

Independent Component Analysis

Eva Bujnošková

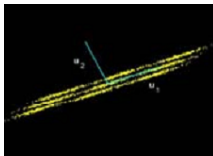
PA055

Analýza nezávislých komponent

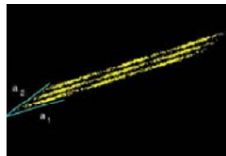
- smysl, rozdíl oproti PCA

- statistická metoda
- dekompozice dat na zdroje
- hledání nezávislých komponent – podobná faktorové analýze, vyžaduje negaussovská data

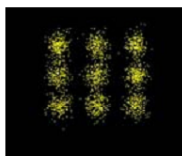
vysvětlená
variabilita



PCA

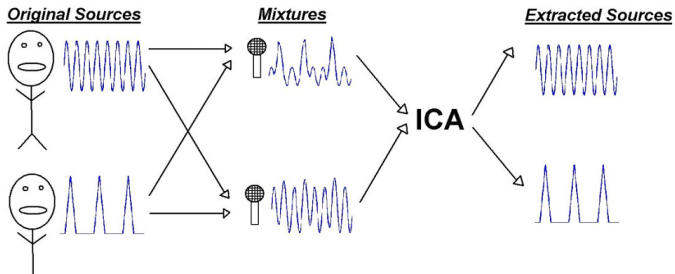
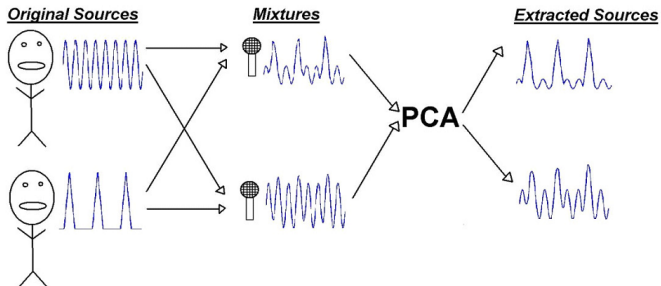


nezávislost
zdrojů

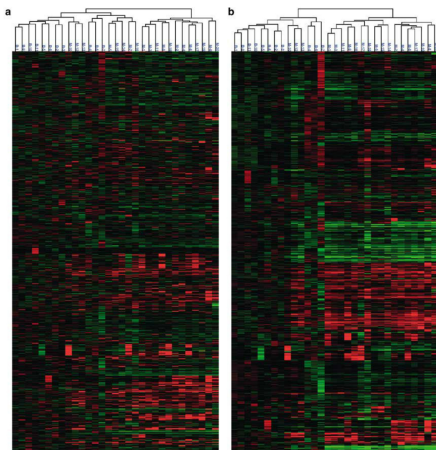
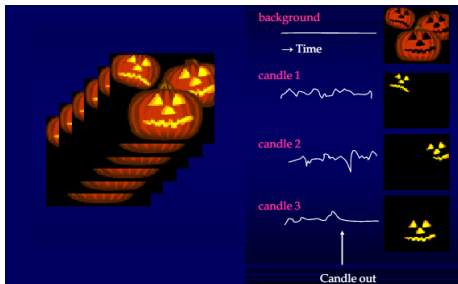


ICA

PCA vs. ICA



Příklady použití ICA



- separace artefaktů, hledání skrytých faktorů ve finančních datech
- redukce šumu v obrazech, separace signálů v telekomunikacích
- analýza fMRI dat, microarrays (pro clustering genů)

rovnice

- dekompozice dat na zdroje – lineární kombinace statisticky nezávislých komponent

$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

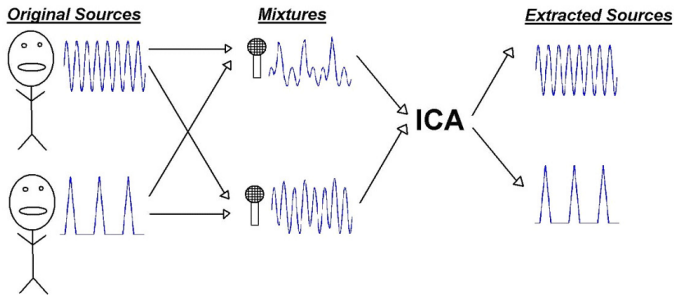
- zdroje (komponenty) \mathbf{S} neznámé
- mixážní matice \mathbf{A} neznámá
- známé pouze odezvy (měřená data) \mathbf{X} na senzorech (min N senzorů pro zjištění N zdrojů)

požadavky na data

- statistická nezávislost zdrojů
- negaussovitost – data mají jiné než normální rozložení

- odečtení průměru: $E(x) = 0$
- bělení pro zajištění nekorelovaných komponent
- PCA pro snížení dimenzionality dat
- kroky závislé na aplikaci – př. filtrování

- iterativní výpočet
- uživatel volí počet komponent → příliš moc nebo příliš málo
- metody měření statistické nezávislosti:
 - - vzájemná informace
 - negaussovitost zdrojů
 - odhad věrohodnosti (maximum likelihood)
- ICALAB – toolbox Matlabu
- konkrétní algoritmy např. FastICA nebo InfoMax



—

- pouhý rozklad signálu – slepá separace signálů
- interpretace uživatelem
- nelze použít na testování hypotéz
- ICA neodhadne pořadí komponent
- ICA neurčí energii komponenty a znaménko

+

- dekompozice směsi dat na nezávislé signály
- nevyžaduje expertní znalost o datech
- častěji vhodnější a smysluplnější komponenty než v případě PCA



HYVÄRINEN, A. and E. OJA

Independent Component Analysis: algorithms and applications.
Neural Networks 13 (2000) 411–430



LANGLOIS, D. and S. CHARTIER and D. GOSSELIN

An Introduction to Independent Component Analysis: InfoMax and FastICA algorithms.
Tutorials in Quantitative Methods of Psychology 6 (2010) 31–38



SAIDI, S. A. and C. M. HOLLAND et al.

Independent component analysis of microarray data in the study of endometrial cancer.
Oncogene 23 (2004) 6677–6683

Děkuji za pozornost.