



# APRENDIZADO EM MÁQUINAS — 1A

*Machine Learning — 1A*

*Prof. Dr. Waldemar Bonventi Jr.*

*mar2020*



# MEDIR DISTÂNCIAS PELOS ATRIBUTOS

$p$  atributos  $X$

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

registros  
dos  $N$  objetos

$$X = \begin{bmatrix} x_{11} & x_{12} \dots & x_{1p} \\ x_{21} & x_{22} \dots & x_{2p} \\ \dots & \dots \dots & \dots \\ x_{N1} & x_{Np} \dots & x_{Np} \end{bmatrix}$$

X1

X2

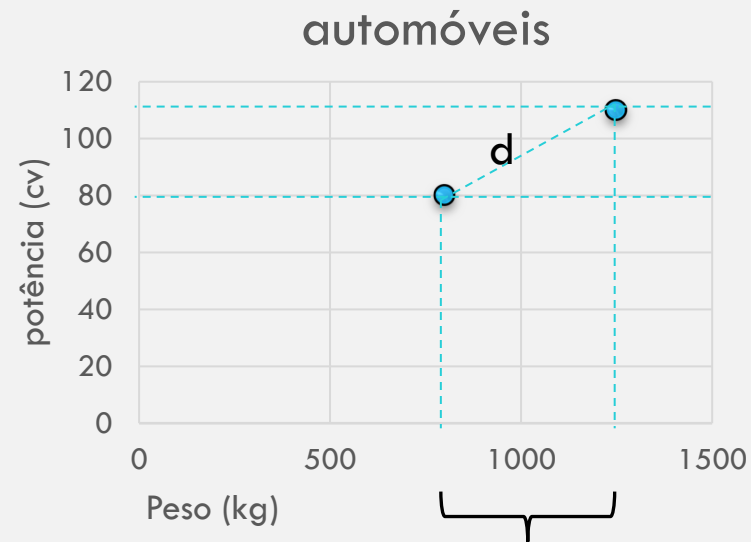
X3

Rótulos / labels



# DISTÂNCIAS ENTRE DOIS OBJETOS

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



Pitágoras:

$$d^2 = (\Delta x)^2 + (\Delta y)^2$$

$$= 450^2 + 30^2 = 203400$$

$$d = \sqrt{203400} = 451$$

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

$$D(\text{Marge}, \text{Lisa}) = 8$$

$$D(\text{Lisa}, \text{Lisa}) = 0$$



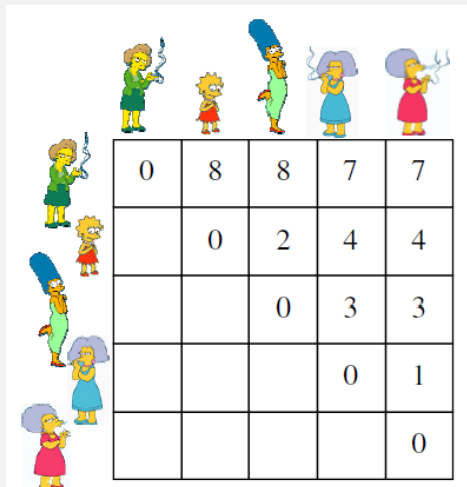
# DISTÂNCIAS ENTRE DOIS OBJETOS: TRÊS ATRIBUTOS











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


Pitágoras “generalizado”:

$$\begin{aligned}d^2 &= (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = \\&= (1250-800)^2 + (110-80)^2 + (650-300)^2 \\&= 325900\end{aligned}$$

$$d = \sqrt{325900} = 570,9$$



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


$$\begin{aligned}D(\text{Mr. Burns}, \text{Homer Simpson}) &= 8 \\D(\text{Lisa Simpson}, \text{Bart Simpson}) &= 1\end{aligned}$$

Em geral:

$$d^2 = \sum (\Delta x_i)^2$$

(Distância  
Euclidiana)



# DISTÂNCIAS ENTRE DOIS OBJETOS: MATRIZ DE DISTÂNCIAS OU DISSIMILARIDADES

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

$$d_{AB}^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (1250-800)^2 + (110-80)^2 + (650-300)^2 = 325900$$

$$d_{AB} = \sqrt{325900} = 570,9$$

$$d_{AC}^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (1250-900)^2 + (110-90)^2 + (650-450)^2 = 162900$$

$$d_{AC} = \sqrt{162900} = 403,6$$

E assim por diante: **AD, AE, AF, BC, BD, BE, BF, CD, CE, CF, DE, DF, EF**

**Pense:** por quê “faltaram” BA, BB, CA, CB, CC, DA, DB, DC, DD, .... ???

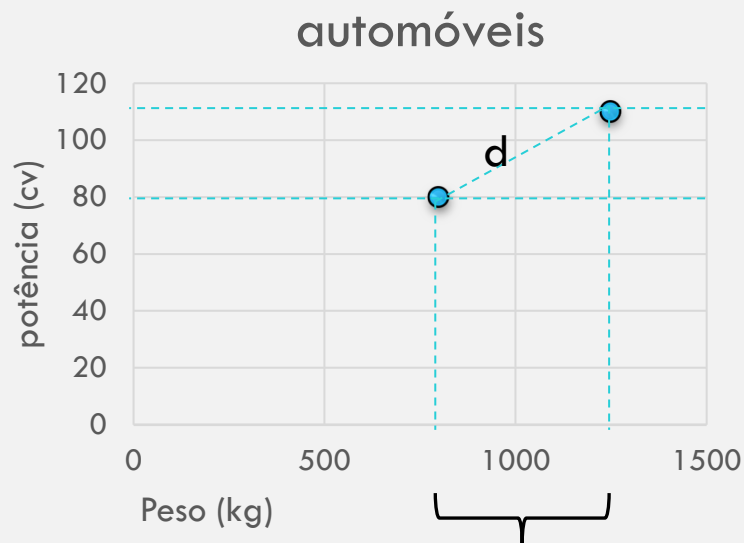
	A	B	C	D	E	F
A	0	570,9	403,6	...	...	...
B		0	...	...	...	...
C			0	...	...	...
D				0	...	...
E					0	...
F						0

‘quantidade’ de cálculos =  $N(N-1)/2$

No caso  $6.(6-1)/2 = 15$



# MAS, HOUSTON, TEMOS UM PROBLEMA...



$$\Delta y = 110 - 80 = 30$$

$$\Delta x = 1250 - 800 = 450$$

Pitágoras:

$$d^2 = (\Delta x)^2 + (\Delta y)^2 \\ = 450^2 + 30^2 = 203400$$

$$d = \sqrt{203400} = 451$$

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

O atributo 'potência' “pesa” pouco na distância  
A distância  $d$  está praticamente determinada  
pelo atributo 'peso'



# PRÓXIMO CAPÍTULO: COMO RESOLVER ?

