SOUČINOVÉ DISTRIBUČNÍ SMĚSI

I. část: EM algoritmus

Jiří Grim

Ústav teorie informace a automatizace AV ČR

Oddělení rozpoznávání obrazů

Květen 2008

Přednáška je volně k dispozici na adrese http://www.utia.cas.cz/RO

ÚTIA

Outline

Využití směsí: aproximace neznámé distribuce

Řešení založené na odhadu rozložení pravděpodobnosti z dat:
problém rozpoznávání, problém predikce, statistické modely, ...

Výchozí informace:

soubor nezávislých pozorování $\mathcal{S}$ (trénovací data) s nějakým neznámým rozložením pravděpodobnosti $P^*(\mathbf{x})$

$$\mathcal{S} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(K)}\}, \qquad \mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_N^{(k)}) \in \mathcal{X}$$

Princip metody součinových směsí:

aproximace neznámého rozložení pravděpodobnosti $P^*(\mathbf{x})$ pomocí distribuční směsi součinových komponent

$$P(\mathbf{x}) = \sum_{m=1}^{M} F(\mathbf{x}|m) f(m), \quad \sum_{m=1}^{M} f(m) = 1, \quad F(\mathbf{x}|m) = \prod_{n=1}^{N} f_n(x_n|m)$$

$f_n(x_n|m)$: hustoty nebo diskrétní distribuce $\left(\sum_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x}|m) = 1\right)$

Součinové směsi jako semiparametrický model

parametrický přístup: např. odhad normální hustoty pravděpodobnosti

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det A}} \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{c})^T A^{-1}(\mathbf{x}-\mathbf{c})\}$$

průměr: $\mathbf{c} = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} \mathbf{x}$, kovariační matice: $A = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} (\mathbf{x}-\mathbf{c})(\mathbf{x}-\mathbf{c})^T$

neparametrický přístup: jádrový (Parzenův) odhad

Theorem (Parzen, 1962)

$$P(\mathbf{x}) = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{y}\in\mathcal{S}} \prod_{n=1}^{N} \frac{1}{\sqrt{(2\pi)}\sigma_n} \exp\left\{\frac{(x_n-y_n)^2}{2\sigma_n^2}\right\}$$

Součinové směsi představují "semiparametrický přístup":

- kompromis mezi jednoduchostí parametrického modelu a obecností neparametrických odhadů
- umožňuje snadné odvození marginálních hustot Odvození marginál
- efektivní odhad parametrů směsi pomocí EM algoritmu

Výpočet odhadu směsi pomocí EM algoritmu

EM algoritmus: maximalizuje věrohodnostní funkci (kritérium)

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \log P(\mathbf{x}) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \log[\sum_{m=1}^{M} f(m)F(\mathbf{x}|m)]$$

Iterační rovnice: $(\mathbf{x} \in \mathcal{S}, \quad \mathcal{S} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(K)}\})$

$$q(m|\mathbf{x}) = \frac{f(m)F(\mathbf{x}|m)}{\sum_{j=1}^{M} f(j)F(\mathbf{x}|j)}, \qquad f^{'}(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})$$

$$F^{'}(.|m) = \arg\max_{F(.|m)} \Big\{ \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log F(\mathbf{x}|m) \Big\}, \qquad m = 1, 2, \dots, M$$

---

pro součinové komponenty: $F(\mathbf{x}|m) = \prod_{n=1}^{N} f_n(x_n|m)$

$$\Rightarrow \qquad f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \Big\{ \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log f_n(x_n|m) \Big\}, \quad n = 1, 2, \dots, N$$

POZN. Počet komponent směsi je nutné zvolit předem.

