

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

IV107 Bioinformatika I

Přednáška 11

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2011

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Algoritmy v bioinformatice

- ▶ Vyhledávání v řetězcích
 - ▶ Boyer-Moore
 - ▶ pomocí konečného automatu
 - ▶ suffixové stromy
 - ▶ suffixové pole
- ▶ Tandemové opakování
- ▶ Palindromy
- ▶ DP - varianty algoritmu Needleman-Wunsch

Statistika více proměnných, ale na rozdíl např. od analýzy rozptylu a regresní analýzy pozorované proměnné nejsou spojité. Data patří do diskrétních tříd.

- ▶ Shlukování
- ▶ Diskriminační analýza
- ▶ Rozhodovací stromy
- ▶ Umělé neuronové sítě
- ▶ SVM

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Rozdělení dat do skupin podle společných vlastností, resp. blízkosti v nějakém prostoru

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

- ▶ Hierarchické
 - ▶ zespodu
 - ▶ shora
- ▶ Nehierarchické (např. centroidní "k-means")

Typy shlukování - podle vzdálenostní funkce

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

- ▶ L_p -normy ($D = (dx^p + dy^p)^{1/p}$)
 - ▶ Manhattan ($D = dx + dy$)
 - ▶ euklidovská ($D = \sqrt{dx^2 + dy^2}$)
- ▶ Čebyševova ($D = \max(dx, dy)$)
- ▶ vektorový součin ($D = X \cdot Y$)
- ▶ korelační koeficient
- ▶ editační
 - ▶ Levinshtein
 - ▶ Hamming

Data můžou být definovaná vlastnostmi (souřadnice)

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

A = (3,1,"white")

B = (4,1,"blue")

C = (3,2,"white")

D = (3,0,"red")

Data mohou být definovaná pomocí párové vzdálenosti

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | | | |
| B | 3 | 0 | | |
| C | 2 | 4 | 0 | |
| D | 3 | 1 | 2 | 0 |

Vzdálenostní matice

Typy shlukování - podle určování vzdálenosti

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

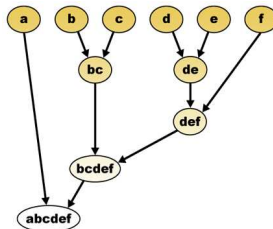
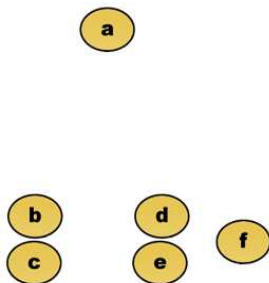
Bayesova teorie

Příště

Zkouška

- ▶ nejbližší soused
- ▶ nejvzdálenější soused
- ▶ centroidní
- ▶ párová (podobné centroidnímu, když centroid nelze spočítat)

Hierarchické aglomerativní shlukování



Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

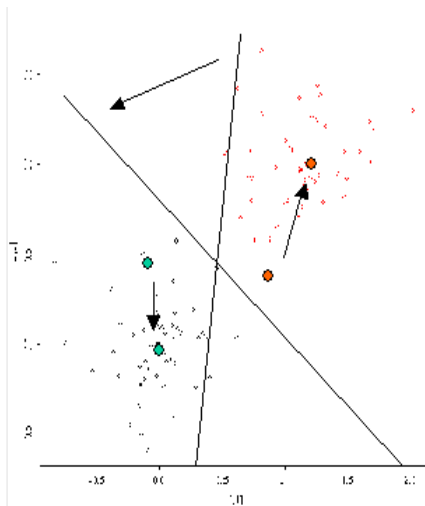
PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška



- ▶ expresní profily genů
- ▶ sekvence genů/proteinů
- ▶ proteiny podle schopnosti interakce
- ▶ struktury proteinů (CATH)
- ▶ segmentace

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shlukování genů podle expresních profilů

Shluková analýza

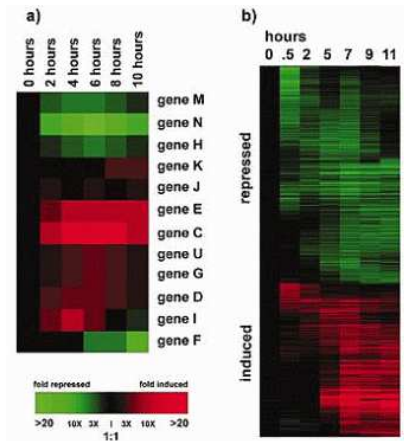
PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška



Shlukování genů podle expresních profilů
vede ke skupinám s podobnou funkcí.
Podobných výsledků lze dosáhnout
použitím informací o interakcích mezi
proteiny

