

PA055 Vizualizace komplexních dat

Týden 2

Katedra informačních technologií
Masarykova Univerzita Brno

Podzim 2020

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Souřadnicová (mnohorozměrná) data - $X = [x_1, x_2]$

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

	Col1	Col2	Col3
E.coli	24.7	26.0	13.2
S.lutea	13.4	37.1	12.1
S.cerevisae	31.3	18.7	17.4
H.sapiens	30.9	19.9	20.0

sloupce - proměnné

řádky - pozorování

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

	E.coli	S.lutea	S.cerevisiae
E.coli	24.7	26.0	13.2
S.lutea	13.4	37.1	12.1
S.cerevisiae	31.3	18.7	17.4

matice číselné hodnoty vyjádřují podobnost

Zjednodušení vzdálenostních dat - shlukování

Souřadnice a vzdálenosti

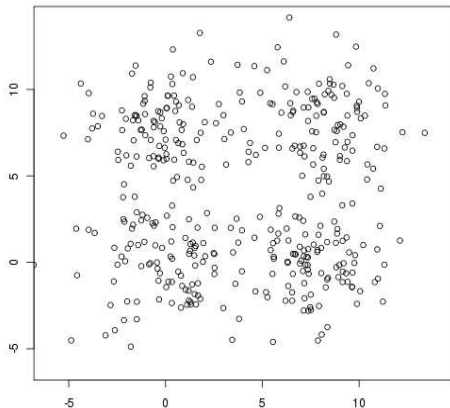
Vizualizace síťových dat

Kostra grafu

Substrát

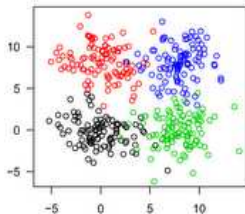
Příklady

True Groups

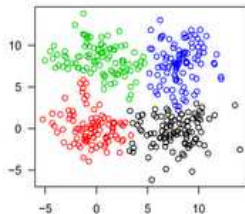


Zjednodušení vzdálenostních dat - shlukování

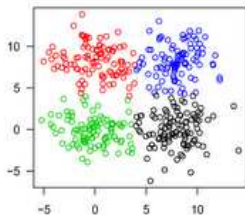
True Groups



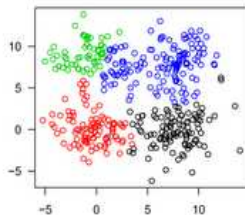
hclust Euclidean ward



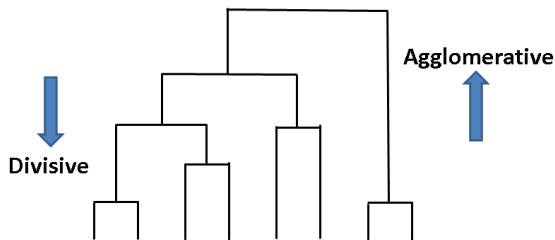
kmeans



hclust Euclidean complete



Zjednodušení vzdálenostních dat - shlukování



Převzato z <https://towardsdatascience.com/hierarchical-clustering-explained-e58d2f936323>

Zjednodušení vzdálenostních dat - shlukování

Souřadnice a vzdálenosti

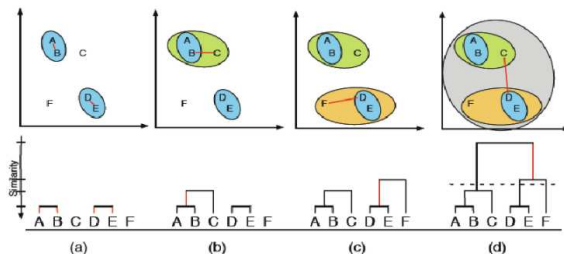
Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

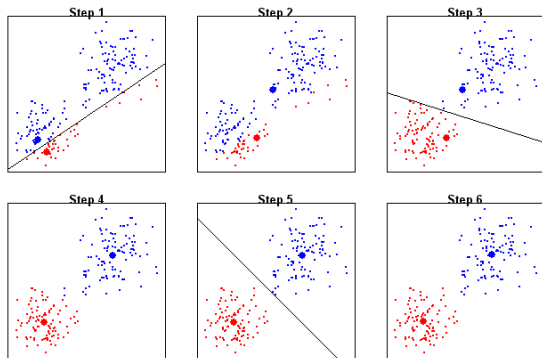
Example: Hierarchical Agglomerative Clustering



Převzato z

<http://machinelearningstories.blogspot.com/2017/09/hierarchical-clustering-bottom-up.html>

Zjednodušení vzdálenostních dat - shlukování



Převzato z

<http://sherrytowers.com/2013/10/24/k-means-clustering/>

Zjednodušení souřadnicových dat - PCA, LDA

Souřadnice a vzdálenosti

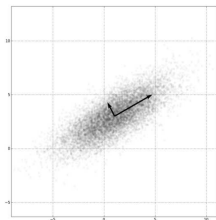
Vizualizace síťových dat

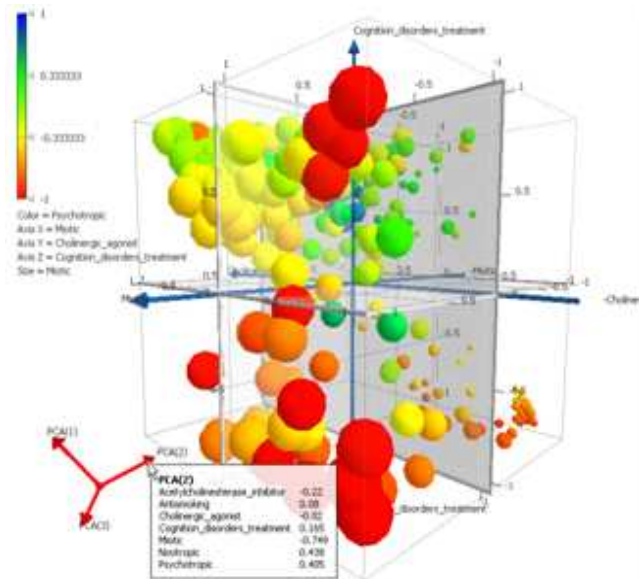
Kostra grafu

Substrát

Příklady

Principal component analysis - používá se pro mnohorozměrná data obsahující vzájemně korelované proměnné. Nalezne takovou transformaci dat do nového ortogonálního systému souřadnic, že první souřadnice vysvětluje nejvíce rozptylu dat, druhá nejvíce reziduálního rozptylu, atd. Jde o nejinformativnější průmět dat.





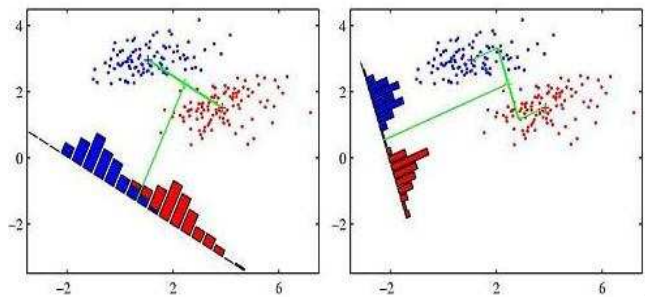
Souřadnice a vzdálenosti

Vizualizace síťových dat

- Kostra grafu
- Substrát
- Příklady

[Souřadnice a vzdálenosti](#)[Vizualizace síťových dat](#)[Kostra grafu](#)[Substrát](#)[Příklady](#)

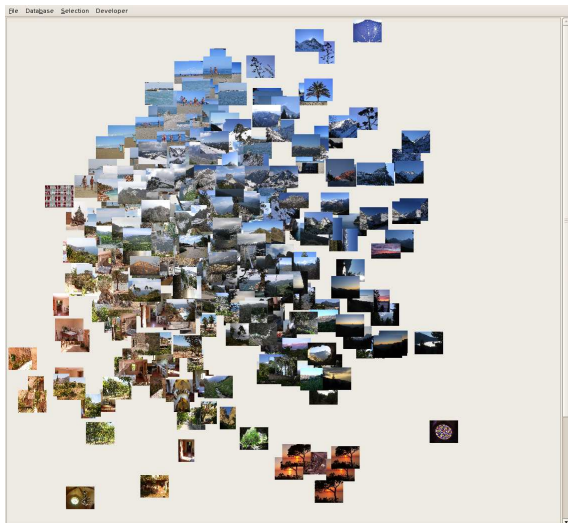
Linear discriminant analysis - příbuzná metoda PCA, hledá lineární kombinaci proměnných, která co nejlépe separuje data do skupin. Často se používá jako příprava ke shlukování za účelem snížení počtu rozměrů.



Převod $d(x,y) \rightarrow [x_1,x_2]$ - multidimensional scaling

Vychází ze známých vzdáleností a nalézá reprezentaci dat, která tyto vzdálenosti co nejlépe respektuje. Jedná se vesměs o hledání minima nějaké funkce popisující odchylku dané reprezentace dat od ideálních vzdáleností

Převod $d(x,y) \rightarrow [x1,x2]$ - MDS



Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Příklad efektivního nasazení MDS

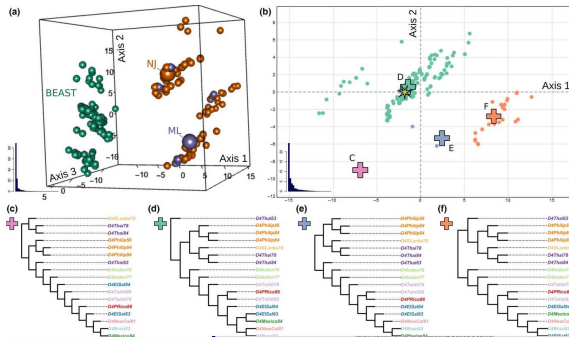
Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady



Převzato z

<https://onlinelibrary.wiley.com/doi/full/10.1111/1755-0998.12676>

- ▶ přístup
 - ▶ dělicí
 - ▶ aglomerativní
- ▶ vzorec
 - ▶ Euklidovská vzdálenost
 - ▶ Manhattan
 - ▶ skalární součin (data jako vektory)
 - ▶ Hammingova vzdálenost (počet rozdílů)
 - ▶ grafová
- ▶ data
 - ▶ střed zhluku
 - ▶ nejbližší
 - ▶ nejvzdálenější
 - ▶ suma všech vzdáleností

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Souřadnice a vzdálenosti

Vizualizace síťových dat

Kostra grafu

Substrát

Příklady

Příloha

Příloha

For Further Reading

For Further Reading

Příloha

For Further Reading

X