Regressão e redes neurais rasas e profundas

Juvenal J. Duarte

Ementa

- Serão abordados nesta disciplina os seguintes tópicos:
  - Revisão de Cálculo Diferencial e Álgebra Linear
  - Introdução ao aprendizado de máquina
  - Regressão Linear
    - Definição de Função de Perda e Função de Custo (loss/cost)
    - Erro Quadrático Médio / Mean Squared Error (MSE)
    - Otimização por Gradiente Descendente (Batch Gradient Descent)

  - Regressão Logística e perceptron
    - Funções de ativação: Sigmoid e ReLU
    - Entropia Cruzada / Cross Entropy
    - Gradiente Descendente: Mini Batch e Estocástico

  - Multilayer Perceptron (MLP)
    - Algoritmo de feed-forward.
    - Backpropagation.

Revisão

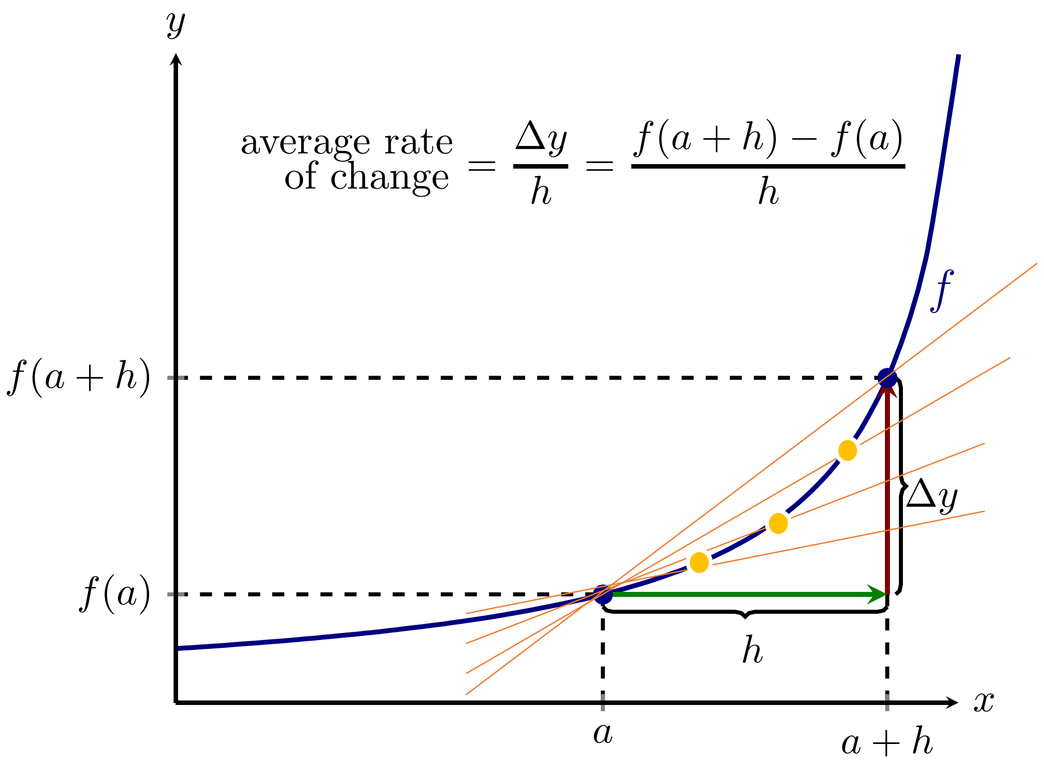
Cálculo diferencial

Juvenal J. Duarte

Cálculo diferencial

Derivada pela definição

$$\frac{dy}{dx}(x_i) = f'(x_i) = \lim_{h \to 0} \frac{f(x_i+h) - f(x_i)}{h}$$



Cálculo diferencial

Derivada de uma constante

$$\frac{d}{dx}(c) = 0$$

Derivada da potência

$$\frac{d}{dx}(x^n) = n.x^{n-1}$$

Portanto:

$$\frac{d}{dx}(x) = 1$$

Soma / Subtração

$$\frac{d}{dx}(u \pm v) = \frac{du}{dx} \pm \frac{dv}{dx}$$

Produto por uma constante

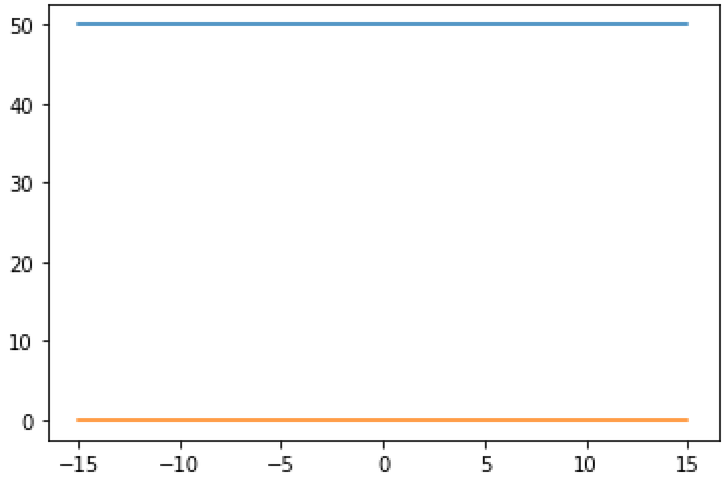
$$\frac{d}{dx}(cu) = c\frac{du}{dx}$$

Derivada do produto

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

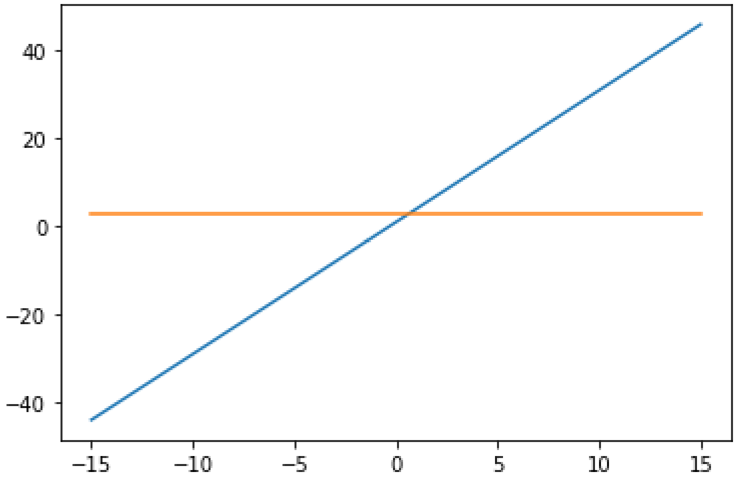
Cálculo diferencial: exemplos

Função constante



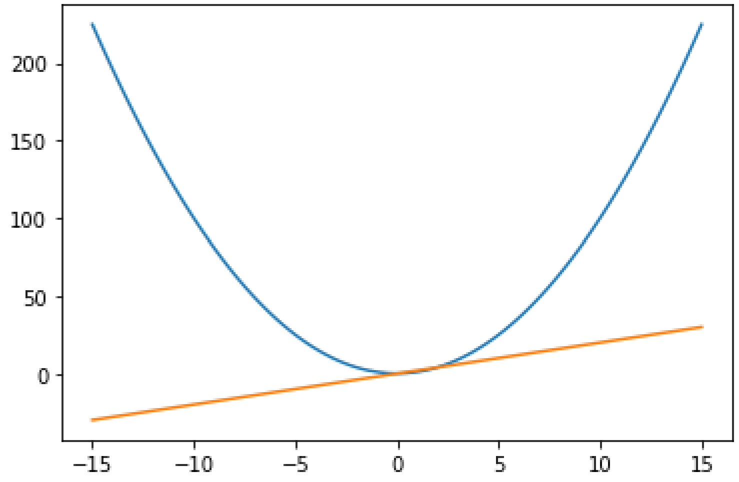
$$f(x) = 50$$
$$f'(x) = \frac{\partial y}{\partial x} = 0$$

Função linear



$$f(x) = 3x + 1$$
$$f'(x) = \frac{\partial y}{\partial x} = 3$$

Função quadrática



$$f(x) = x^2$$
$$f'(x) = \frac{\partial y}{\partial x} = 2x$$

Cálculo diferencial: exercícios

(1) $f(x) = 11$

(2) $f(x) = 4x + 20x + 2$

(3) $f(x) = 4x^3 + x + 1$

(4) Se a população mundial é dada pela função $f(x)$, qual interpretação pode ser dada à sua derivada?

Derivadas Parciais

Em funções de múltiplas variáveis, é possível estabelecer a derivada em relação a cada variável particular

$$V = \frac{r^2 h\pi}{3}$$

Derivada em relação à “r”

$$\frac{\partial V}{\partial r} = \frac{2rh\pi}{3}$$

Derivada em relação à “h”

$$\frac{\partial V}{\partial h} = \frac{r^2\pi}{3}$$

Revisão
Álgebra linear

Juvenal J. Duarte

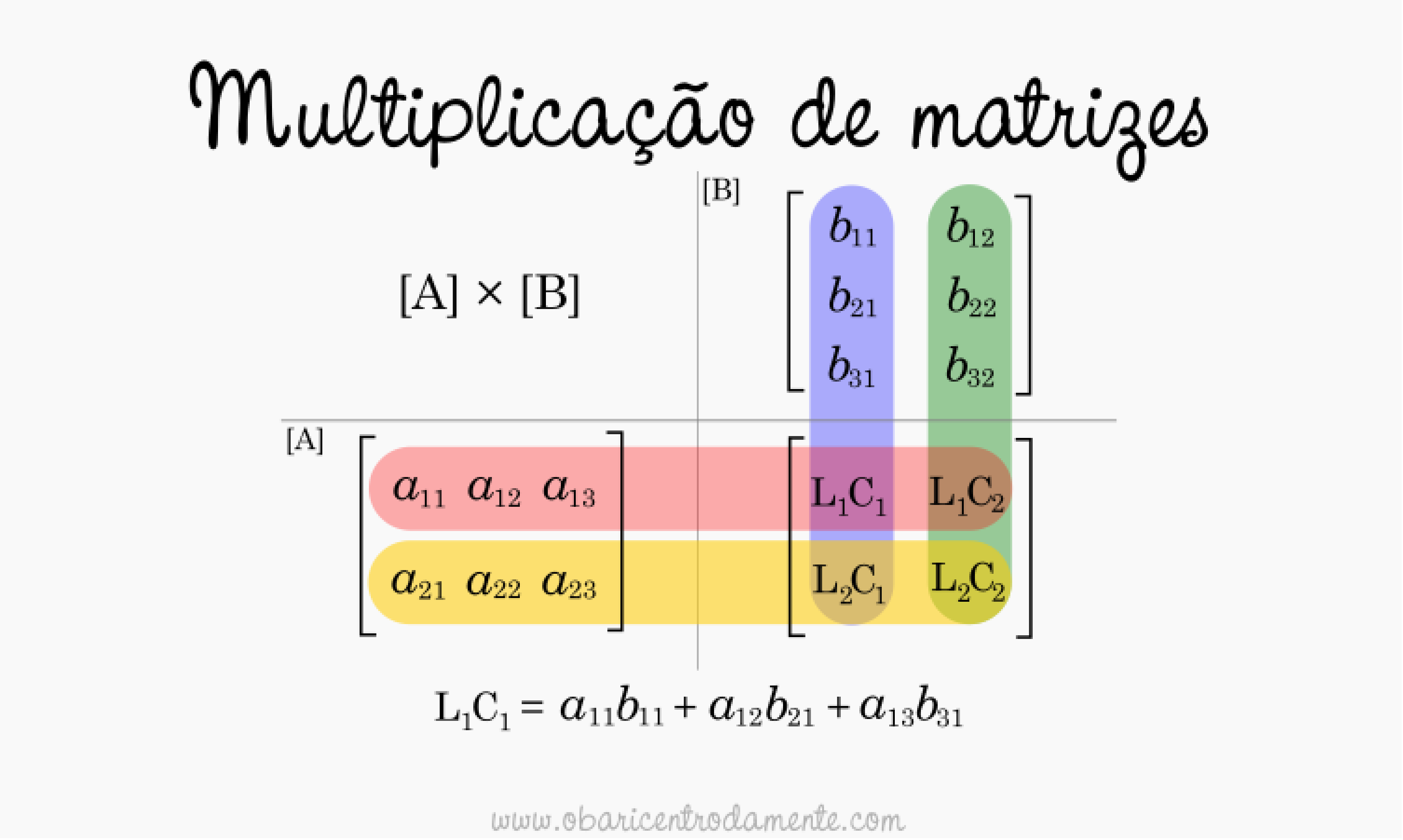
Álgebra linear

Multiplicação por constante

$$3 \cdot \begin{bmatrix} 2 & 4 & 7 \end{bmatrix} = \begin{bmatrix} 3.2 & 3.4 & 3.7 \end{bmatrix} = \begin{bmatrix} 6 & 12 & 21 \end{bmatrix}$$

$$½ \cdot \begin{bmatrix} 4 & 6 \\ 10 & 1 \end{bmatrix} = \begin{bmatrix} 4.1/2 & 6.1/2 \\ 10.1/2 & 1.1/2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & \frac{1}{2} \end{bmatrix}$$

Álgebra linear



Multiplicação exemplo:

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 0 & 2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 1 \end{bmatrix}$$

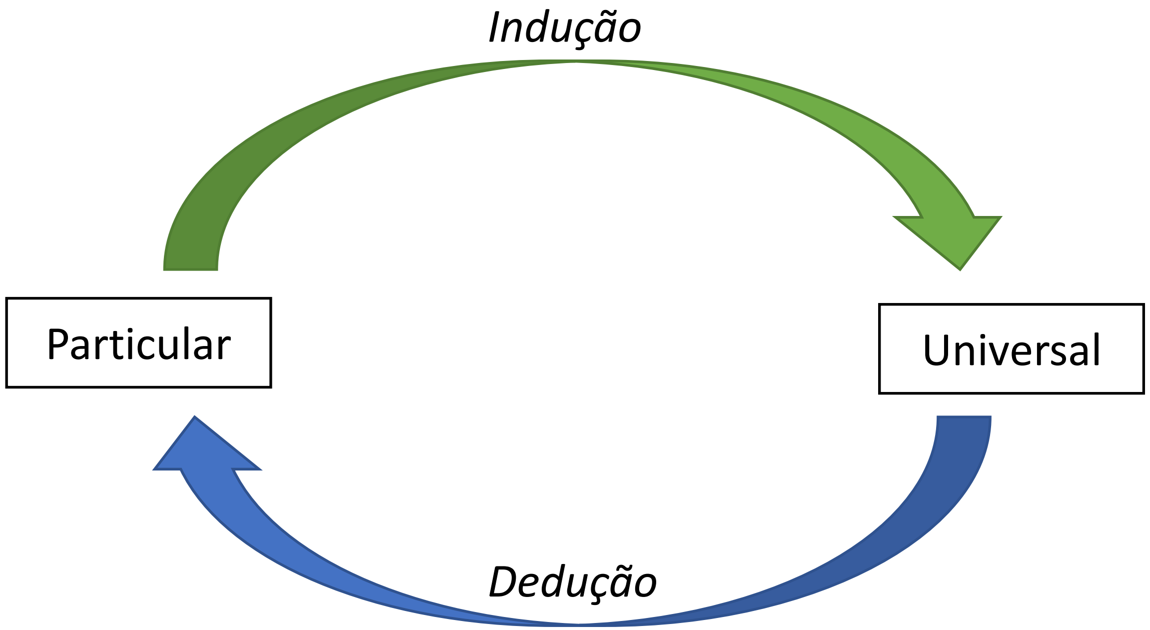
$$\mathbf{A}_{2x3} \cdot \mathbf{B}_{3x2} = \mathbf{C}_{2x2}$$

