1

↳ K-nearest neighbor → K elementos mais próximos (dependente da métrica)

↳ leave-one-out schema → "fingimos" que não temos acesso ao output de X_i, iterando para todos os i's.

a) i)

Euclidean distance:

$$\|X_i - X_j\| = \sqrt{\sum_K (X_{iK} - X_{jK})^2}$$

$\|X_1 - X_2\| = \underline{1}$ B

$\|X_1 - X_3\| = \sqrt{5}$

$\|X_1 - X_4\| = \sqrt{8}$

$\|X_1 - X_5\| = \underline{\sqrt{2}}$ B

$\|X_1 - X_6\| = \underline{1}$ A

Escolha: menores valores

$$\hat{z}_1 = \text{mode}(y_{3,2}, y_{3,5}, y_{3,6})$$

$$= \text{mode}(B, A, A)$$

$$= A$$

ii)

Hamming distance : # elements in which inputs differ

$H(X_1, X_2) = \underline{1}$ B

$H(X_1, X_3) = 2$

$\hat{z}_1 = \text{mode}(B, A, A)$

$H(X_1, X_4) = 2$

$= A$

$H(X_1, X_5) = \underline{1}$ A

$H(X_1, X_6) = \underline{1}$ A

b)

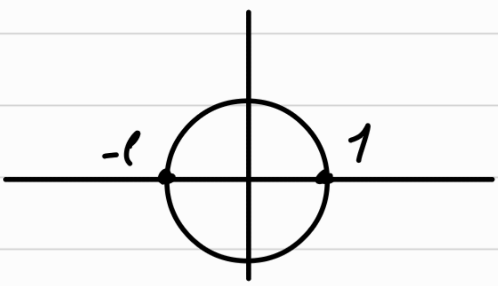
Cos similarity :

produto interno

$$\cos(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$$

$$\cos(X_1, X_2) = \frac{\overbrace{(1,1) \cdot (2,1)}^{1\times2+1\times1}}{\sqrt{2} \times \sqrt{5}} \simeq \underline{0{,}95} \quad 0{,}5$$

$\cos(X_1, X_3) \simeq \underline{0{,}98}$ 2



$\cos(X_1, X_4) = \underline{1}$ 2.2

Escolha: valores mais próximos de 1.

$\cos(X_1, X_5) = 0{,}70$

$\cos(X_1, X_6) \simeq 0{,}86$

$\hat{z}_1 = \text{mean}(0{,}5; 2; 2{,}2) = \underline{1{,}567}$

c) ① $\|X_1 - X_2\| = 1$ B 0,5

$\|X_1 - X_3\| = \sqrt{5}$ B 2

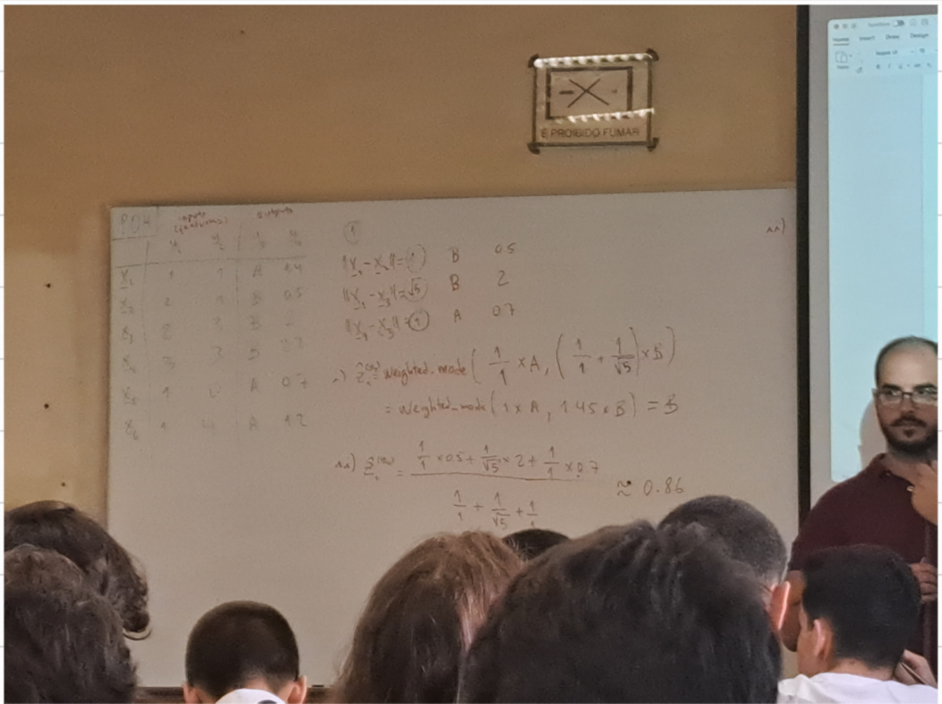
$\|X_1 - X_5\| = 1$ A 0,7

i)

$$\hat{z}_1 = \text{weighted\_mode}\left(\frac{1}{1} \times A, \left(\frac{1}{1} + \frac{1}{\sqrt{5}}\right) \times B\right)$$

$$= \text{weighted\_mode}(1 \times A; 1{,}45 \times B) = \underline{B}$$

ii)



