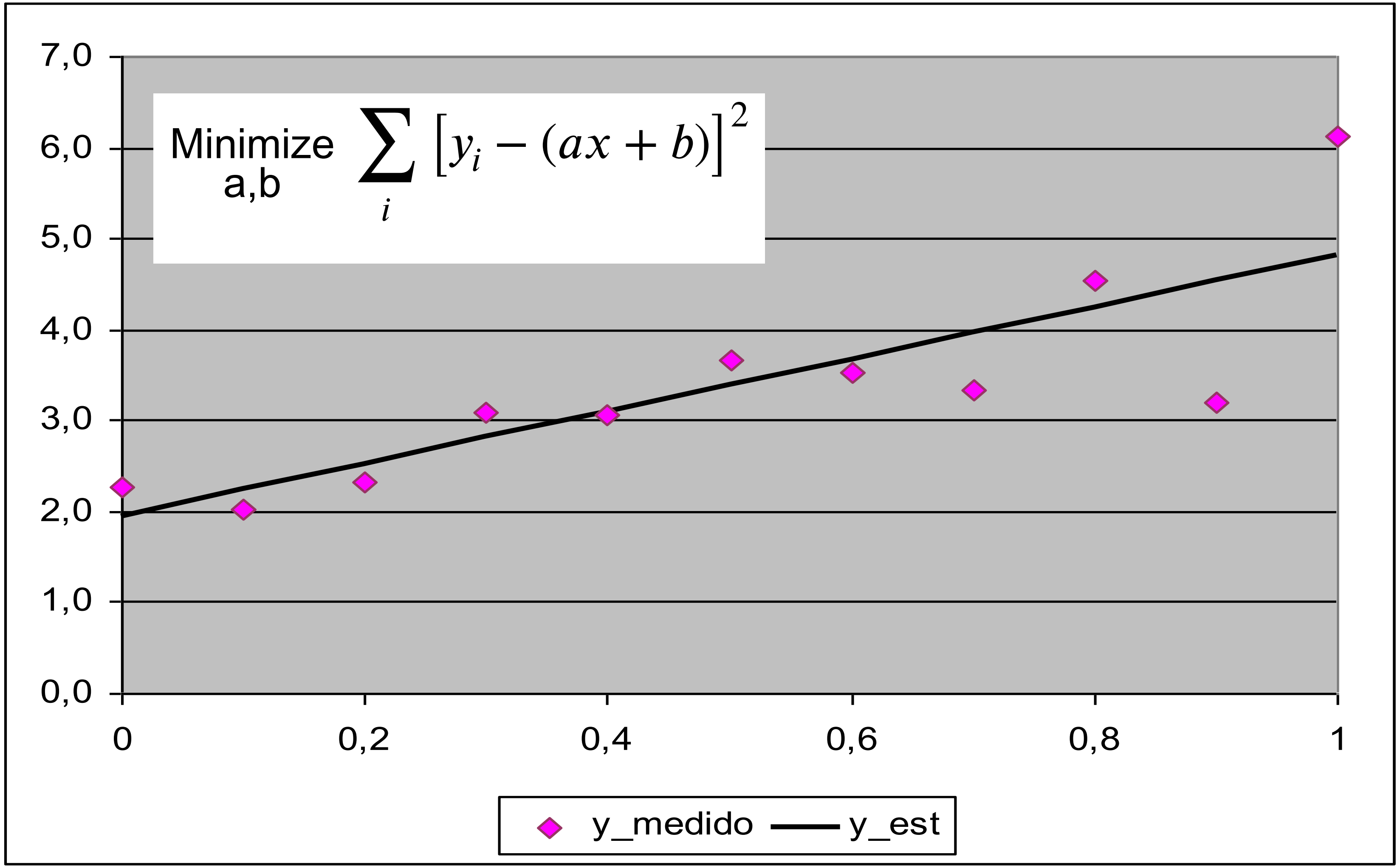
Gradiente Descendente

por

Marcelo Gattass

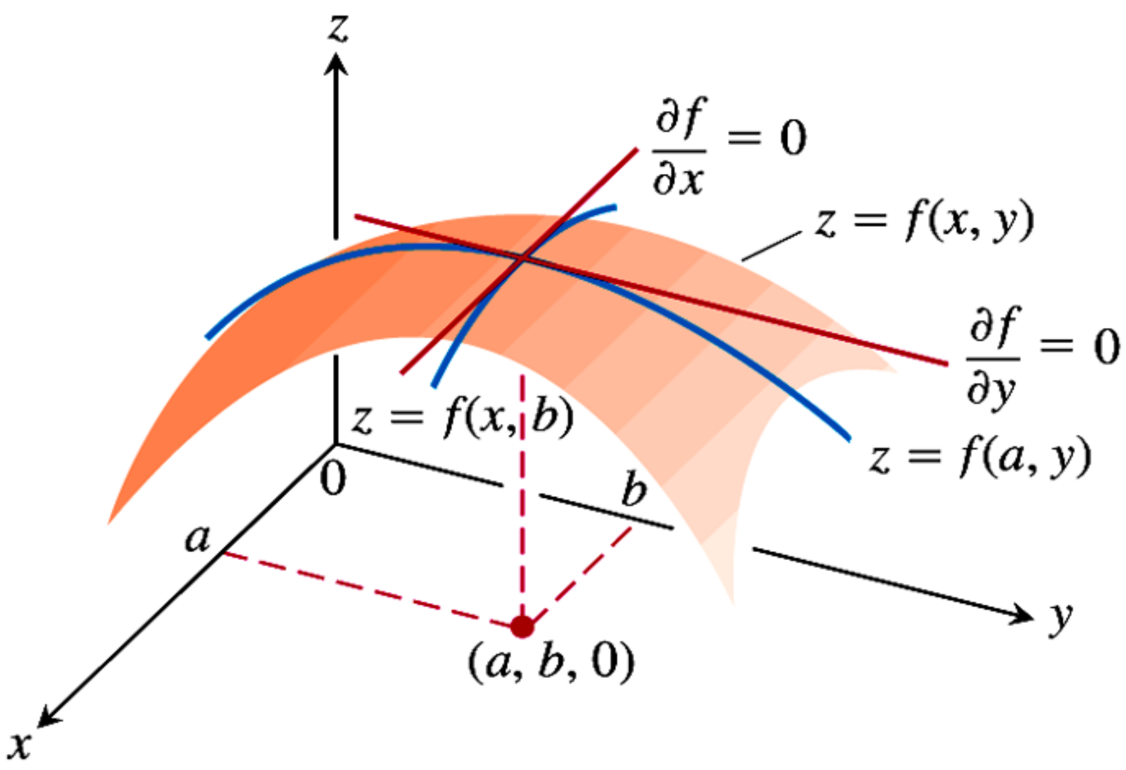
2022.1

Considere o problema



ajuste de curvas (Modelo)

