

Shluková hierarchická analýza – příklad

Bylo provedeno měření objemu hipokampu a mozkových komor (v cm^3) u 5 pacientů se schizofrenií. Naměřené hodnoty objemu hipokampu a mozkových komor byly zaznamenány do matice :

Určete podobnost pěti pacientů na základě naměřených charakteristik pomocí hierarchické shlukové analýzy, použijte metodu nejbližšího a nejvzdálenějšího souseda.

Řešení:

Shluková analýza se snaží o identifikaci shluků objektů ve vícerozměrném prostoru a následnou kategorizaci objektů do zjištěných shluků. Vzájemnou pozici objektů ve vícerozměrném prostoru lze popsat jejich vzdáleností. Dle vzdálenosti objektů je následně možné objekty slučovat do shluků.

Nejdříve je tedy potřeba matici převést na matici vzdálenosti. V tomto případě použijeme Euklidovskou vzdálenost, jako vhodnou míru k vyjádření vzdáleností mezi jednotlivými objekty. Euklidovská vzdálenost je dána vztahem:

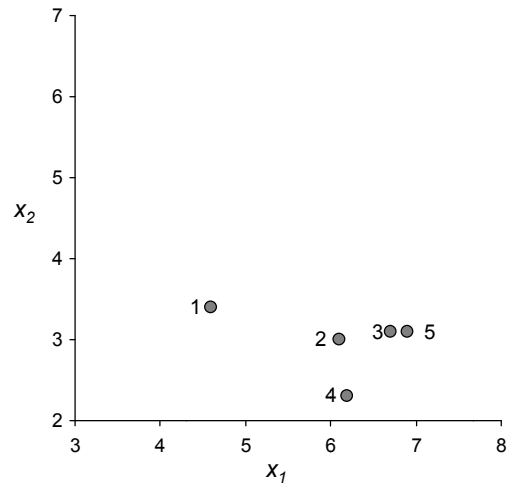
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Po dosazení hodnot do vzorce získáme asociační matici obsaženou v tabulce 1.

Tabulka 1 Matice vzdáleností mezi objekty založena na Euklidovské vzdálenosti.

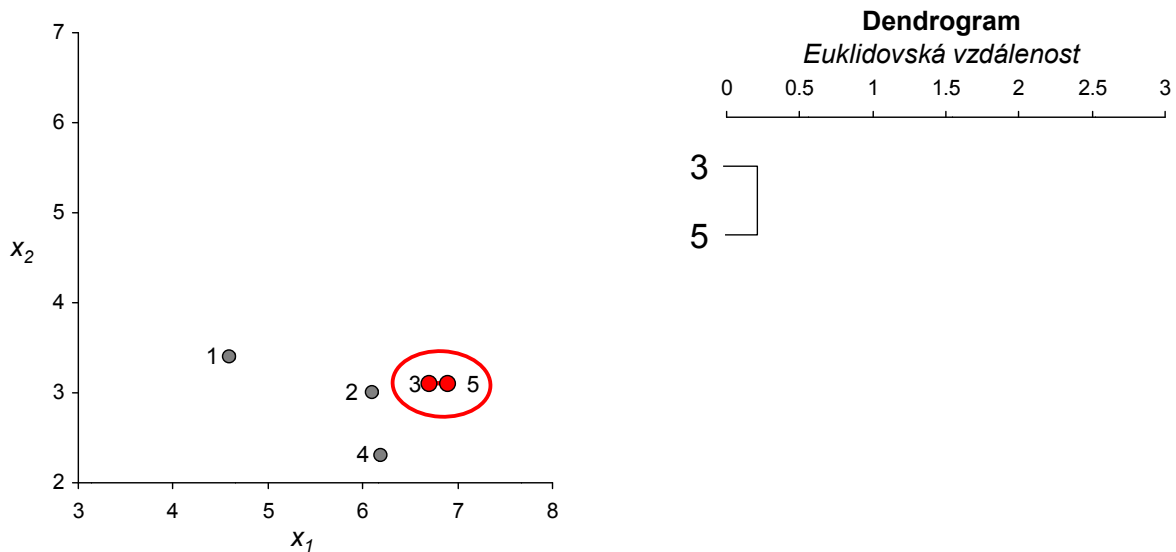
	1	2	3	4	5
1	0,0	1,6	2,1	1,9	2,3
2	1,6	0,0	0,6	0,7	0,8
3	2,1	0,6	0,0	0,9	0,2
4	1,9	0,7	0,9	0,0	1,1
5	2,3	0,8	0,2	1,1	0,0

Pro snadnější představu postupu výpočtu si jednotlivé objekty vykreslíme do jednoduchého xy grafu.



