



# APRENDIZADO EM MÁQUINAS - 1

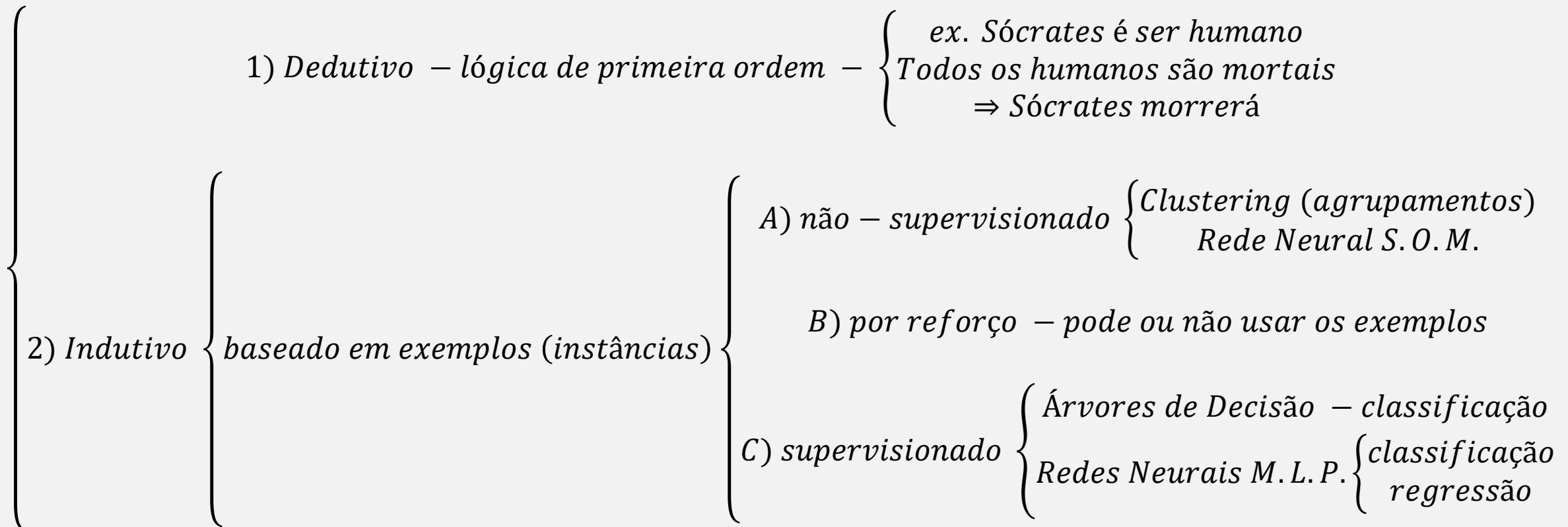
Machine Learning – 1

Prof. Dr. Waldemar Bonventi Jr.