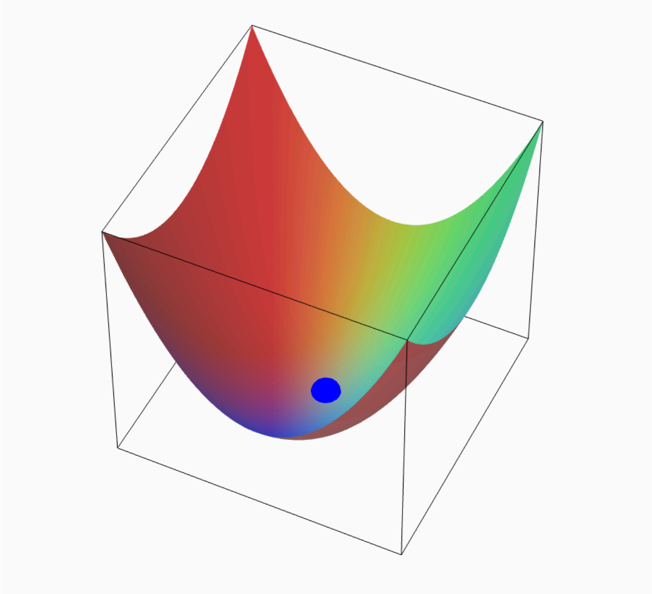
Máximos e mínimos



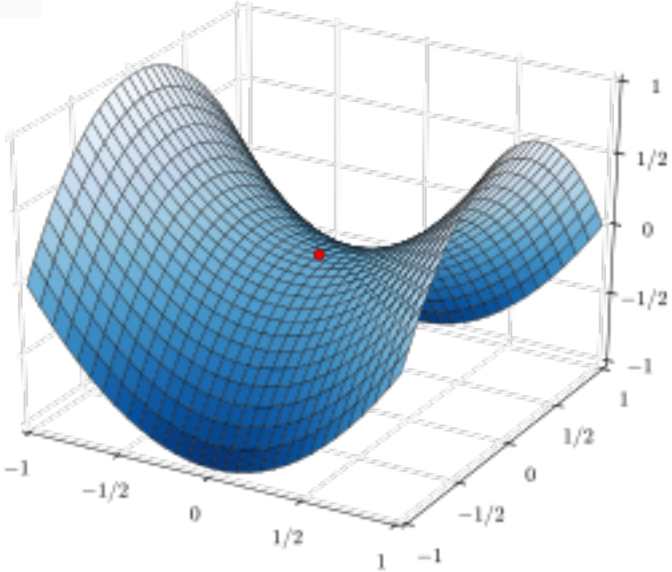
Num ponto extremo (máximo ou mínimo) a o gradiente da função é nulo.

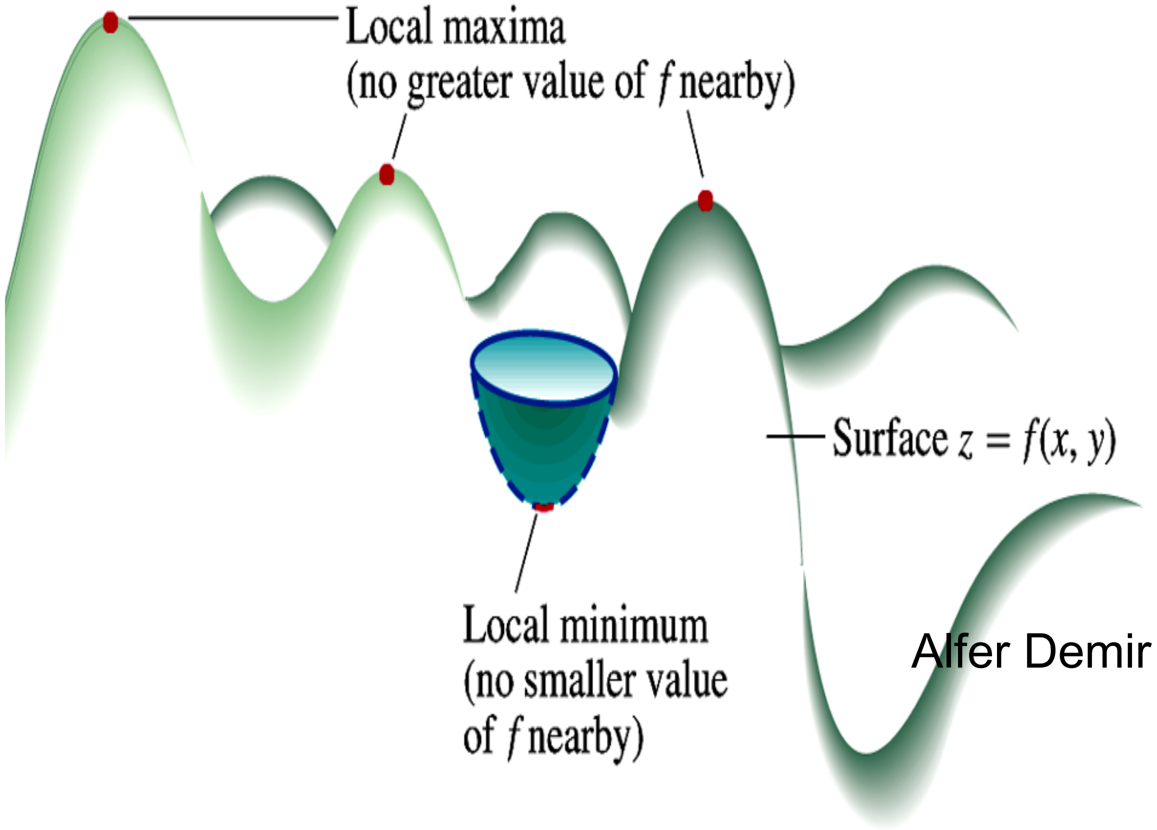
$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

Condição necessária, mas não suficiente



$$\nabla f = (0{,}0)$$





Alfer Demir

Ajuste por mínimos quadrados

$$\underset{a,b}{\text{Minimize}} \sum_i \left[y_i - (ax+b)\right]^2$$

$$\frac{\partial\, erro}{\partial a} = \frac{\partial}{\partial a}\left(\sum_{i=1}^{n}(y_i - b - ax_i)^2\right) = 0 \qquad 2\sum_{i=1}^{n}(y_i - b - ax_i)(-x_i) = 0$$

$$\frac{\partial\, erro}{\partial b} = \frac{\partial}{\partial b}\left(\sum_{i=1}^{n}(y_i - b - ax_i)^2\right) = 0 \qquad 2\sum_{i=1}^{n}(y_i - b - ax_i)(-1) = 0$$

$$-\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = 0$$

$$-\sum_{i=1}^{n} y_i + bn + a\sum_{i=1}^{n} x_i = 0$$

$$\begin{bmatrix} \sum_i x_i^2 & \sum_i x_i \\ \sum_i x_i & n \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_i x_i y_i \\ \sum_i y_i \end{bmatrix}$$

$$a = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - \sum_i x_i \sum_i x_i}$$

$$b = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \sum_i x_i y_i}{n\sum_i x_i^2 - \sum_i x_i \sum_i x_i}$$

