Fundamentos Matemáticos e Computacionais de Machine Learning

Especialização em Machine Learning e Big Data

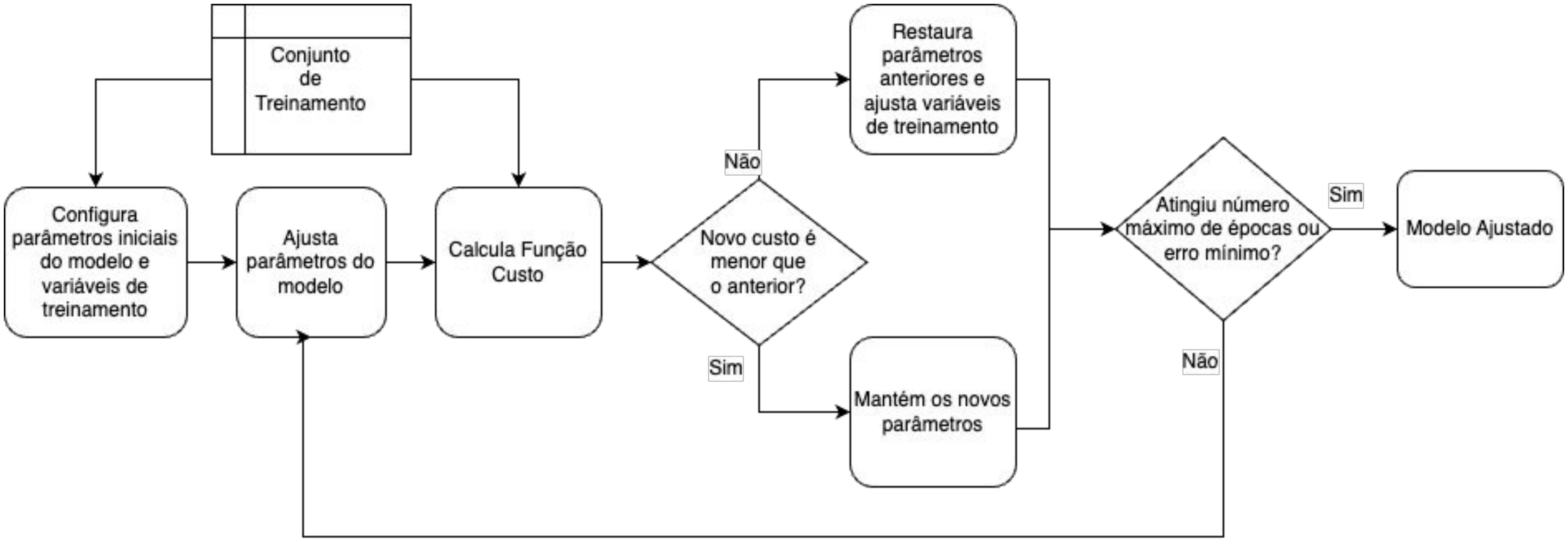
UEL

Profa. Dra. Juliana Felix

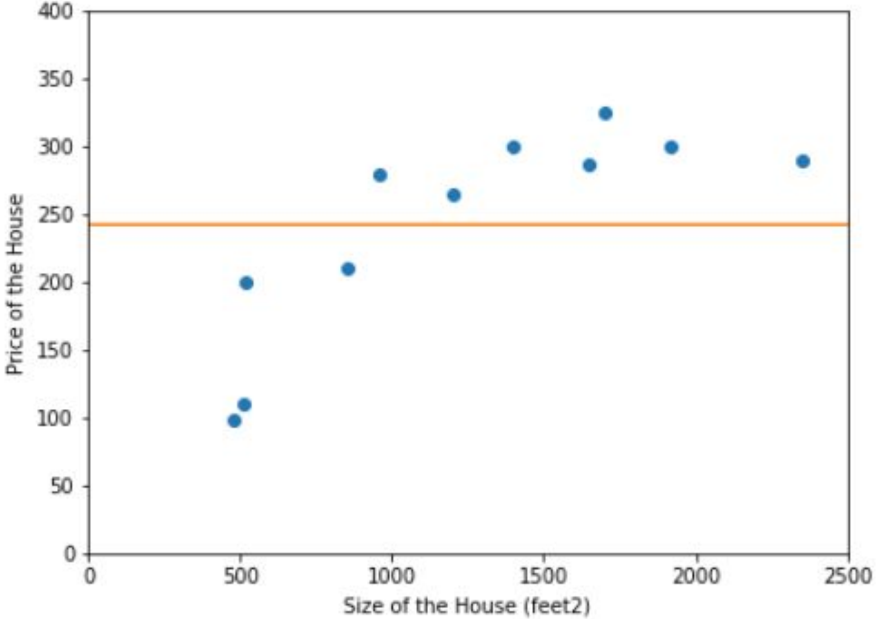
jufelix16@uel.br

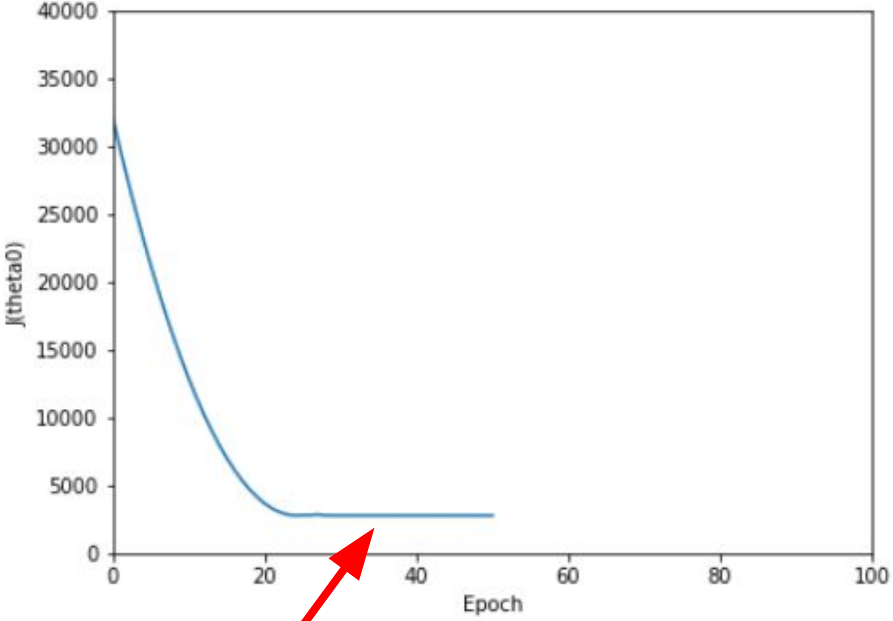
Critérios de Parada

Treinamento do Modelo



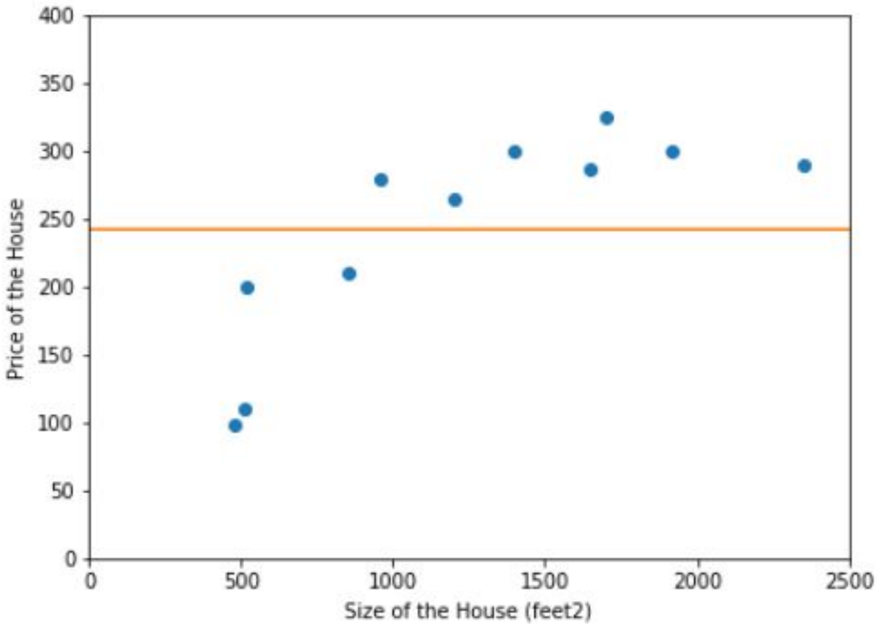
Número de épocas

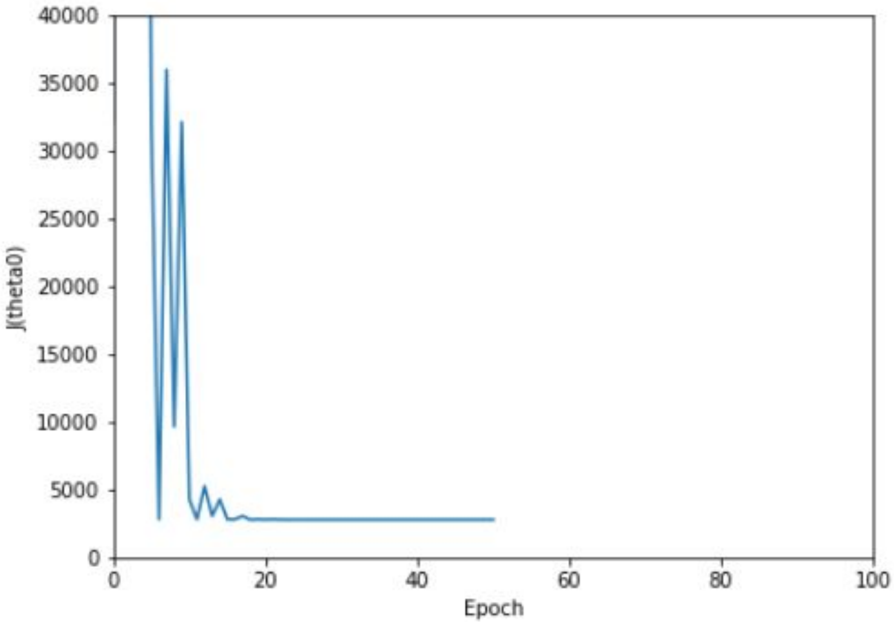