mar2020

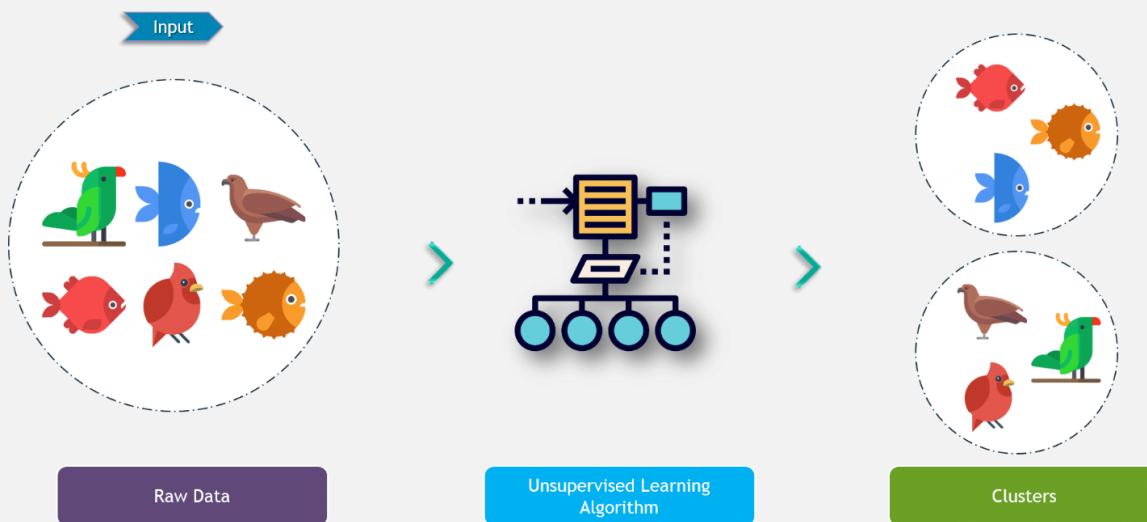


# FORMAS DE APRENDIZADO EM MÁQUINAS

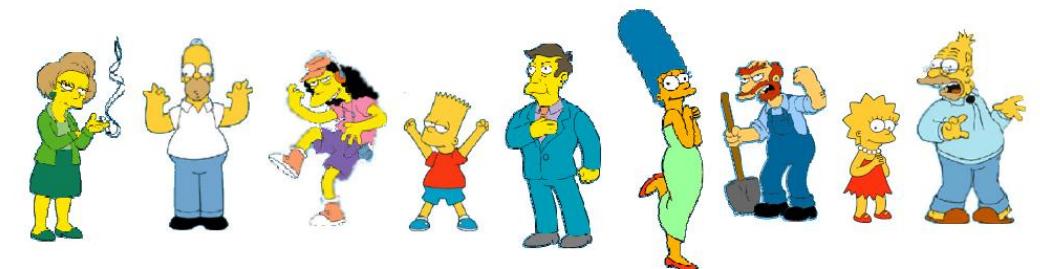


Classificação – separar os exemplos em categorias (discretas, não contínuas)  
Regressão – associar um valor de saída (contínuo) ao exemplo de entrada

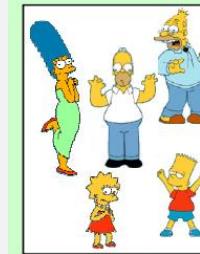
# APRENDIZADO NÃO-SUPERVISIONADO CLUSTERING (AGRUPAMENTOS)



What is a natural grouping among these objects?



Clustering is subjective



Simpson's Family



School Employees



Females



Males



# COMO AGRUPAR OS ANIMAIS?



terra

água



Sem bico

Com bico

mamífero

ovíparo



# COMO DEFINIR SEMELHANÇA?

Semelhante ≠ igual



# MEDINDO AS CARACTERÍSTICAS (ATRIBUTOS) DE CADA OBJETO

e verificando as diferenças entre elas, como um todo

0	8	8	7	7
0	2	4	4	
	0	3	3	
		0	1	
			0	

Matriz de distâncias (ou diferenças)

onde

$$\begin{aligned} D(\text{Marge 1}, \text{Bart}) &= 8 \\ D(\text{Marge 2}, \text{Marge 3}) &= 1 \end{aligned}$$

As propriedades que definem uma medida de similaridade são 3

$$d(x,y) = d(y,x), \text{ simetria}$$

$$d(x,y) \geq 0$$

$$d(x,x) = 0$$

Além destas propriedades, também valem

$$d(x,y) = 0, \text{ se e somente se } x = y$$

$d(x,y) \leq d(x,z) + d(z,y)$ , também conhecida como desigualdade triangular



# TRABALHAR COM BASE DE DADOS

Matriz de dados

contendo os dados de  $N$  objetos, cada qual com  $p$  atributos

$$X = \begin{bmatrix} x_{11} & x_{12} \dots & x_{1p} \\ x_{21} & x_{22} \dots & x_{2p} \\ \dots & \dots & \dots \\ x_{N1} & x_{Np} \dots & x_{Np} \end{bmatrix} \quad \xleftarrow{\hspace{1cm}} \quad x_i = [x_{i1} \ x_{i2} \dots x_{ip}]$$

Cada objeto dessa matriz é denotado por um vetor  $x_i$

Exemplo: concessão de crédito

Cliente	Renda (R\$)	Restrições ao crédito	Número de dependentes	Moradia
A	5000	Sim	3	Própria
B	1500	Sim	2	Aluguel
C	600	Não	2	Aluguel
D	1000	Não	1	Aluguel
E	3500	Sim	2	Própria

automóvel	peso(kg)	potencia(cv)	bagagem(litros)
A	1250	110	650
B	800	80	300
C	900	90	450
D	750	100	400
E	1100	90	350
F	1050	90	600



**PRÓXIMA: COMO MEDIR AS DISTÂNCIAS,  
A PARTIR DOS ATRIBUTOS**

