

# Aprendizado Supervisionado

Prof. Leandro Alvim, D.Sc.

Julho, 2016

# Agenda

1. O que é
2. Classificação e Regressão
3. Terminologias
4. O Problema de Otimização
5. Exercícios

- É fornecido um **gabarito** para cada **exemplo**
- Tipos de Problemas
  - Classificação
  - Regressão

# Problema de Classificação

## Conceito de **Classe**

- Email: Spam ou Não?
- Identificação de Caracteres Manuscritos (a,b, ..1,2,..)
- Tumor: Maligno ou Benigno?
- Análise de Sentimentos

## Exemplo...

Densidade	Tamanho	Câncer?
0,10	1,2	0
0,15	1,3	0
0,3	2	0
0,4	3	1
0,8	8	1

Tabela: Conjunto de Dados

Dados extraídos a partir de imagens de tumores e diagnosticados manualmente por um especialista

# Problema de Regressão

Nosso **alvo** agora é **contínuo**

Exemplos...

- Estimar preços de ações
- Estimar preços de casas

## Exemplo...

Tamanho ( $m^2$ )	Preço (R\$)
55	250.000
70	450.000
40	120.000
80	500.000
110	780.000
..	..

**Tabela:** Conjunto de Dados

Nosso **alvo** é o preço (contínuo).

# Aprendizado Supervisionado

- É fornecido um **gabarito** para cada **exemplo**
- Tipos de Problemas
  - Classificação (**discreto**)
  - Regressão (**contínuo**)



# Terminologia

Definindo alguns conceitos...

Densidade	Tamanho	Câncer?
0,10	1,2	0
0,15	1,3	0
0,3	2	0
0,4	3	1
0,8	8	1
..	..	..

Tabela: Conjunto de Dados

**Exemplo:** linha da matriz

**Característica:** coluna da matriz

**Alvo:** o que desejamos aprender

# Terminologia

Densidade	Tamanho	Câncer?
0,10	1,2	0
0,15	1,3	0
0,3	2	0
0,4	3	1
0,8	8	1
..	..	..

Tabela: Conjunto de Dados

**Exemplo:** linha da matriz

**Característica:** coluna da matriz

**Alvo:** o que desejamos aprender é **discreto**

classe 0 : benigno

classe 1 : maligno

# Visualizando os Dados

Exemplos, atributos e gabaritos...

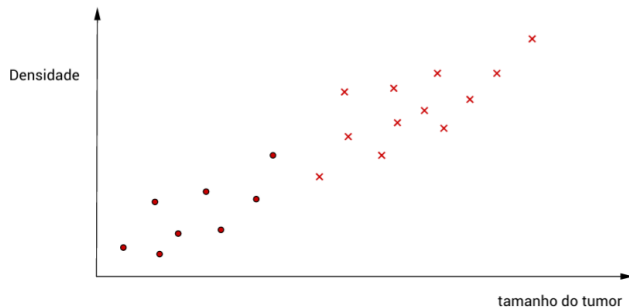


Figura: Classificação de tumores

$$x_i = \begin{bmatrix} 1 \\ \text{tam}_i \\ \text{den}_i \end{bmatrix}, y_i = \{0, 1\}$$

# Classificação - Objetivo

Encontrando um **separador ou modelo**  $\theta$  ideal...

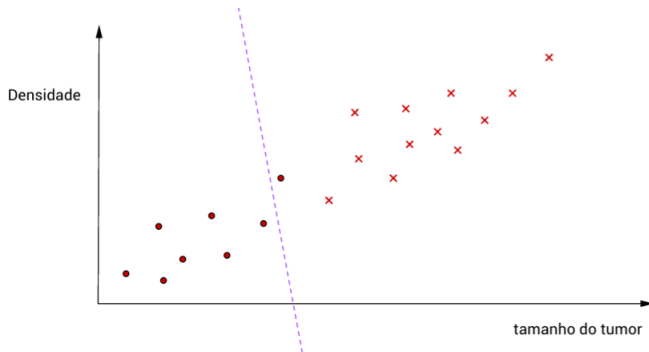


Figura: Classificação de tumores

$$\theta_i^T = [\theta_0 \ \theta_1 \ \theta_2]$$

# Classificação - Objetivo

Encontrando um **separador** ou **modelo**  $\theta$  ideal...

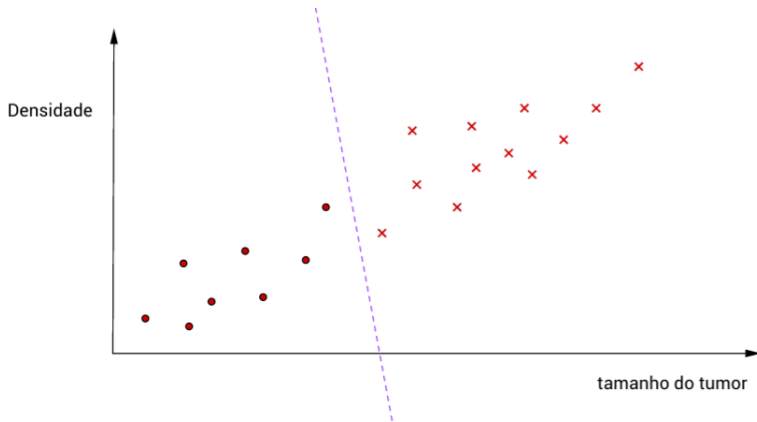
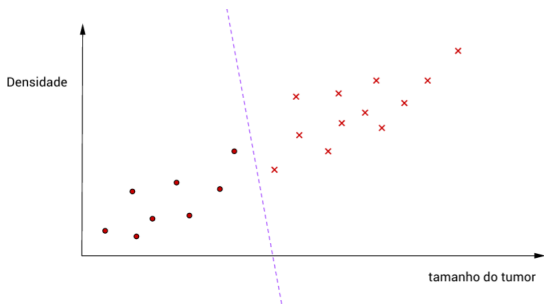


Figura: Classificação de tumores

# Classificação - Objetivo

Nossa hipótese  $h_\theta$  é...



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

Limiar de classificação em 0:

Se  $h_\theta(x) \geq 0$ , estimar "y=1"

Caso contrário, estimar "y=0"

# Regressão - Objetivo

Encontrando um **modelo**  $\theta$  ideal...

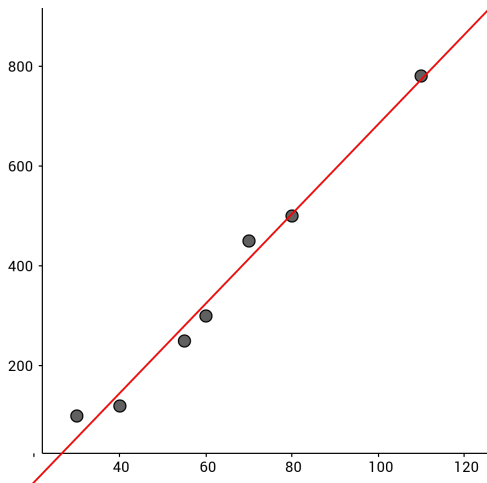


Figura: Regressão Linear ( $h_{\theta} = 8.95x_1 - 212$ )

# Regressão - Objetivo

Encontrando um **modelo**  $\theta$  ideal...

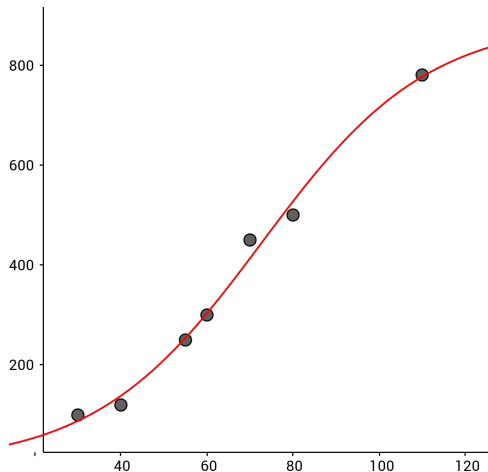
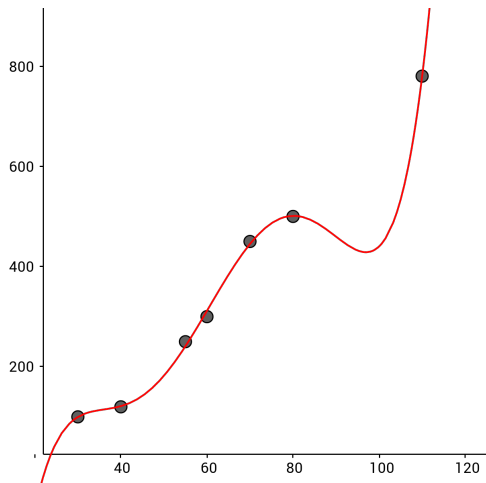


Figura: Regressão Logística ( $h_{\theta} = \frac{889.83}{1+42.88e^{-0.05x_1}}$ )



# Regressão - Objetivo

Encontrando um **modelo**  $\theta$  ideal...



**Figura:** Regressão Polinomial

$$(h_{\theta} = 0x^5 - 0x^4 + 0.24x^3 - 12.75x^2 + 326.22x - 3091.71)$$

# Formalizando

Seja um conjunto de treino

$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  em que  $y_i$  discreto ou contínuo

$$\text{tal que, } x_i = \begin{bmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ a_d \end{bmatrix}, y_i$$

em que  $a_i$  corresponde ao  $i$ -ésimo atributo do exemplo  $x_i \in \mathbb{R}^{d+1}$ .

# Formalizando

Nosso objetivo é:

$$\operatorname{argmin}_{\theta} J(\theta)$$

em que,

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n \text{custo}(h_{\theta}(x_i, y_i))$$

# Exercícios

- Elabore uma função de custo para o problema de regressão
- Como podemos melhorar a qualidade do modelo de regressão de preços de casas?
- O que aconteceria se descartássemos uma característica do conjunto de dados de tumores?
- Dentre os modelos de regressão vistos, qual você acha que possui maior qualidade?