Obrázek 1 Vykreslení jednotlivých objektů v xy grafu

Nyní přistoupíme k výpočtu shlukové analýzy. Nejdříve si ukážeme postup dle metody nejbližšího souseda. V prvním kroku nalezneme v asociační matici dva objekty, které mají mezi sebou nejmenší vzdálenost a definujeme tak počáteční shluk. V našem případě to budou objekty 3 a 5, vzdálenost mezi nimi je 0,2 (Obrázek 2).



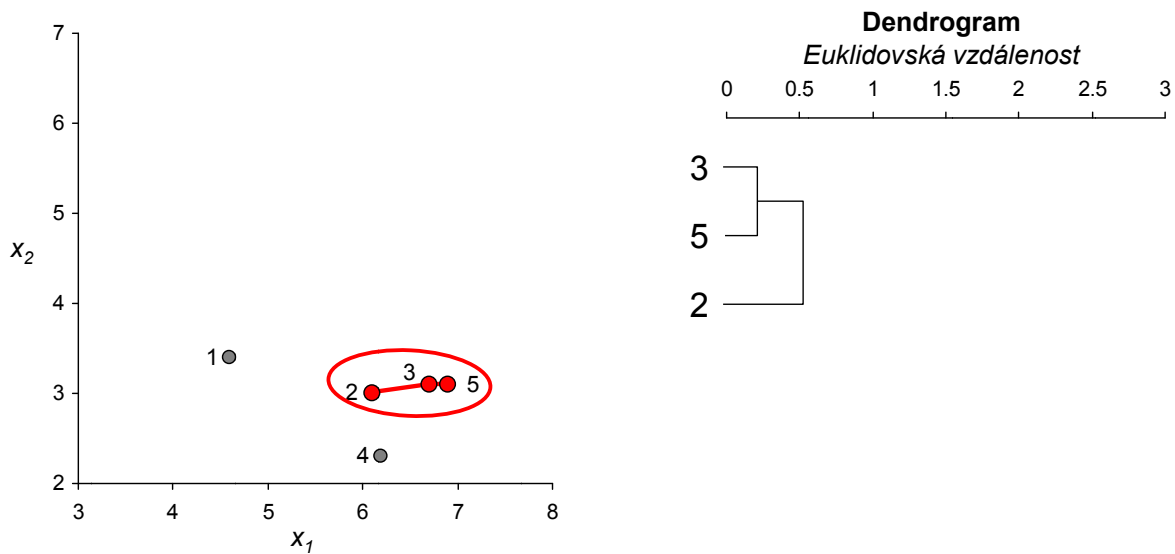
Obrázek 2. První krok shlukové analýzy. Sloučení dvou nejbližších objektů, definován počáteční shluk.

Následně agregujeme řádky asociační matice pro objekty 3 a 5. V přepočítané matici objekty 3 a 5 vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána nejmenší vzdáleností od jeho členů (3, 5).

Tabulka 2 Asociační matice s agregovaným řádkem pro objekty 3 a 5, kde sloučené objekty vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána nejmenší vzdáleností od jeho členů.

	1	2	3 - 5	4
1	0,0	1,6	2,1	1,9
2	1,6	0,0	0,6	0,7
3 - 5	2,1	0,6	0,0	0,9
4	1,9	0,7	0,9	0,0

V dalším kroku postupujeme analogicky ke kroku předchozímu. Opět vybereme dva objekty, které jsou si nejbližší, a sloučíme je. V tomto kroku se jedná o objekt 2 s již vytvořenou dvojicí 3 a 5 z kroku předchozího, definujeme tak shluk 2-(3-5) na vzdálenosti 0,6 (Obrázek 3).



