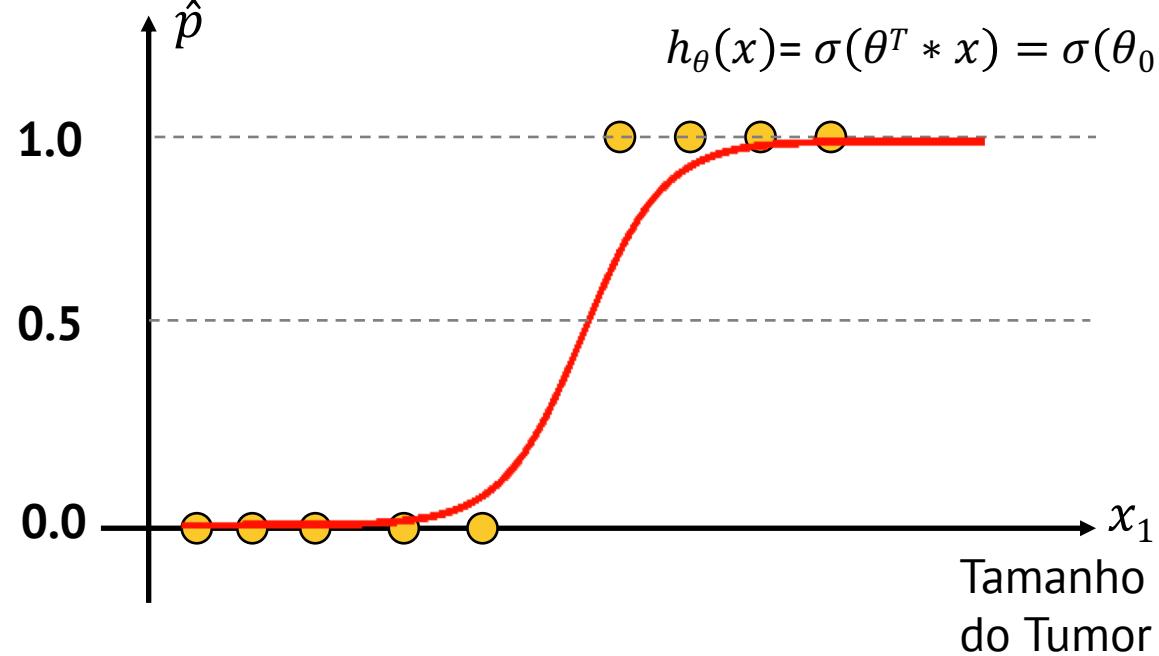
Decision Boundary

Probabilidade
Tumor ser Maligno



Decision Boundary

Probabilidade
Tumor ser Maligno



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

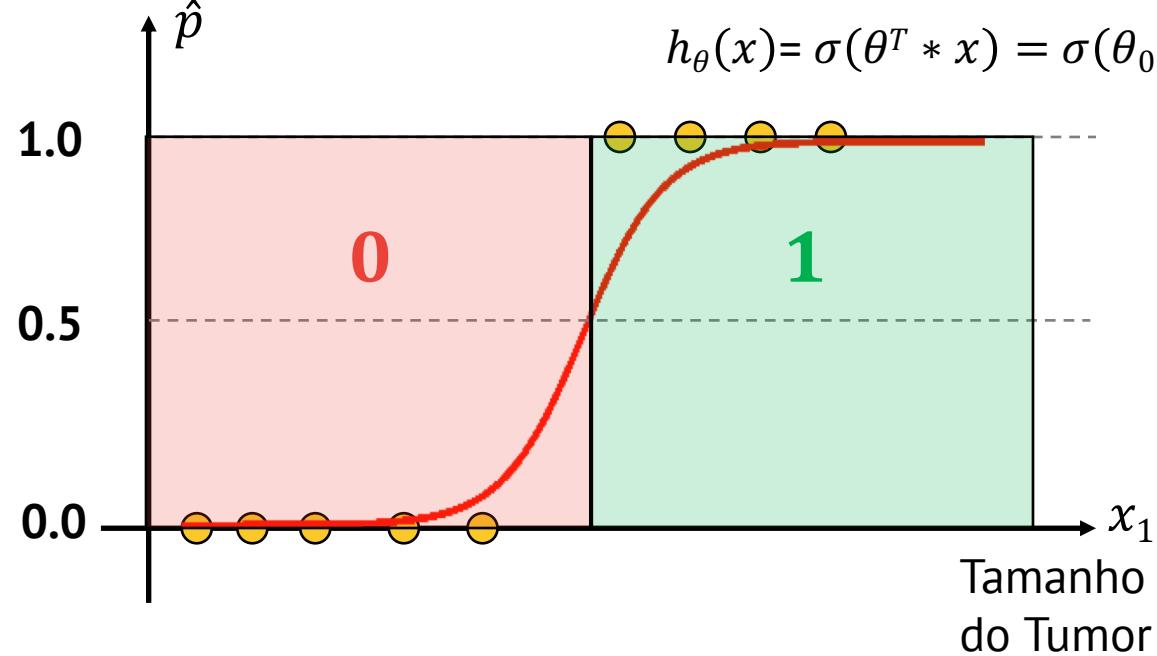
$\sigma(t) < 0.5$ quando $t < 0$ (**classe negativa**)

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

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

Decision Boundary

Probabilidade
Tumor ser Maligno



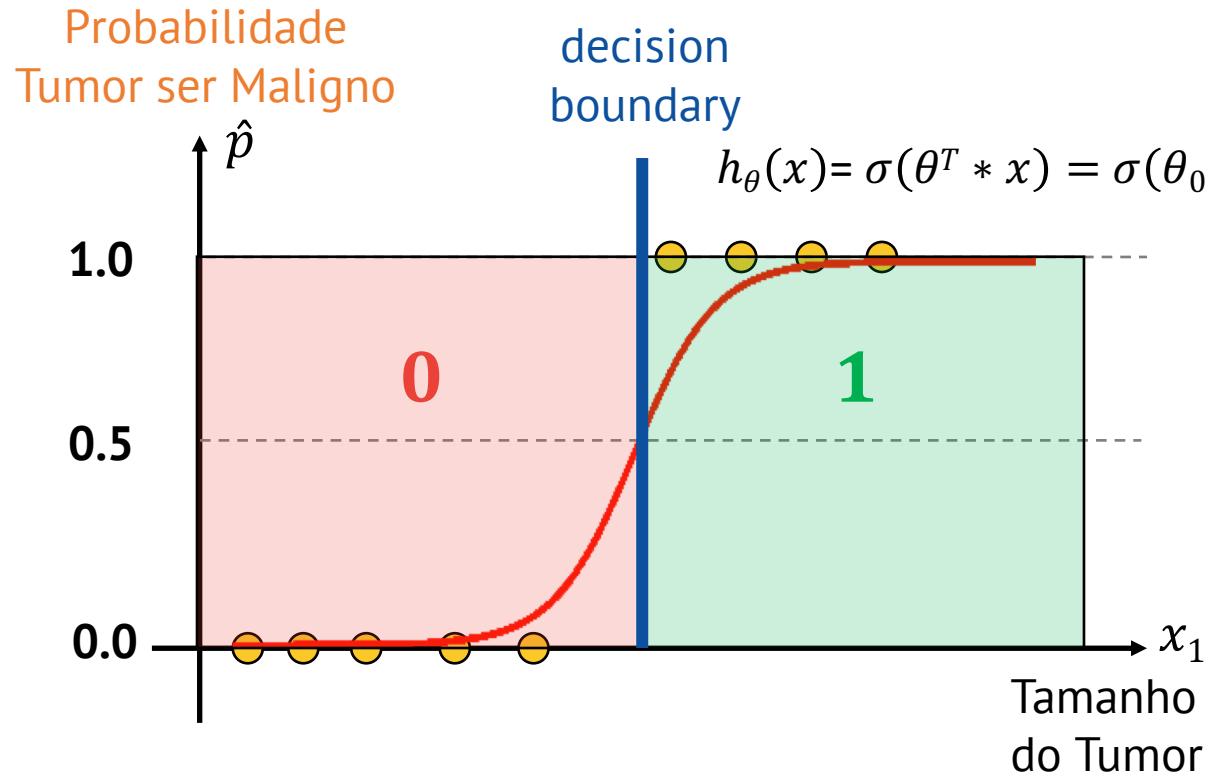
$\sigma(t)$ retorna um número entre **0** e **1**