Ajuste pelo Gradiente Descendente

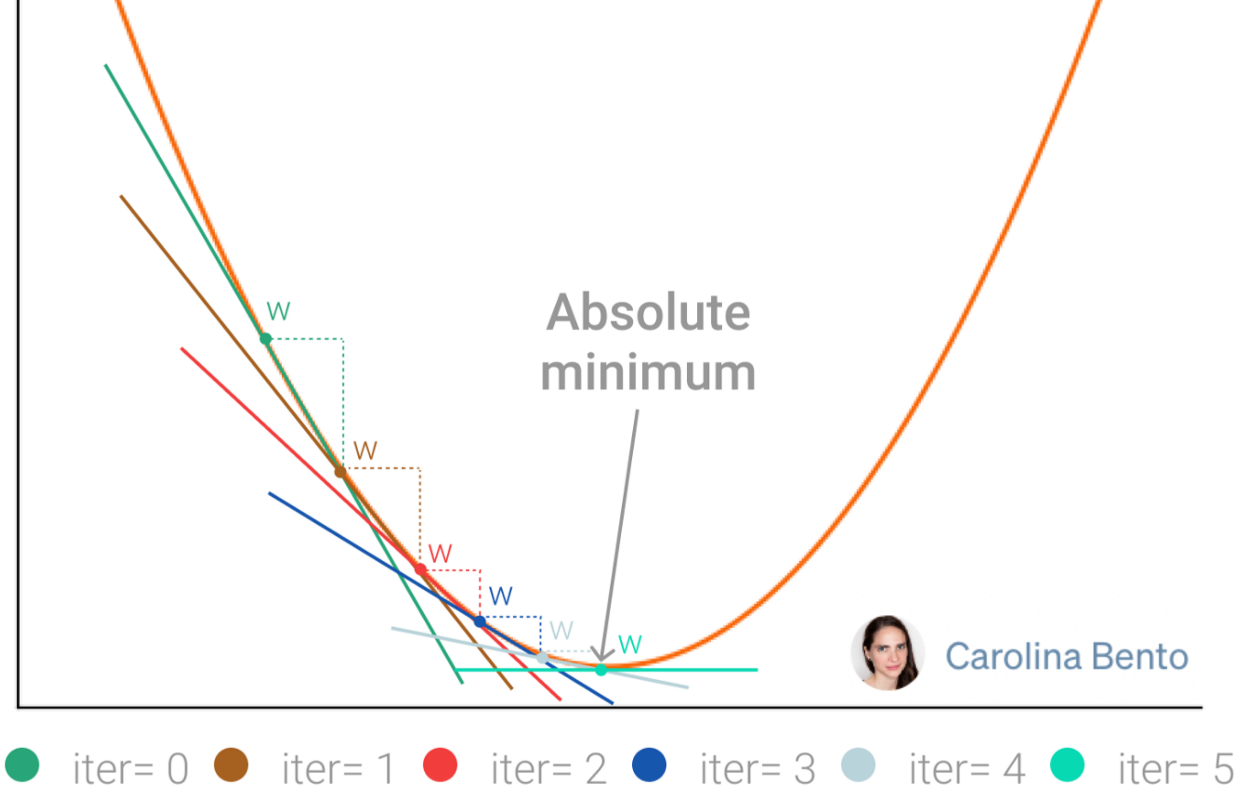
atribua um valor inicial para a e b
itere até convergir
    a = a – taxa_de_aprendizado*da
    b = b – taxa_de_aprendizado*db