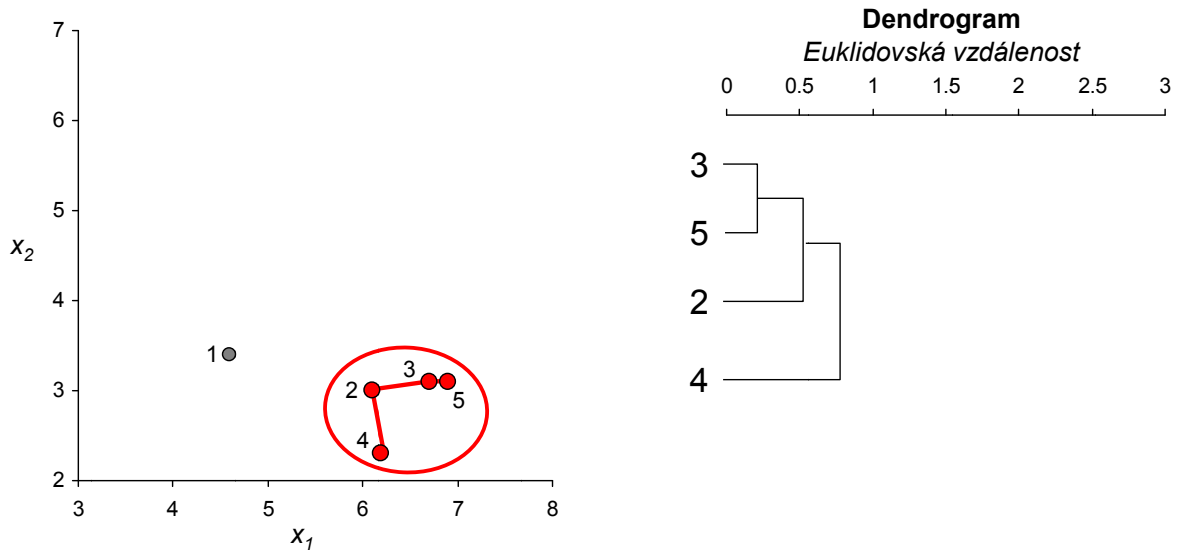
Obrázek 3 Následující krok shlukové analýzy, kde došlo ke sloučení objektu 2 s již vytvořenou dvojicí z předchozího kroku a byl tak definován shluk 2-(3 - 5).

A nyní opět agregujeme příslušné řádky asociační matice (Tabulka 3).

Tabulka 3 Asociační matice se sloučeným řádkem pro objekty 2 – (3 – 5), kde sloučené objekty vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána nejmenší vzdáleností od jeho členů.

	1	2-3-5	4
1	0,0	1,6	1,9
2-3-5	1,6	0,0	0,7
4	1,9	0,7	0,0

Na základě nové asociační matice, nyní definujeme shluk 4-(2-(3-5)), který vznikl na vzdálenosti 0,7 (Obrázek 4).

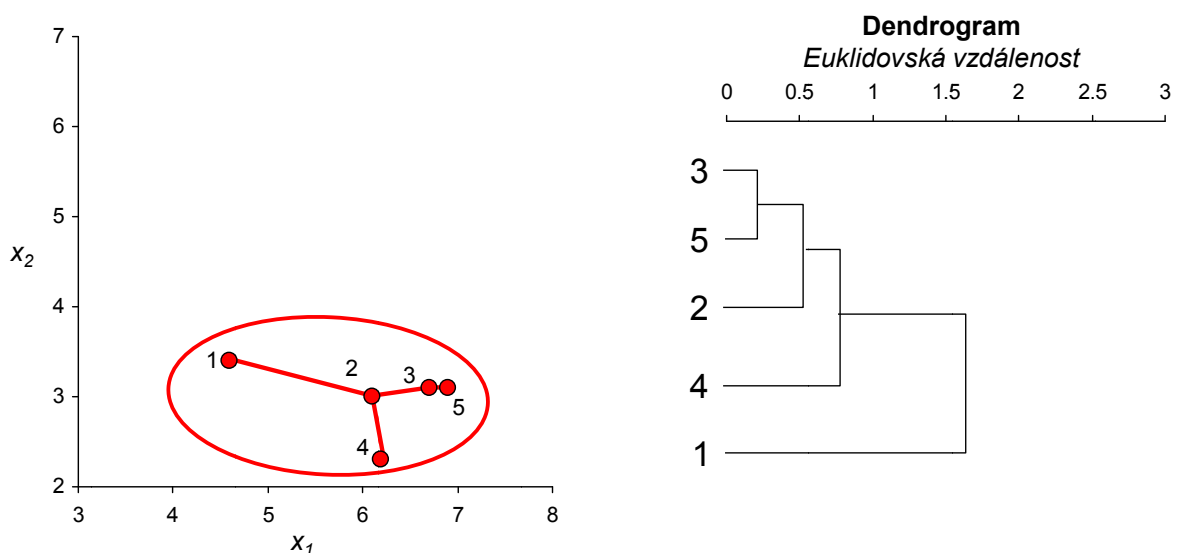


Obrázek 4 Definice shluku 4-(2-(3-5)), který vznikl na vzdálenosti 0,7

Tabulka 4 Asociační matice se sloučeným řádkem pro objekty 4 –(2 – (3 – 5)), kde sloučené objekty vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána nejmenší vzdáleností od jeho členů