Monotónní vlastnost EM algoritmu (Schlesinger, 1968)

Posloupnost hodnot věrohodnostní funkce $\{L^{(t)}\}_{t=0}^{\infty}$ je neklesající:

$$L^{(t+1)} - L^{(t)} \geq 0, \quad t = 0, 1, 2, \ldots$$

▸ Důkaz monotónní vlastnosti ▸ Alternativní důkaz

a pokud je shora omezená, konverguje monotónně k lokálnímu nebo globálnímu maximu (nebo sedlovému bodu):

$$\lim_{t \to \infty} L^{(t)} = L^* < \infty.$$

Z existence konečné limity $L^* < \infty$ plynou následující nutné podmínky konvergence: ▸ Důkaz

$$\lim_{t \to \infty} (L^{(t+1)} - L^{(t)}) = 0 \quad \Rightarrow$$

$$\Rightarrow \quad \lim_{t \to \infty} ||f^{(t+1)}(\cdot) - f^{(t)}(\cdot)|| = 0, \quad \lim_{t \to \infty} ||q^{(t+1)}(\cdot|\mathbf{x}) - q^{(t)}(\cdot|\mathbf{x})|| = 0$$

POZN. Z konvergence posloupnosti $\{L^{(t)}\}_{t=0}^{\infty}$ neplyne automaticky konvergence posloupností odhadovaných parametrů směsi!

Příklad EM algoritmu - obecná normální směs

data: $\mathcal{S} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}\}, \ \mathbf{x}^{(k)} \in \boldsymbol{R}^N$

$$F(\mathbf{x}|\mathbf{c}_m, A_m) = \frac{1}{\sqrt{((2\pi)^N \det A_m)}} \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{c}_m)^T A_m^{-1}(\mathbf{x}-\mathbf{c}_m)\}$$

$$L = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} \log[\sum_{m=1}^{M} F(\mathbf{x}|\mathbf{c}_m, A_m) f(m)]$$

Iterační rovnice: $(\mathbf{x} \in \mathcal{S})$

$$q(m|\mathbf{x}) = \frac{f(m)F(\mathbf{x}|\mathbf{c}_m, A_m)}{\sum_{j=1}^{M} f(j)F(\mathbf{x}|\mathbf{c}_j, A_j)}, \quad m = 1, 2, \ldots, M$$

$$f^{'}(m) = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}), \qquad \mathbf{c}_m^{'} = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\mathbf{x}$$

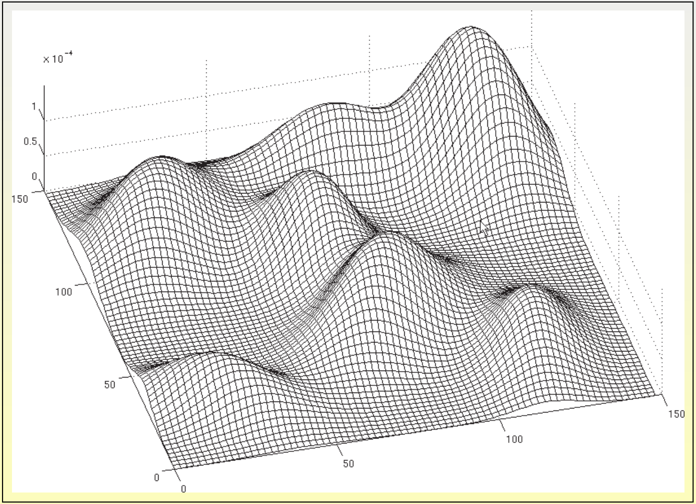
$$A_m^{'} = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})(\mathbf{x}-\mathbf{c}_m^{'})(\mathbf{x}-\mathbf{c}_m^{'})^T,$$

Explicitní řešení

POZN. Možné výpočetní problémy: $\{\det A_m \to 0\} \Rightarrow \{L \to \infty\}$