$$\mathbf{C} = \begin{bmatrix} 6 & 12 \\ 7 & 4 \end{bmatrix}$$

Introdução:
Tipos de aprendizado

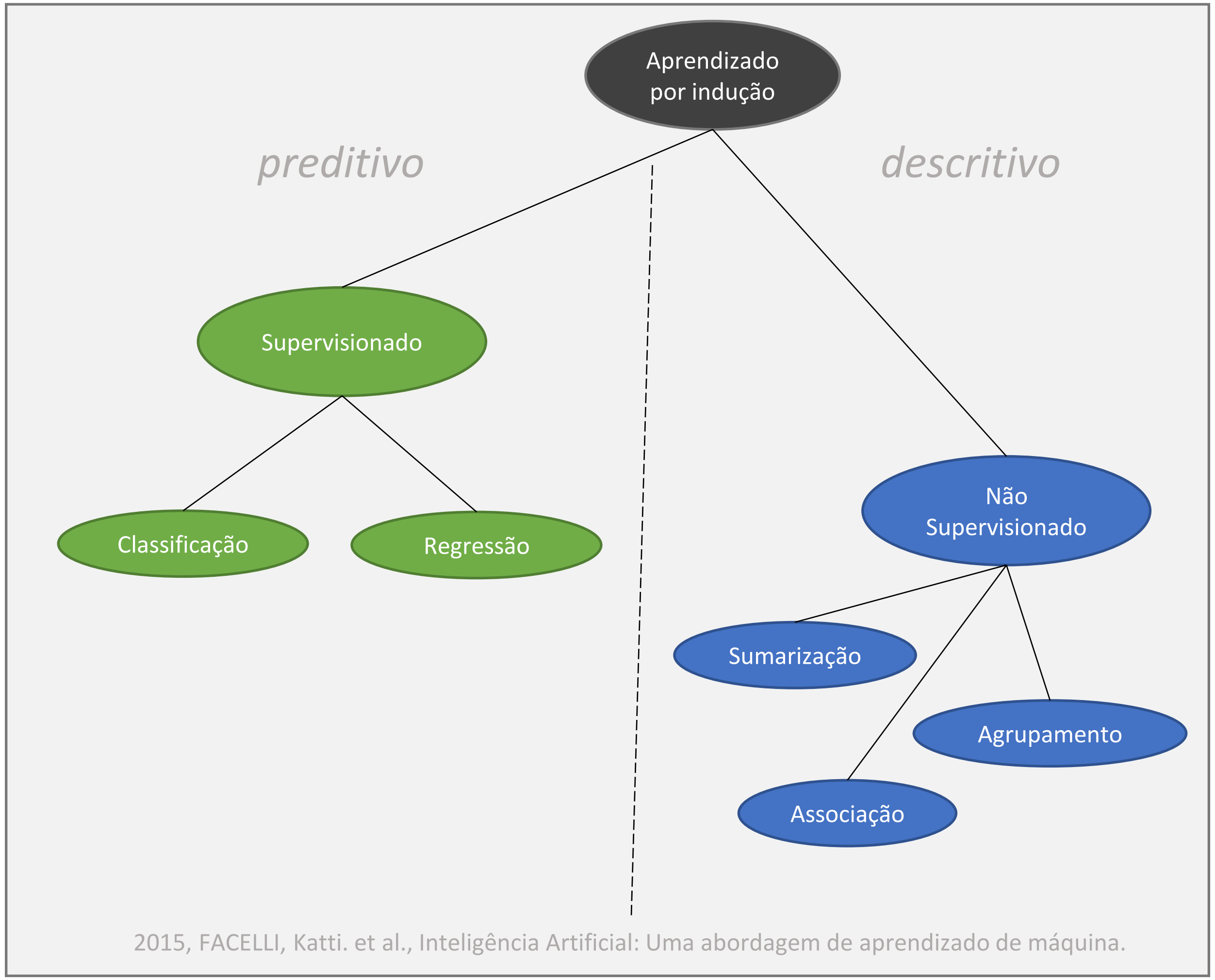
Juvenal J. Duarte

Aprendizado por indução



Aprendizado por indução

- Principais tarefas são:
  - Predizer (Apr. Supervisionado):
    - Diagnóstico de doenças.
    - Reconhecimento facial.
    - Predição de fraudes.
    - Predição de desistência (*churn*).
    - SAC: Qual será o volume de ligações?
    - Qual o preço justo de um imóvel?
    - Algum outro caso interessante?
  - Descrever (Apr. não Supervisionado):
    - Sistemas de recomendação.
    - Segmentação de clientes.
    - Agrupamento de documentos similares.
    - Algum outro exemplo?



Aprendizado por indução

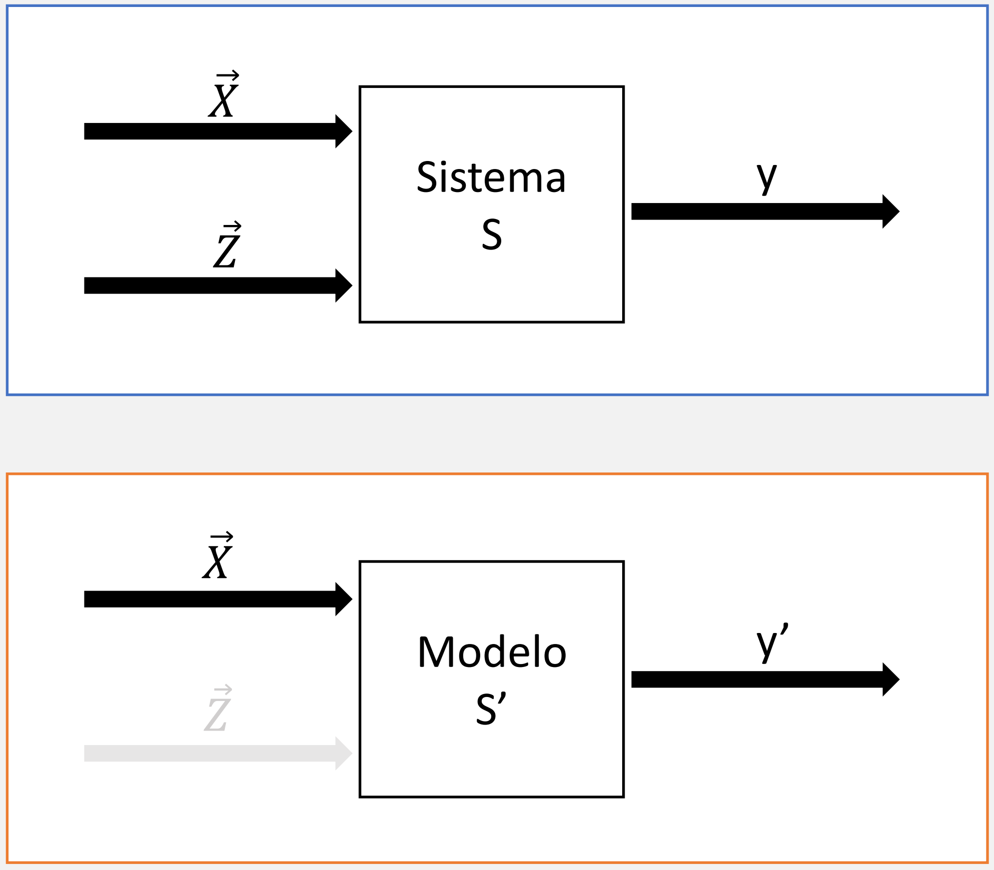
- Aprendizado supervisionado:
  - São fornecidos o conjunto de atributos independentes ($\vec{X}$) e o atributo alvo (y). A partir dos exemplos o algoritmo gera um modelo, por *indução*, capaz predizer novas amostras por *dedução*.

- Aprendizado não supervisionado:
  - É fornecido apenas o conjunto de atributos ($\vec{X}$). Os algoritmos buscam padrões descritivos nos dados através da análise das características dos exemplos e suas relações.

Aprendizado supervisionado: Formalização

- Existe um sistema S, com entradas observadas $\vec{X}$ e entradas não observadas $\vec{Z}$, que produz os resultados y baseado nos inputs.

- Existe um modelo S’, cuja função busca imitar o comportamento do Sistema S, mapeando o conjunto de entradas $\vec{X}$ a uma saída y’.



2007, CHERKASSKY, V.; MULIER, F, Learning from data: Concepts, Theory and Methods

Aprendizado supervisionado

- S1 = 2, 4, 6, 8, 10 … ?
- S2 = 1, 4, 9, 16, 25 … ?
- S3 = 0, 1, 1, 2, 3, 5, 8, 13, 21 … ?
- S4 = ●, □, Δ, ●, □, Δ … ?

Aprendizado supervisionado

- S1 = 2, 4, 6, 8, 10 … ?
- S2 = 1, 4, 9, 16, 25 … ?
- S3 = 0, 1, 1, 2, 3, 5, 8, 13, 21 … ?
- S4 = ●, □, Δ, ●, □, Δ … ?

S1:

$$f(n) = 2n, \forall\, n > 0$$

S2:

$$f(n) = n^2, \forall\, n > 0$$

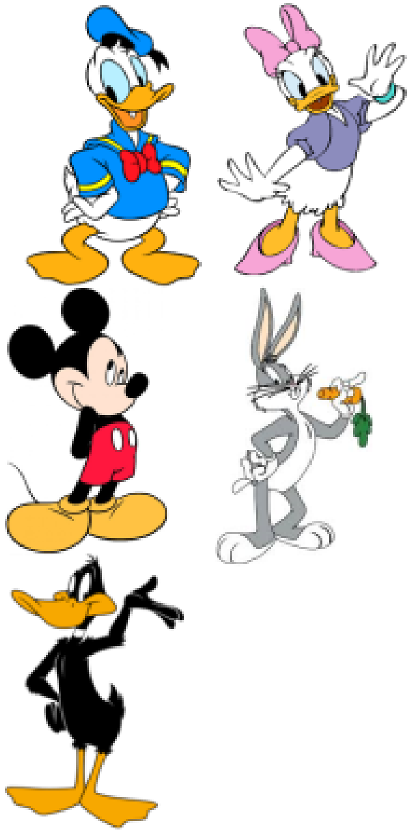
S3:

$$f(n) = \begin{cases} f(n-1) + f(n-2), \forall\, n > 1 \\ 1, n = 1 \\ 0, n = 0 \end{cases}$$