	1	4-2-3-5
1	0,0	1,6
4-2-3-5	1,6	0,0

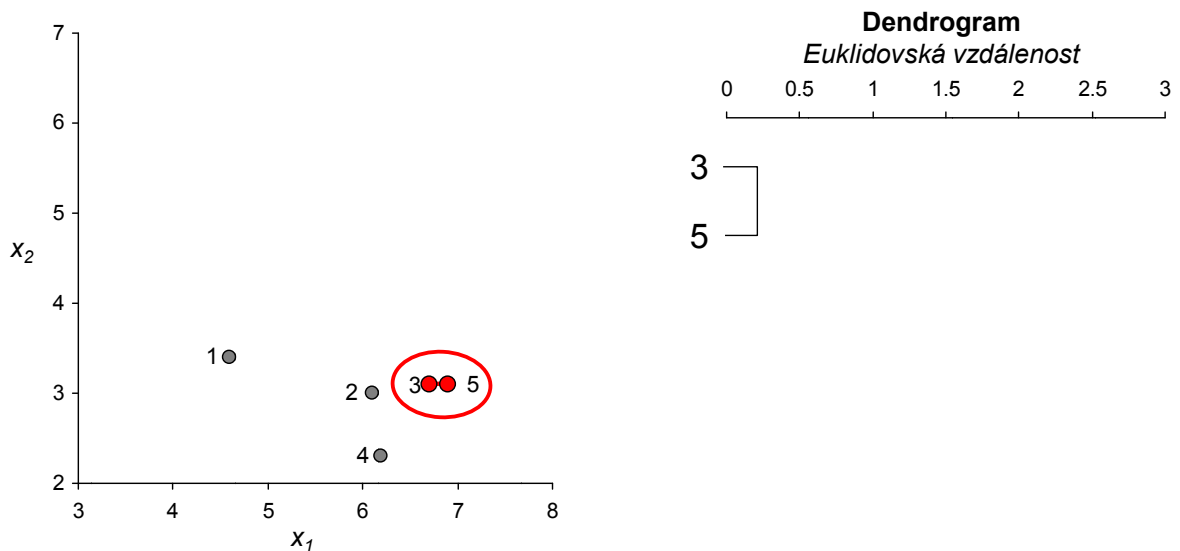
Nyní spojíme poslední objekt 1 s ostatními objekty, definujeme tak shluk 1 – (4 –(2 – (3 – 5))), který nyní zahrnuje všechny sledované objekty, tímto je algoritmus ukončen (Obrázek 5).



Obrázek 5 Definice shluku 1-(4-(2-(3-5))), který vznikl na vzdálenosti 1,6. Nyní jsou všechny objekty sloučeny, algoritmus je ukončen.

Analogicky budeme postupovat v případě metody nejvzdálenějšího souseda, jen s tím rozdílem, že při přepočtu asociační matice, kde právě sloučené objekty budou vystupovat jako jeden objekt, použijeme vzdálenost od ostatních objektů, která je dána největší vzdáleností od jeho členů.

Při výpočtu shlukové analýzy dle nejvzdálenějšího souseda vyházíme z původní asociační matice (Tabulka 1). První krok shlukování podle nejvzdálenějšího souseda je stejný jako v případě nejbližšího souseda: nalezneme v asociační matici dva objekty, které mají mezi sebou nejmenší vzdálenost a definujeme tak počáteční shluk. Opět to budou objekty 3 a 5 na vzdálenosti 0,2 (Obrázek 6).