$\sigma(t) < 0.5$ 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**

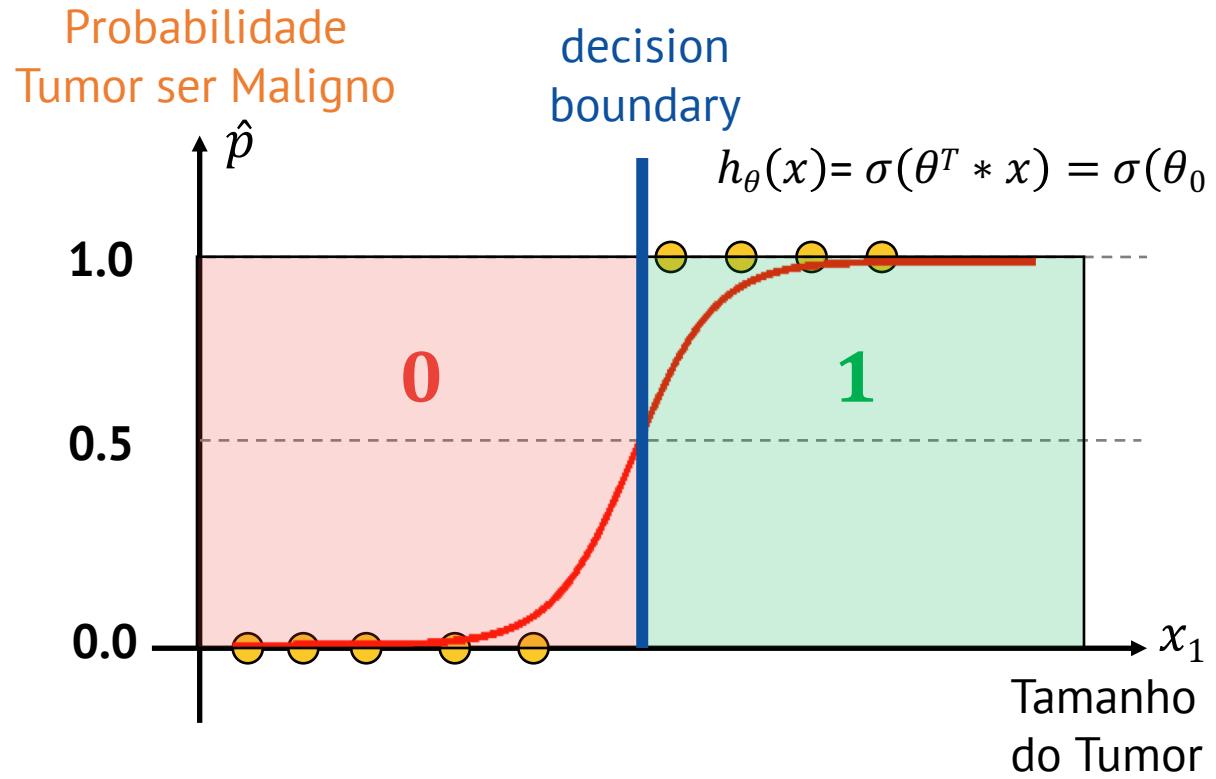
$\sigma(t) < 0.5$ quando $t < 0$ (**classe negativa**)

$\sigma(t) \geq 0.5$ 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



$\sigma(t)$ 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**)