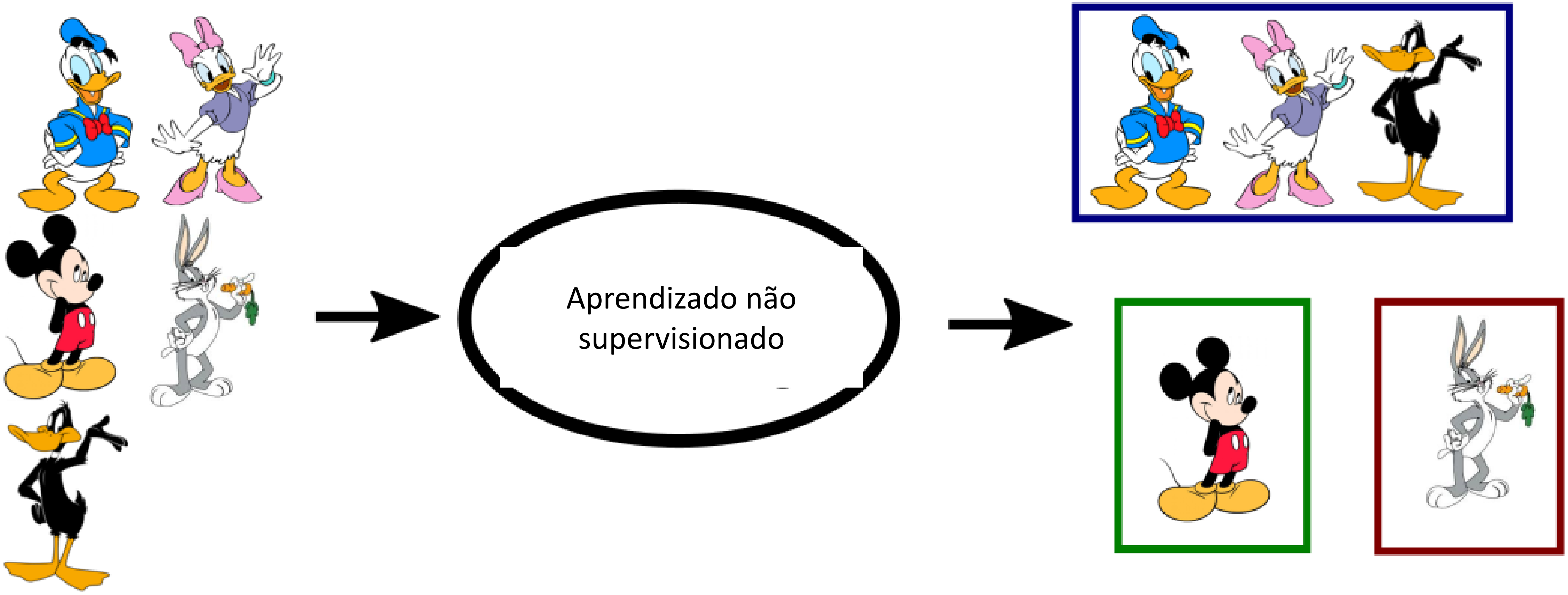
S4:

$$f(n) = \begin{cases} f(n-3), \forall\, n > 2 \\ \Delta, n = 2 \\ \square, n = 1 \\ \bullet, n = 0 \end{cases}$$

Aprendizado não supervisionado



Aprendizado não supervisionado



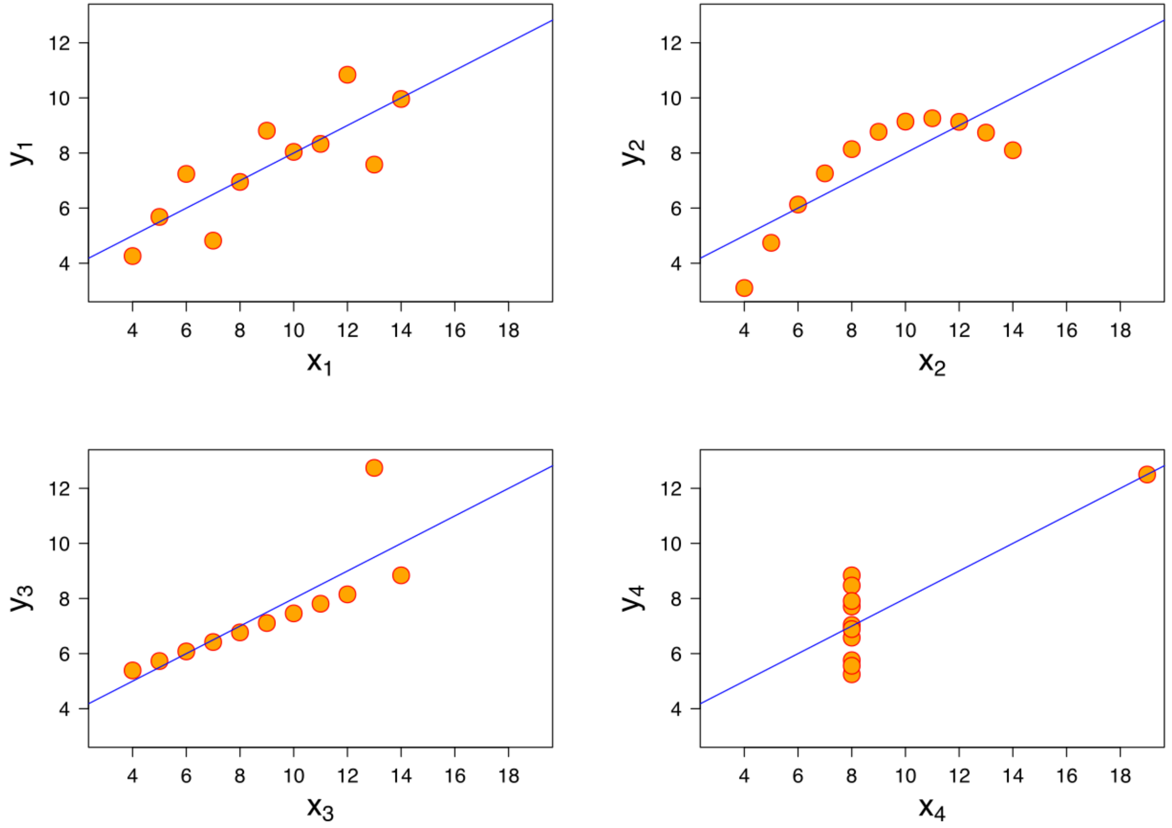
2018, Devin Soni, acessado em:
towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d

Regressão linear

Juvenal J. Duarte

Regressão linear

Investiga relações lineares entre variáveis. Qual das relações abaixo apresenta relação mais linear?



Wikipedia: https://en.wikipedia.org/wiki/Linear_regression

Regressão linear univariada

O que é uma relação de dependência linear?

$$y = \beta x + \alpha$$

Notação alternativa:

$$y = mx + b$$

Regressão linear univariada

O que é uma relação de dependência linear?

Variável dependente

Bias

$$y = \beta x + \alpha$$

Scale factor
(curve slope)

Variável independente

Regressão linear multivariada

Como generalizer a formula para o caso de n atributos?

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n + \alpha$$

$$y = \alpha + \sum_{i=1}^{n} x_i \beta_i$$

Regressão linear multivariada

Cálculo vetorial

Quantidade de linhas / registros / amostras: m

Quantidade de colunas / atributos / characteristicas: n

Conjunto de entradas: $\vec{X}_{m,n} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$

Conjunto de pesos: $\vec{B}_{n,1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_n \end{bmatrix}$

Bias: α

(1) Predição (formula vetorial):

$$\hat{y}_{m,1} = \vec{X}_{m,n} \vec{B}_{n,1} + \alpha$$

Regressão linear: função de custo

Como estimar os parâmetros e saber se o modelo é bom o suficiente?

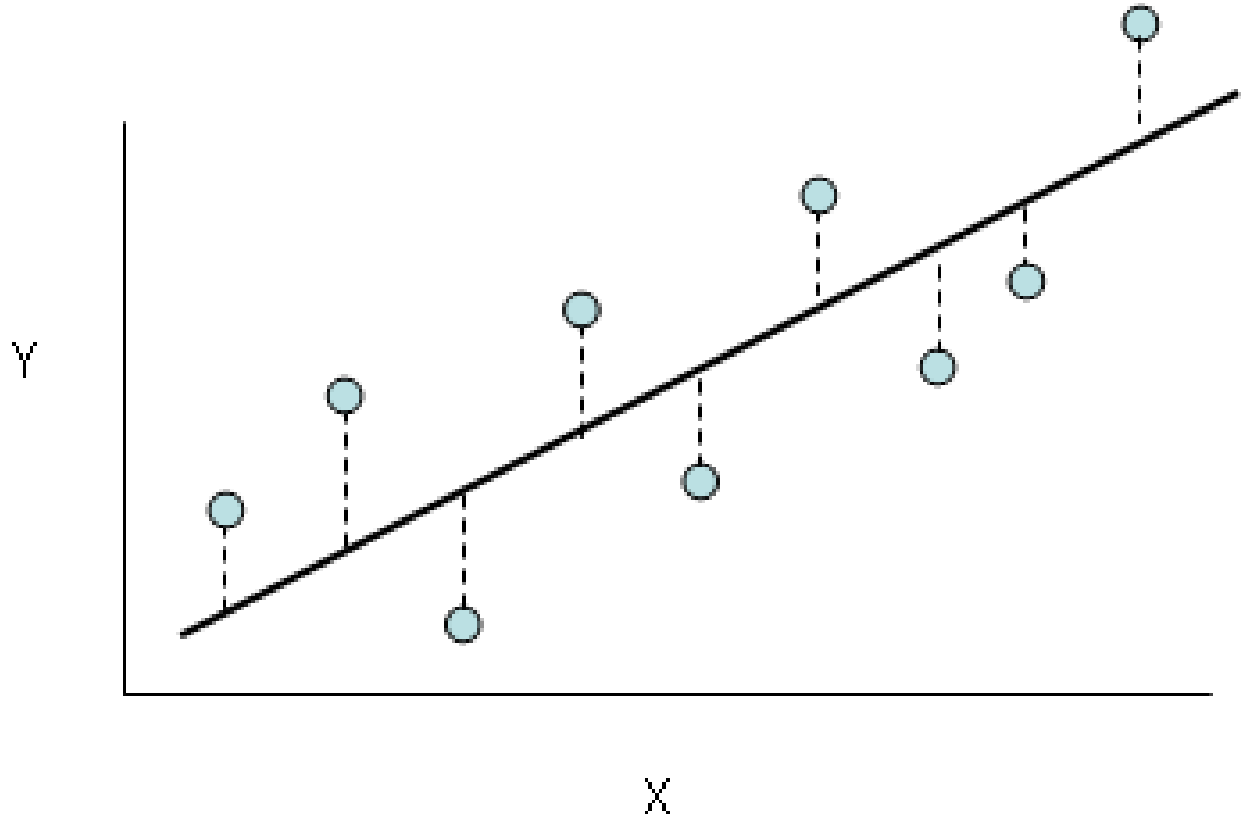
(2)

$$loss(i) = y_i - \hat{y}_i$$

Erro Quadrático Médio (MSE):