$$\frac{\partial\, erro}{\partial a} = 2\sum_{i=1}^{n}(y_i - b - ax_i)(-x_i)$$

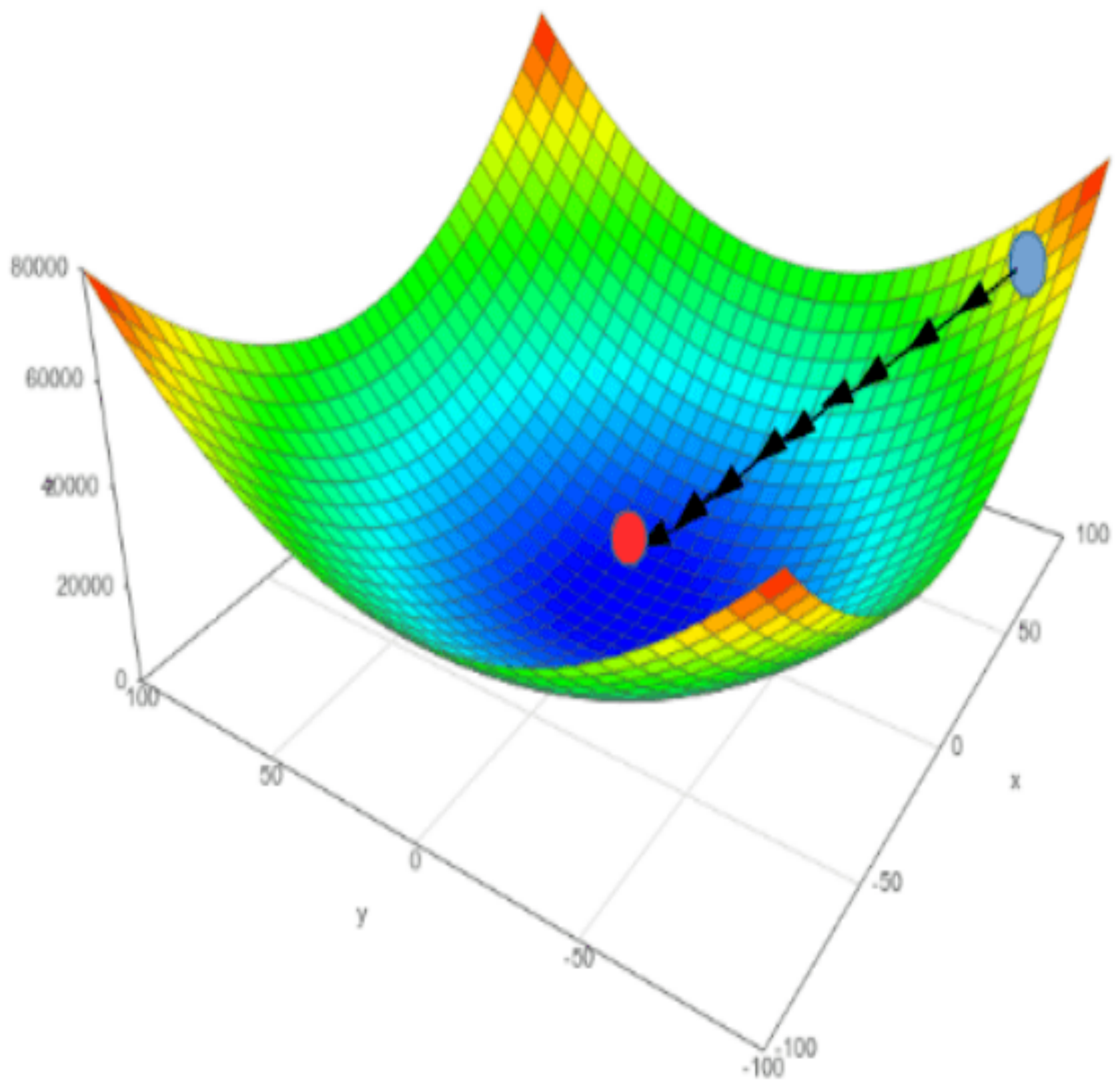
$$\frac{\partial\, erro}{\partial b} = 2\sum_{i=1}^{n}(y_i - b - ax_i)(-1)$$

