Exemplo: $D = \{ x^1 = \binom{0}{0}, x^2 = \binom{1/2}{0}, x^3 = \binom{2}{1}, x^4 \binom{2}{2.5}, x^5 = \binom{0}{3} \}$

$\mathcal{A} = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$

$d =$ Manhatten

Single linkage

distância mínima entre os dos 2 clusters.

$$\mathcal{P}_0 = \{ \underset{C_1}{\{x^1\}}, \underset{C_2}{\{x^2\}}, \underset{C_3}{\{x^3\}}, \dots, \underset{C_N}{\{x^N\}} \}$$

$T_0 \rightarrow$ MATRIZ DE LINKAGE

		C_1 $\{x^1\}$	C_2 $\{x^2\}$	C_3 $\{x^3\}$	C_4 $\{x^4\}$	C_5 $\{x^5\}$
C_1	$\{x^1\}$	0	1/2	3	4.5	3
C_2	$\{x^2\}$	—	0	2.5	4	3.5
C_3	$\{x^3\}$	—	—	0	1.5	4
C_4	$\{x^4\}$	—	—	—	0	2.5
C_5	$\{x^5\}$	—	—	—	—	0

→ só é necessário calcular elementos acima da diagonal já que $d(C^i, C^k) = dd(C^k, C^i)$, $i \neq k$

Pretende-se encontrar os 2 subconjuntos de $\mathcal{P}(0)$ com linkage mínimo, ou seja

$$\boxed{\underset{\substack{c, c' \in \mathcal{P}(0) \\ c \neq c'}}{\operatorname{argmin}} \; dd(c, c')}$$

dd – single linkage

onde a métrica usada em d é a de Manhattan.

$$\boxed{dd(C^1, C^2) = |x_1^1 - x_1^2| + |x_2^1 - x_2^2| = |0 - 1/2| + |0 - 0| = 1/2}$$
$$dd(C^1, C^3) = |x_1^1 - x_1^3| + |x_2^1 - x_2^3| = |0 - 2| + |0 - 1| = 3$$
$$dd(C^1, C^4) = |x_1^1 - x_1^4| + |x_2^1 - x_2^4| = |0 - 2| + |0 - 2.5| = 4.5$$
$$dd(C^1, C^5) = |x_1^1 - x_1^5| + |x_2^1 - x_2^5| = |0 - 0| + |0 - 3| = 3$$

$dd(C^2, C^3) = |x_1^2 - x_1^3| + |x_2^2 - x_2^3| = |1/2 - 2| + |0 - 1| = 2.5$

$dd(C^2, C^4) = |x_1^2 - x_1^4| + |x_2^2 - x_2^4| = |1/2 - 2| + |0 - 2.5| = 4$

$dd(C^2, C^5) = |x_1^2 - x_1^5| + |x_2^2 - x_2^5| = |1/2 - 0| + |0 - 3| = 3.5$

$dd(C^3, C^4) = |x_1^3 - x_1^4| + |x_2^3 - x_2^4| = |2 - 2| + |1 - 2.5| = 1.5$

$dd(C^3, C^5) = |x_1^3 - x_1^5| + |x_2^3 - x_2^5| = |2 - 0| + |1 - 3| = 4$

$dd(C^4, C^5) = |x_1^4 - x_1^5| + |x_2^4 - x_2^5| = |2 - 0| + |2.5 - 3| = 2.5$

Calculamos o mínimo valor de linkage, que é $dd(C^1, C^2) = 1/2$.
Vamos então fazer o merge dos 2 conjuntos C^1 e C^2.

$$P(1) = \{\overbrace{\{x^1, x^2\}}^{C_1}, \overbrace{\{x^3\}}^{C_2}, \overbrace{\{x^4\}}^{C_3}, \overbrace{\{x^5\}}^{C_4}\}$$

$e(1) = 1/2$

↓
energia fusão

Voltamos a ver o linkage mínimo entre conjuntos de $P(1)$ usando single linkage.

Π – matriz de linkage

	C_1	C_2	C_3	C_4
C_1	0	2.5	4	3
C_2	–	0	1.5	4
C_3	–	–	0	2.5
C_4	–	–	–	0

$dd(C_1, C_2) = \min(d(x^1, x^3), d(x^2, x^3)) = \min(3, 2.5) = 2.5$

$d(x^1, x^3) = |x_1^1 - x_1^3| + |x_2^1 - x_2^3| = |0 - 2| + |0 - 1| = 3$

$d(x^2, x^3) = |x_1^2 - x_1^3| + |x_2^2 - x_2^3| = |1/2 - 2| + |0 - 1| = 2.5$

$dd(C_1, C_3) = \min(d(x^1, x^4), d(x^2, x^4)) = \min(4.5, 4) = 4$

$d(x^1, x^4) = |0-2| + |0-2.5| = 4.5$

$d(x^2, x^4) = |1/2-2| + |0-2.5| = 4$

$dd(C_1, C_4) = \min(d(x^1, x^5), d(x^2, x^5)) = \min(3, 3.5) = 3$

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

$d(x^2, x^5) = |1/2-0| + |0-3| = 3.5$

$dd(C_2, C_3) = d(x^3, x^4) = |2-2| + |1-2.5| = 1.5$

$dd(C_2, C_4) = d(x^3, x^5) = |2-0| + |1-3| = 4$

$dd(C_3, C_4) = d(x^4, x^5) = |2-0| + |2.5-3| = 2.5$

Calculamos o mínimo que é $dd(C_2, C_3)$.
Vamos fazer o merge de C_2 com C_3 obtendo uma nova partição $P(2)$.

$$P(2) = \{ \underset{C_1}{\{x^1, x^2\}}, \underset{C_2}{\{x^3, x^4\}}, \underset{C_3}{\{x^5\}} \}$$

$e(2) = 1.5$ (energia de fusão)

Como $P(2)$ ainda tem mais do que 1 elemento $- (|P(2)| = 3)$, repetimos o processo.