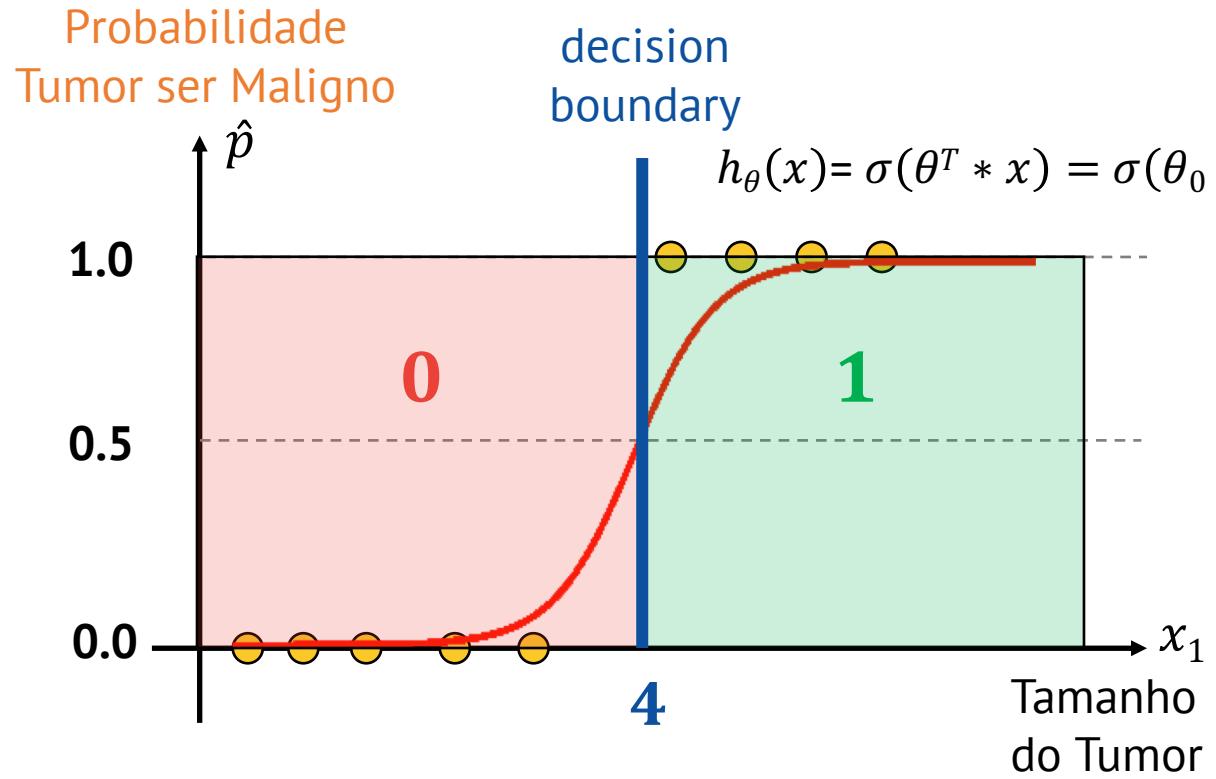
$\sigma(t) = 0.5$ quando $t = 0$

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

Suponha que o melhor vetor θ encontrado foi:

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

Decision Boundary



Suponha que o melhor vetor θ encontrado foi:

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

$\sigma(t)$ 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**)

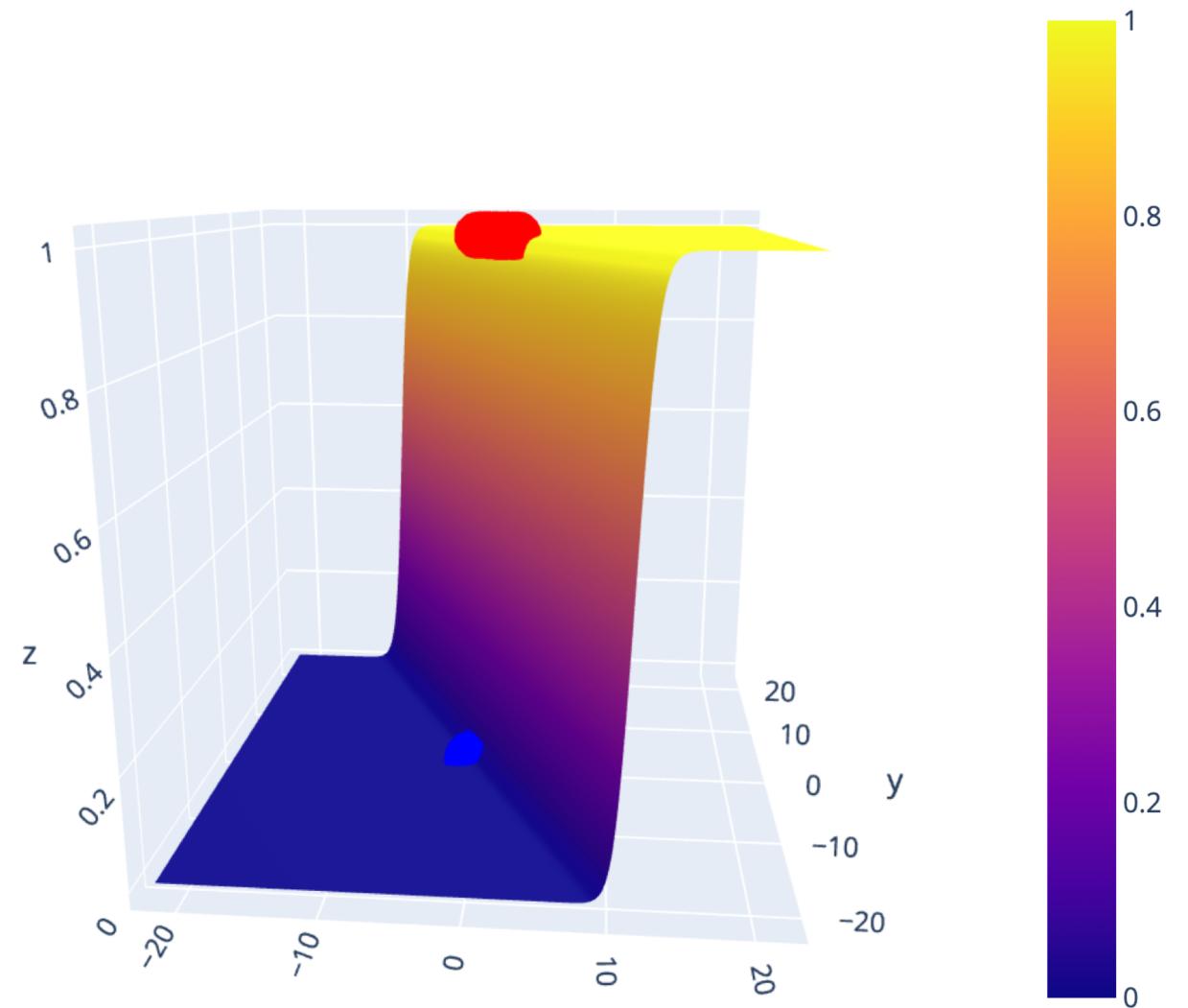
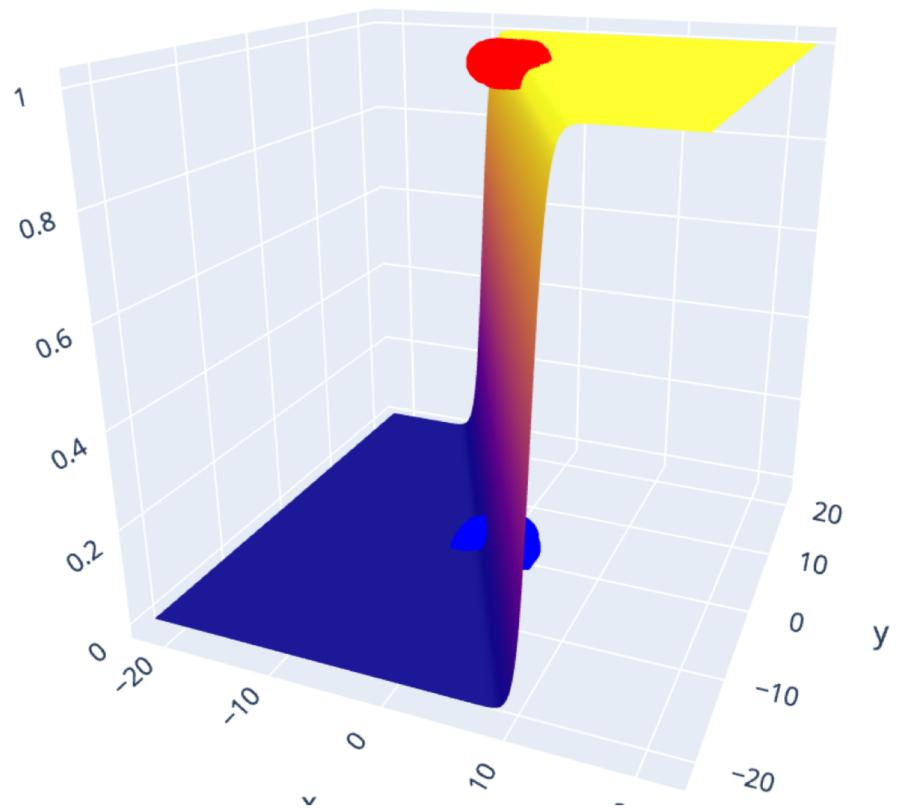
$\sigma(t) = 0.5$ quando $t = 0$