taxa_de_aprendizado

$$erro = \sum_{i=1}^{n}(y_i - b - ax_i)^2$$

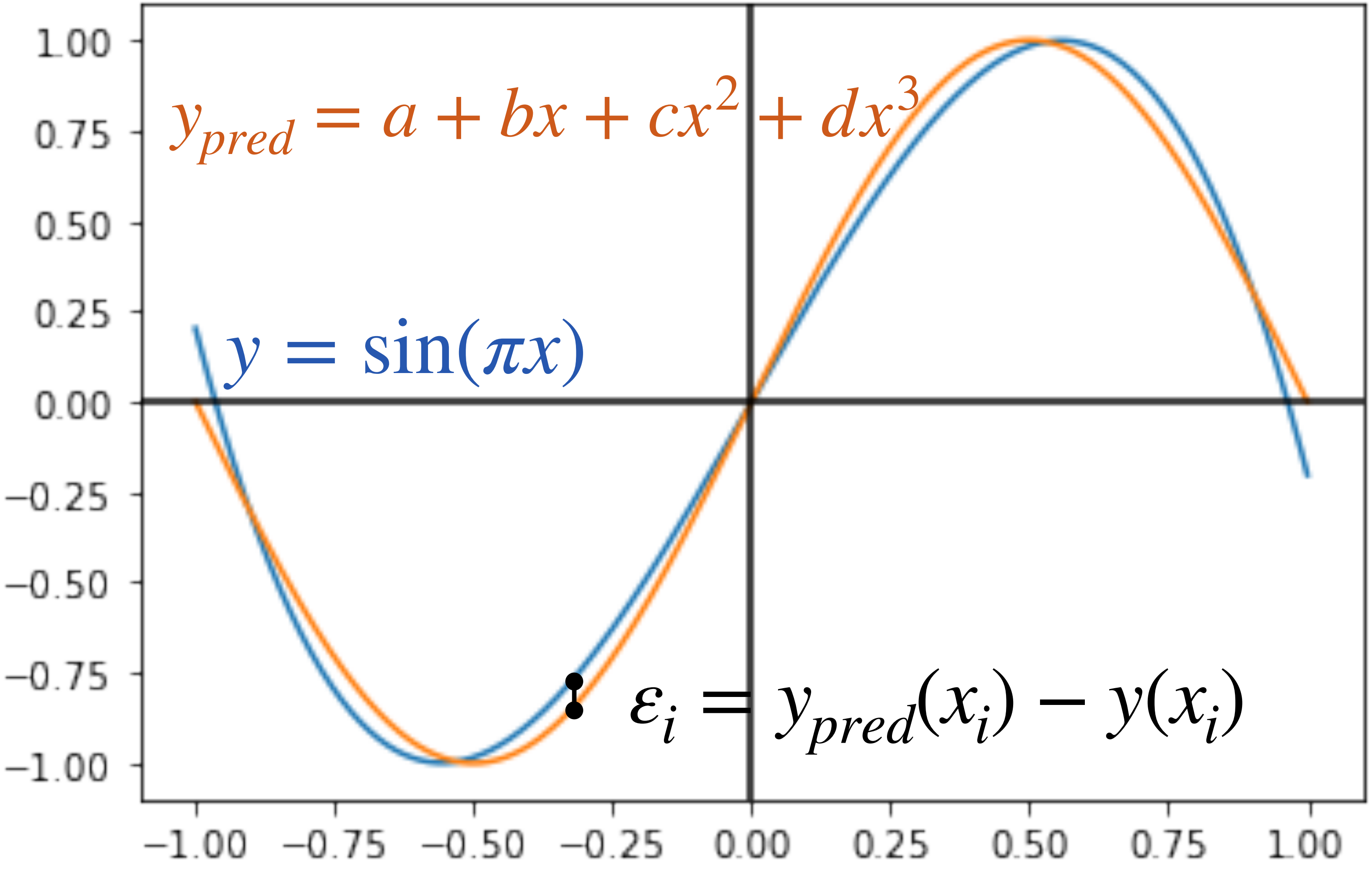


ref


80000
60000
40000
20000
0
100
50
0
-50
-100
y
100
50
0
-50
-100
x

Problema 1: Ajuste um polinômio cúbico

A uma função seno



$$\text{loss} = \bar{\varepsilon} = \sum_{i=0}^{n} \varepsilon_i^2$$

$$\bar{\varepsilon} = \sum_{i=0}^{n} \left(y_{pred}(x_i) - y(x_i)\right)^2$$

$$x \in [-1, +1]$$

Gradiente em função dos pontos

$$\nabla \bar{\varepsilon} = \nabla \left(\sum_{i=0}^{n} \varepsilon_i^2 \right)$$