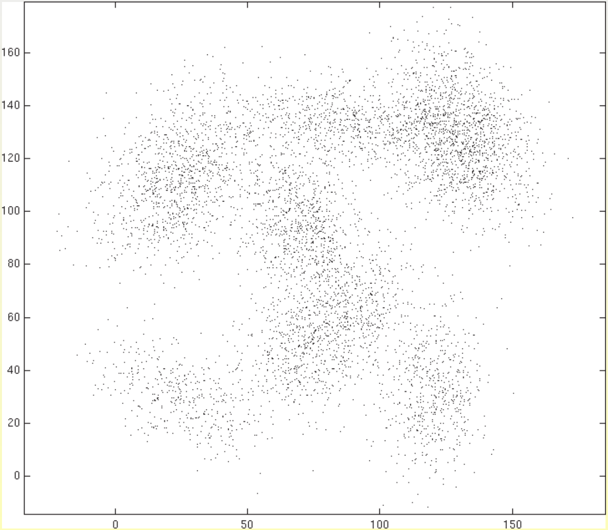
ÚTIA

Příklad normální směsi



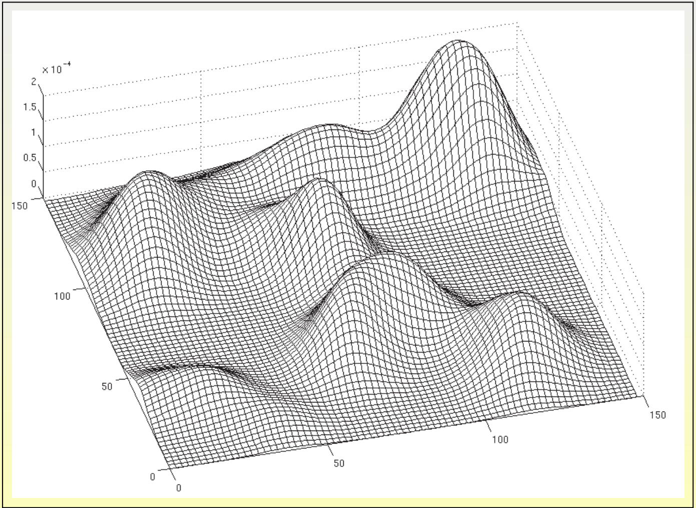
dimenze distribuční směsi $N = 2$, počet komponent směsi $M = 7$

Data náhodně generovaná podle normální směsi (M=7)



počet bodů: $|\mathcal{S}| = 6000$

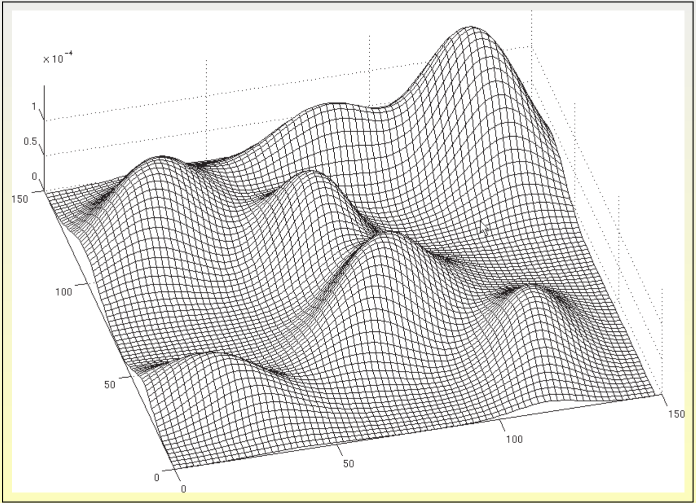
Příklad odhadu normální směsi (M=28)



použitý počet komponent směsi $M = 28$ ($\neq 7$)

(SROVNÁNÍ: jádrový odhad)

Příklad normální směsi



dimenze distribuční směsi $N = 2$, počet komponent směsi $M = 7$

Příklad EM algoritmu - součinová normální směs

data: $\mathcal{S} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(K)}\}, \ \mathbf{x}^{(k)} \in \boldsymbol{R}^N$

$$F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{mn}} \exp\left\{ -\frac{(x_n - \mu_{mn})^2}{2\sigma_{mn}^2} \right\}, \quad m = 1, 2, \dots, M$$

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \log[\sum_{m=1}^{M} f(m) F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m)]$$

iterační rovnice:

$$q(m|\mathbf{x}) = \frac{f(m)F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m)}{\sum_{j=1}^{M} f(j)F(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j)}, \qquad \mathbf{x} \in \mathcal{S}, \ m = 1, 2, \dots, M$$

$$f^{'}(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}), \qquad \mu^{'}_{mn} = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})} \sum_{\mathbf{x}\in\mathcal{S}} x_n q(m|\mathbf{x})$$

$$(\sigma^{'}_{mn})^2 = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})} \sum_{\mathbf{x}\in\mathcal{S}} (x_n - \mu^{'}_{mn})^2 q(m|\mathbf{x}), \quad n = 1, 2, \dots, N$$

