

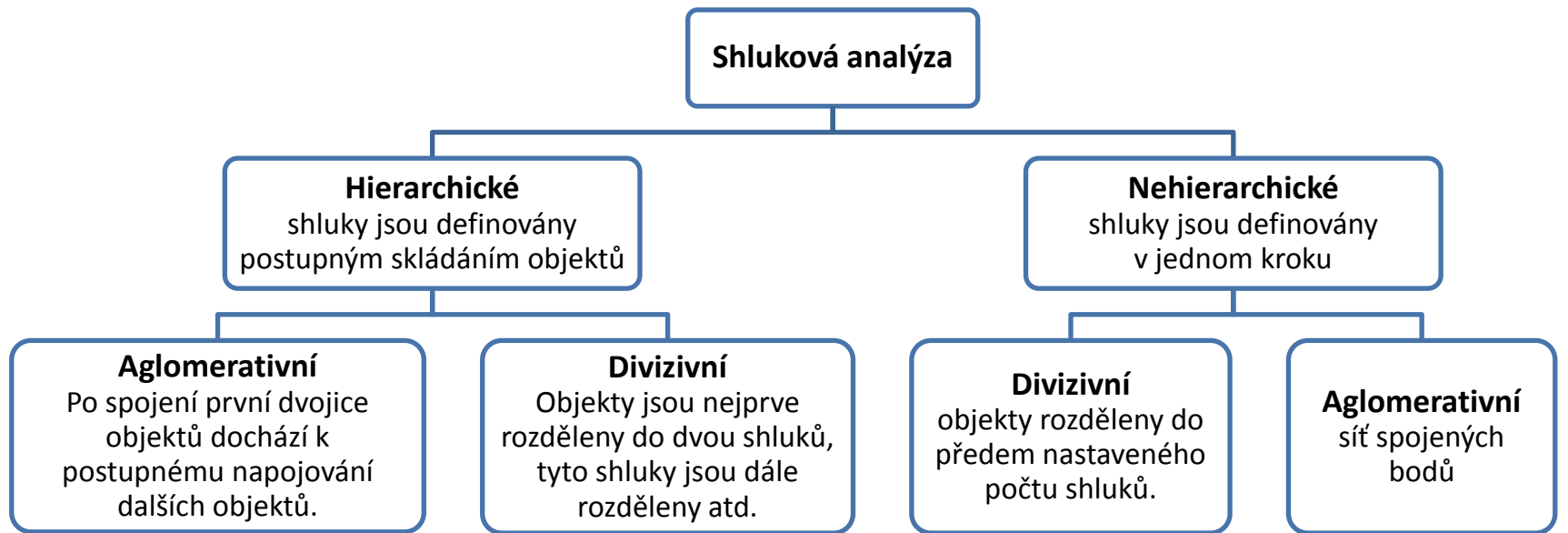
# Vícerozměrné metody - cvičení



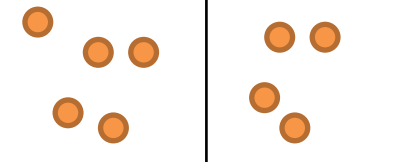
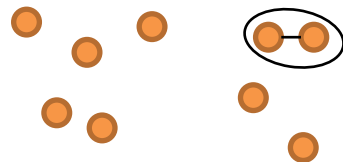
RNDr. Eva Janoušová

Podzim 2015

# Shluková analýza – typy metod – opakování



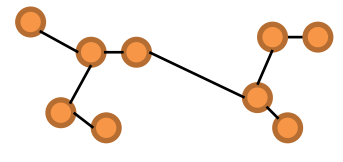
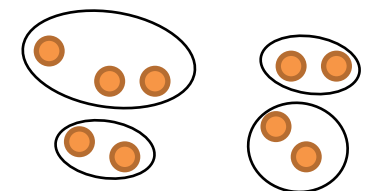
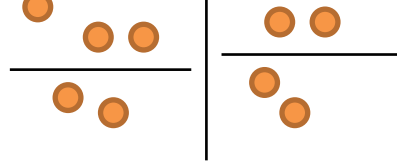
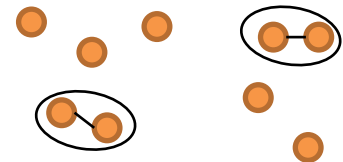
1. Krok



Kolik shluků chceme  
definovat? Například 4

Minimum spanning  
tree, Prime network

2. Krok



X. Krok

Atd.

Atd.

Výpočet ukončen

Výpočet ukončen

# Příklad 1

V experimentu byla u 5 buněčných linií zjišťována kvantita membránových markerů popisujících jejich citlivost k chemoterapii. V přiložené tabulce naleznete změřené hodnoty standardizované na referenční buněčnou linii.

Buněčná linie	Marker 1	Marker 2
A	2	4
B	2	8
C	6	10
D	10	14
E	11	13

Vztahy mezi liniemi jsou vyjádřeny následující asociační maticí:

	A	B	C	D	E
A	0.0	4.0	7.2	12.8	12.7
B	4.0	0.0	4.5	10.0	10.3
C	7.2	4.5	0.0	5.7	5.8
D	12.8	10.0	5.7	0.0	1.4
E	12.7	10.3	5.8	1.4	0.0

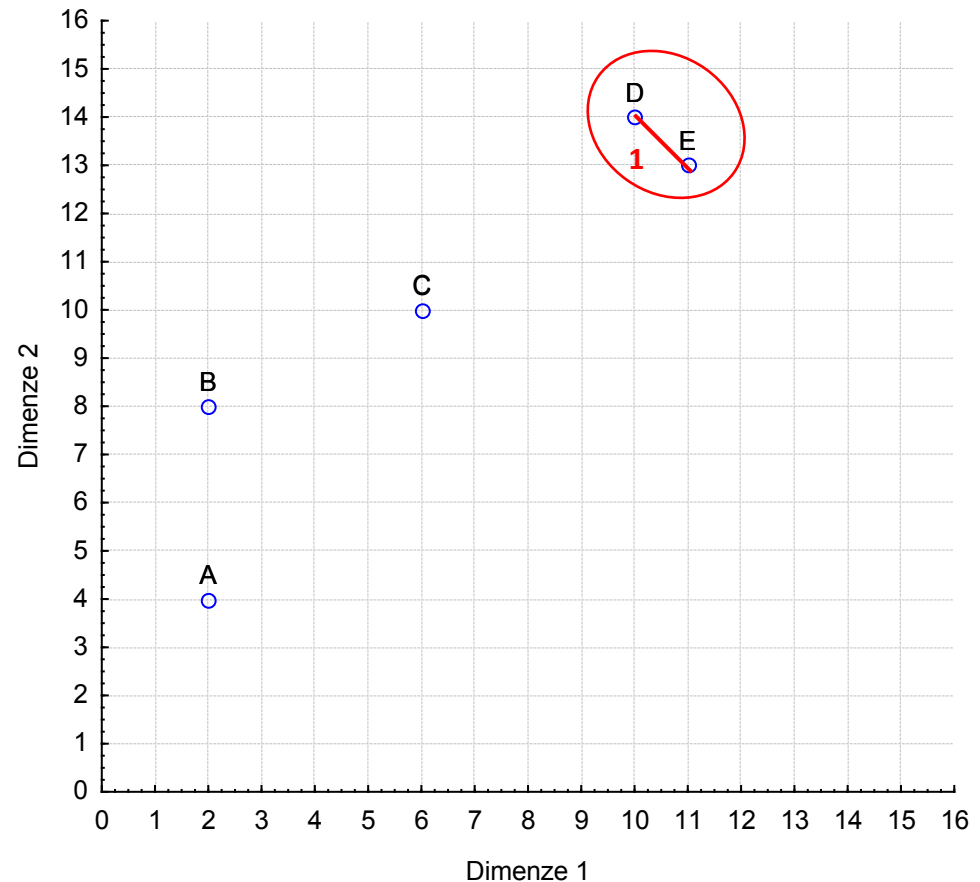
1. Výše uvedená asociační matice vyjadřuje podobnost nebo vzdálenost? A proč?
2. K výpočtu prvků asociační matice byl použit Jaccardův koeficient, Gowerův koeficient, Euklidova metrika nebo Hammingova (manhattanská) metrika?
3. Zdůvodněte vhodnost či nevhodnost použití tohoto koeficientu či metriky v případě těchto dat.
4. Vytvořte dendrogram pomocí algoritmu nejbližšího a nejvzdálenějšího souseda, rozepište jednotlivé kroky výpočtu.

# Metoda nejblížešího souseada: 1. krok výpočtu

- Je vypočtena asociační matice

	A	B	C	D	E
A	0.0	4.0	7.2	12.8	12.7
B	4.0	0.0	4.5	10.0	10.3
C	7.2	4.5	0.0	5.7	5.8
D	12.8	10.0	5.7	0.0	1.4
E	12.7	10.3	5.8	1.4	0.0

- Je definován shluk dvou nejblížeších objektů  
**D-E**



# Metoda nejbližšího souseda: 2. krok výpočtu

- Je vypočtena asociační matice, kde objekty D-E již vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána **nejmenší vzdáleností od jeho členů (D, E)**