$$\varepsilon_i = y_{pred}(x_i) - y(x_i) = a + bx_i + cx_i^2 + dx_i^3 - y_i$$

$$\nabla \bar{\varepsilon} = \begin{pmatrix} \frac{\partial \bar{\varepsilon}}{\partial a} \\ \frac{\partial \bar{\varepsilon}}{\partial b} \\ \frac{\partial \bar{\varepsilon}}{\partial c} \\ \frac{\partial \bar{\varepsilon}}{\partial d} \end{pmatrix} = \sum_{i=0}^{n} \begin{pmatrix} \frac{\partial \varepsilon_i^2}{\partial a} \\ \frac{\partial \varepsilon_i^2}{\partial b} \\ \frac{\partial \varepsilon_i^2}{\partial c} \\ \frac{\partial \varepsilon_i^2}{\partial d} \end{pmatrix} = \sum_{i=0}^{n} \begin{pmatrix} 2\varepsilon_i \frac{\partial \varepsilon_i}{\partial a} \\ 2\varepsilon_i \frac{\partial \varepsilon_i}{\partial b} \\ 2\varepsilon_i \frac{\partial \varepsilon_i}{\partial c} \\ 2\varepsilon_i \frac{\partial \varepsilon_i}{\partial d} \end{pmatrix} = \sum_{i=0}^{n} \begin{pmatrix} 2\varepsilon_i \\ 2\varepsilon_i x_i \\ 2\varepsilon_i x_i^2 \\ 2\varepsilon_i x_i^3 \end{pmatrix}$$

Condição de máximo e mínimo $\nabla \bar{\varepsilon} = 0$

$$\sum_{i=0}^{n} \begin{pmatrix} 2\varepsilon_i \\ 2\varepsilon_i x_i \\ 2\varepsilon_i x_i^2 \\ 2\varepsilon_i x_i^3 \end{pmatrix} = \begin{pmatrix} 0. \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\sum_{i=0}^{n} \begin{pmatrix} 2(a + bx_i + cx_i^2 + dx_i^3 - y_i) \\ 2(a + bx_i + cx_i^2 + dx_i^3 - y_i)x_i \\ 2(a + bx_i + cx_i^2 + dx_i^3 - y_i)x_i^2 \\ 2(a + bx_i + cx_i^2 + dx_i^3 - y_i)x_i^3 \end{pmatrix} = 0$$

Condição de máximo e mínimo $\nabla \bar{\varepsilon} = 0$

$$\sum_{i=0}^{n} (a + bx_i + cx_i^2 + dx_i^3 - y_i) = 0$$

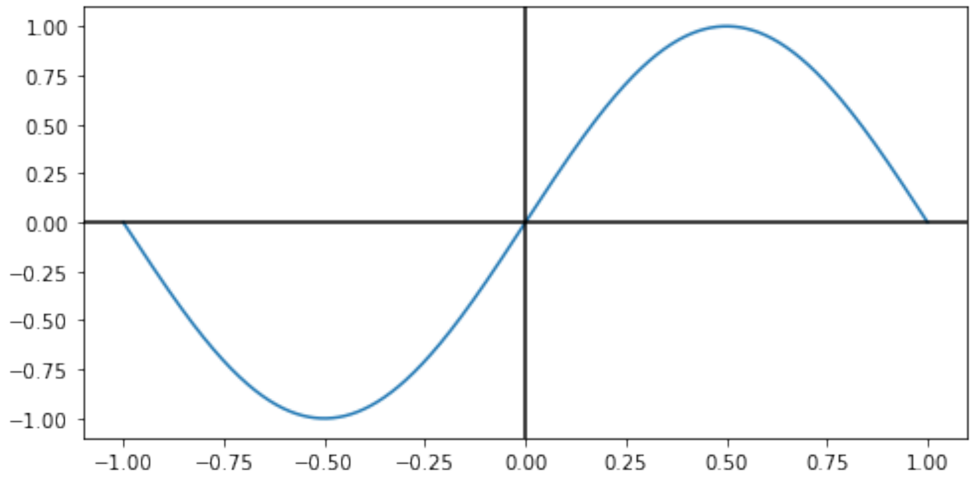
$$\sum_{i=0}^{n} (a + bx_i + cx_i^2 + dx_i^3 - y_i)x_i = 0$$

$$\sum_{i=0}^{n} (a + bx_i + cx_i^2 + dx_i^3 - y_i)x_i^2 = 0$$

$$\sum_{i=0}^{n} (a + bx_i + cx_i^2 + dx_i^3 - y_i)x_i^3 = 0$$

$$\begin{bmatrix} \sum_{i=0}^{n} 1 & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 \\ \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \sum_{i=0}^{n} x_i^5 \\ \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \sum_{i=0}^{n} x_i^5 & \sum_{i=0}^{n} x_i^6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^2 y_i \\ \sum_{i=0}^{n} x_i^3 y_i \end{bmatrix}$$