POZN. Diagonální kov. matice $\approx$ jednodušší a stabilnější výpočet

Příklad EM algoritmu - obecná diskrétní součinová směs

$\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathcal{X}, \quad \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_N$
$x_n \in \mathcal{X}_n, \; |\mathcal{X}_n| < \infty$ $\approx$ diskrétní proměnné s konečným počtem hodnot
(např. dotazníky, popisy herních situací, obrázky textur $x_n \in [0, 255]$)

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \log[\sum_{m=1}^{M} f(m) F(\mathbf{x}|m)], \quad F(\mathbf{x}|m) = \prod_{n=1}^{N} f_n(x_n|m)$$

iterační rovnice: $(\mathbf{x} \in \mathcal{S}, \quad \mathcal{S} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}\})$

$$q(m|\mathbf{x}) = \frac{f(m) F(\mathbf{x}|m)}{\sum_{j=1}^{M} f(j) F(\mathbf{x}|j)}, \qquad f'(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x})$$

$$f_n'(\xi|m) = \frac{1}{\sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x})} \sum_{\mathbf{x} \in \mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x})$$

Podrobnější odvození

POZN. Diskrétní součinová směs není identifikovatelná Důkaz
(problém při shlukové analýze $\times$ výhoda při aproximaci)

ÚTIA

Příklad EM algoritmu - směs Bernoulliho rozložení

$x_n \in \{0,1\}, \quad \mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathcal{X}, \quad \mathcal{X} = \{0,1\}^N \quad \approx \quad$ binární data
(např. číslice na binárním rastru, výsledky biochemických testů a pod.)

$$F(\mathbf{x}|m) = F(\mathbf{x}|\boldsymbol{\theta}_m) = \prod_{n=1}^{N} f_n(x_n|m) = \prod_{n=1}^{N} \theta_{mn}^{x_n}(1-\theta_{mn})^{1-x_n}$$

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \log[\sum_{m=1}^{M} f(m) F(\mathbf{x}|\boldsymbol{\theta}_m)], \qquad \mathcal{S} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}\}$$

iterační rovnice:

$$q(m|\mathbf{x}) = \frac{f(m)F(\mathbf{x}|\boldsymbol{\theta}_m)}{\sum_{j=1}^{M} f(j)F(\mathbf{x}|\boldsymbol{\theta}_j)}, \quad \mathbf{x}\in\mathcal{S}, \quad m = 1, 2, \ldots, M$$

$$f^{'}(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}), \qquad \theta_{mn}^{'} = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})} \sum_{\mathbf{x}\in\mathcal{S}} x_n q(m|\mathbf{x})$$

POZN. Možný problém přesnosti součinů velkého počtu parametrů θ_{mn}.

Poznámky k implementaci EM algoritmu

"Ideální" postup implementace EM algoritmu:

generování umělých dat podle směsi + reidentifikace parametrů

- implementace EM algoritmu: jako cyklus přes data (pro $|\mathcal{S}| >> 1$)

$$\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \to f^{'}(m), \qquad \sum_{\mathbf{x}\in\mathcal{S}} x_n \, q(m|\mathbf{x}) \to \mu^{'}_{mn}, \theta^{'}_{mn}$$

- podmínka ukončení výpočtu: $(L^{'} - L)/L < \epsilon, \quad (\epsilon \approx 10^{-3} - 10^{-5})$ (v konečné fázi výpočtu obvykle dochází k nadměrnému přizpůsobení odhadovaných parametrů k datům, tzv. "overpeaking")
- EM algoritmus spontánně potlačuje váhy přebytečných komponent, z rozložení vah lze posoudit potřebný počet komponent
- pro data s velkou dimenzí ($N >> 1$) je typický malý "překryv" komponent, tj. poměrně vysoké hodnoty $\bar{q}_{max} \approx 0.85 - 0.99$

$$q_{max}(\mathbf{x}) = \max_{m\in\mathcal{M}} \{q(m|\mathbf{x})\}, \qquad \bar{q}_{max} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q_{max}(\mathbf{x})$$

problém přesnosti reprezentace čísel

ÚTIA

Identifikace směsi × aproximace pomocí směsi

Problém identifikace směsi (např. shluková analýza)

