

# Shlukovací metody

## Metody shlukovací analýzy

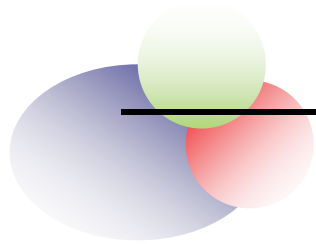
Dělicí metody

Hierarchické metody

Metody založené na hustotě

Mřížkové metody

Další metody...



# Shlukovací metody

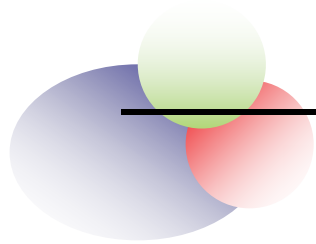
## Dělicí metody

Cíl: Rozdělit  $n$  objektů do  $k$  shluků

Princip: Dělení podle kriteria – podobnostní funkce

Možnosti:

- Výpočet všech možných rozdělení
- Použití heuristik:
  - K-Means – shluk je reprezentován těžištěm
  - K-Medoids – shluk je reprezentován objektem nejblíže k těžišti



# Shlukovací metody

## Dělicí metody

Algoritmus :

Vstup: Počet shluků  $k$  a databáze  $n$  objektů

Výstup:  $k$  shluků

Postup:

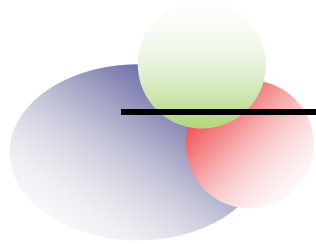
Vyber  $k$  objektů jako středy shluků

repeat

    přepočítej těžiště shluků

    přiřaď každý objekt do shluku na základě  
    vzdálenosti od těžiště

until Žádná\_změna

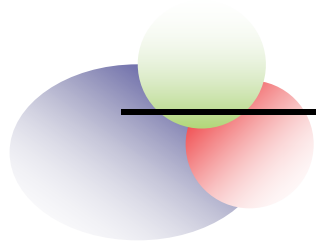


# Shlukovací metody

## Dělicí metody

K-Means, K-Medoids algoritmy

- Ideální při výběru malého počtu shluků z velkého počtu objektů
- Málo odolná vzhledem k nekonzistentním vzorkům dat a šumu
- Nutnost volit počáteční počet shluků  $k$



# Shlukovací metody

## Metody shlukovací analýzy

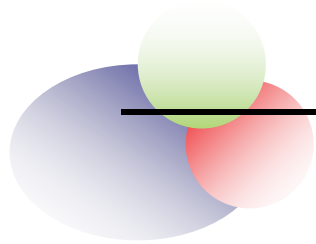
Dělicí metody

**Hierarchické metody**

Metody založené na hustotě

Mřížkové metody

Další metody...



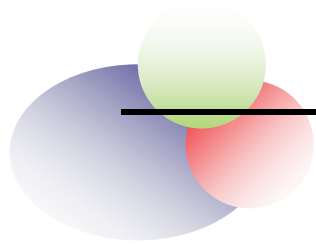
# Shlukovací metody

## Hierarchické metody

Princip: Vytváření stromové struktury shluků

Možnosti: Přístup **aglomerativní**

- Strategie **bottom-up**
- Každý objekt = 1 shluk
- Postupně slučujeme shluky dokud nejsou všechny objekty v jediném shluku nebo nejsou splněna kritéria ukončení (typicky: počet shluků)
- Většina metod >> liší se pouze v porovnávání shluků



# Shlukovací metody

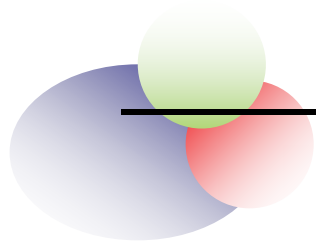
## Hierarchické metody

Přístup **divisivní**

- Strategie **top-down**
- Všechny objekty v jediném shluku
- Postupně dělí shluky na menší dokud netvoří každý objekt jeden shluk nebo nejsou splněna kritéria ukončení

Výpočet odlišnosti shluků:

- Minimální vzdálenost
- Maximální vzdálenost
- Střední vzdálenost
- Průměrná vzdálenost



# Shlukovací metody

## Metody shlukovací analýzy

Dělicí metody

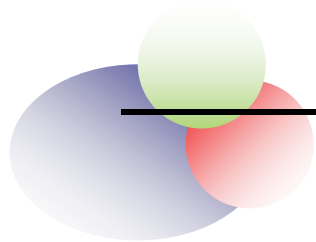
Hierarchické metody

Metody založené na hustotě

Mřížkové metody

Další metody...





## To nejdůležitější

- Shluk je množina objektů podobných si navzájem v rámci shluků a odlišná mezi jednotlivými shluky
- Proces vytváření shluků je **shlukování**
- Kvalita shlukování je úměrná schopností **porovnat odlišnosti** jednotlivých objektů
- Umíme porovnat a změřit míru podobnosti objektů popsaných **různými typy** dat (intervalové, binární, ... proměnné)
- Shlukování je významný nástroj při dolování dat
- Bylo vyvinuto mnoho algoritmů, které lze rozčlenit na Dělicí, Hierarchické, Pracující s hustotou, Mřížkové ...
- Velmi důležitá je:
  - **volba vstupních parametrů**
  - **interpretace výsledků**