

Quiz 1

Nome: Hélio Zaira Francisco

1) Apresente os valores preditos para $x = 0.5$, 2.1 e 4 pelo método KNN nas seguintes situações:

$$x=1; y=1$$

$$x=0,5$$

$$y=1$$

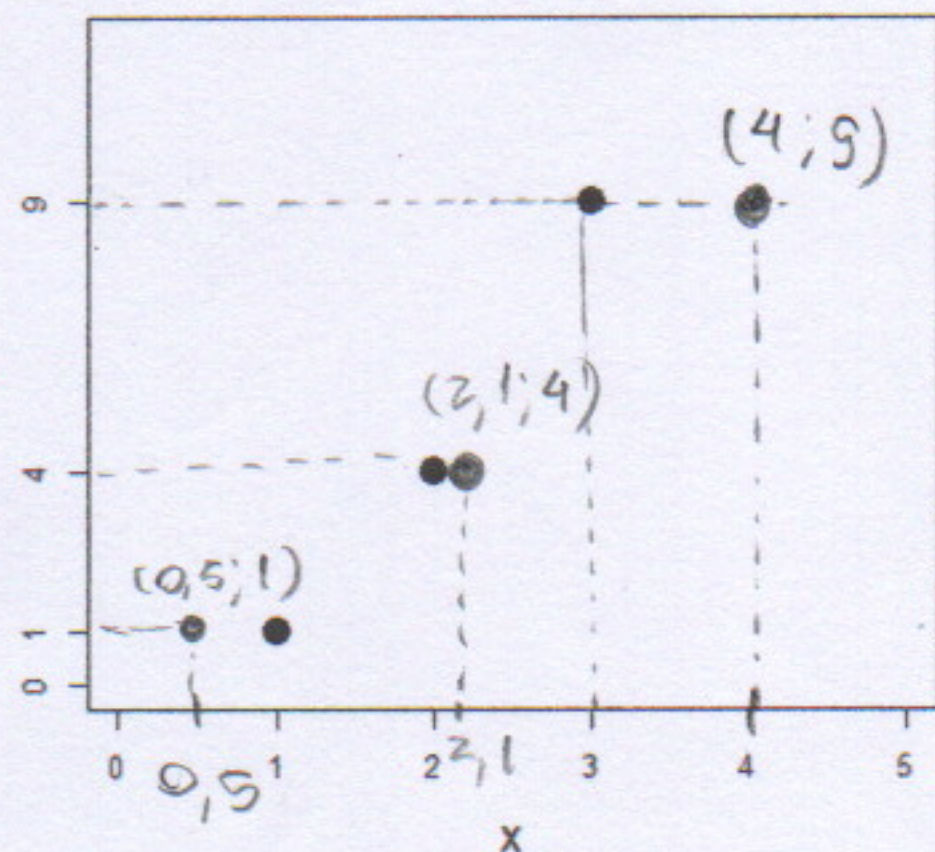
$$x=2,1 \rightarrow x=2; y=4$$

$$x=2,1; y=4$$

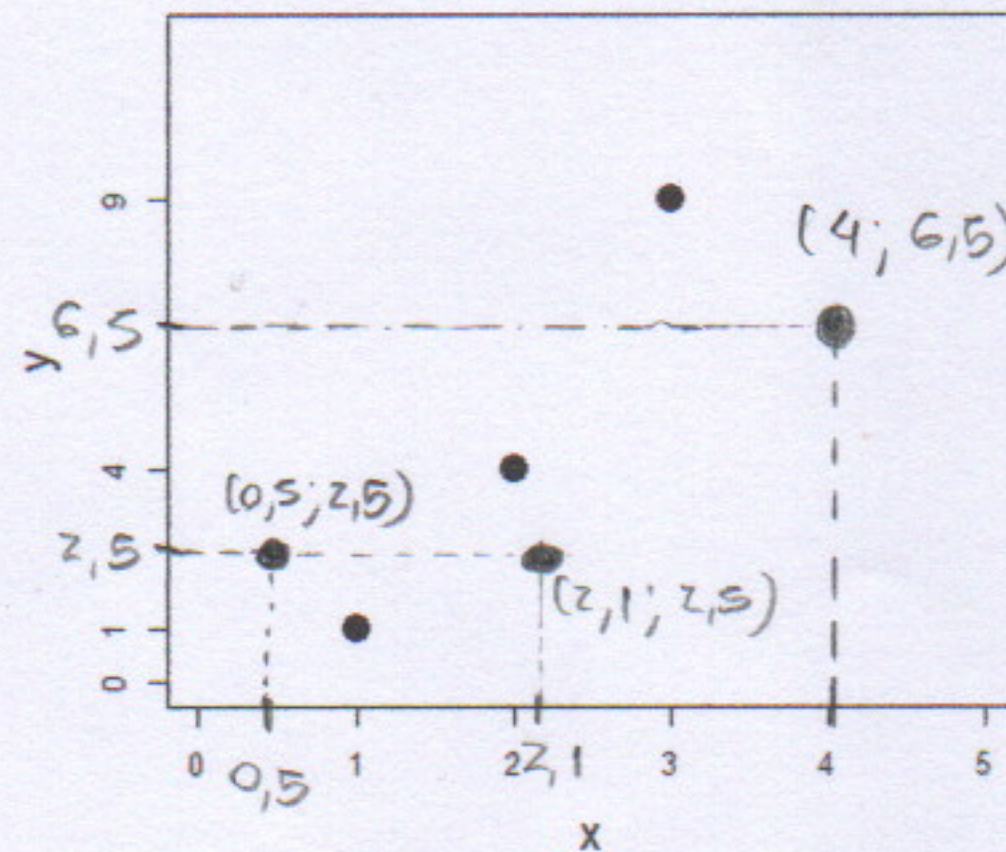
$$x=4 \rightarrow x=3; y=9$$

$$x=4; y=9$$

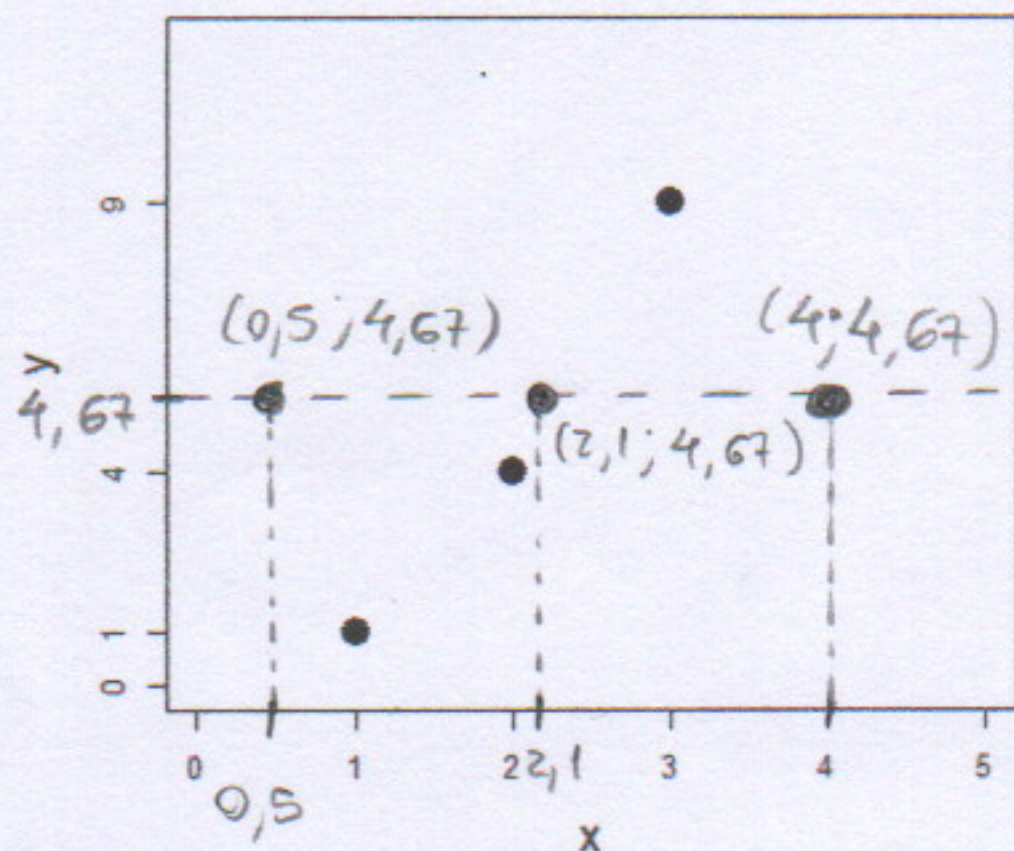
K=1



K=2



K=3



$$x=0,5 \rightarrow (x=1; y=1) (x=2; y=4)$$

$$y = \frac{1+4}{2} = 2,5$$

$$x=2,1 \rightarrow (1; 1) (2; 4)$$

$$y = \frac{1+4}{2} = 2,5$$

$$x=4 \rightarrow (3; 9) (2; 4)$$

$$y = \frac{9+4}{2} = 6,5$$