(3)

$$cost = J = \frac{1}{m}\sum_{i=}^{m} loss(y_i, \hat{y}_i)^2$$

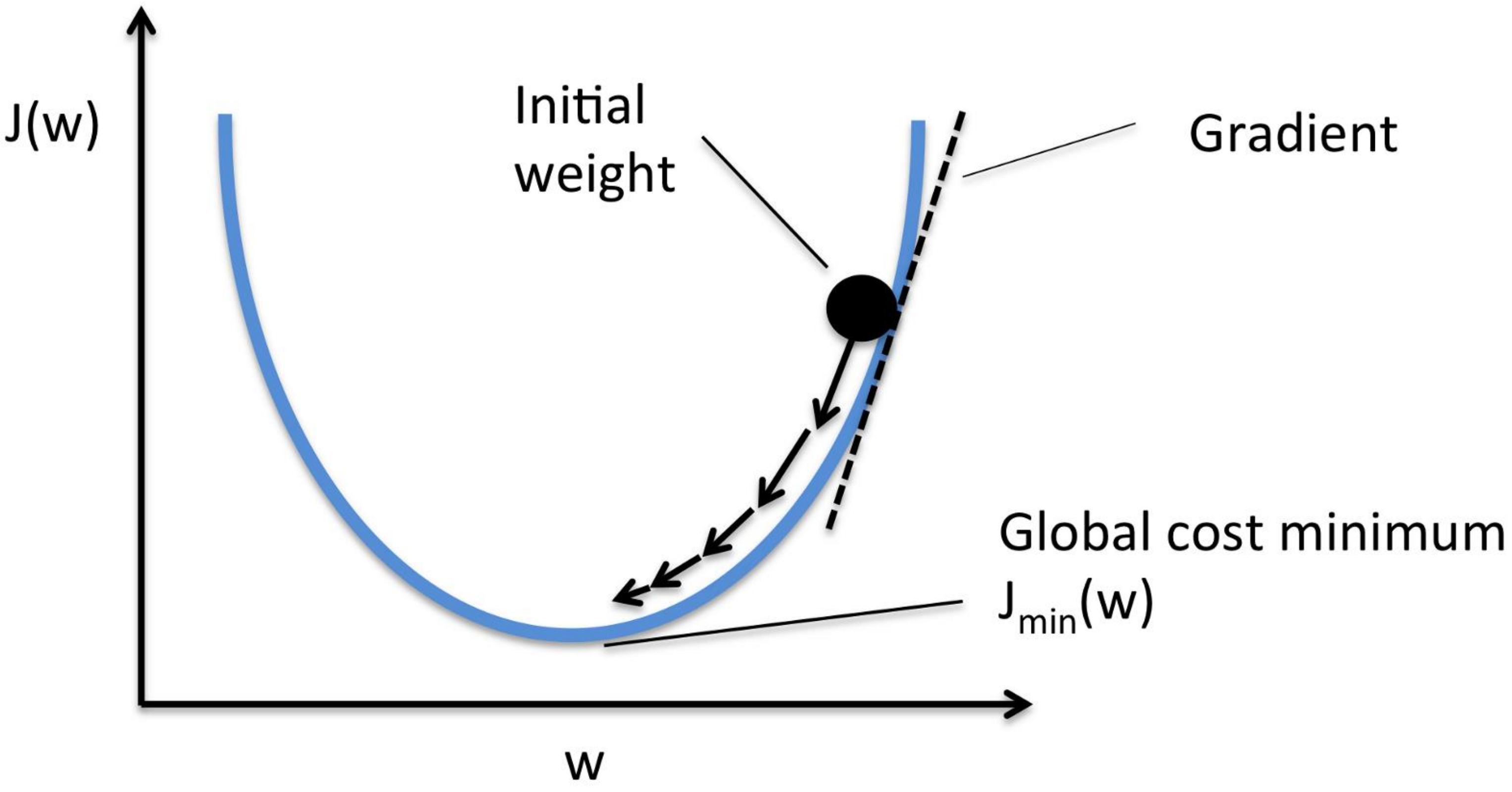


Regressão linear:
Gradiente Descendente

Juvenal J. Duarte

Gradiente descendente

Parametros do modelo sao corrigidos iterativamente, com ajuste dado pela derivada da função de custo.



Regressão linear: gradiente descendente

(3)

$$J = \frac{1}{m}\sum_{i=}^{m} loss(y_i\,, \hat{y}_i)^2$$

$$\frac{\partial J}{\partial \beta} = \frac{1}{m}\sum_{i=}^{m} -2\vec{x_i}(y_i \;\; - (\vec{x_i}\beta + \; \alpha)$$

$$\frac{\partial J}{\partial \alpha} = \frac{1}{m}\sum_{i=}^{m} -2(y_i \;\; - (\vec{x_i}\beta + \; \alpha)$$

Regressão linear: gradiente descendente

(3)

$$J = \frac{1}{m}\sum_{i=}^{m} loss(y_i\,, \hat{y}_i)^2$$

(4)

$$\frac{\partial J}{\partial \beta} = -\frac{2}{m} \times\ transp(\vec{X}_{m,n}) \times \left[y_{m,1}\ - \left(\vec{X}_{m,n} \times \vec{B}_{n,1} +\ \alpha\right)\right]$$

(5)

$$\frac{\partial J}{\partial \alpha} = -\frac{2}{m} \times \sum (y_{m,1}\ - \left(\vec{X}_{m,n} \times \vec{B}_{n,1} +\ \alpha\right))$$

Hiper-Parametros

Learning Rate = Alpha:

Define a velocidade do aprendizado. Quando baixa, exige mais iteracoes ate a solucao otima, quando alta dificulta a convergencia do algoritmo.

Epochs:

Define quantas iteracoes de ajustes nos coeficientes serao executadas.

Ajuste dos pesos / coeficientes

(6)

$$\vec{B}_{n,1} = \vec{B}_{n,1} - alpha \times \frac{\partial J}{\partial \beta}$$

(7)

$$\alpha = \alpha - alpha \times \frac{\partial J}{\partial \alpha}$$