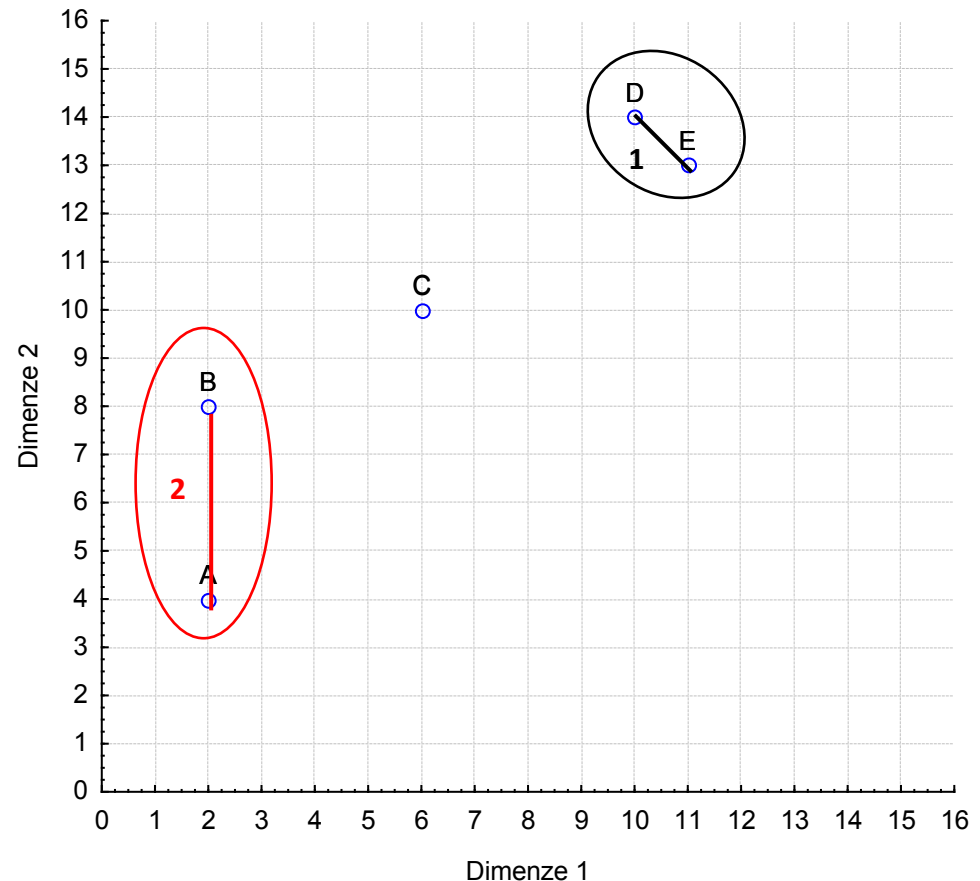
	A	B	C	D	E
A	0.0	4.0	7.2	12.8	12.7
B	4.0	0.0	4.5	10.0	10.3
C	7.2	4.5	0.0	5.7	5.8
D	12.8	10.0	5.7	0.0	1.4
E	12.7	10.3	5.8	1.4	0.0



	A	B	C	D+E
A	0.0	4.0	7.2	12.7
B	4.0	0.0	4.5	10.0
C	7.2	4.5	0.0	5.7
D+E	12.7	10.0	5.7	0.0

- Je definován shluk dvou nejbližších objektů

**A-B**



# Metoda nejbližšího souseda: 3. krok výpočtu

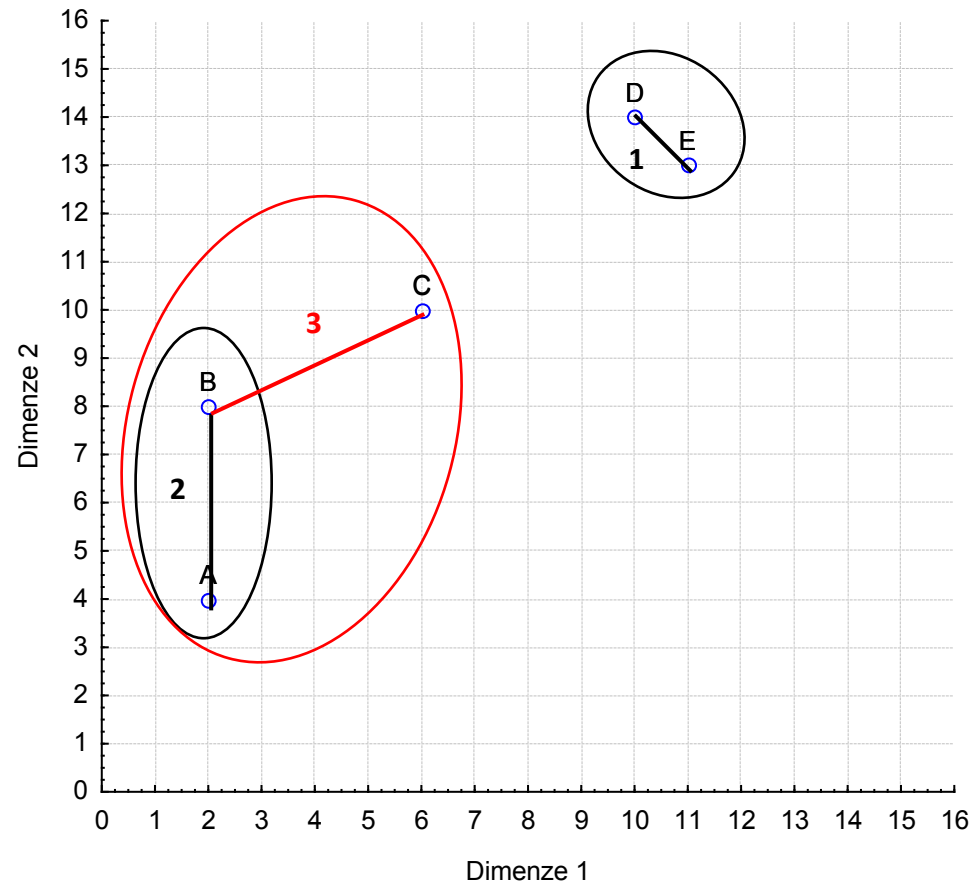
- Je vypočtena asociační matice, kde objekty A-B již vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána **nejmenší vzdáleností od jeho členů (A, B)**

	A	B	C	D+E
A	0.0	4.0	7.2	12.7
B	4.0	0.0	4.5	10.0
C	7.2	4.5	0.0	5.7
D+E	12.7	10.0	5.7	0.0



	A+B	C	D+E
A+B	0.0	4.5	10.0
C	4.5	0.0	5.7
D+E	10.0	5.7	0.0

- Je definován shluk dvou nejbližších objektů (A-B)-C



# Metoda nejbližšího souseda: 4. krok výpočtu

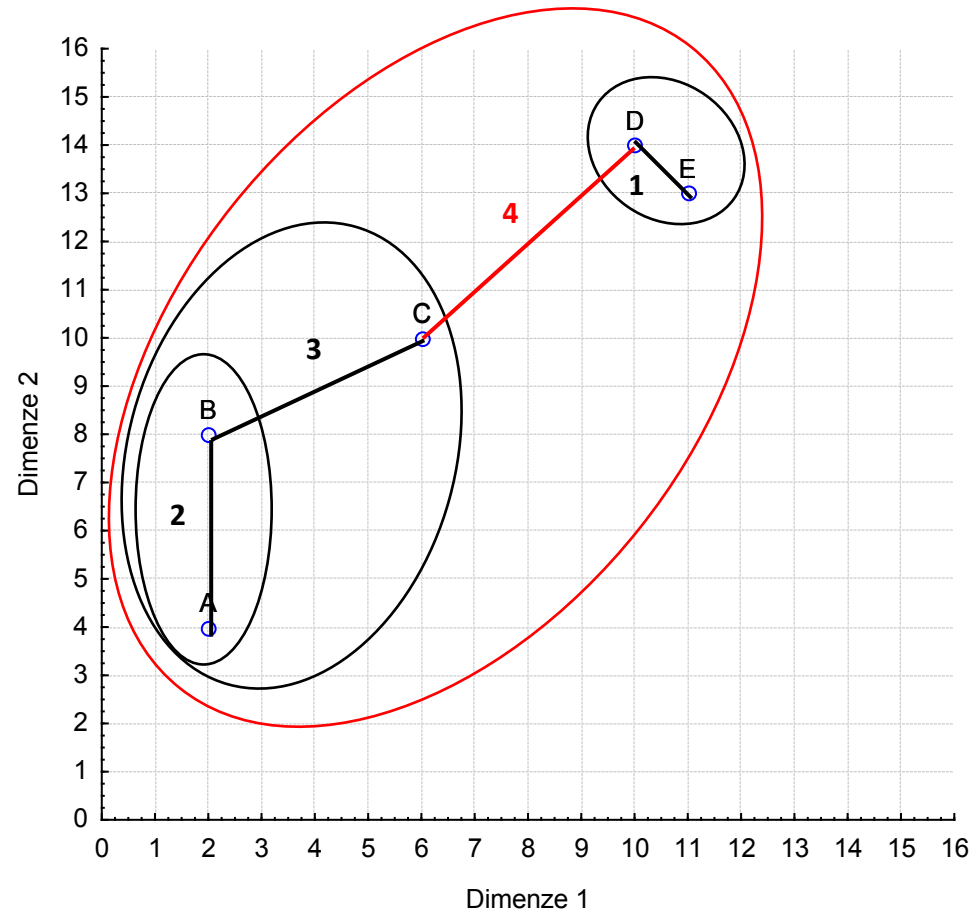
- Je vypočtena asociační matice, kde objekty (A-B)-C již vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána **nejmenší vzdáleností od jeho členů (A, B, C)**

	A+B	C	D+E
A+B	0.0	4.5	10.0
C	4.5	0.0	5.7
D+E	10.0	5.7	0.0



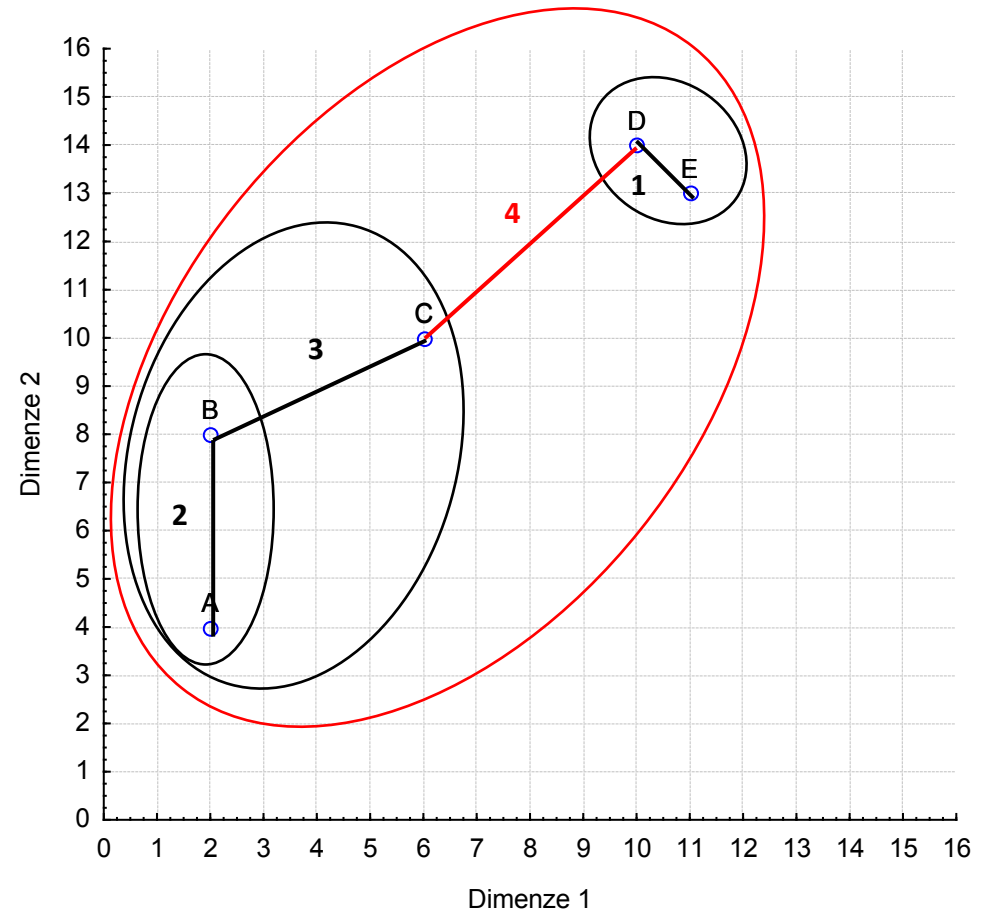
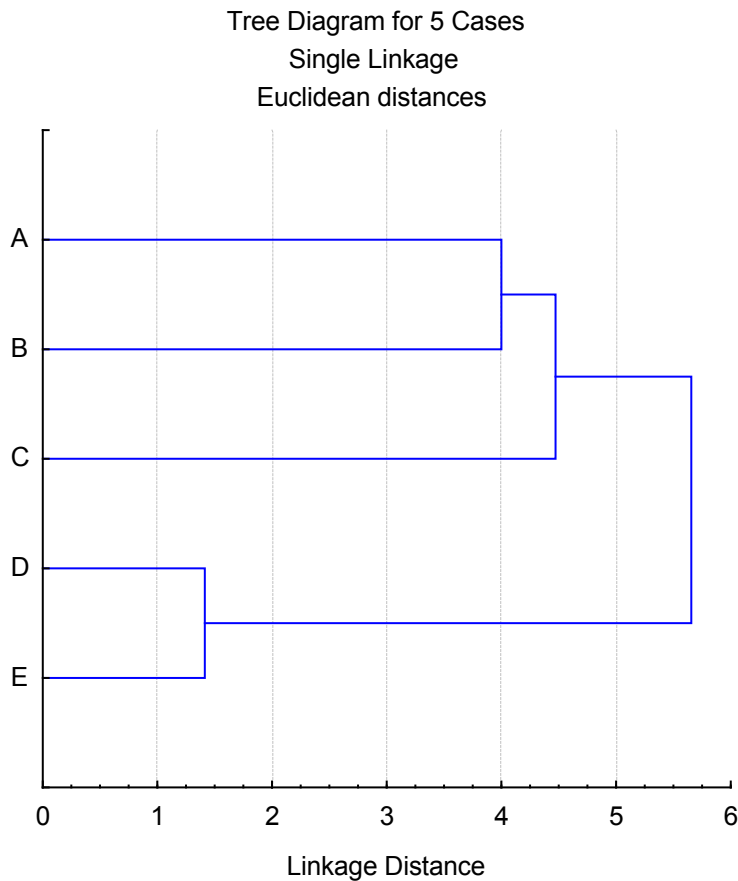
	A+B+C	D+E
A+B+C	0.0	5.7
D+E	5.7	0.0

- Je definován shluk dvou nejbližších objektů **((A-B)-C)-(D-E)**
- Všechny objekty jsou spojeny, algoritmus je ukončen



# Metoda nejbližšího souseda: výsledek analýzy

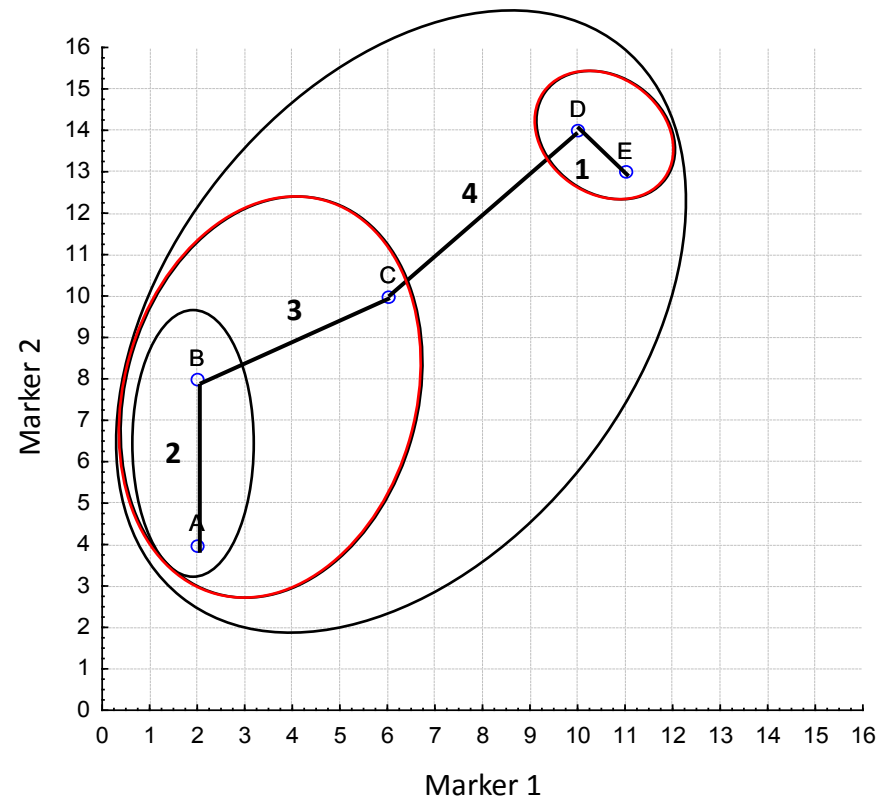
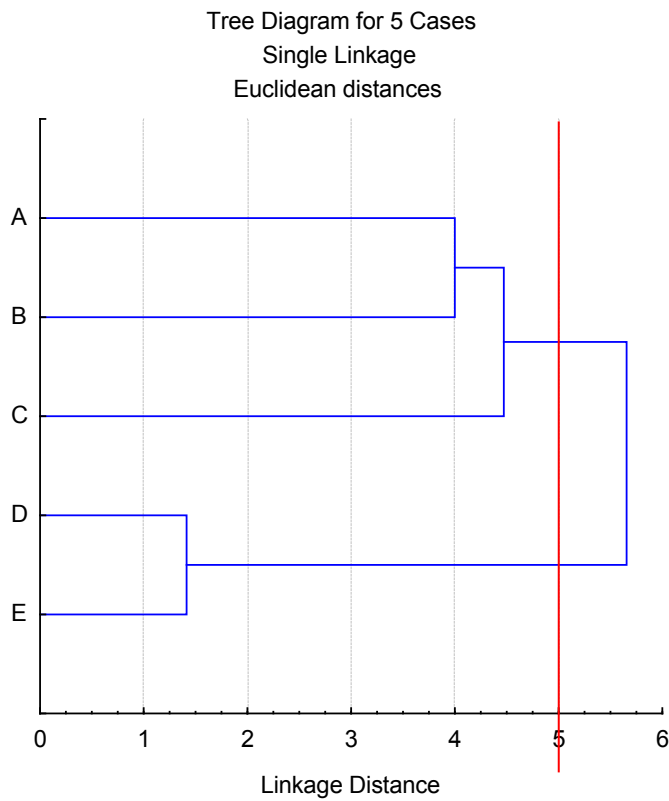
- Výsledek analýzy je vizualizován ve formě dendrogramu





# Metoda nejbližšího souseda: výsledek analýzy

Pokud bychom v dendrogramu provedli řez na podobnosti/vzdálenosti 5, kolik dostaneme shluků? Které buněčné linie budou v jednotlivých shlucích? Výsledek interpretejte.



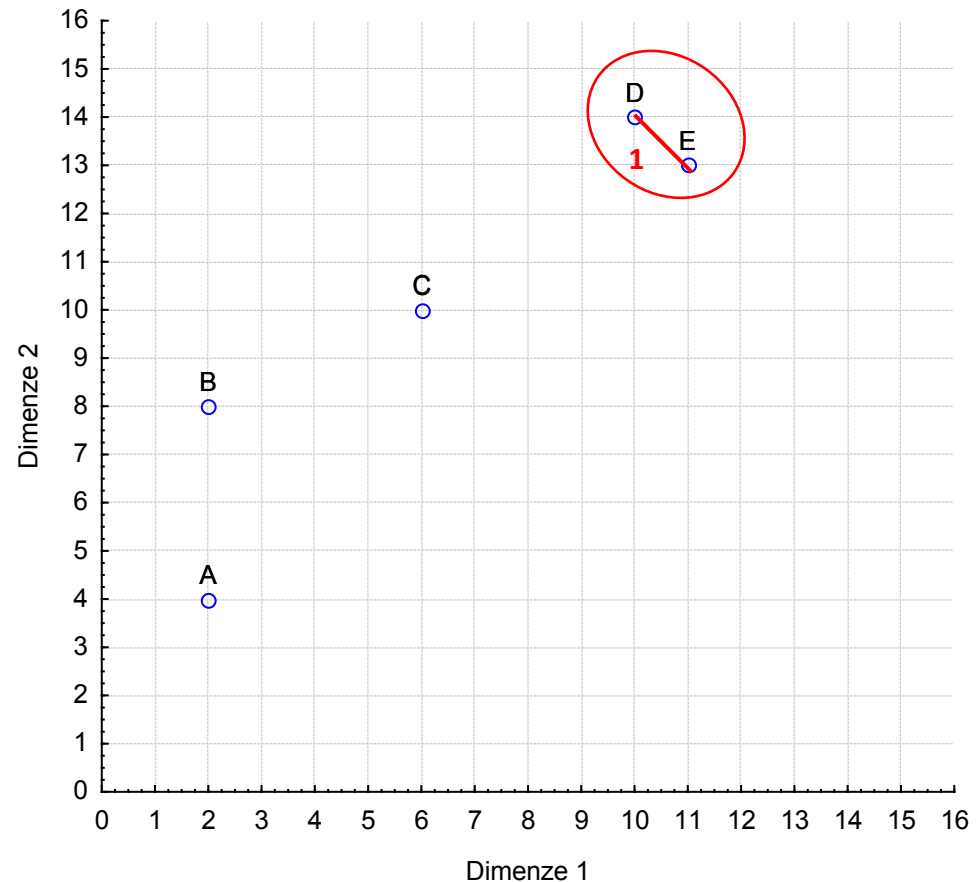
→ dostaneme 2 shluky: (A+B+C) a (D+E); přičemž linie D a E mají mnohem vyšší hodnoty obou markerů než linie A, B a C

# Metoda nejvzdálenějšího souseda: 1. krok výpočtu

- Je vypočtena asociační matice

	A	B	C	D	E
A	0.0	4.0	7.2	12.8	12.7
B	4.0	0.0	4.5	10.0	10.3
C	7.2	4.5	0.0	5.7	5.8
D	12.8	10.0	5.7	0.0	1.4
E	12.7	10.3	5.8	1.4	0.0

- Je definován shluk dvou nejblížeších objektů  
**D-E**



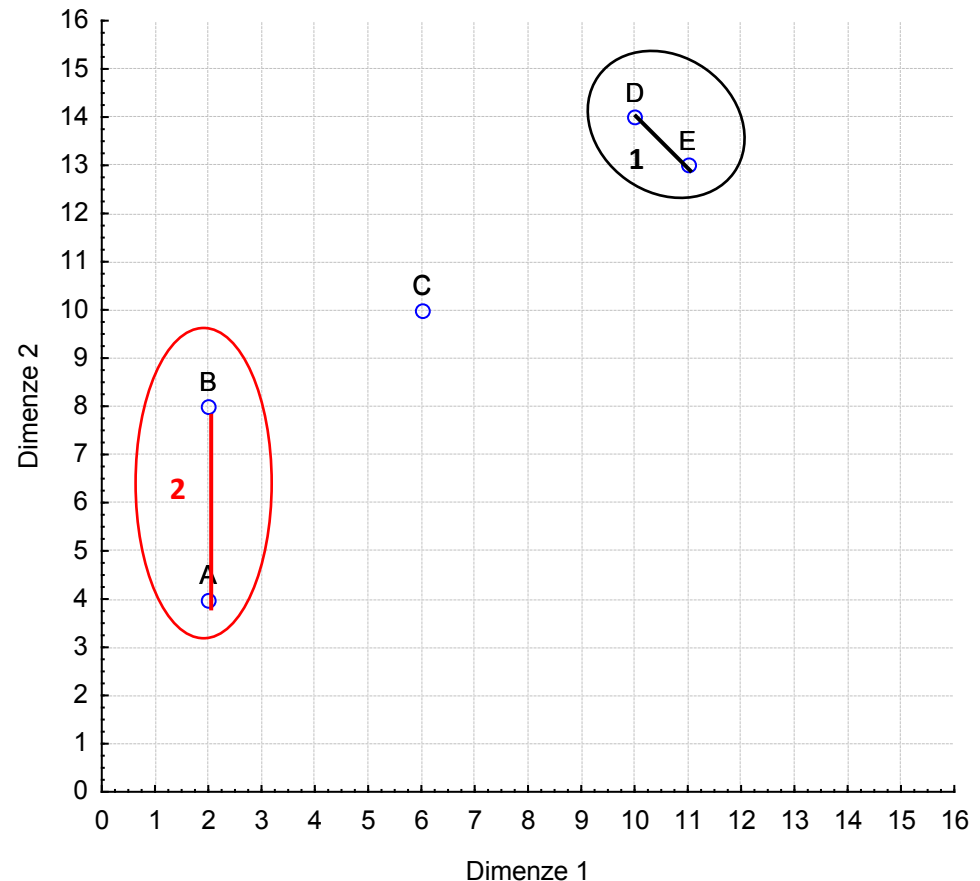
# Metoda nejvzdálenějšího souseda: 2. krok výpočtu

- Je vypočtena asociační matice, kde objekty D-E již vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána **největší vzdáleností od jeho členů (D, E)**

	A	B	C	D	E
A	0.0	4.0	7.2	12.8	12.7
B	4.0	0.0	4.5	10.0	10.3
C	7.2	4.5	0.0	5.7	5.8
D	12.8	10.0	5.7	0.0	1.4
E	12.7	10.3	5.8	1.4	0.0



	A	B	C	D+E
A	0.0	4.0	7.2	12.8
B	4.0	0.0	4.5	10.3
C	7.2	4.5	0.0	5.8
D+E	12.8	10.3	5.8	0.0



- Je definován shluk dvou nejbližších objektů

**A-B**

# Metoda nejvzdálenějšího souseda: 3. krok výpočtu

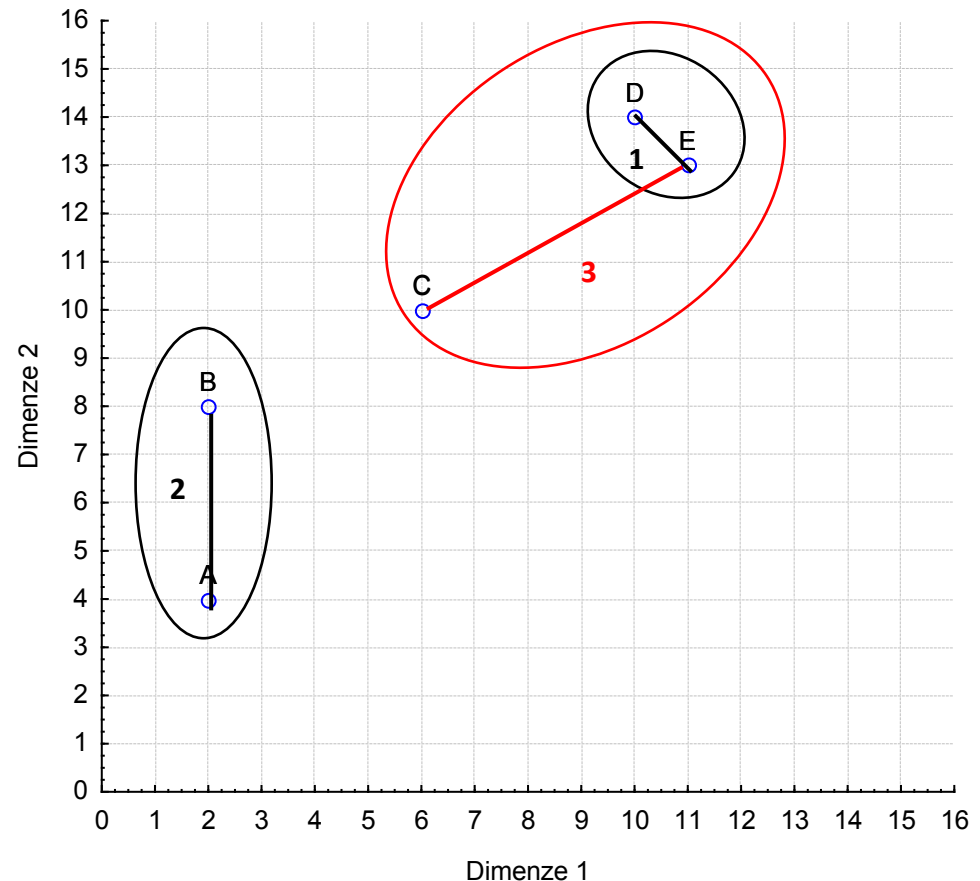
- Je vypočtena asociační matice, kde objekty A-B již vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána **největší vzdáleností od jeho členů (A, B)**

	A	B	C	D+E
A	0.0	4.0	7.2	12.8
B	4.0	0.0	4.5	10.3
C	7.2	4.5	0.0	5.8
D+E	12.8	10.3	5.8	0.0



	A+B	C	D+E
A+B	0.0	7.2	12.8
C	7.2	0.0	5.8
D+E	12.8	5.8	0.0

- Je definován shluk dvou nejblížeších objektů **(D-E)-C**



# Metoda nejvzdálenějšího souseda: 4. krok výpočtu

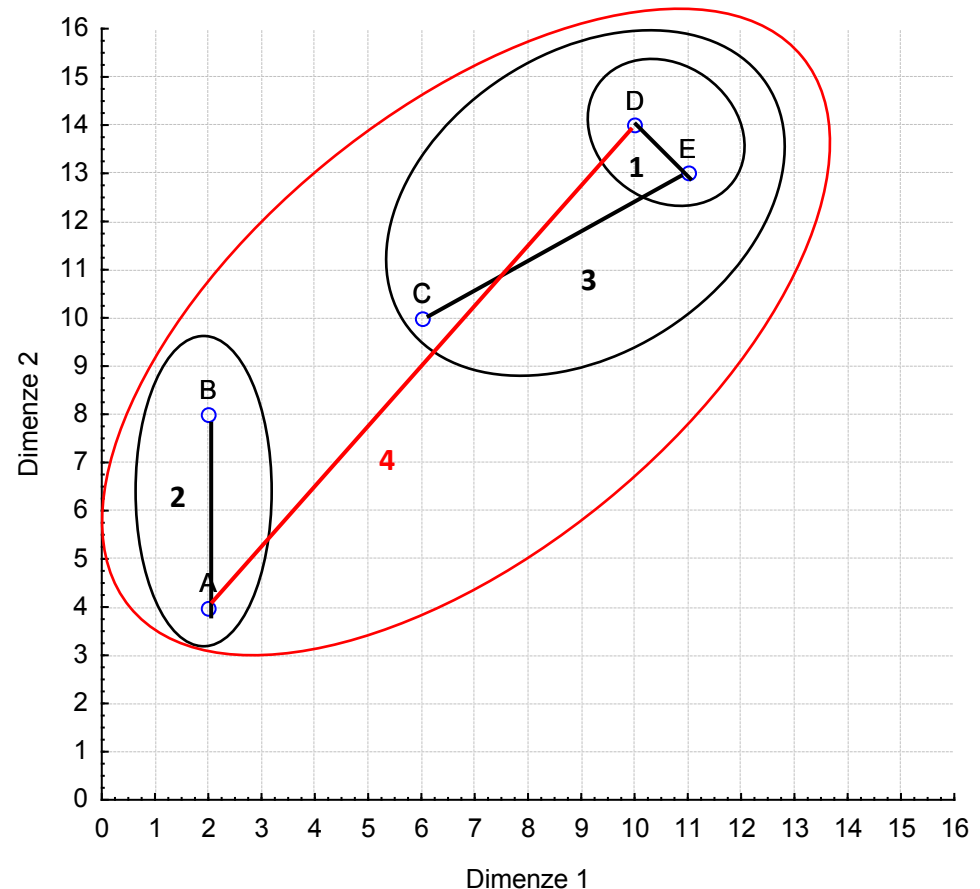
- Je vypočtena asociační matice, kde objekty (D-E)-C již vystupují jako jeden objekt, jehož vzdálenost od ostatních objektů je dána **největší vzdáleností od jeho členů (D, E, C)**

	A+B	C	D+E
A+B	0.0	7.2	12.8
C	7.2	0.0	5.8
D+E	12.8	5.8	0.0



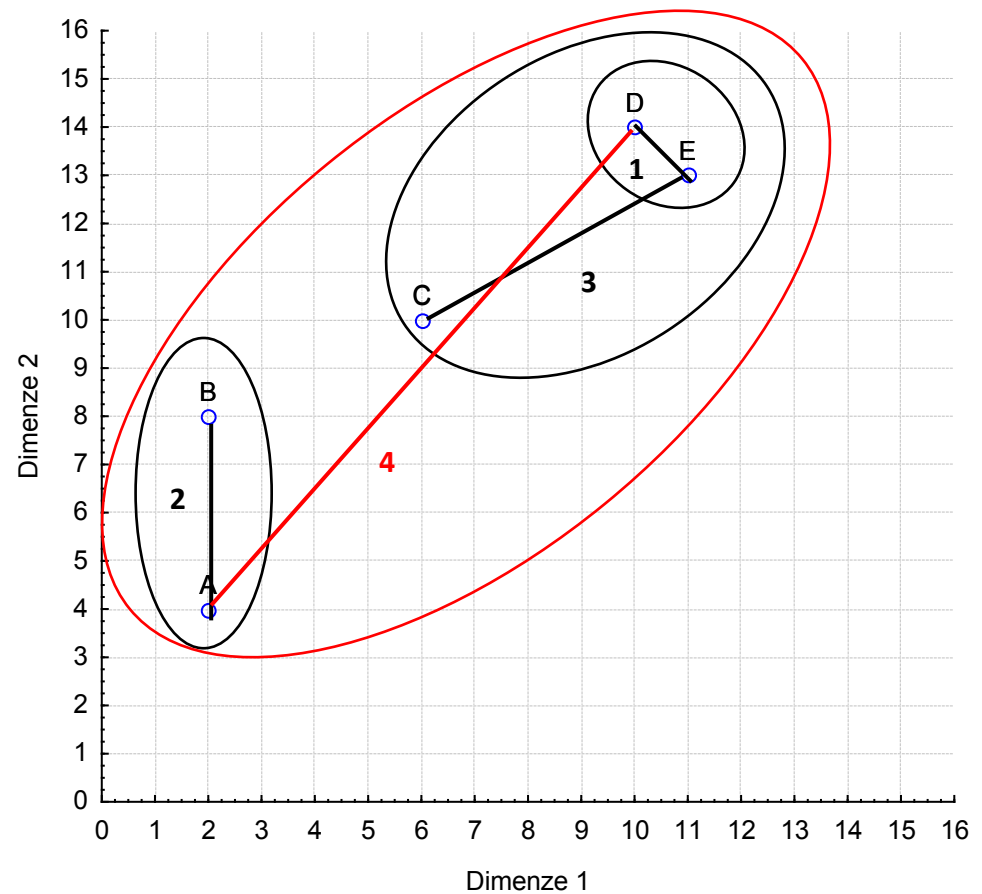
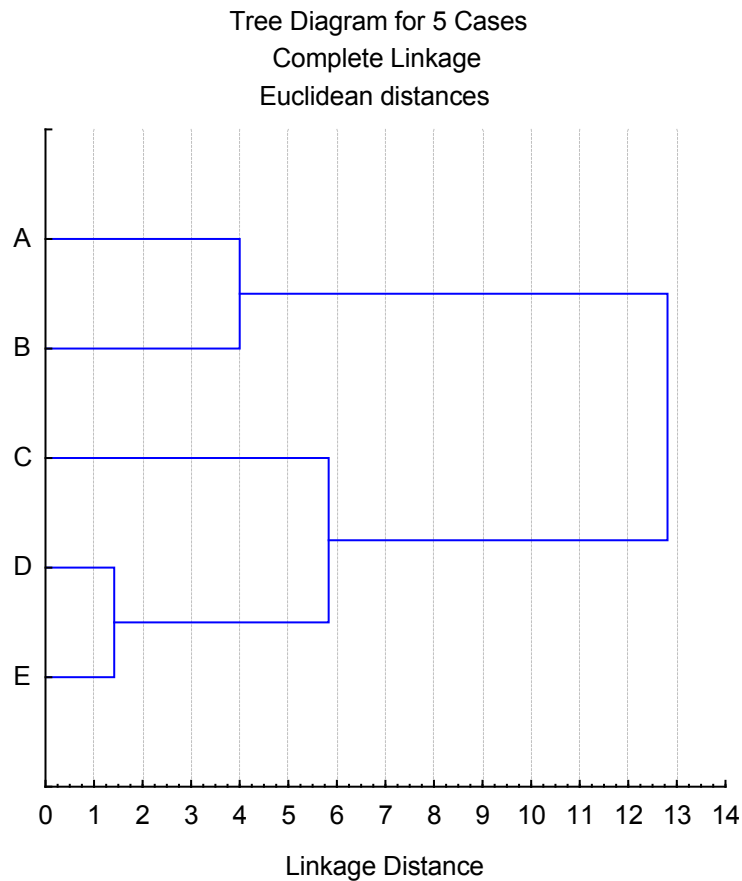
	A+B	D+E+C
A+B	0.0	12.8
D+E+C	12.8	0.0

- Je definován shluk dvou nejblíže objektů **((D-E)-C)-(A-B)**
- Všechny objekty jsou spojeny, algoritmus je ukončen



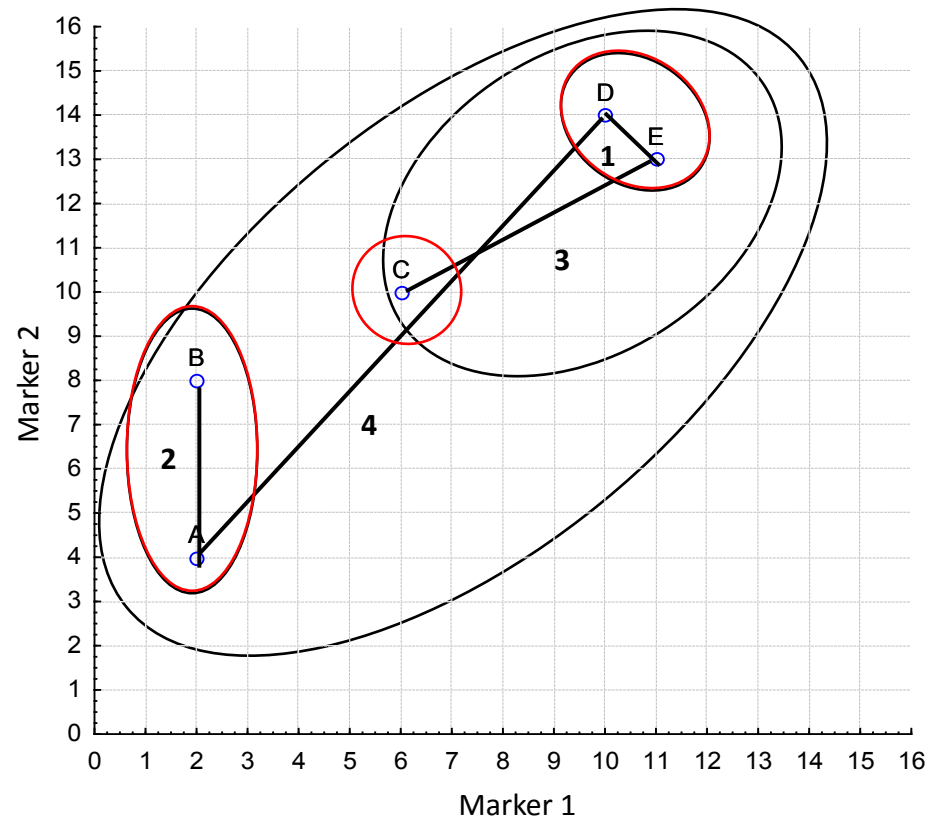
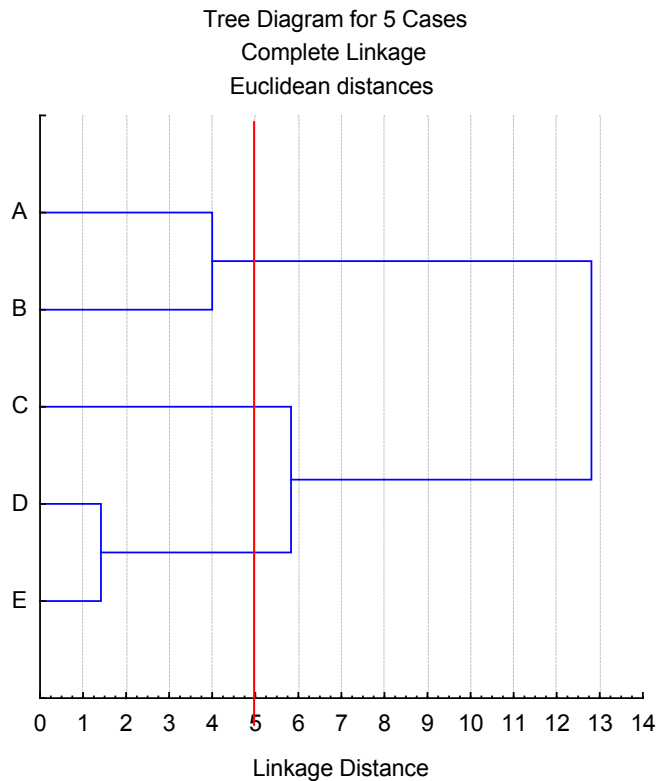
# Metoda nejvzdálenějšího souseda: výsledek analýzy

- Výsledek analýzy je vizualizován ve formě dendrogramu



# Metoda nejvzdálenějšího souseda: výsledek analýzy

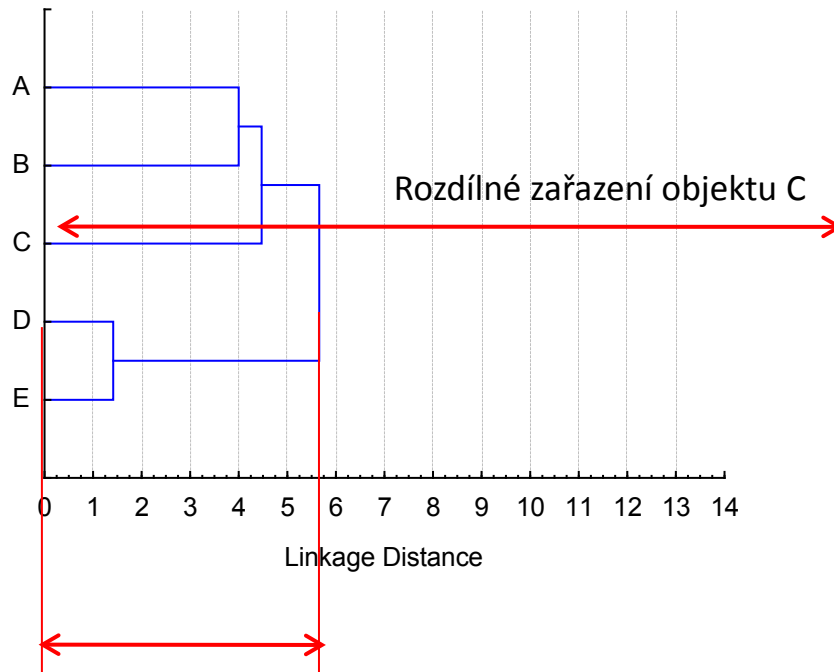
Pokud bychom v dendrogramu provedli řez na podobnosti/vzdálenosti 5, kolik dostaneme shluků? Které buněčné linie budou v jednotlivých shlucích? Výsledek interpretujte.



→ dostaneme 3 shluky: (A+B), (C) a (D+E); přičemž linie D a E mají vysoké hodnoty obou markerů, A a B mají nízké hodnoty obou markerů a linie C má střední hodnoty markerů

# Metoda nejbližšího a nejvzdálenějšího souseda – interpretace výsledků

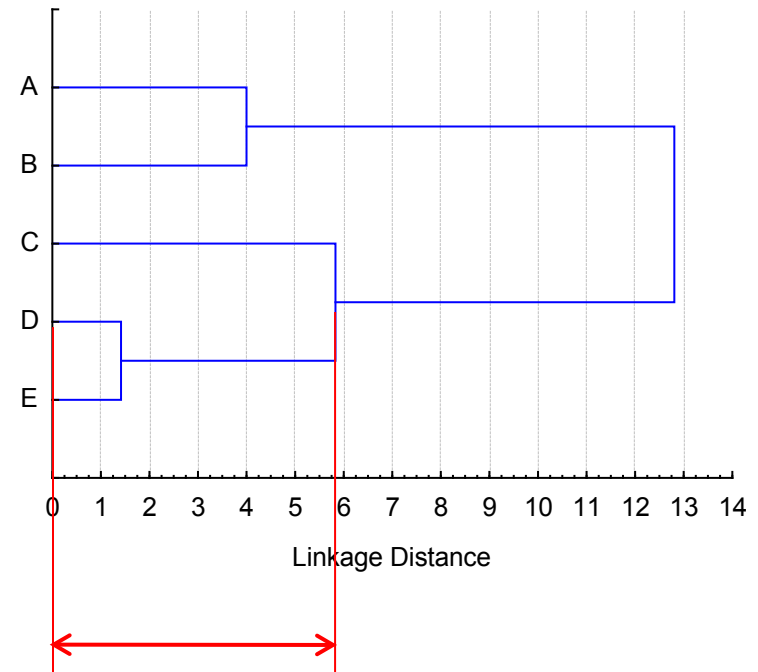
## Metoda nejbližšího souseda



**Vzdálenost, na níž došlo ke spojení shluku:**

- u metody nejbližšího souseda znamená nejmenší vzdálenost objektů shluku, tedy ve shluku mohou existovat objekty s větší vzdáleností

## Metoda nejvzdálenějšího souseda



**Vzdálenost, na níž došlo ke spojení shluku:**

- u metody nejvzdálenějšího souseda znamená největší vzdálenost objektů shluku, tedy objekty ve shluku už mohou být k sobě pouze blíže nebo stejně vzdálené jako je tato vzdálenost



## Příklad 2

Bylo provedeno měření objemu hipokampu a mozkových komor (v cm<sup>3</sup>) u 5 pacientů se schizofrenií. Naměřené hodnoty objemu hipokampu a mozkových komor byly zaznamenány do matice  $\mathbf{X}_D$ :

$$\mathbf{X}_D = \begin{bmatrix} 4,6 & 3,4 \\ 6,1 & 3,0 \\ 6,7 & 3,1 \\ 6,2 & 2,3 \\ 6,9 & 3,1 \end{bmatrix}.$$

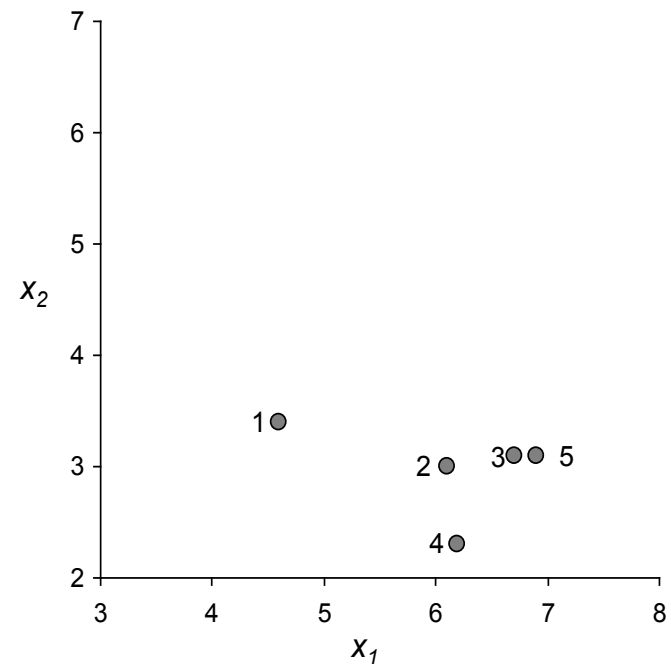
Určete podobnost pěti pacientů na základě naměřených charakteristik pomocí hierarchické shlukové analýzy, použijte metodu nejbližšího a nejvzdálenějšího souseda.

# Příklad 2 – asociační matice

Nejprve vypočteme matici vzdáleností mezi objekty založenou 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.

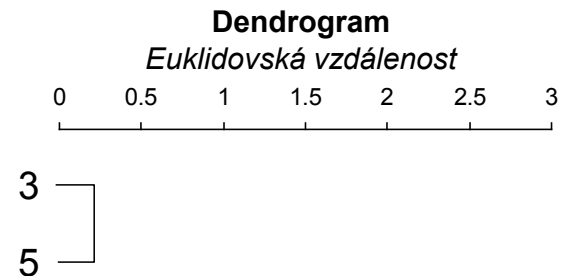
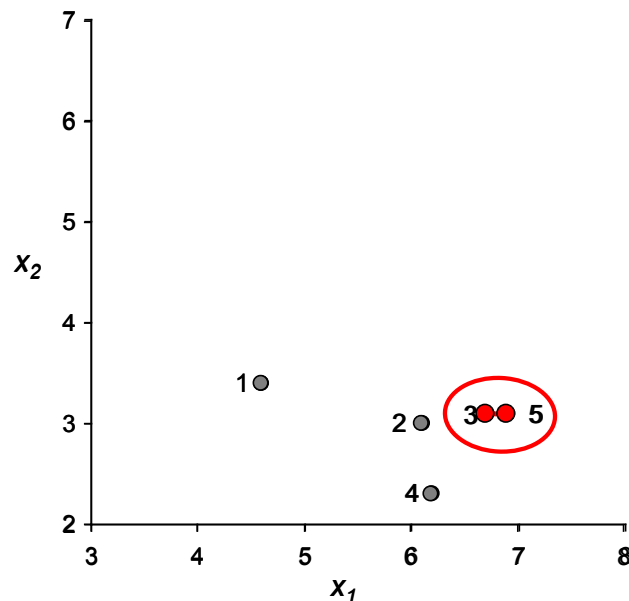


# Metoda nejbližšího souseda – krok 1

	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



	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

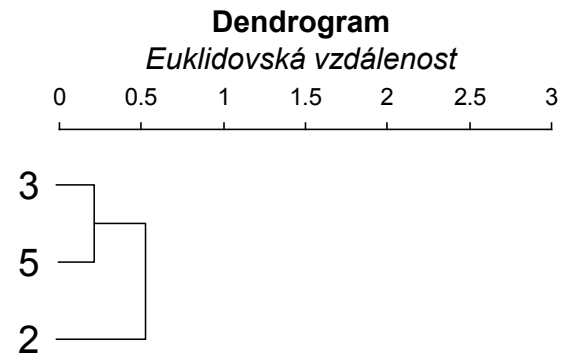
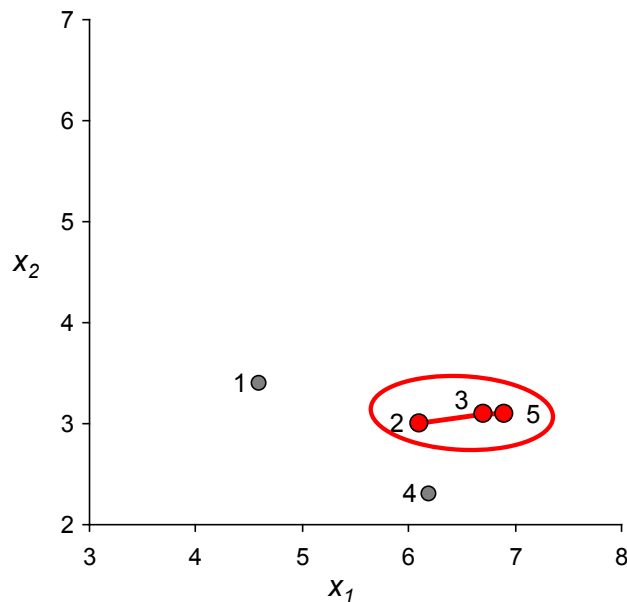


# Metoda nejbližšího souseda – krok 2

	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



	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

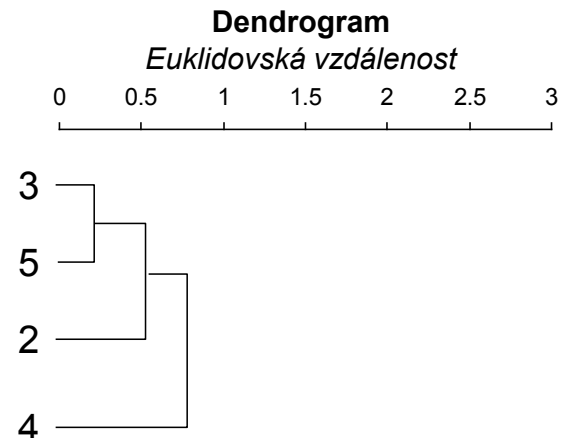
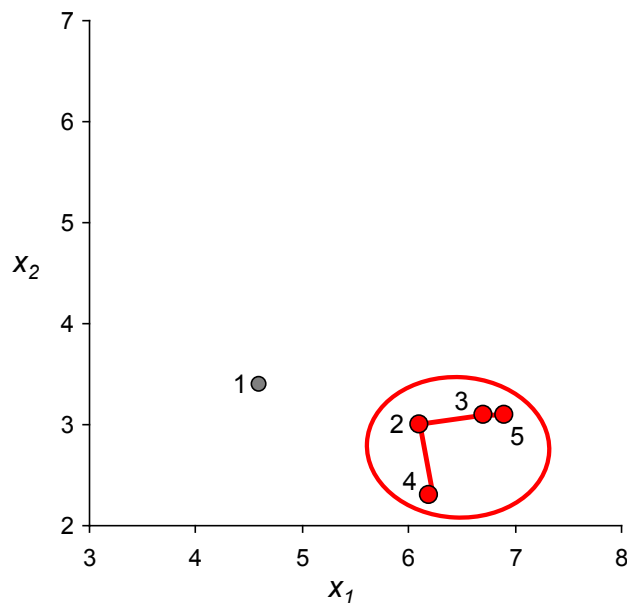


# Metoda nejbližšího souseda – krok 3

	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

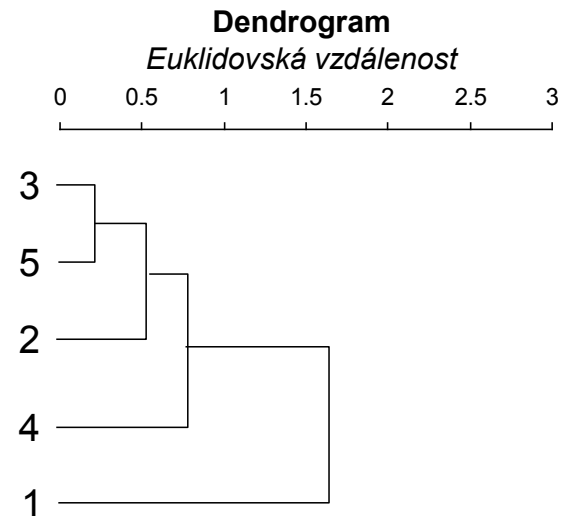
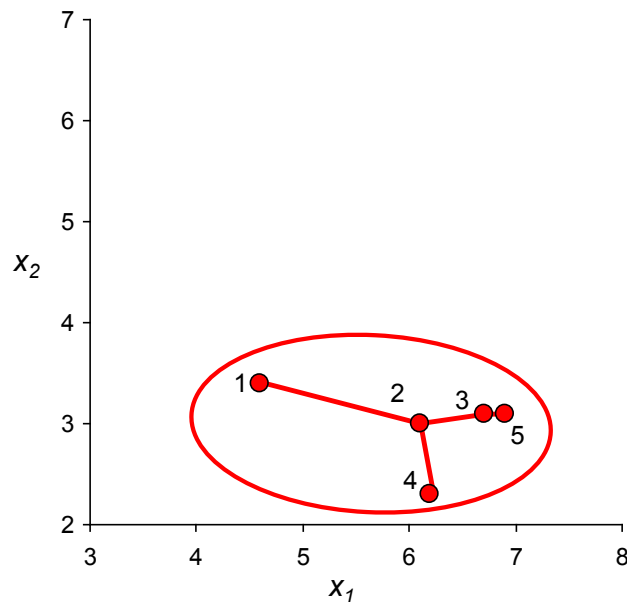


	1	4+2+3+5
1	0,0	1,6
4+2+3+5	1,6	0,0



# Metoda nejbližšího souseda – krok 4

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

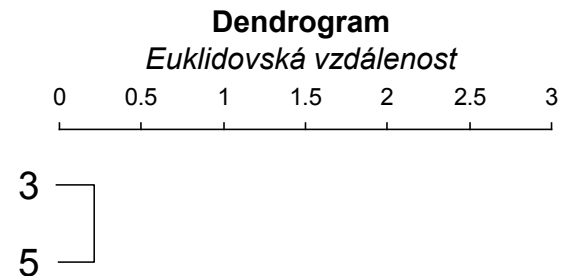
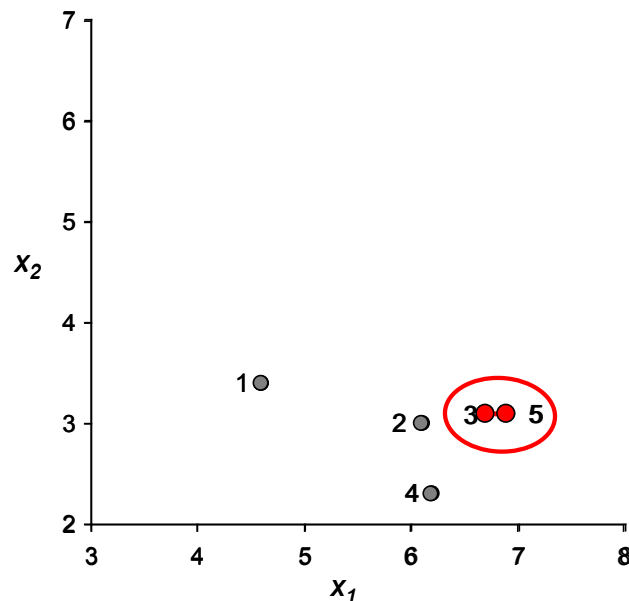


# Metoda nejvzdálenějšího souseda – krok 1

	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



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

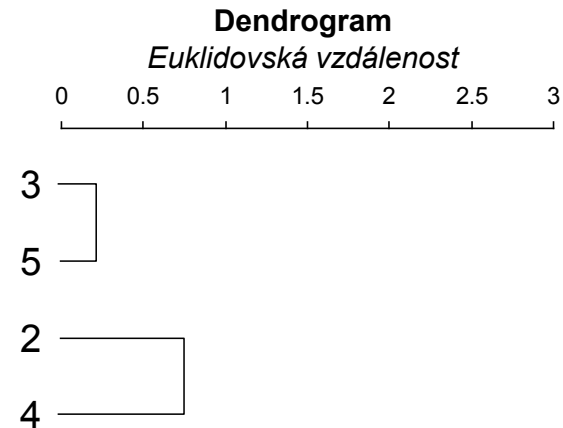
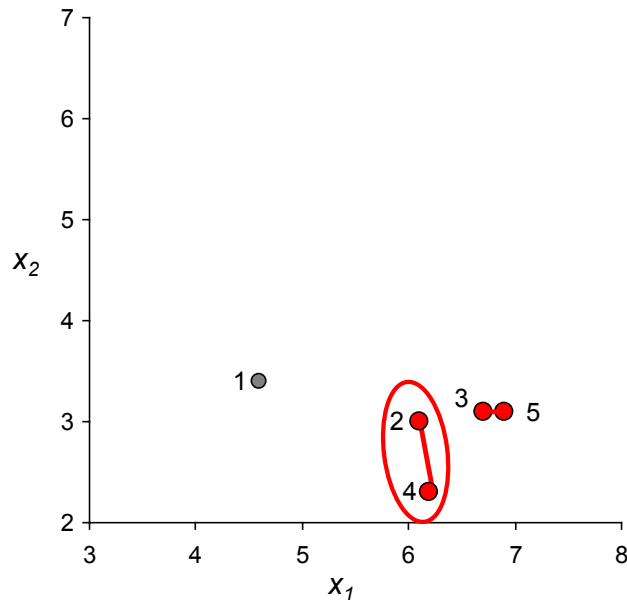