T_2 – Matriz de linkage

	C_1	C_2	C_3
C_1	0	2.5	3
C_2	–	0	2.5
C_3	–	–	0

85

Notar que $C^1 = \{x^1, x^2\}$, $C^2 = \{x^3, x^4\}$, $C^3 = \{x^5\}$

$$dd(C^1, C^2) = \min\left(d(x^1,x^3), d(x^1,x^4), d(x^2,x^3), d(x^2,x^4)\right)$$

$$dd(C^1,C^2) = 2.5$$

$$d(x^1,x^3) = |x_1^1 - x_1^3| + |x_2^1 - x_2^3| = |0-2| + |0-1| = 3$$
$$d(x^1,x^4) = |x_1^1 - x_1^4| + |x_2^1 - x_2^4| = |0-2| + |0-2.5| = 4.5$$
$$d(x^2,x^3) = |x_1^2 - x_1^3| + |x_2^2 - x_2^3| = |1/2-2| + |0-1| = 2.5$$
$$d(x^2,x^4) = |x_1^2 - x_1^4| + |x_2^2 - x_2^4| = |1/2-2| + |0-2.5| = 4$$

$$dd(C^1, C^3) = \min\left(d(x^1,x^5), d(x^2,x^5)\right) = 3$$
$$d(x^1,x^5) = |x_1^1 - x_1^5| + |x_2^1 - x_2^5| = |0-0| + |0-3| = 3$$
$$d(x^2,x^5) = |x_1^2 - x_1^5| + |x_2^2 - x_2^5| = |1/2-0| + |0-3| = 3.5$$

$$dd(C^2, C^3) = \min\left(d(x^3,x^5), d(x^4,x^5)\right)$$
$$d(x^3,x^5) = |x_1^3 - x_1^5| + |x_2^3 - x_2^5| = |2-0| + |1-3| = 4$$
$$d(x^4,x^5) = |x_1^4 - x_1^5| + |x_2^4 - x_2^5| = |2-0| + |2.5-3| = 2.5$$

O mínimo de todas estas métricas ocorre entre C^1, C^2 e C^2, C^3 e vale 2.5. Podíamos fazer qq destas fusões. Vamos escolher fundir C^2 e C^3.

$$P(3) = \{\underbrace{\{x^1, x^2\}}_{C_1}, \underbrace{\{x^3, x^4, x^5\}}_{C_2}\}$$

$e(3) = 2.5$ (energia de fusão)

Falta apenas juntar as classes C_1 e C_2.
Para saber a energia de fusão e_4 temos que determinar o [illegible] mínimo entre elementos de C_1 e de C_2

$C_1 = \{x^1, x^2\}$; $C_2 = \{x^3, x^4, x^5\}$

$$dd(C_1, C_2) = \min\left(d(x^1,x^3), d(x^1,x^4), d(x^1,x^5), d(x^2,x^3), d(x^2,x^4), d(x^2,x^5)\right)$$

já tínhamos feito estes cálculos atrás.

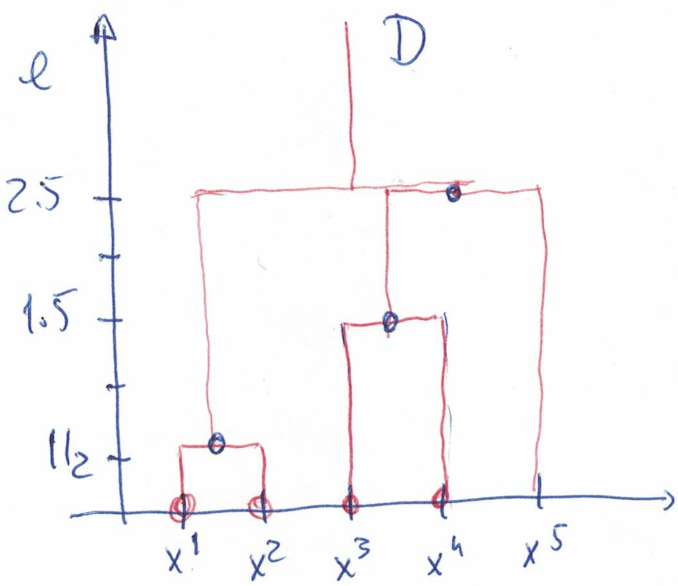
$$dd(C_1, C_2) = \min(3, 4.5, 3, 2.5, 4, 3.5) = 2.5$$

$\therefore$

$$P(4) = \{x^1, x^2, x^3, x^4, x^5\}$$

$e(4) = 2.5$ (energia de fusão)

Podemos construir o dendograma.



O objectivo não é apresentar $P(4)$ como solução do problema, aí estamos a agrupar todos os dados.

Um critério de paragem do procedimento descrito poderá ser especificar o nº de subconjuntos a ter em P ou por exemplo parar o procedimento qd $e >$ valor especificado.

O gráfico da energia de fusão é: