

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 2008

Předchozí týden

Algoritmy v bioinformatice

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

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Statistické učení a klasifikace

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

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

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Outline

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

Shlukování

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

Typy shlukování

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

Typy shlukování - podle vzdálenostní funkce

- ▶ 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
 - ▶ Levinshstein
 - ▶ Hamming

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Data můžou být definovaná vlastnostmi (souřadnice)

A = (3,1,"white")

B = (4,1,"blue")

C = (3,2,"white")

D = (3,0,"red")

Data mohou být definovaná pomocí párové vzdálenosti

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

Vzdálenostní matice

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

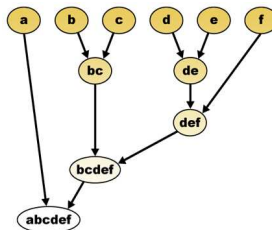
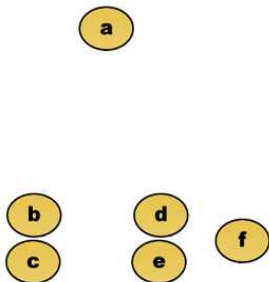
Příště

Zkouška

Typy shlukování - podle určování vzdálenosti

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

Hierarchické aglomerativní shlukování



Shluková analýza

PCA - Principal component analysis

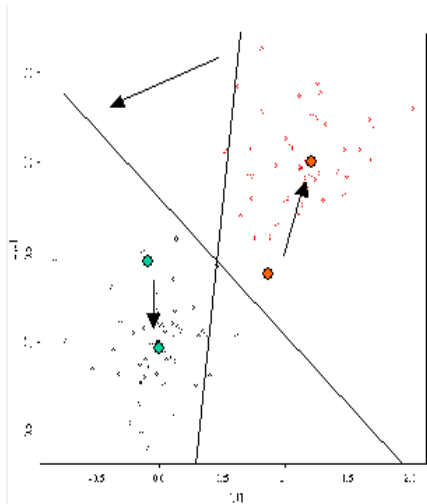
LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Nehierarchické centroidní 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

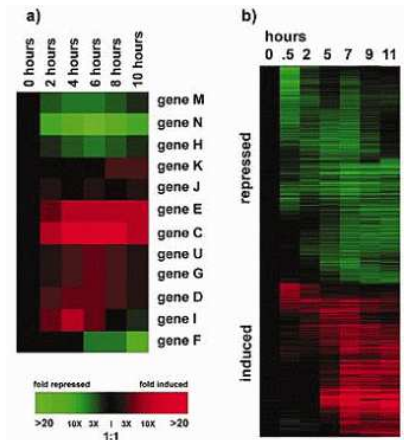
Příště

Zkouška

Aplikace shlukování v bioinformatice

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

Shlukování genů podle expresních profilů



Shluková analýza

PCA - Principal component analysis

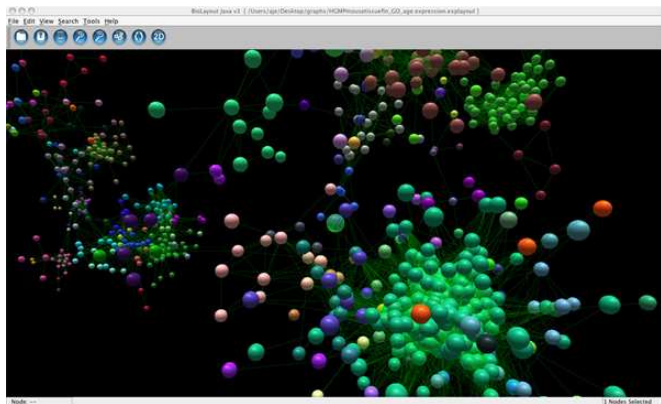
LDA - Linear Discriminant Analysis

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



<http://cgg.ebi.ac.uk/services/biayout/>

Shluková analýza

PCA - Principal component analysis

LDA - Linear Discriminant Analysis

Bayesova teorie

Příště

Zkouška

Outline

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

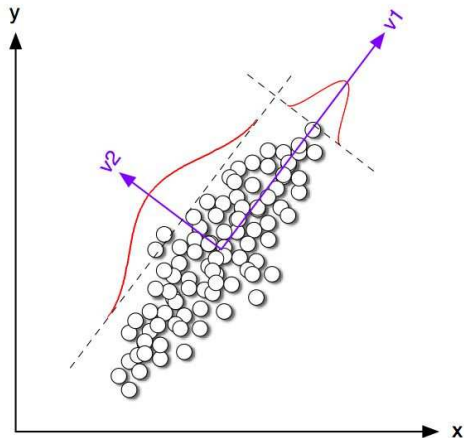
Příště

Zkouška

PCA

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

Outline

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

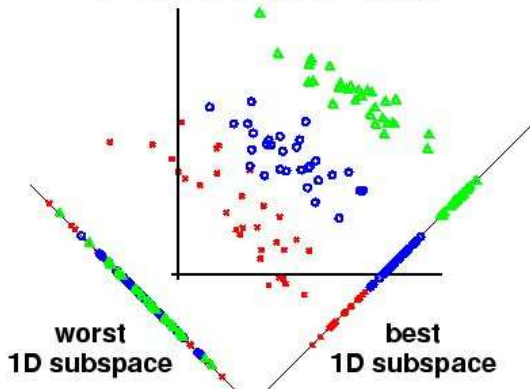
Zkouška

LDA

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

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

3-class feature data



Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Outline

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

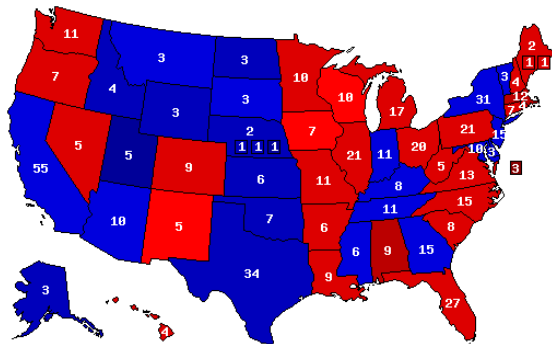
Bayesova veta

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

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



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

Outline

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

Příště

Zkouška B204

27.5.2008 8:30

3.6.2008 11:00

19.6.2008 13:00

V ISe bonus za identifikáciu osobností z oblasti genetiky,
molekul'arnej biológie a bioinformatiky.

Shluková analýza

PCA - Principal component
analysis

LDA - Linear Discriminant
Analysis

Bayesova teorie

Příště

Zkouška

Outline

Dodatek

Dodatek

For Further Reading

Dodatek

For Further Reading

For Further Reading
X