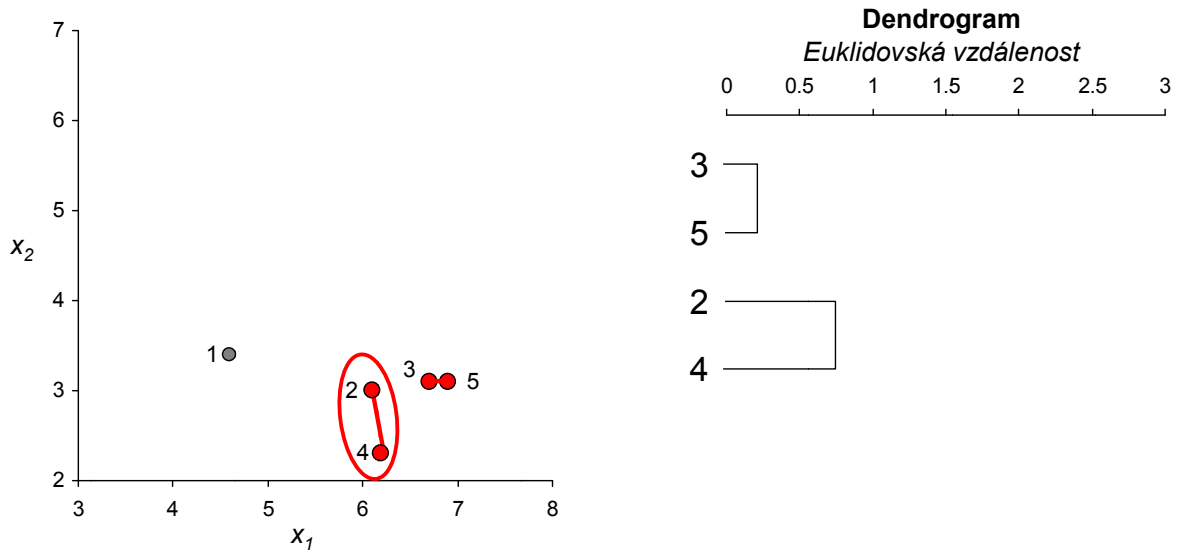
Obrázek 6. První krok shlukové analýzy dle metody nejvzdálenějšího souseda. Sloučení dvou nejbližších objektů, definován počáteční shluk.

Opět agregujeme řádky asociační matice pro objekty 3 a 5. A zde v případě shlukování pole metody nejvzdálenějšího souseda, v přepočítané matici, kde objekty 3 a 5 vystupují jako jeden objekt, je jeho vzdálenost od ostatních objektů dána právě největší vzdáleností od jeho členů (3, 5) (Tabulka 5).

Tabulka 5 Asociační matice s agregovaným řádkem pro objekty 3 a 5, kde sloučené objekty vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána největší vzdáleností od jeho členů.

	1	2	4	3 - 5
1	0,0	1,6	1,9	2,3
2	1,6	0,0	0,7	0,8
4	1,9	0,7	0,0	1,1
3 - 5	2,3	0,8	1,1	0,0

V dalším kroku postupujeme analogicky ke kroku předchozímu. Opět vybereme dva objekty, které jsou si nejbližší, a sloučíme je. V tomto kroku se jedná o objekt 2 a 4, definujeme tak shluk 2 – 4 na vzdálenosti 0,7 (Obrázek 7).



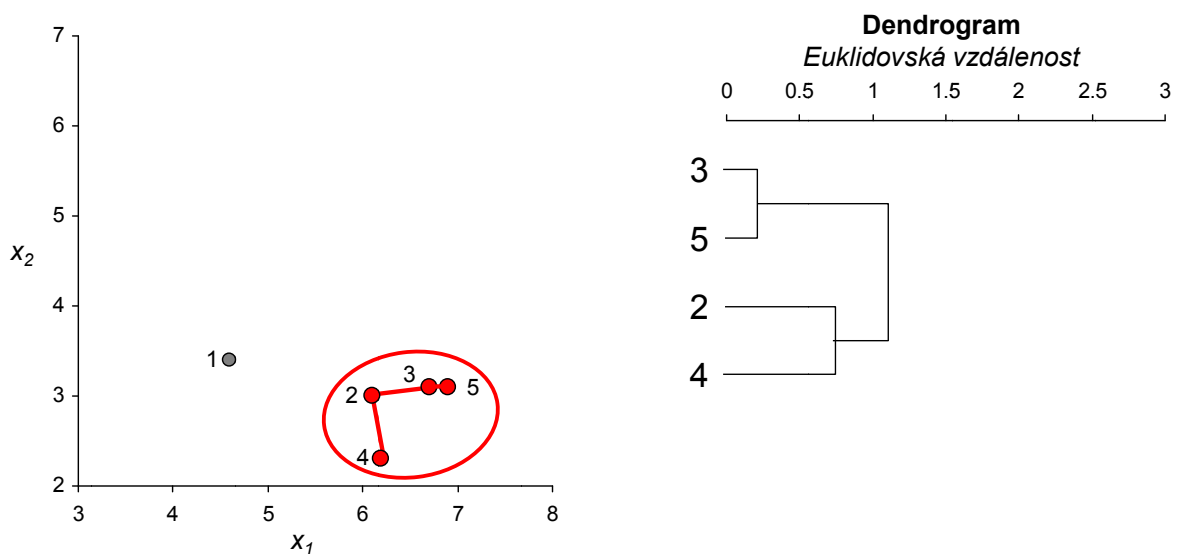
Obrázek 7 Následující krok shlukové analýzy, kde došlo ke sloučení objektu 2 a 4, byl tak definován shluk 2 - 4.

