AMOSTRA DE EXEMPLO: $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in M_{n \times 1}(\mathbb{R})$

MODELO: família de funções parametrizadas

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

FUNÇÃO DE ERRO: serve para julgar a qualidade de $h_\theta(x)$ como preditor

exemplo:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$\hat{y}^{(i)}$: y predito da i-ésima amostra de treino

$y^{(i)}$: y real da i-ésima amostra de treino

UM PROBLEMA DE MACHINE LEARNING REQUER:

i) os dados
ii) o modelo
iii) a função de erro

MODELO LINEAR:

$$\hat{y} = h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

$$X_{train} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \quad y_{train} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad \Leftrightarrow$$

CONVERTENDO O MODELO P/ FORMA MATRICIAL

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

$$\Updownarrow$$

$$\hat{y} = \theta_0 \cdot 1 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

ADOTAR:

$$x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$$

NESTA SITUAÇÃO:

$$\hat{y}^{(i)} = \left(x^{(i)}\right)^T \cdot \theta$$

POIS:

$$\left(x^{(i)}\right)^T \cdot \theta = \underbrace{\begin{bmatrix} 1 & x_1^{(i)} & x_2^{(i)} & \dots & x_n^{(i)} \end{bmatrix}}_{\left(x^{(i)}\right)^T} \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \Rightarrow$$

$$\Rightarrow \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \dots + \theta_n x_n^{(i)} = \hat{y}^{(i)}$$

COM ESTA EQUAÇÃO, PODEMOS:

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix} \cdot \theta \Rightarrow y = X\theta \quad \blacksquare$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & & \vdots \\ 1 & x_1^{(m)} & \dots & x_n^{(m)} \end{bmatrix}$$

MSE NA FORMA MATRICIAL:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 \Leftrightarrow \frac{1}{m}(\hat{y} - y)^T \cdot (\hat{y} - y)$$

PROVANDO:

$$\frac{1}{m}(\hat{y} - y)^T \cdot (\hat{y} - y) = \frac{1}{m} \begin{bmatrix} (\hat{y}^{(1)} - y^{(1)}) \\ (\hat{y}^{(2)} - y^{(2)}) \\ \vdots \\ (\hat{y}^{(m)} - y^{(m)}) \end{bmatrix}^T \begin{bmatrix} (\hat{y}^{(1)} - y^{(1)}) \\ (\hat{y}^{(2)} - y^{(2)}) \\ \vdots \\ (\hat{y}^{(m)} - y^{(m)}) \end{bmatrix} \Rightarrow$$

$$\Rightarrow \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 \quad \blacksquare$$

EXPANDINDO:

$$MSE = \frac{1}{m}(\hat{y} - y)^T \cdot (\hat{y} - y)$$

$$= \frac{1}{m}(X\theta - y)^T \cdot (X\theta - y)$$

COMO: $(A+B)^T = A^T + B^T$, $(AB)^T = B^T A^T$

$$= \frac{1}{m}(\theta^T X^T - y^T) \cdot (X\theta - y)$$

$$= \frac{1}{m}\left(\theta^T X^T X\theta - \theta^T X^T y - y^T X\theta + y^T y\right)$$

NOTE QUE $\theta^T X^T y \in M_{(1)}(\mathbb{R})$ E $y^T X\theta$ TAMBÉM

$$\Rightarrow \theta^T X^T y = y^T X\theta$$

LOGO:

$$MSE = \frac{1}{m}\left(\theta^T X^T X\theta - 2\theta^T X^T y + y^T y\right)$$

PARA ACHAR θ ÓTIMO:

$$\frac{\partial MSE}{\partial \theta_0} = 0, \quad \frac{\partial MSE}{\partial \theta_1} = 0, \quad \dots, \quad \frac{\partial MSE}{\partial \theta_n} = 0 \Rightarrow \nabla_\theta MSE = 0$$

ASSIM: $MSE = \frac{1}{m}\left(\theta^T X^T X\theta - 2\theta^T X^T y + y^T y\right) \Rightarrow \nabla_\theta MSE = \frac{1}{m}\left(2X^T X\theta - 2X^T y\right) = 0 \Rightarrow 2X^T X\theta = 2X^T y$

$$\Rightarrow \theta = (X^T X)^{-1} X^T y$$

CUSTO COMPUTACIONAL DA EQUAÇÃO NORMAL:

$$\theta = (X^T X)^{-1} X^T y$$

i) CUSTO DE $X^T X$:

$$\left.\begin{matrix} \underbrace{\begin{bmatrix} X^T \end{bmatrix}}_{(n+1)\times m} \underbrace{\begin{bmatrix} X \end{bmatrix}}_{m\times(n+1)} = \begin{bmatrix} O(m) & \cdots & O(m) \\ \vdots & & \\ O(m) & \cdots & O(m) \end{bmatrix} & \begin{matrix} \rightarrow \text{Transposta} \Rightarrow O(n^2) \\ \\ \rightarrow \text{Operações} \Rightarrow O(m) \end{matrix} \end{matrix}\right\} O(n^2 m)$$

ii) CUSTO DE $(\cdot)^{-1} \Rightarrow O(n^3)$

iii) CUSTO DE $X^T y$

$$\left.\underbrace{\begin{bmatrix} X^T \end{bmatrix}}_{(n+1)\times m} \cdot \underbrace{\begin{bmatrix} y \end{bmatrix}}_{m\times 1} = \begin{bmatrix} O(m) \\ \vdots \\ O(m) \end{bmatrix}\right\} O(nm)$$

iv) CUSTO DO PRODUTO FINAL $\underbrace{(X^T X)^{-1}}_{(n+1)\times(n+1)} \underbrace{X^T y}_{(n+1)\times 1} \Rightarrow O(n^2)$

JUNTANDO TUDO: $O(n^2 m + n^3)$