ntra exemplo:

19 $r = 1/4$; $x = (0,0)$, $x' = (0,1)$, $x''(1,1)$

$$d_{1/4}((0,0),(0,1)) = \left[(|0-0|)^{1/4} + (|0-1|)^{1/4} \right]^4 = 1$$

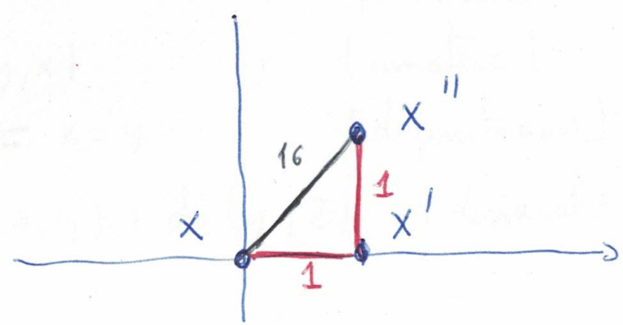
$$d_{1/4}((0,1),(1,1)) = \left[(|0-1|)^{1/4} + (|1-1|)^{1/4} \right]^4 = 1$$

$$d_{1/4}((0,0),(1,1)) = \left[(|0-1|)^{1/4} + (|0-1|)^{1/4} \right]^4 = 2^4 = 16$$

$$d_{1/4}((0,0),(1,1)) \leq d((0,0),(0,1)) + d((0,1),(1,1)) \Leftrightarrow$$

$$\Leftrightarrow 16 \leq 1 + 1$$

Falso!



↳ pode não ser boa métrica para clustering!

[DI]STÂNCIA DE MANHATTAN $(r=1)$

20 $d_1(x,x') = \sum_{i=1}^{I} |x_i - x'_i|$

Exemplo $x, x' \in \mathbb{R}^2$. Seja $x = (1,1)$ e $x' = (2,2)$

$$d_1(x,x') = \sum_{i=1}^{2} |x_i - x'_i| = |1-2| + |1-2| = 2$$

DISTÂNCIA Euclidiana $(r=2)$

$$d_2(x,x') = \sqrt{\sum_{i=1}^{I} |x_i - x'_i|^2}$$

Usando o mesmo exemplo acima, $d_2(x,x') = \sqrt{|1-2|^2 + |1-2|^2} = \sqrt{2}$

Exercício: verifique que a métrica euclidiana é uma distância.

Temos que provar para qq $x, y \in A$:

a) $d_2(x,y) \geq 0$ (positividade)
b) $d_2(x,y) = d_2(y,x)$ (simetria)
c) $d_2(x,y) = 0 \iff x = y$ (definiteness)
d) $d_2(x,z) \leq d_2(x,y) + d_2(y,z)$ (desigualdade triangular)

$\boxed{Ex2}$

a) $d_2(x,x') \geq 0$ - raiz quadrada dum nº positivo ou zero é sempre positivo ou zero.

$x = x' \longrightarrow d_2(x,x) = \sqrt{|x_1-x'_1|^2 + |x_2-x'_2|^2 + \cdots + |x_I-x'_I|^2} = 0$

$x \neq x' \longrightarrow d_2(x,y) =$ raiz quadrada da soma de nºs positivos ou zero.

b) $d_2(x,x') = d_2(x',x)$

$$\sqrt{(x_1-x'_1)^2 + (x_2-x'_2)^2 + \cdots + (x_I-x'_I)^2} = \sqrt{(x'_1-x_1)^2 + (x'_2-x_2)^2 + \cdots + (x'_I-x_I)^2}$$

facilmente se vê que é verdade pq cada parcela debaixo da raiz quadrada (na mesma posição) é numericamente igual, i.e $(x_i-x'_i)^2 = (x'_i-x_i)^2$.

c) $d_2(x,x') = 0 \iff x = x'$

- $\sqrt{(x_1-x'_1)^2 + (x_2-x'_2)^2 + \cdots + (x_I-x'_I)^2} = 0 \iff (x_1-x'_1)^2 + (x_2-x'_2)^2 + \cdots + (x_I-x'_I)^2 = 0.$

Vemos que a igualdade anterior só pode ser verdadeira se $x_1 = x'_1$ e $x_2 = x'_2$ e ... e $x_I = x'_I$, ou seja $x = x'$.

- Se $x = x' \longrightarrow d_2(x,x') = 0$

$x = x' \iff x_1 = x'_1 \wedge x_2 = x'_2 \wedge \cdots \wedge x_I = x'_I$.

Então, $d(x,x') = \sqrt{(x_1-x_1)^2 + (x_2-x_2)^2 + \cdots + (x_I-x_I)^2} = 0.$

d) falta provar desigualdade triangular! Fica como desafio!

$$d(x,x') \leq d(x,x'') + d(x'',x')$$

SEMELHANÇA DE ROOK

$\mathcal{X} = \mathbb{R}^2$

$x = (x_1, x_2)$

$\boxed{23}$ $$d_{rook} = \begin{cases} |x_2 - x'_2| , & x_1 = x'_1 \\ |x_1 - x'_1| , & x_2 = x'_2 \\ +\infty , & \text{noutros casos} \end{cases}$$

Ex: Verificar propriedades da métrica de Rook com $x \in \mathbb{R}^2$.