Função anti-simétrica com domínio em torno do zero

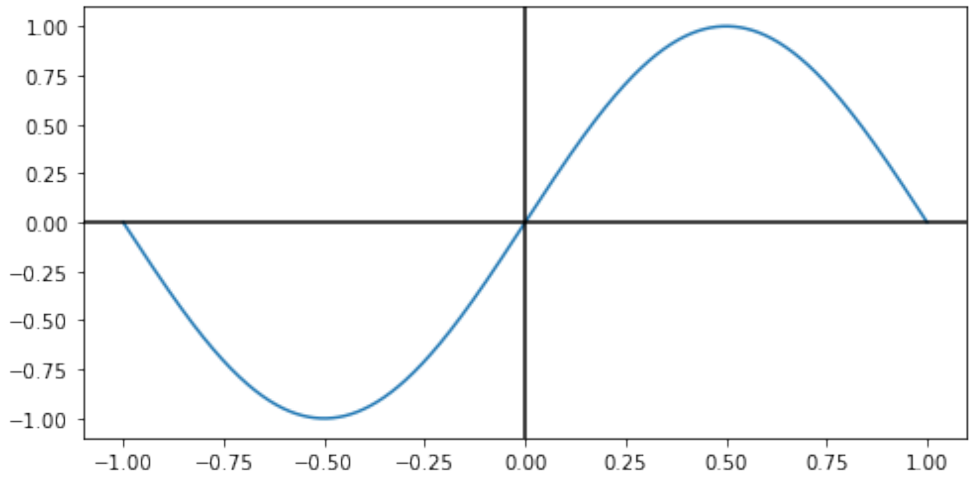


$$\begin{bmatrix} \sum_{i=0}^{n} 1 & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 \\ \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \sum_{i=0}^{n} x_i^5 \\ \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \sum_{i=0}^{n} x_i^5 & \sum_{i=0}^{n} x_i^6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^2 y_i \\ \sum_{i=0}^{n} x_i^3 y_i \end{bmatrix}$$

Se $\begin{cases} x_{\frac{n-1}{2}-i} = -x_{\frac{n-1}{2}+i} \\ y(x_i) = -y(-x_i) \end{cases} \Rightarrow \begin{bmatrix} n & 0 & \sum_{i=0}^{n} x_i^2 & 0 \\ 0 & \sum_{i=0}^{n} x_i^2 & 0 & \sum_{i=0}^{n} x_i^4 \\ \sum_{i=0}^{n} x_i^2 & 0 & \sum_{i=0}^{n} x_i^4 & 0 \\ 0 & \sum_{i=0}^{n} x_i^4 & 0 & \sum_{i=0}^{n} x_i^6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{i=0}^{n} x_i y_i \\ 0 \\ \sum_{i=0}^{n} x_i^3 y_i \end{bmatrix}$

$$a = c = 0$$

Função anti-simétrica com domínio em torno do zero



$$\Rightarrow \begin{bmatrix} n & 0 & \sum_{i=0}^{n} x_i^2 & 0 \\ 0 & \sum_{i=0}^{n} x_i^2 & 0 & \sum_{i=0}^{n} x_i^4 \\ \sum_{i=0}^{n} x_i^2 & 0 & \sum_{i=0}^{n} x_i^4 & 0 \\ 0 & \sum_{i=0}^{n} x_i^4 & 0 & \sum_{i=0}^{n} x_i^6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{i=0}^{n} x_i y_i \\ 0 \\ \sum_{i=0}^{n} x_i^3 y_i \end{bmatrix}$$

$$y_{pred}(x_i) = b x_i + d x_i^3$$

$$\Rightarrow \begin{bmatrix} \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^4 \\ \sum_{i=0}^{n} x_i^4 & \sum_{i=0}^{n} x_i^6 \end{bmatrix} \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^3 y_i \end{bmatrix}$$

Um notebook exemplo

$$erro = \sum_{i=1}^{n} (y_i - b - ax_i)^2$$

$$\frac{\partial\, erro}{\partial a} = -2 \sum_{i=1}^{n} x_i (y_i - b - ax_i) \qquad \frac{\partial\, erro}{\partial b} = -2 \sum_{i=1}^{n} (y_i - b - ax_i)$$

$$\begin{bmatrix} \sum_i x_i y_i \\ \sum_i y_i \end{bmatrix} = \begin{bmatrix} \sum_i x_i^2 & \sum_i x_i \\ \sum_i x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$a = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - \sum_i x_i \sum_i x_i} \qquad b = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \; \sum_i x_i y_i}{n \sum_i x_i^2 - \sum_i x_i \sum_i x_i}$$