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

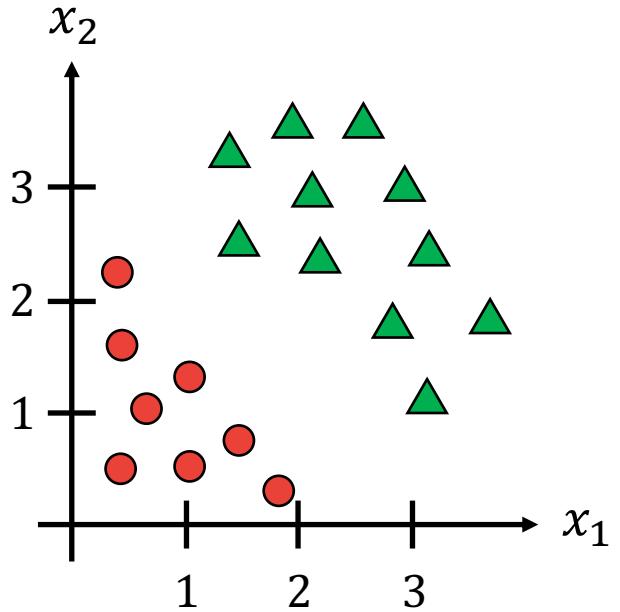
$$\begin{aligned} \theta^T * x &= 0 \\ -4 + 1 * x_1 &= 0 \\ x_1 &= 4 \end{aligned}$$