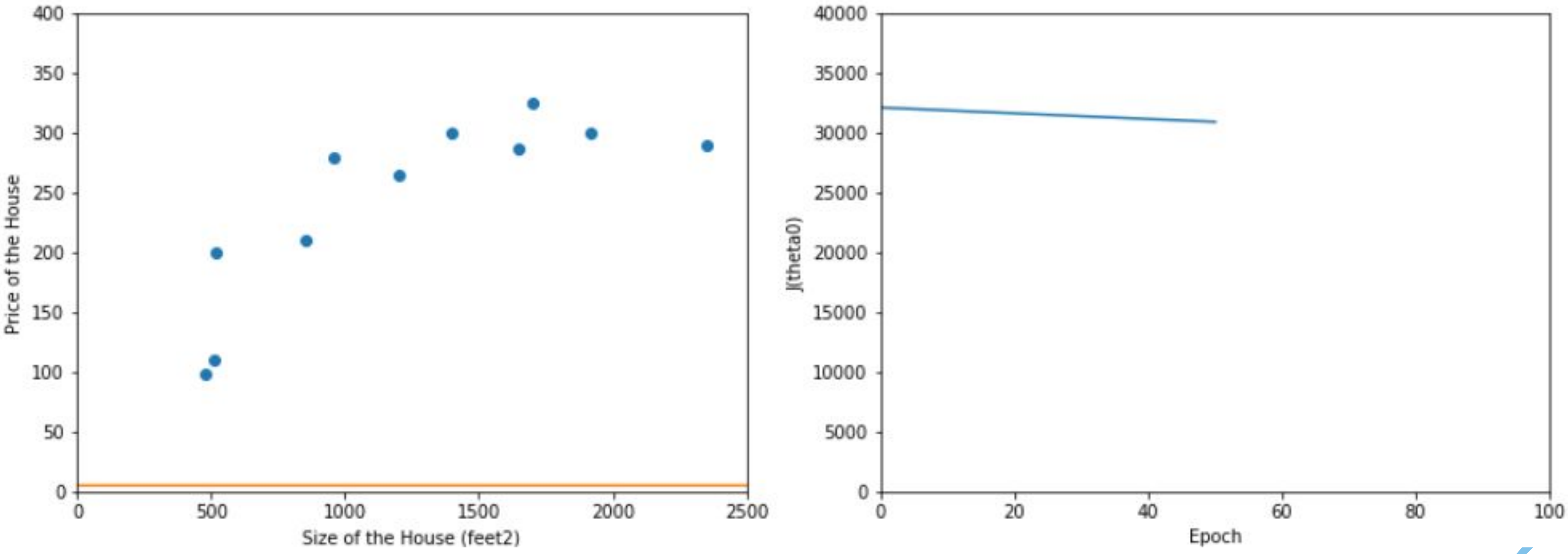
Custo não se altera depois de x épocas

Usando um valor muito alto para alpha

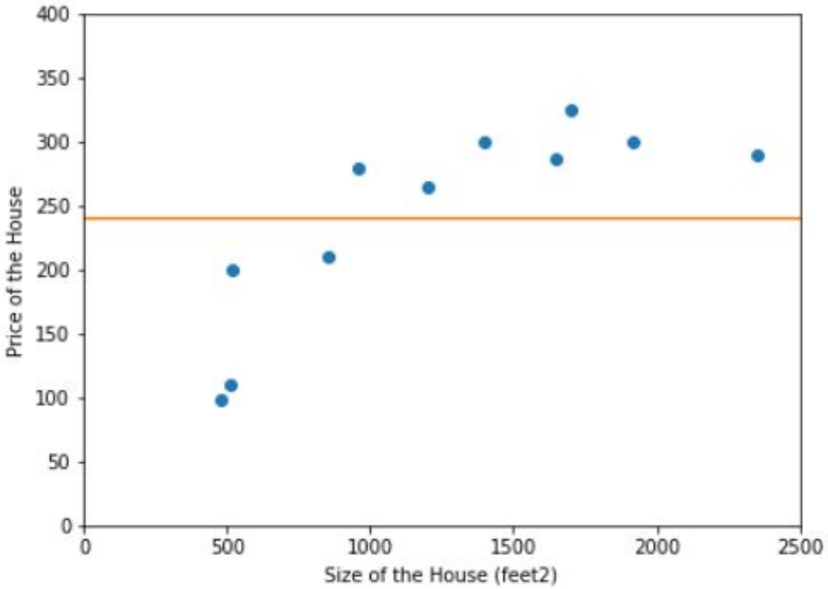


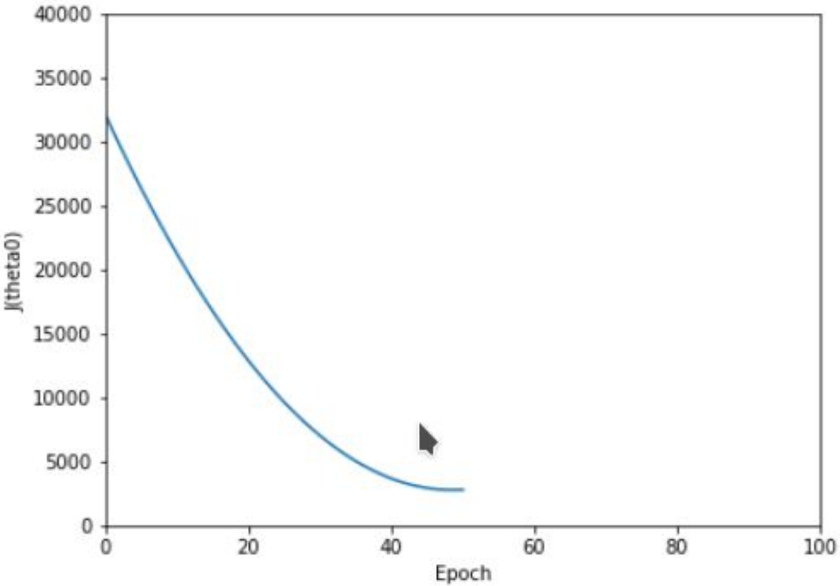


Usando um valor baixo para o alpha



Usando um "bom" alpha





Regressão Linear Múltipla

Múltiplas entradas (features)

Área (m^2)	Quartos	Andares	Idade	Valor (R$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...	...	...	...	...

Entrada/Saída dos dados

Neste caso, cada **amostra** possui dados de diversas variáveis:

- Entrada:
    - Área, x_1
    - Quartos, x_2
    - Andares, x_3
    - Idade, x_4
- Saída:
    - Valor, y

Entrada/Saída dos dados

Neste caso, cada **amostra** possui dados de diversas variáveis:

- Entrada:
  - Área, $\mathbf{x_1}$
  - Quartos, $\mathbf{x_2}$
  - Andares, $\mathbf{x_3}$
  - Idade, $\mathbf{x_4}$
- Saída:
  - Valor, $\mathbf{y}$

$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_3 + \theta_4 \cdot x_4$$

Modelo geral

$$h_\theta(x) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \cdots + \theta_n \cdot x_n$$

Função custo

O objetivo continua o mesmo:

$$J(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = \left[\frac{1}{2m}\sum_{i=1}^{m}\left(y^{(i)} - h\left(x^{(i)}\right)\right)^2\right]$$

$$\underset{\Theta}{min} J(\Theta) = \left[\frac{1}{2m}\sum_{i=1}^{m}\left(y^{(i)} - h\left(x^{(i)}\right)\right)^2\right]$$

Gradiente descendente

```
repeat{
    for (j=0;j<n;j++)
```

$$new\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\Theta)$$

```
    for (j=0;j<n;j++)
```

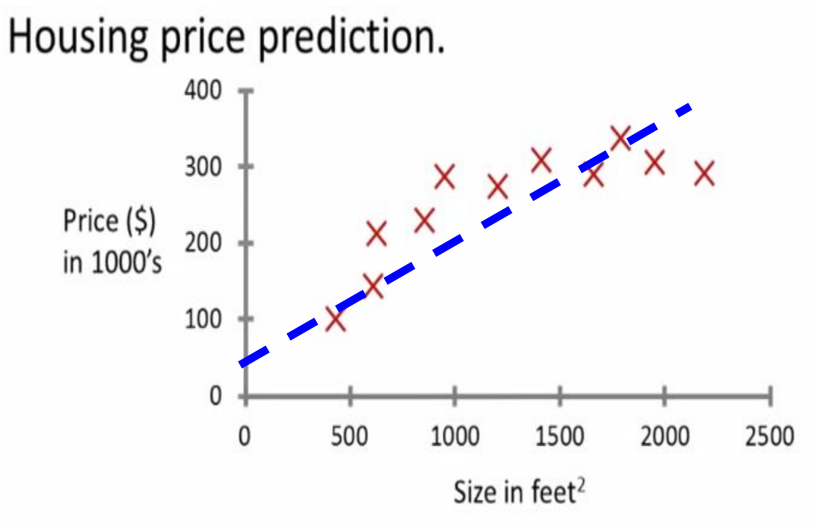
$$\theta_j = new\theta_j$$

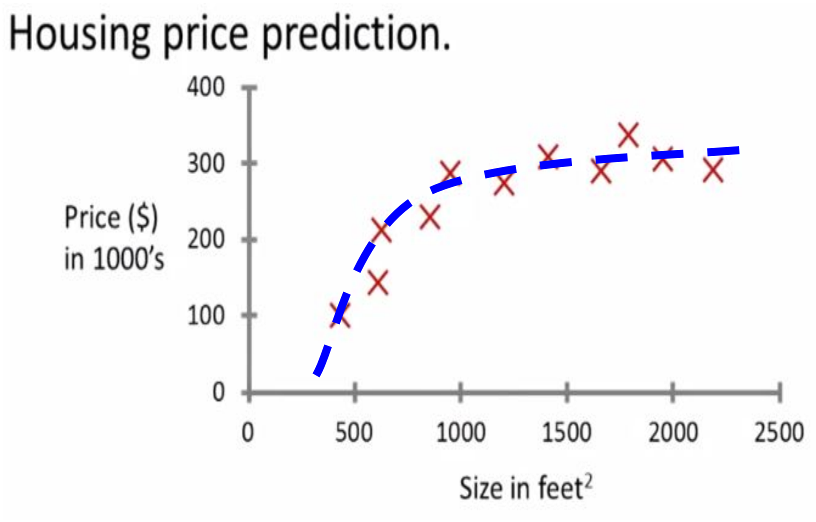
```
}until(stop_condition)
```

$$J(\Theta) = \left[\frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)} - h\left(x^{(i)} \right) \right)^2 \right]$$

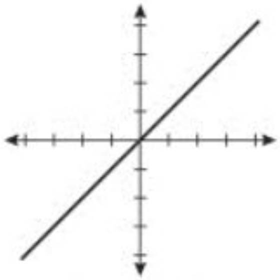
Regressão Linear?

Regressão Linear

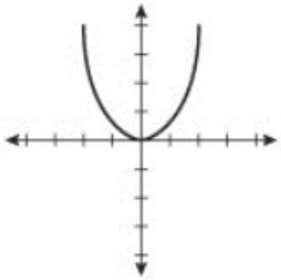




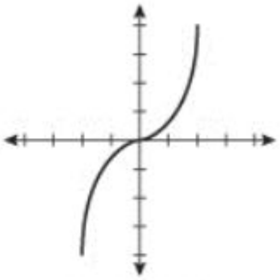
Regressão Linear



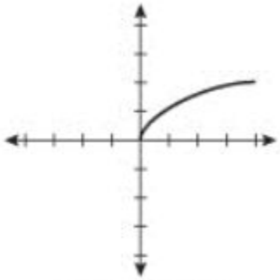
$y = f(x) = x$



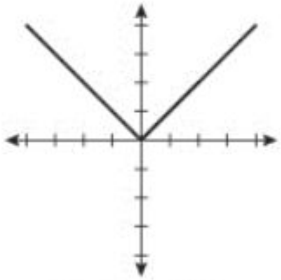
$y = f(x) = x^2$



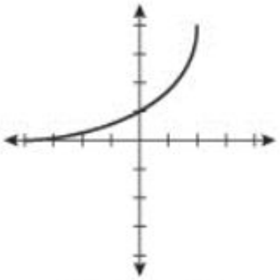
$y = f(x) = x^3$



$y = f(x) = \sqrt{x}$



$y = f(x) = |x|$



$y = f(x) = 2^x$

Outras funções que podem representar os problemas...

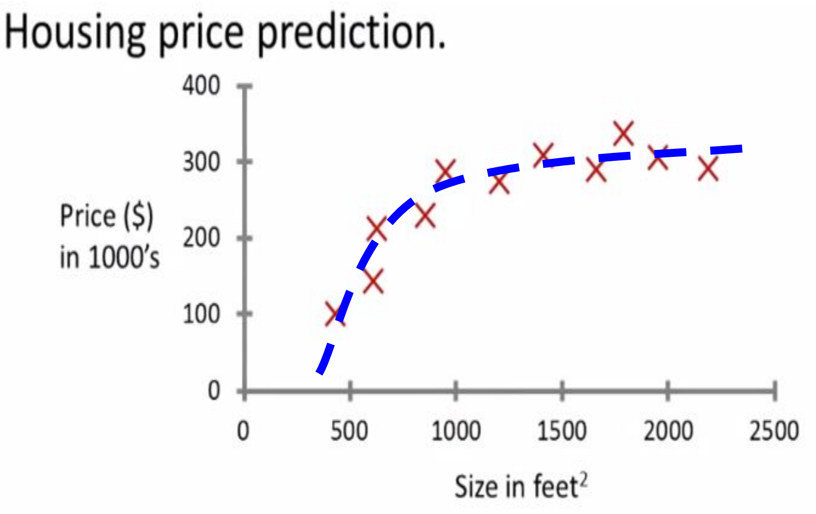
$$h_\theta = \theta_0 + \theta_1 \cdot x + \theta_2 x^2$$

$$h_\theta = \theta_0 + \theta_1 \cdot x + \theta_2 x^2 + \theta_3 x^3$$

$$h_\theta = \theta_0 + \theta_1 \cdot x + \theta_2 \sqrt{x}$$

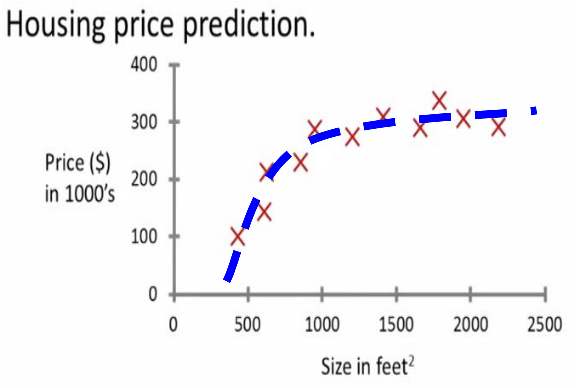
Regressão Linear

Neste exemplo, a função $h_\theta = \theta_0 + \theta_1 \cdot x + \theta_2\sqrt{x}$ parece representar os dados de forma melhor que uma reta.



Regressão Linear

$$h_\theta = \theta_0 + \theta_1 \cdot x + \theta_2 \sqrt{x}$$



$$\begin{aligned} x_1 &= x \\ x_2 &= \sqrt{x} \end{aligned}$$

$$h_\theta = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2$$

Regressão Linear

$$h_\theta = \theta_0 + \theta_1 \cdot x + \theta_2 \cdot x^2 + \theta_3 \cdot x^3$$

Temos uma **transformação**, e as features continuam sendo lineares

$$\begin{aligned} x_1 &= x \\ x_2 &= x^2 \\ x_3 &= x^3 \end{aligned}$$

$$h_\theta = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_3$$

Problemas de Classificação

Aprendizado supervisionado

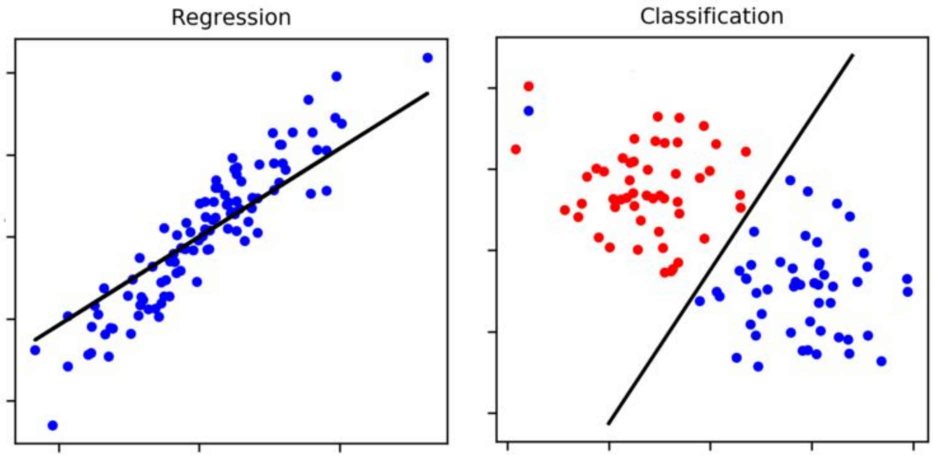
Técnica na qual o algoritmo de aprendizado recebe um conjunto de dados rotulados que define aquilo que deverá ser buscado pelo algoritmo

Classificação

Classificação é uma subcategoria do **aprendizado supervisionado**, cujo objetivo é **predizer se** uma instância (conjunto de features) **pertence a uma determinada classe**.

Classificação

- Diferente da regressão, não estamos interessados em construir um modelo que seja capaz de aproximar os valores de uma função.
- O interesse da classificação é o de decidir se algo pertence ou não a um grupo (classe)



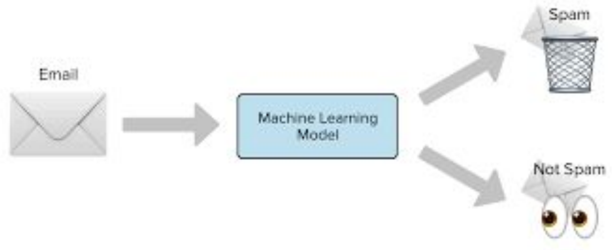
Classificação

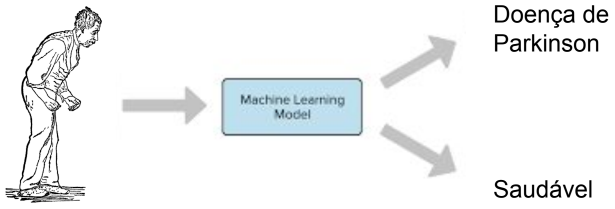
Existem dois tipos de problemas de classificação:

- Classificação **Binária**
- Classificação **Multiclasse**

Classificação Binária

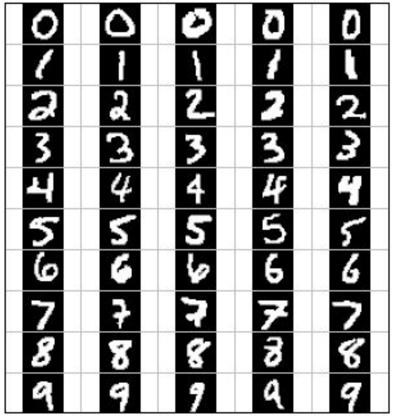
- Decidir se um email é um "spam" ou não
- Diagnosticar se um paciente tem uma determinada doença ou não
- Identificar se um animal é um gato ou cachorro
- Classe "positiva" vs. "negativa"
    - A saída esperada y é um valor dentro um conjunto de 2 elementos previamente definidos, por exemplo, $y \in \{0, 1\}$

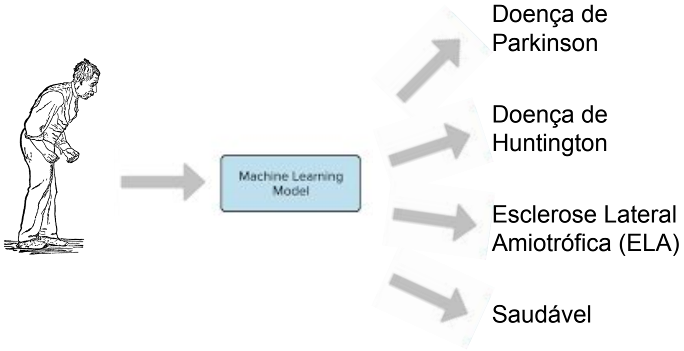




Classificação Multiclasse

- Reconhecimento de dígitos (0–9), 10 classes
- Diagnosticar se um paciente tem uma entre uma lista de doenças
- Identificar se um animal é um gato, cachorro, peixe, ou ave
- Categorizar resenhas de filme como "positivas", "negativas", ou "neutras"
- A saída esperada y é um valor dentro um conjunto de n elementos previamente definidos, por exemplo, $y \in \{1, 2, 3, \dots, n\}$