# Metoda nejvzdálenějšího souseda – krok 2

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



	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



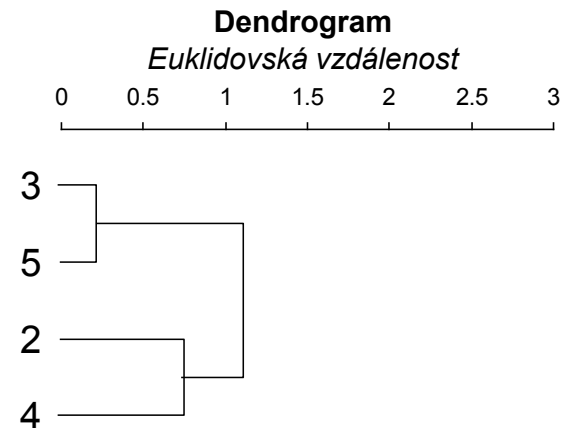
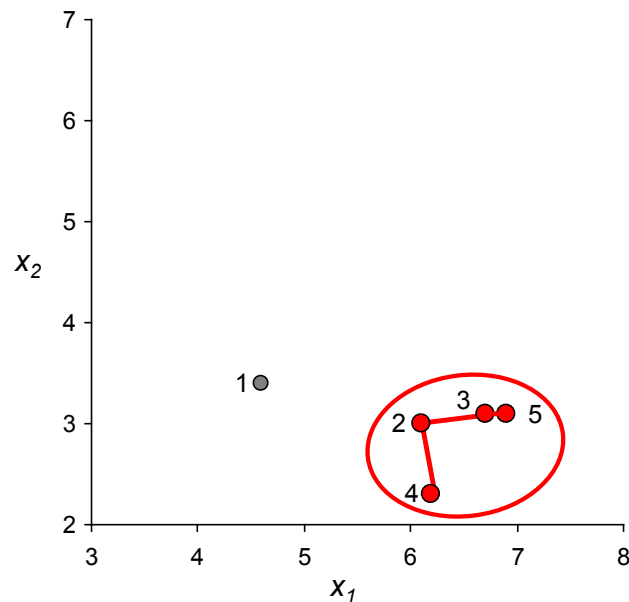


# Metoda nejvzdálenějšího souseda – krok 3

	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

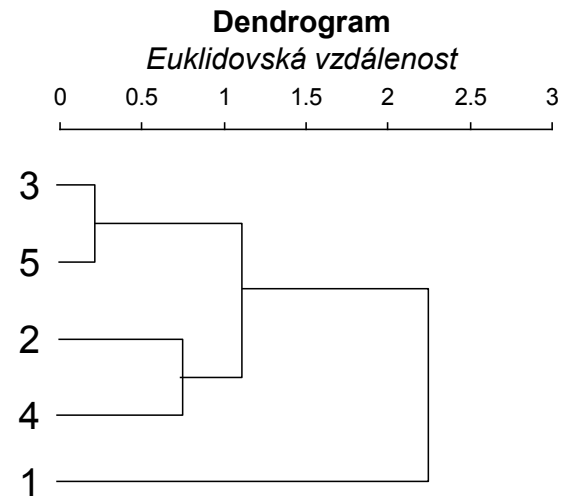
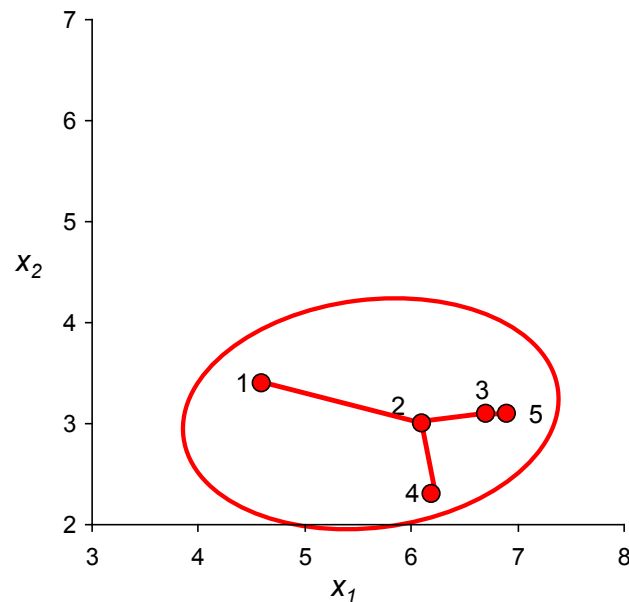


	1	4+2+3+5
1	0,0	2,3
4+2+3+5	2,3	0,0

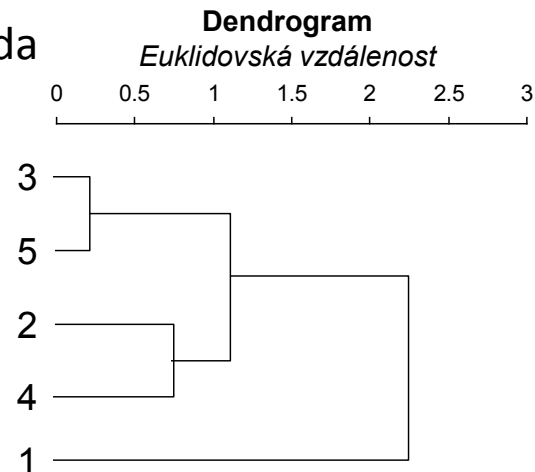
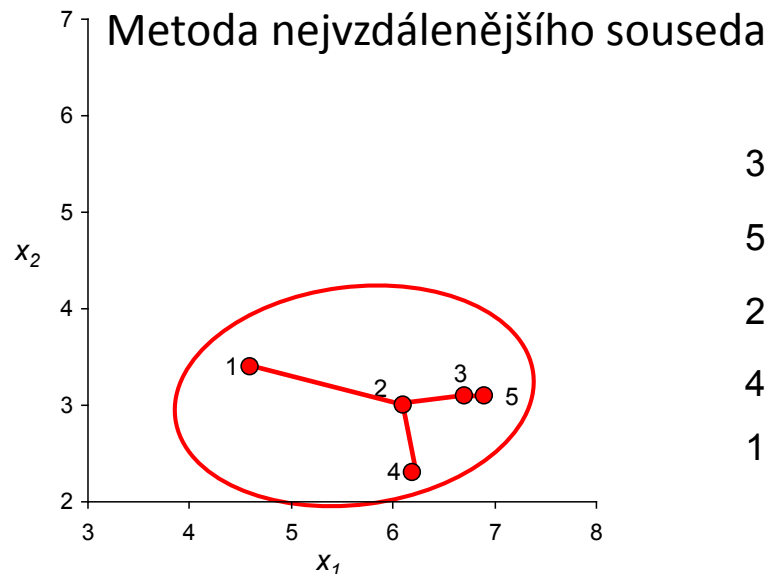
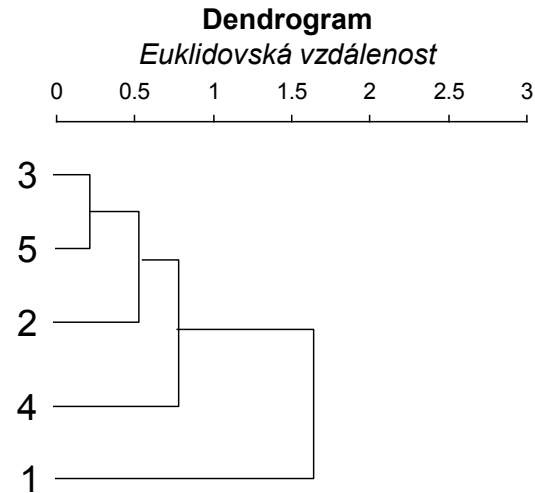
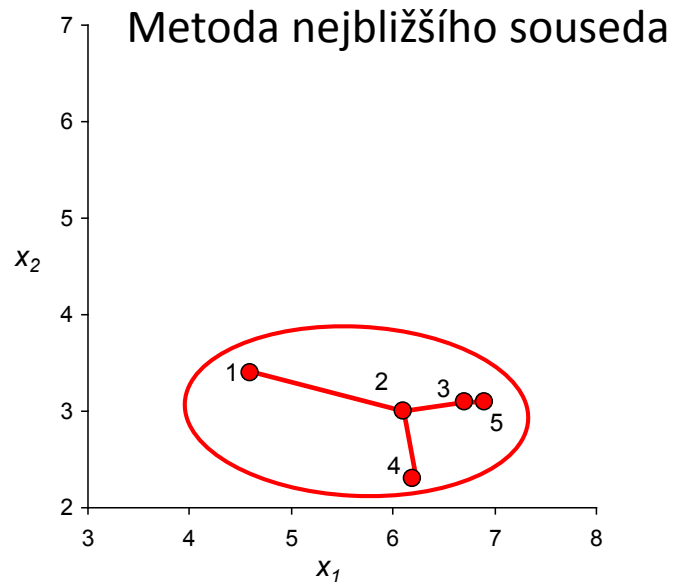


# Metoda nejvzdálenějšího souseda – krok 4

	1	4+2+3+5
1	0,0	2,3
4+2+3+5	2,3	0,0



# Srovnání metody nejbližšího a nejvzdálenějšího souseda



→ metoda nejbližšího souseda má tendenci vytvářet protáhlé shluky