$$x=0,5 \rightarrow (1; 1) (2; 4) (3; 9)$$

$$y = \frac{1+4+9}{3} = 4,67$$

$$x=2,1 \rightarrow (1; 1) (2; 4) (3; 9)$$

$$y = \frac{1+4+9}{3} = 4,67$$

$$x=4 \rightarrow (1; 1) (2; 4) (3; 9)$$

$$y = \frac{1+4+9}{3} = 4,67$$

PROGRAMA AVANÇADO EM DATA SCIENCE E DECISÃO

- 2) Para prever se o cliente realizará a compra de um produto com base em um índice de satisfação e um índice de engajamento, foi determinada a utilização do método dos k vizinhos mais próximos. Nesse exercício você deverá avaliar o erro de previsão considerando $k = 1$ e $k = 3$ vizinhos mais próximos com um processo de validação cruzada em 4 lotes. Na tabela a seguir é apresentado o banco de dados que você deverá utilizar nessa tarefa. Note que você deverá sortear o lote de cada observação e preencher a coluna fold com o respectivo lote variando de 1 a 4.

ID	Índice de satisfação	Índice de engajamento	Comprou o produto	Fold	k = 1	k = 3
a	0.71	0.89	sim	2	NÃO	NÃO
b	0.25	0.57	não	1	Sim	NÃO
c	0.39	0.59	sim	3	NÃO	NÃO
d	0.09	0.37	não	1	NÃO	NÃO
e	0.96	0.36	sim	2	NÃO	Sim
f	0.01	0.59	sim	3	NÃO	NÃO
g	0.57	0.87	não	4	Sim	Sim
h	0.76	0.68	sim	1	NÃO	Sim
i	0.87	0.14	não	2	Sim	Sim
j	0.04	0.55	não	4	Sim	NÃO
k	0.66	0.68	não	3	Sim	Sim
l	0.88	0.53	sim	4	Sim	Sim

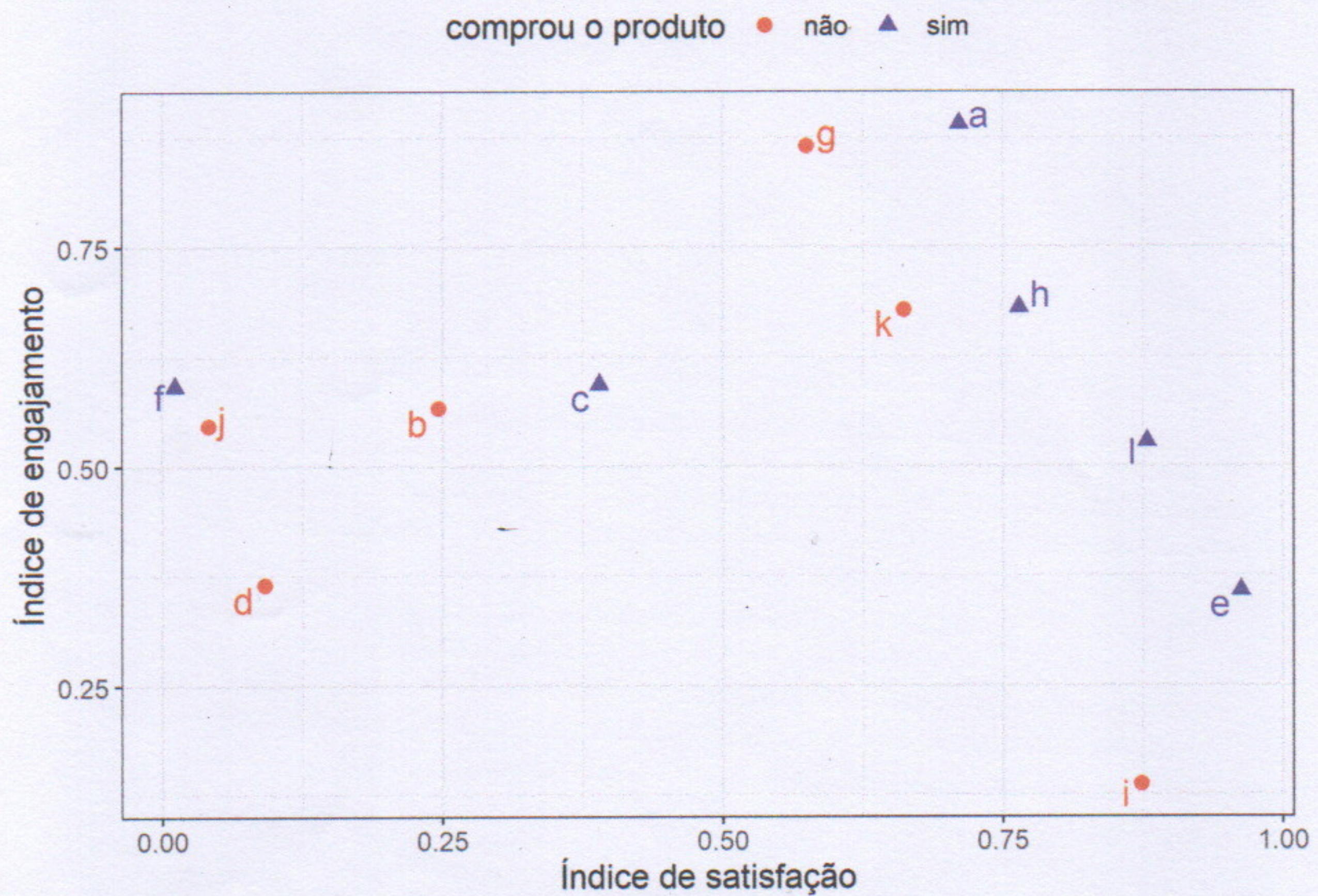
Para auxiliar nessa tarefa, apresentamos um gráfico de dispersão dos índices de acordo com a compra do produto e uma matriz com as distâncias dois a dois entre os elementos.

	A	B	C	D	E	F	G	H	I	J	K	L	Eqm k=1	Eqm k=3
Lote 1	A	(B)	C	(D)	E	F	G	(H)	I	J	K	L	1/3	3/3
Lote 2	(A)	B	C	D	(E)	F	G	H	(I)	J	K	L	9/3	1/3
Lote 3	A	B	(C)	D	E	(F)	G	H	I	J	(K)	L	0/3	0/3
Lote 4	A	B	C	D	E	F	(G)	H	I	(J)	K	(L)	1/3	2/3

$$Eqm \ k=1 \Rightarrow \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{6}$$

$$Eqm \ k=3 \Rightarrow \frac{6}{3} \cdot \frac{1}{4} = \frac{1}{2}$$

PROGRAMA AVANÇADO EM DATA SCIENCE E DECISÃO



	a	b	c	d	e	f	g	h	i	j	k	l
a	0											
b	0.566	0										
c	0.437	0.146	0									
d	0.812	0.254	0.376	0								
e	0.590	0.746	0.619	0.871	0							
f	0.761	0.237	0.379	0.241	0.980	0						
g	0.138	0.444	0.329	0.696	0.639	0.626	0					
h	0.217	0.531	0.385	0.744	0.379	0.759	0.265	0				
i	0.771	0.760	0.665	0.814	0.238	0.975	0.787	0.554	0			
j	0.752	0.206	0.352	0.189	0.940	0.054	0.621	0.736	0.928	0		
k	0.218	0.430	0.284	0.650	0.440	0.656	0.207	0.103	0.581	0.634	0	
l	0.400	0.633	0.493	0.804	0.190	0.870	0.454	0.190	0.391	0.838	0.264	0