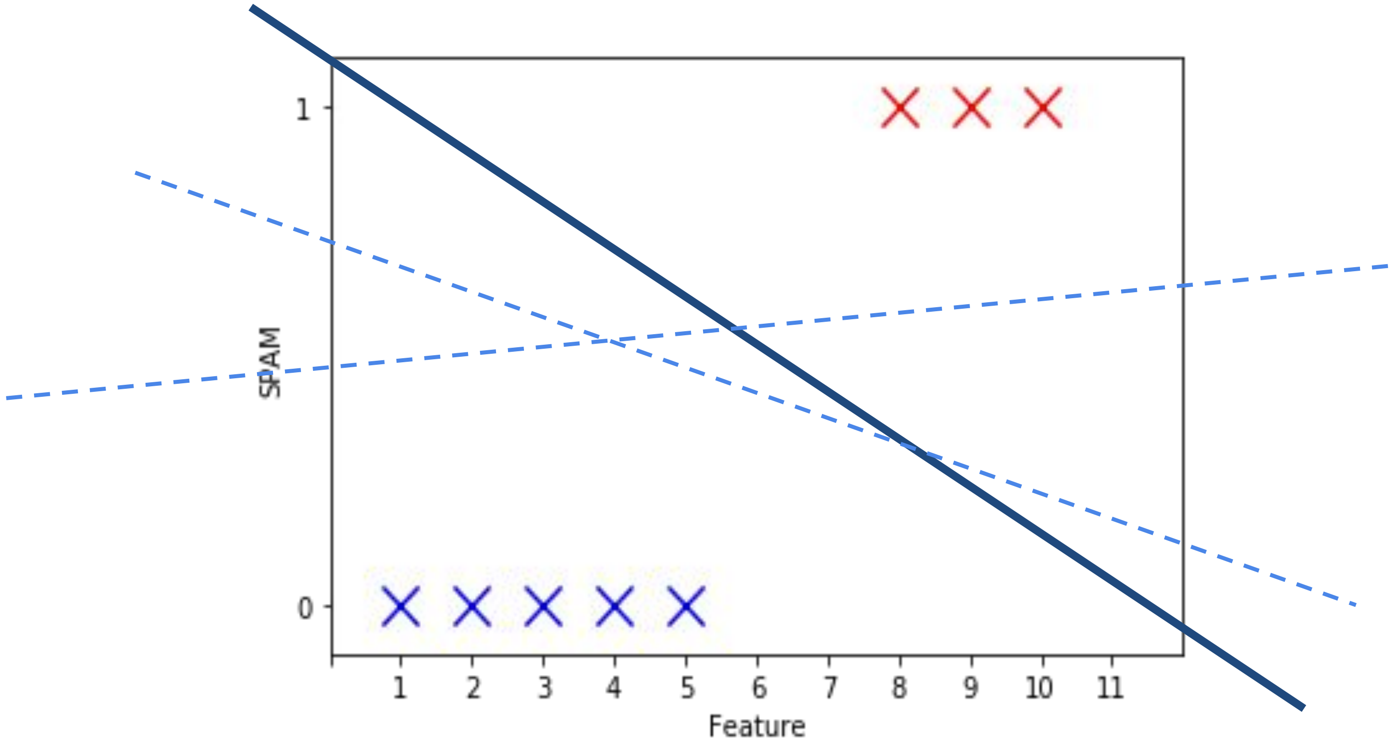


Classificação Multiclasse

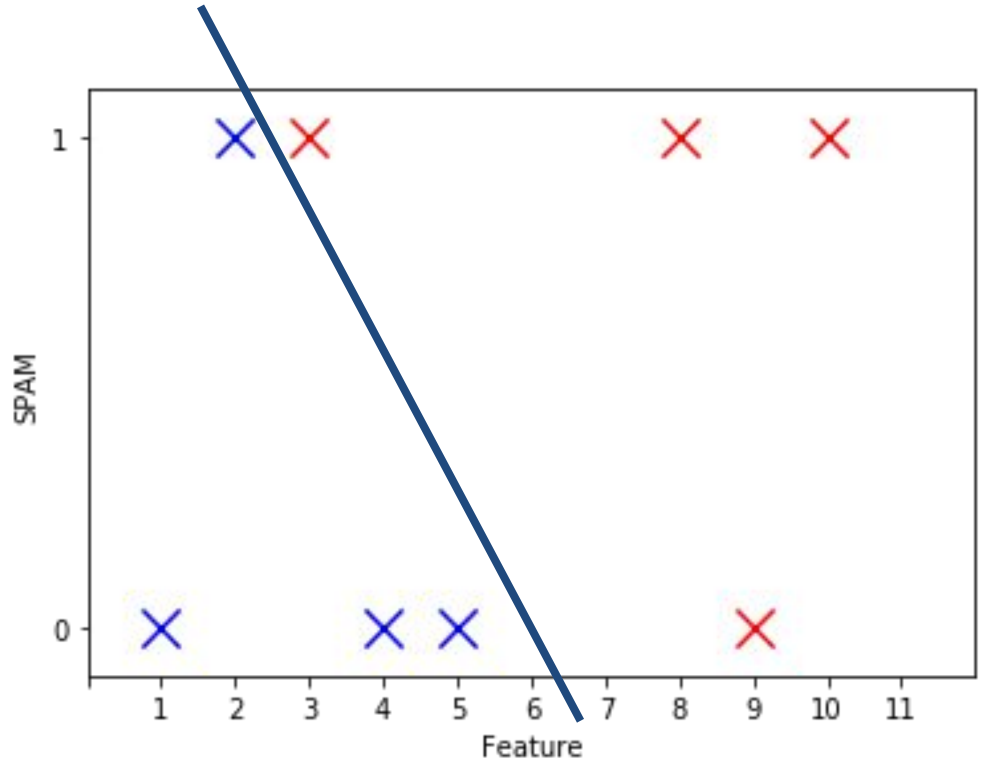
Na prática, ao invés de construir um classificador multiclasse, uma tática bastante utilizada consiste na construção de vários classificadores binários para realizar a mesma tarefa.

- A técnica é conhecida como one-vs-all (um-contra-todos)
- Idealmente, o resultado final consiste na escolha do classificador que retornou "sim" para a classe desejada, enquanto os demais classificadores retornaram "não".