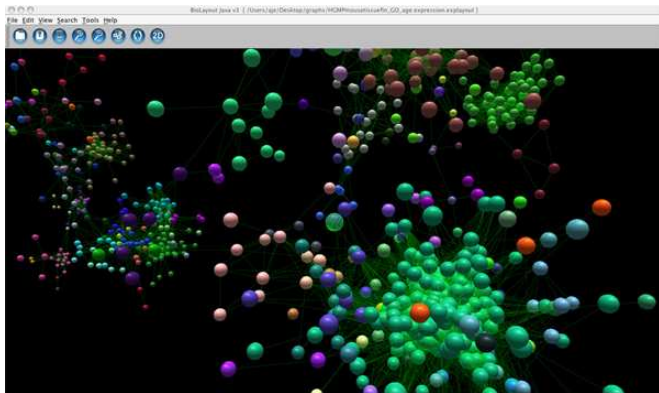
Shluková analýza

PCA - Principal component
analysisLDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška



Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Transformuje data do nového ortogonálního souřadnicového systému tak, že osi nejnižšího řádu pokrývají největší část variability dat.

Transformace souřadnicové soustavy v PCA

Shluková analýza

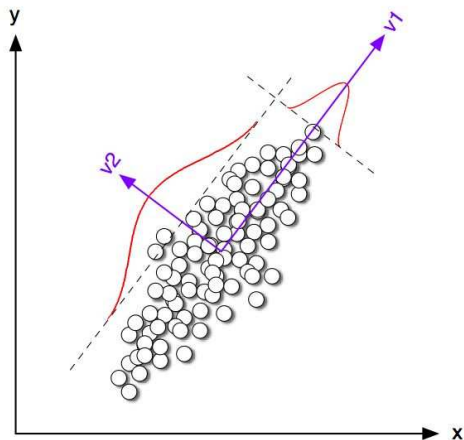
PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška



Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component
analysis

**LDA - Linear Discriminant
Analysis**

Bayesova teorie

Příště

Zkouška

Nalezne lineární transformaci jednotlivých parametrů, která poskytuje nejlepší separaci do tříd.

LDA najde lineární kombinaci vlastností, podle které lze nejlépe data klasifikovat

Shluková analýza

PCA - Principal component
analysis

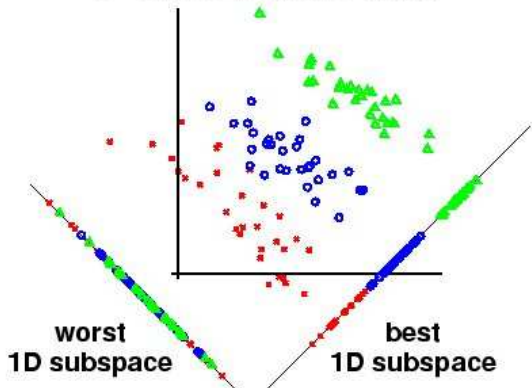
LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

3-class feature data



Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component
analysisLDA - Linear Discriminant
Analysis**Bayesova teorie**

Příště

Zkouška

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Lze odvodit ze vztahu $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$

V případě aplikace v klasifikaci a statistickém učení budou A různé hypotézy o příslušnosti dat k určité třídě. B budou pozorované parametry.

Příklad klasifikace s použitím Bayesovy teorie

Shluková analýza

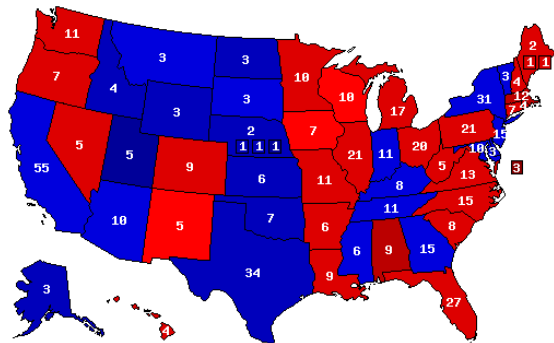
PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška



Hypotéza 1) bydlí na západě 2) bydlí na východě USA
Experiment: zavolám náhodně zvolenému američanovi a zeptám se koho volil, republikány (modře) nebo demokraty (červeně). Výsledek: podle odpovědi se změní pravděpodobnost pravdivosti jednotlivých hypotéz

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Zkouška A107

30.5.2008 14:00

13.6.2008 16:00

20.6.2008 14:00

28.6.2008 10:00 opravny

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Dodatek

For Further Reading

Dodatek

For Further Reading

Dodatek

For Further Reading

X