A nyní opět agregujeme příslušné řádky asociační matice (Tabulka 6).

Tabulka 6 Asociační matice se sloučeným řádkem pro objekty 2 – 4, kde sloučené objekty vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána největší vzdáleností od jeho členů.

	1	2 - 4	3 - 5
1	0,0	1,9	2,3
2 - 4	1,9	0,0	1,1
3 - 5	2,3	1,1	0,0

Na základě nové asociační matice nyní definujeme shluk (2- 4) - (3-5), který vznikl na vzdálenosti 1,1 (Obrázek 8).

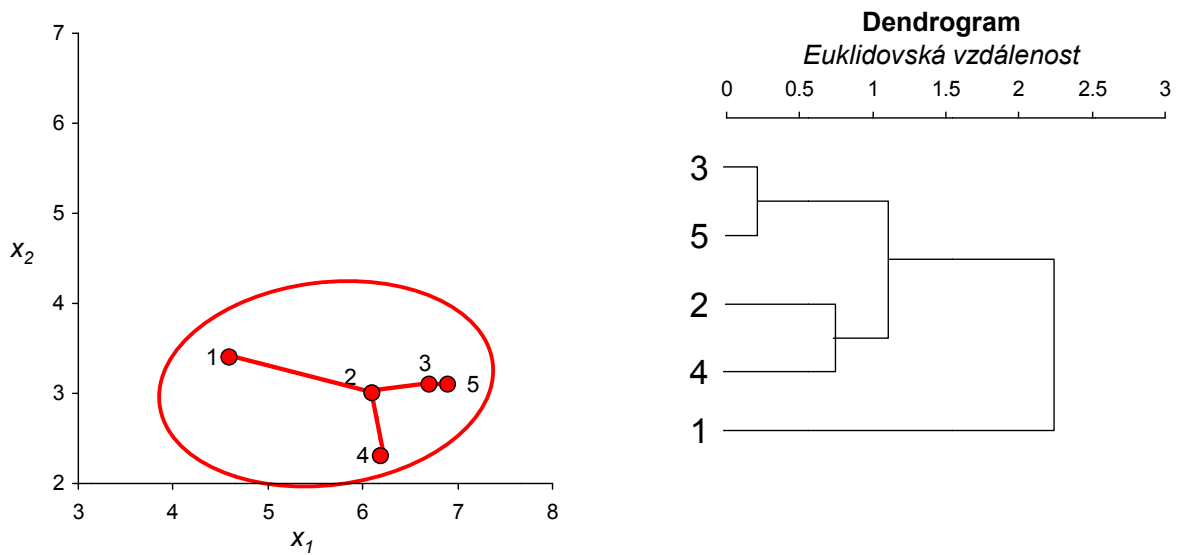


Obrázek 8 Definice shluku (4-2) - (3-5), který vznikl na vzdálenosti 1,1.

Tabulka 7 Asociační matice se sloučeným řádkem pro objekty (2- 4) - (3-5), kde sloučené objekty vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána největší vzdáleností od jeho členů.

	1	4-2-3-5
1	0,0	2,3
4-2-3-5	2,3	0,0

Nyní spojíme poslední objekt 1 s ostatními objekty, definujeme tak shluk 1 - (2- 4) - (3-5), který nyní zahrnuje všechny sledované objekty, tímto je algoritmus ukončen (Obrázek 9).



Obrázek 9 Definice shluku 1-(2- 4) - (3-5), který vznikl na vzdálenosti 2,3. Nyní jsou všechny objekty sloučeny, algoritmus je ukončen.