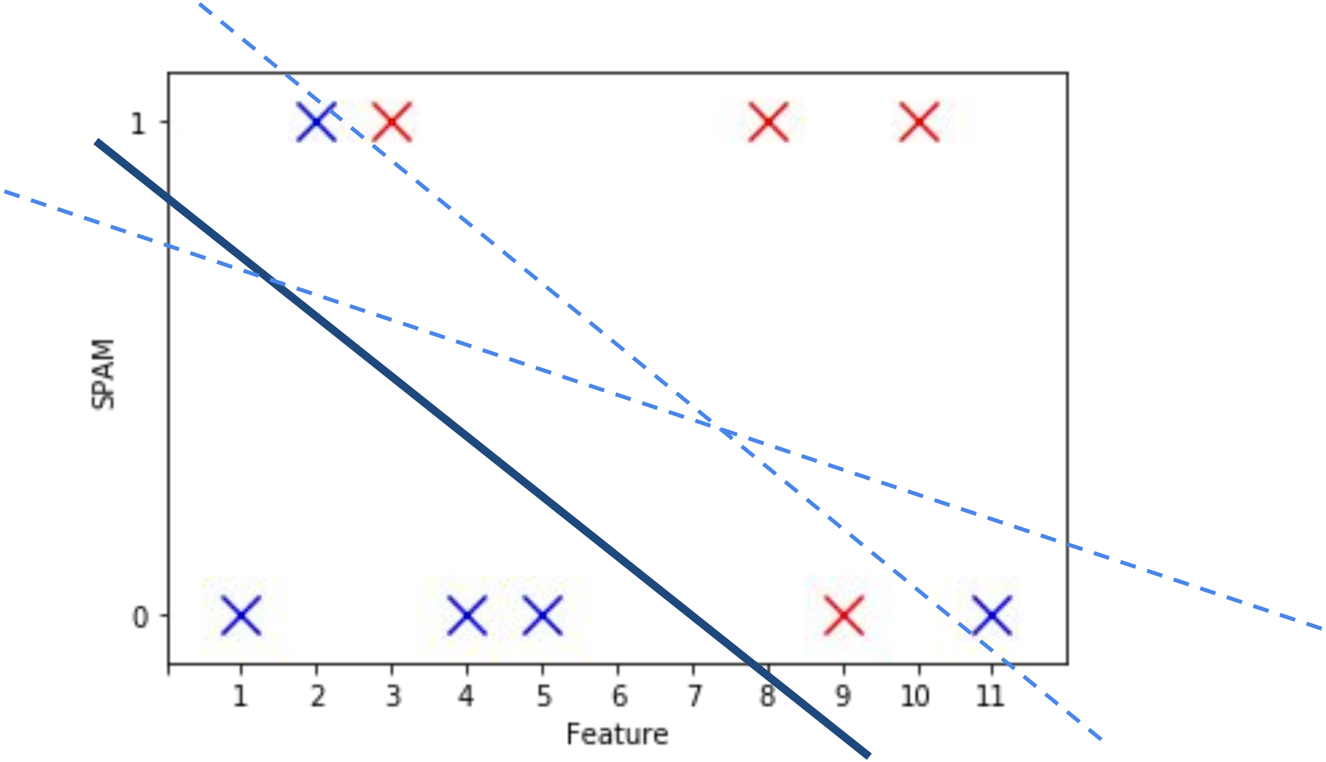
Classificação



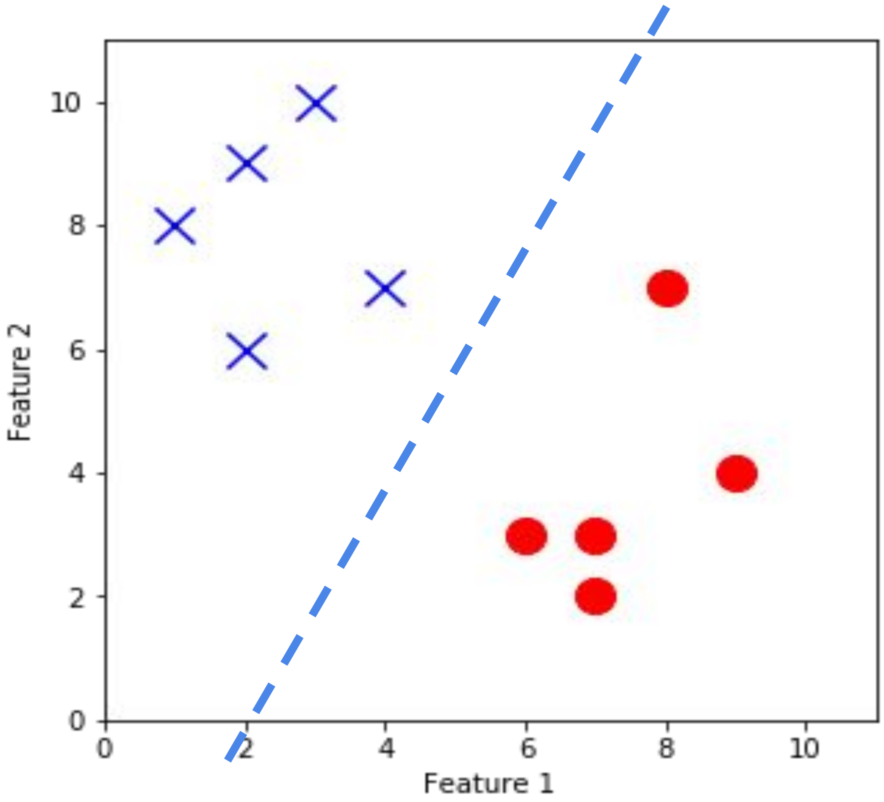
Classificação



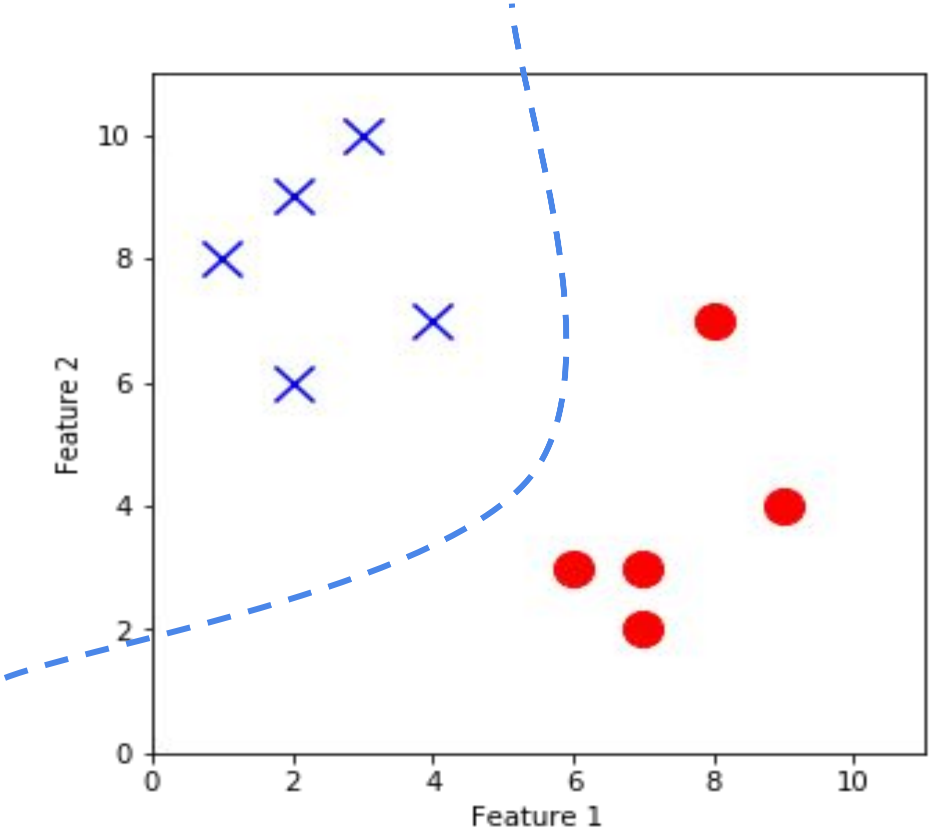
Classificação



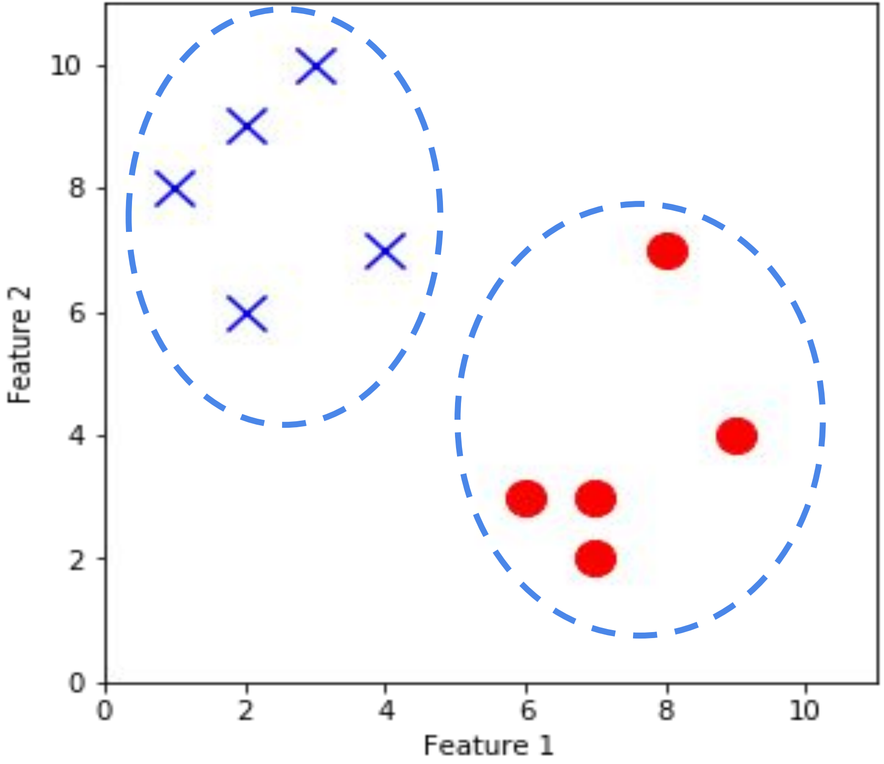
Classificação



Classificação



Classificação



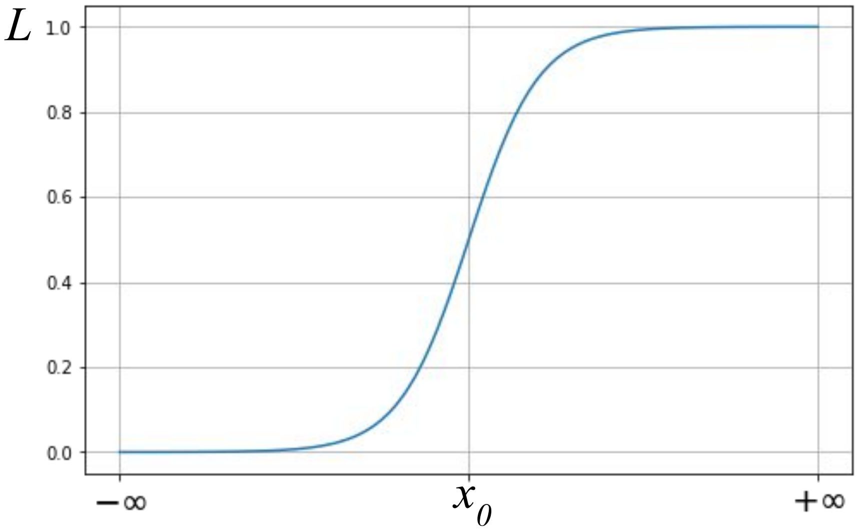
Regressão Logística

Regressão Logística

Função logística

$$f(x) = \frac{L}{1 + \mathrm{e}^{-k(x-x_0)}}$$

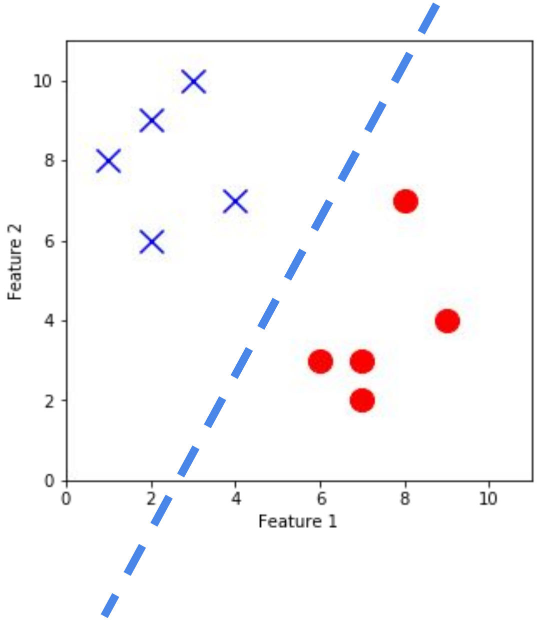
- e = base dos logaritmos naturais (número de Euler)
- x_0 = valor do ponto médio da curva sigmóide
- L = valor máximo da curva
- k = declividade da curva

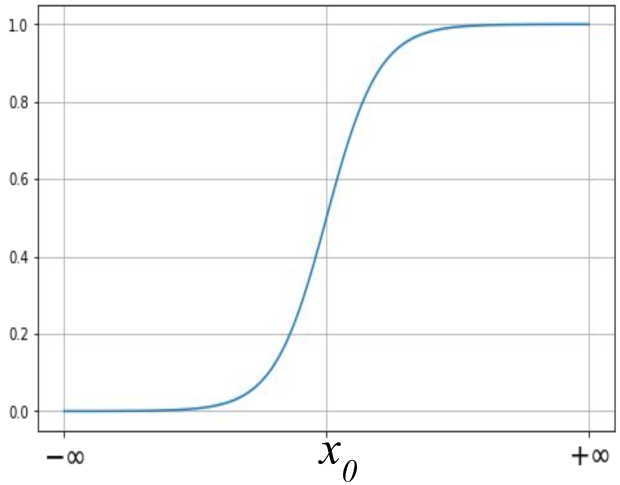


Função Hipótese *h(x)*

$$z = \theta_0 + \theta_1 x$$

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

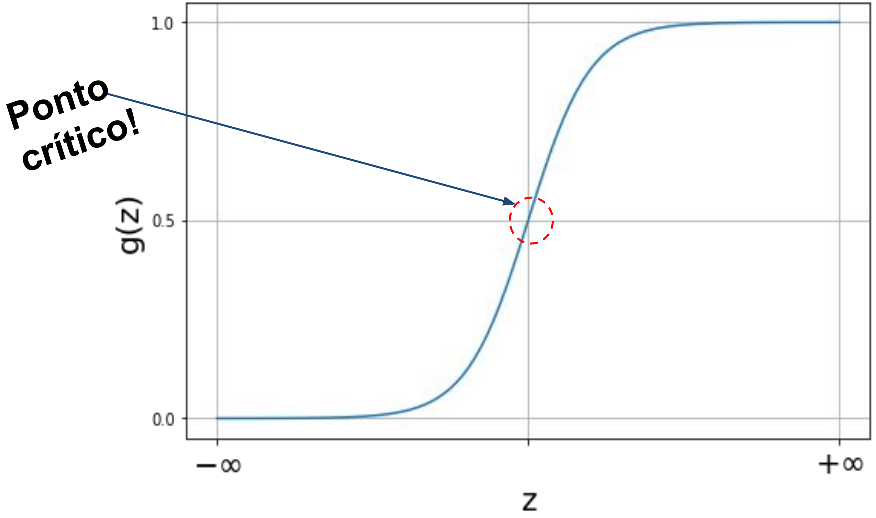




Interpretação de $h(x)$

Temos $h(x) = g(z) = \dfrac{1}{1 + e^{-z}}$ sendo que $z = \theta_0 + \theta_1 x$

- $y = 1$, se $g(z) >= 0.5$
- $y = 0$, se $g(z) < 0.5$
- $g(z) = 0.5$, se $z = 0$



Interpretação de $h(x)$

Em um problema multi característico, supondo $z = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

E assumindo que $\theta_0 = -2$, $\theta_1 = -1$, $\theta_2 = 1$

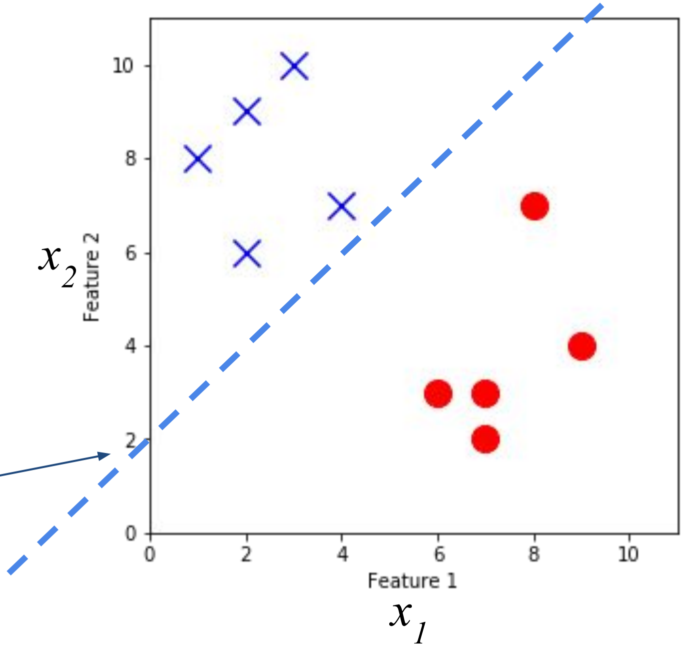
Teremos $z = -2 - x_1 + x_2$

Quando $z = 0$, teremos

$$-2 - x_1 + x_2 = 0$$
$$-x_1 + x_2 = 2$$

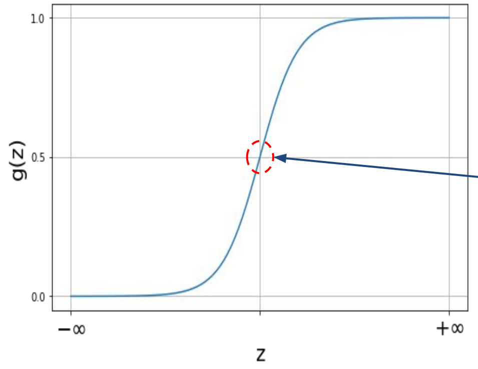
Quando $x_1 = 0$, temos que

$$x_2 = 2 + x_1$$
$$x_2 = 2$$

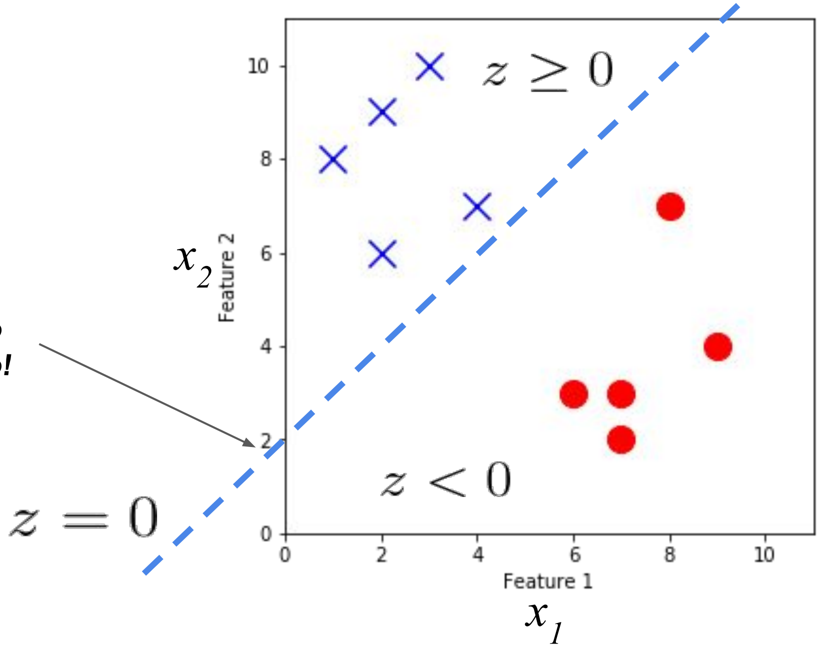


Interpretação de *h(x)*

Temos $h(x) = g(z) = \dfrac{1}{1+e^{-z}}$ sendo que $z = -2 - x_1 + x_2$



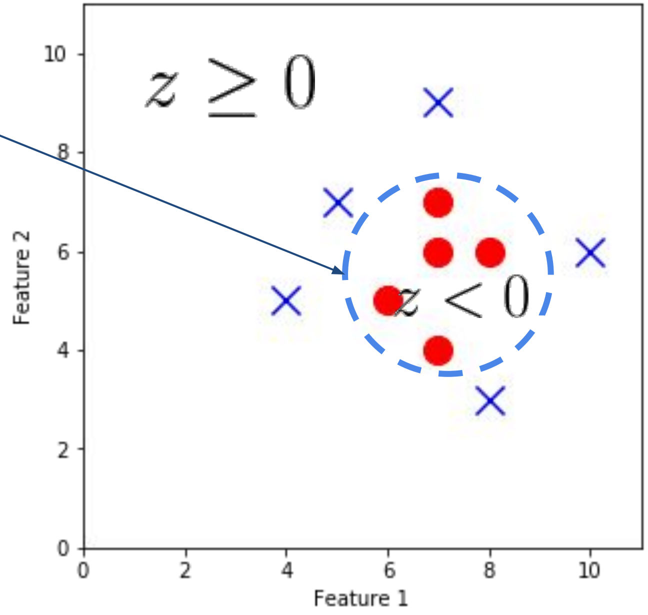
Ponto crítico!



Dados não-lineares

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2$$

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



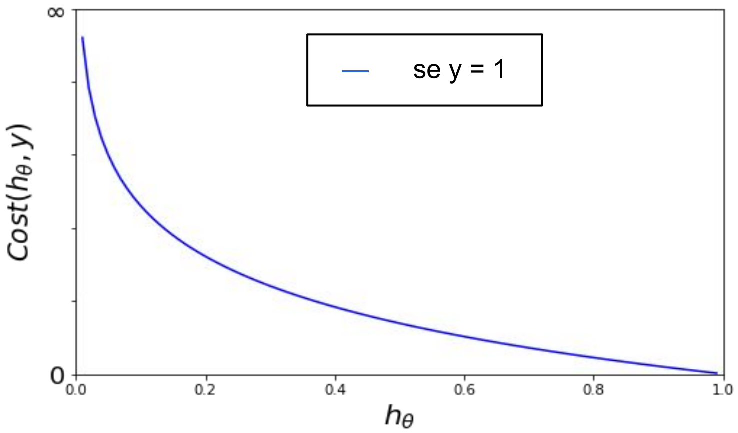
Assim como na regressão linear, a regressão logística permite que o modelo ajuste tanto dados lineares quanto não-lineares

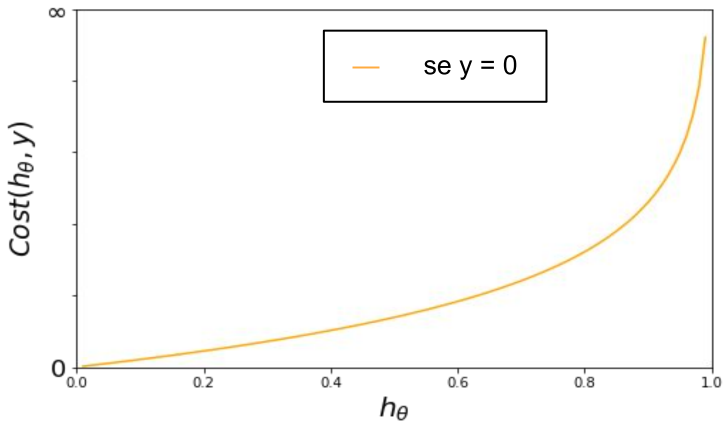
Função Custo

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

$$z = \theta_0 + \theta_1 x$$

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





Função Custo

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

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

$$Cost\left(h_\theta(x^{(i)}), y^{(i)}\right) = -y \cdot log(h_\theta(x)) - (1-y) \cdot log(1-h_\theta(x)))$$

$$\boxed{\min_{\Theta} J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \cdot log(h_\theta(x^{(i)}))) + (1-y^{(i)}) \cdot log(1-h_\theta(x^{(i)})))}$$

Gradiente Descendente

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

```
repeat{
    for (j=0;j<n;j++)
```

$$new\theta_j = \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\Theta)$$

```
    for (j=0;j<n;j++)
```

$$\theta_j = new\theta_j$$

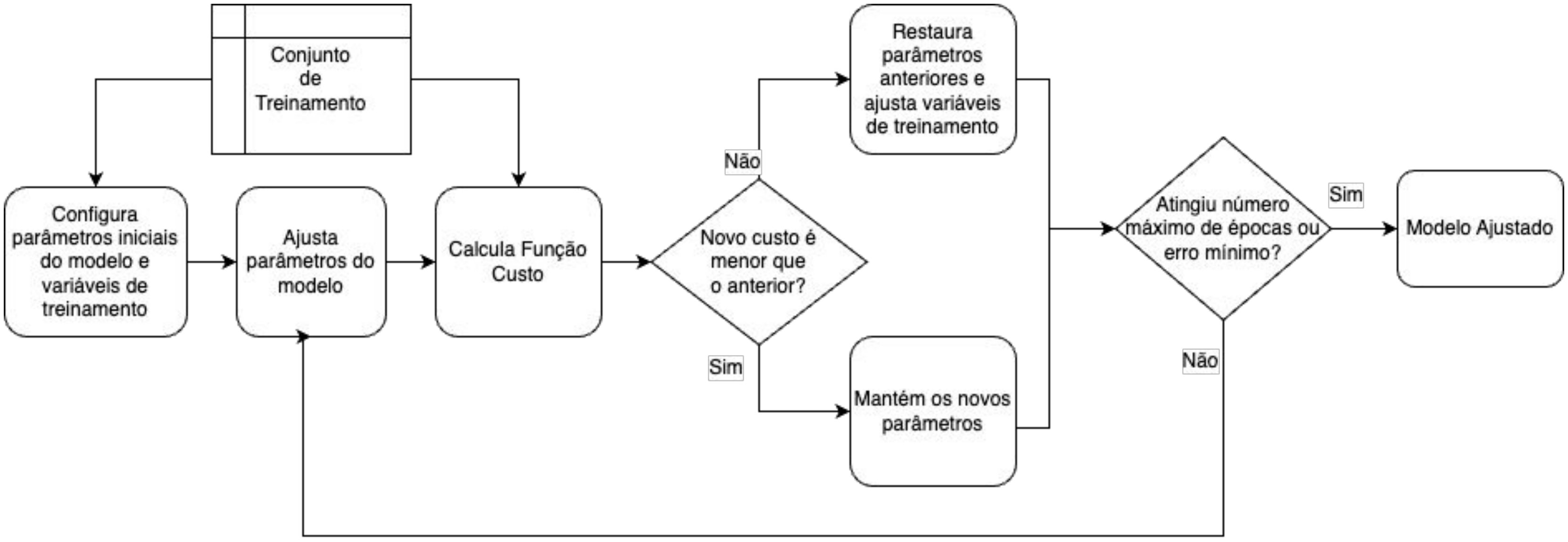
```
}
```

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

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

$$\boxed{z} = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

Treinamento do Modelo



Leitura recomendada: O gradiente (artigo) | Khan Academy

Lista 3

- Regressão Linear Simples, Múltipla e Logística
- Data de entrega: segunda, 22/05, até 23h59