2

a)

$$\hat{z} = \begin{pmatrix} 0,6 \\ 0 \\ 2,4 \\ 2,2 \\ 0,2 \\ 4,2 \end{pmatrix} \qquad z = \begin{pmatrix} 1,4 \\ 0,5 \\ 2 \\ 2,2 \\ 0,7 \\ 1,2 \end{pmatrix}$$

b)

$$z - \hat{z} = \begin{pmatrix} 0,8 \\ 0,5 \\ -0,4 \\ 0 \\ 0,5 \\ 3 \end{pmatrix}$$

Mean Absolute Error

$$\text{MAE} = \frac{1}{n} \sum_{k} |z_k - \hat{z}_k| = \frac{1}{6}\left(|0,8| + (\dots) + |3|\right)$$

$$\simeq 0,867$$

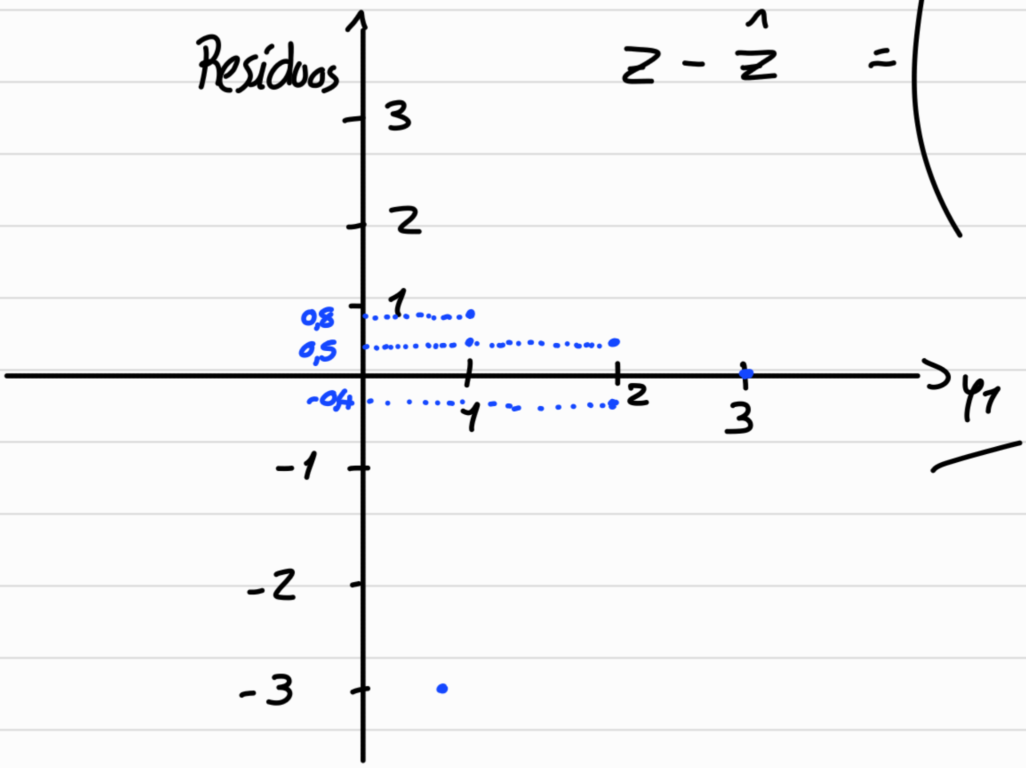
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k} (z_k - \hat{z}_k)^2}$$ "prejudica" desvios grandes

Root Mean Squared Error

$\simeq 1{,}39$

c) y_1:

$$z - \hat{z} = \begin{pmatrix} 0{,}8 \\ 0{,}5 \\ -0{,}4 \\ 0 \\ 0{,}5 \\ 3 \end{pmatrix}$$



y_2:

(semelhante)

③ Considerando o evento A positivo:

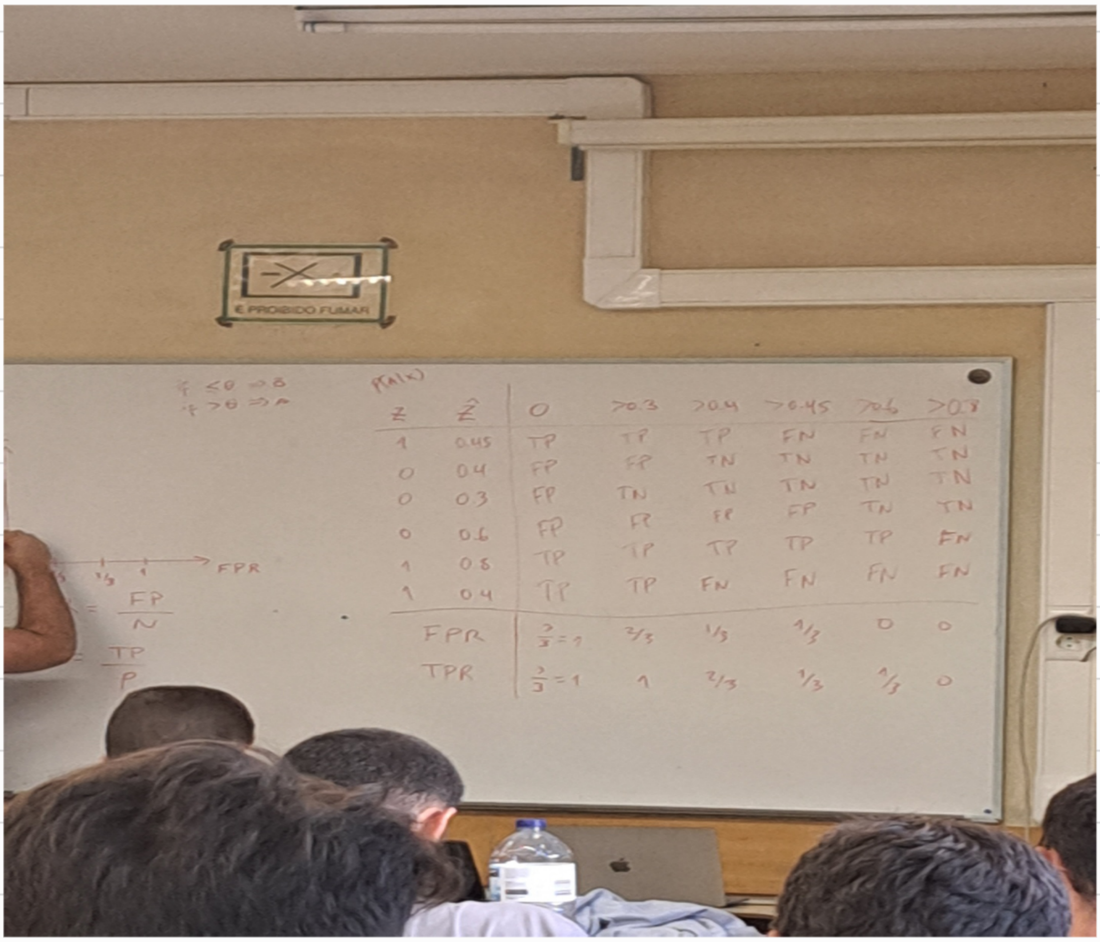
$\frac{TP}{P}$

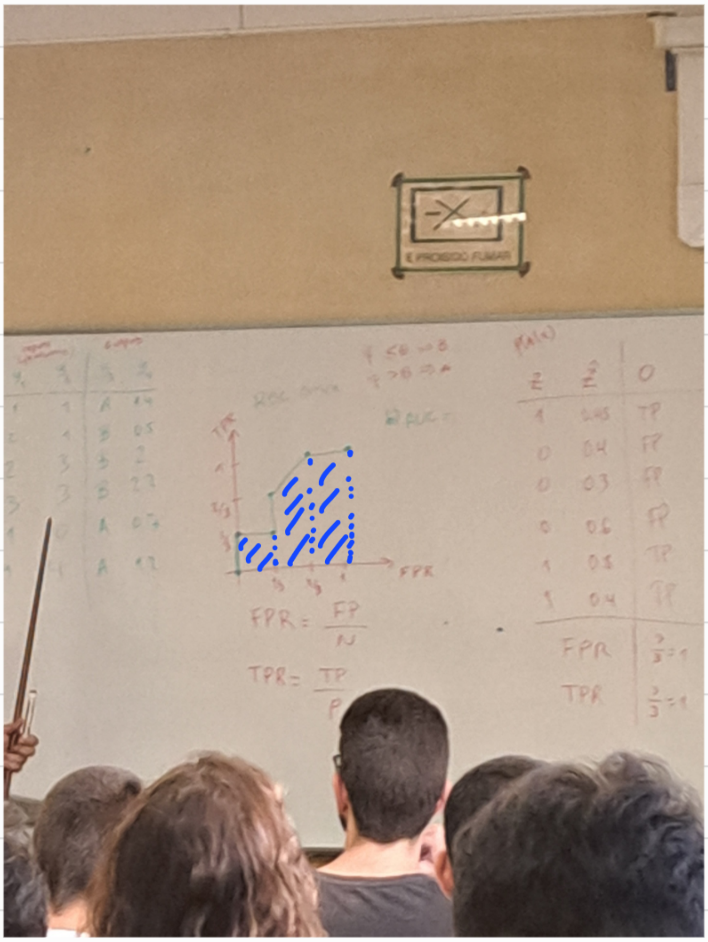
$\frac{FP}{N}$

threshold < maior que o threshold -> A (1)
menor ou igual " " -> B(0)

P(A|X)

z	$\hat{z}$	0	>0,3	>0,4	>0,45	>0,6	>0,8
1	0,45						
0	0,4						
0	0,3						
0	0,6						
1	0,8						
1	0,4						





b) AUC (igual área debaixo da linha)

$$= \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3} \times \frac{2}{3}\right) + \frac{1}{2}\left(\frac{1}{3}\right)^2 + \frac{1}{3} \times 1$$

$$\simeq \underline{0{,}72}$$