decision boundary

Decision Boundary: outro exemplo



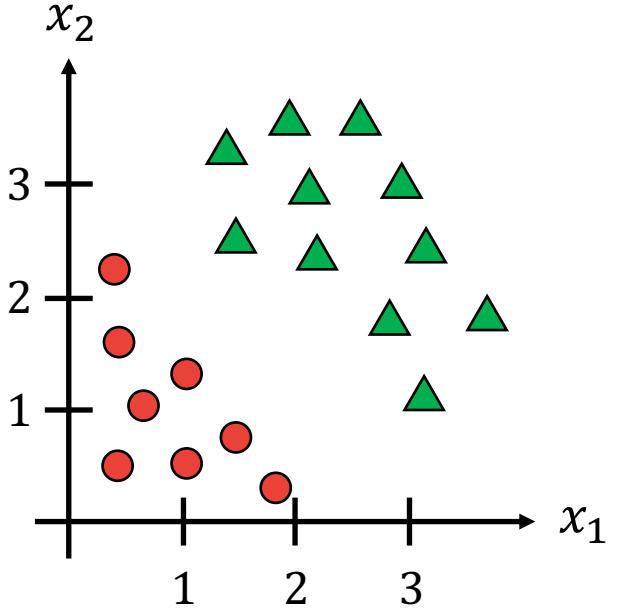
Decision Boundary: outro exemplo



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

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

Decision Boundary: outro exemplo



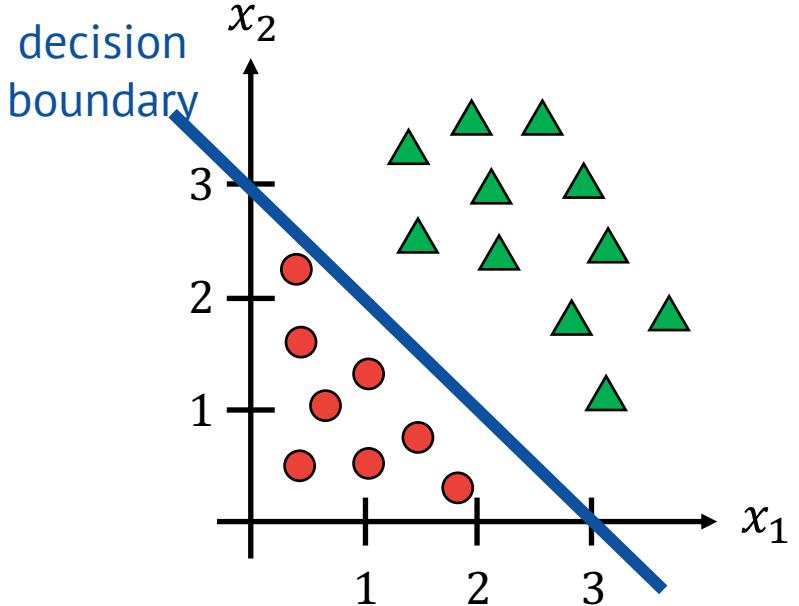
O hiperplano $(\theta^T * x) = 0$ forma a **decision boundary** do **classificador**.

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

Suponha que o melhor vetor θ encontrado foi:

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

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:

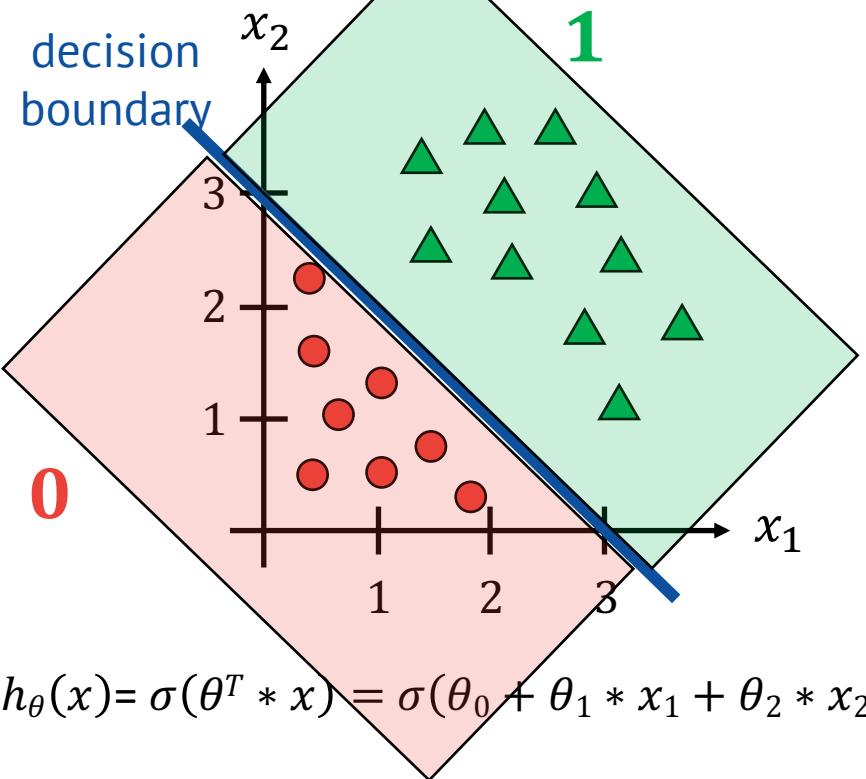
$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \quad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

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



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

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