- $d(x, x') \geq 0$

  $x_1 = x'_1 \longrightarrow d = |x_2 - x'_2| \geq 0$

  $x_2 = x'_2 \longrightarrow d = |x_1 - x'_1| \geq 0$

  $x_1 \neq x'_1 \wedge x_2 \neq x'_2 \rightarrow d = +\infty \geq 0$

- $d(x, x) = 0$

  $x_1 = x'_1 \wedge x_2 = x'_2 \longrightarrow d(x, x) = 0$

- Simetria: $d(x, x') = d(x', x)$

  $$d(x, x') = \begin{cases} |x_2 - x'_2| , & x_1 = x'_1 \\ |x_1 - x'_1| , & x_2 = x'_2 \\ +\infty , & \text{noutros casos} \end{cases} \qquad d(x', x) = \begin{cases} |x'_2 - x_2| , & x'_{\cancel{2}} = x_1 \\ |x'_1 - x_1| , & x'_2 = x_2 \\ +\infty , & \text{noutros casos} \end{cases}$$

As conds das 2 funções são iguais, assim como os outputs por causa do módulo $\therefore$ d é simétrica!

$d(x,x')=0 \iff x=x'$ (definiteness)

$$d(x,x') = \begin{cases} |x_2-x'_2|, & x_1=x'_1 \\ |x_1-x'_1|, & x_2=x'_2 \\ +\infty, & \text{noutros casos} \end{cases}$$

d só pode ser 0 se

- $x_1=x'_1 \longrightarrow d(x,x')=|x_2-x'_2|=0$ então $x_2=x'_2$ Logo $x=x'$
- $x_2=x'_2 \longrightarrow d(x,x')=|x_1-x'_1|=0$ então $x_1=x'_1$ logo $x=x'$

- não verifica desigualdade triangular

$x=(0,0)$ $\quad d(x,x'')=+\infty$

$x'=(0,1)$ $\quad d(x,x')=1$

$x''=(1,1)$ $\quad d(x',x'')=1$

$$\therefore \left.\begin{aligned} d(x,x'') &\leq d(x,x')+d(x',x'') \\ \infty &\leq 1+1 \end{aligned}\right\} \text{Falso!}$$

$\boxed{d \text{ não é adequado para clustering}}$

$\overline{5\frac{1}{3}}$

Ex24

a) $$\bar{m} = \frac{1}{5}\begin{pmatrix} 0+2+2+1-2 \\ 0+1+4-2+4 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 0.6 \\ 1.4 \end{pmatrix}$$

b) $$\bar{m} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

ordenando as 1as componentes temos $\{-2, 0, 1, 2, 2\}$ $\quad N = 5$

" as 2a componentes temos $\{-2, 0, 1, 4, 4\}$

- O representante resulta de calcular a mediana de cada conjunto.

Neste caso, $\overline{m_1} = 1$ e $\overline{m_2} = 1$.

- Notar que $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin C$ pelo que se trata dum representante do tipo centroide.

- Se pretendêssemos determinar um representante do tipo medoide (que pertence a C) teríamos que resolver um problema de optimização com restrições. Pesado computacionalmente!
Um compromisso seria ver qual o ponto de C mais próximo de $\bar{m}$

EXE Assim, vamos calcular todas as distâncias entre 2 pontos de C, usando a métrica

$$d(x,x') = \sum_{i=1}^{I} |x_i - x'_i| .$$

$$d(x^1,x^2) = 2+1 = 3$$
$$d(x^1,x^3) = |0-2| + |0-4| = 6$$
$$d(x^1,x^4) = |0-1| + |0-(-2)| = 3$$
$$d(x^1,x^5) = |0-(-3)| + |0-1| = 4$$

Se o representante for x^1, a função custo é:

$$f(x^1) = \sum_{m=1}^{N} d(x^1, x^m) = 3+6+3+4 = 16$$

x x

$$d(x^2,x^3) = |2-2| + |1-4| = 3$$
$$d(x^2,x^4) = |2-1| + |1+2| = 4$$
$$d(x^2,x^5) = |2+3| + |1-1| = 5$$

- Se o representante for x^2, a função custo vale:

$$f(x^2) = \sum_{m=1}^{N} d(x^2,x^m) = 3 + 3+4+5 = 15$$
$$= d(x^2,x^1) + d(x^2,x^3) + d(x^2,x^4) + d(x^2,x^5)$$

x x

$$d(x^3,x^5) = |2+3| + |4-1| = 8$$
$$d(x^3,x^4) = |2-1| + |4+2| = 7$$

- Se o representante for x^3, a função custo vale

$$f(x^3) = d(x^3,x^1) + d(x^3,x^2) + d(x^3,x^4) + d(x^3,x^5) =$$
$$= 6+3+8+7 = 24$$

x x

$$d(x^4,x^5) = |1+3| + |-2-1| = 7$$

- Se o representante for x^4, a função custo vale

$$f(x^4) = d(x^4,x^1) + d(x^4,x^2) + d(x^4,x^3) + d(x^4,x^5)$$
$$= 3+4+7+7 = 21$$

- Se o representante for x^5, a função custo vale:

$$f(x^5) = d(x^5,x^1) + d(x^5,x^2) + d(x^5,x^3) + d(x^5,x^4) =$$

Então, o valor da funcional é mínimo quando o representante $m = x^2$!

$\bar{m} = (2)$