- cílem je zjistit počet komponent a odhadnout parametry směsi
- je třeba aby odhadovaná směs byla identifikovatelná
- věrohodnostní funkce směsi má téměř vždy lokální maxima (zvláště v mnohorozměrných prostorech a při malém počtu dat)
- $\Rightarrow$ kvalita odhadů vypočtených pomocí EM algoritmu závisí na zvoleném počtu komponent a počátečních hodnotách parametrů

Problém aproximace neznámé pravděpodobnostní distribuce

- cílem je co nejpřesnější aproximace neznámého rozložení pravděpodobnosti pomocí součinové distribuční směsi Problém aproximace
- směs nemusí být identifikovatelná; není-li, naopak se snižuje riziko lokálních extrémů a zvyšuje se přesnost aproximace
- přesný počet komponent směsi není důležitý (× trvání výpočtu, "overfitting")
- při větším počtu komponent je dosažená přesnost aproximace pro různá lokální maxima srovnatelná

Výpočetní vlastnosti směsí s velkým počtem komponent

Předpoklad: dostatečně velký počet dat

- existence lokálních extrémů není důležitá protože při velkém počtu komponent je hodnota kriteria v různých lokálních maximech srovnatelná
- $\Rightarrow$ inicializace parametrů směsi nemá podstatný vliv, počáteční parametry komponent je možné generovat náhodně
- při velkém počtu komponent ($M \approx 10^1 - 10^2$) má značná část komponent obvykle řádově nižší váhu než je maximální váha komponenty (část lze zanedbat $\Rightarrow$ přesný počet není důležitý)
- iterace EM algoritmu v závěrečné pomalé fázi výpočtu mají obvykle malý vliv na hodnotu kritéria a přesnost aproximace a lze je vynechat
- závěrečná fáze výpočtu obvykle zvyšuje riziko "přeurčení" parametrů (overfitting), tzn. včasné ukončení iterací může zlepšit kvalitu řešení

POZN. Uvedené vlastnosti neplatí striktně, závisí na konkrétních datech.

Modifikace EM algoritmu - vážená data

$\gamma(\mathbf{x}) > 0$: relativní četnost výskytu vektoru $\mathbf{x}$ ($\approx$ váha) v posloupnosti $\mathcal{S}$
$\mathcal{M} = \{1, \ldots, M\}, \quad \mathcal{N} = \{1, \ldots, N\} \approx$ indexové množiny

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \log[\sum_{m \in \mathcal{M}} f(m) F(\mathbf{x}|m)] = \sum_{\mathbf{x} \in \mathcal{X}} \gamma(\mathbf{x}) \log[\sum_{m \in \mathcal{M}} f(m) F(\mathbf{x}|m)]$$

$\bar{\mathcal{S}} = \{\mathbf{x} \in \mathcal{S} : \gamma(\mathbf{x}) > 0\}$: sčítání lze omezit na vektory $\mathbf{x} \in \bar{\mathcal{S}}$:

vážené iterační rovnice: $(m \in \mathcal{M}, \ n \in \mathcal{N}, \ \mathbf{x} \in \bar{\mathcal{S}})$

$$q(m|\mathbf{x}) = \frac{f(m) F(\mathbf{x}|m)}{\sum_{j \in \mathcal{M}} f(j) F(\mathbf{x}|j)}, \qquad F(\mathbf{x}|m) = \prod_{n \in \mathcal{N}} f_n(x_n|m)$$

$$f'(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) = \sum_{\mathbf{x} \in \bar{\mathcal{S}}} \gamma(\mathbf{x}) q(m|\mathbf{x})$$

$$f_n'(.|m) = \arg \max_{f_n(.|m)} \left\{ \sum_{\mathbf{x} \in \bar{\mathcal{S}}} \gamma(\mathbf{x}) q(m|\mathbf{x}) \log f_n(x_n|m) \right\}$$

POUŽITÍ: agregace dat, "nekonečný" datový soubor: $\gamma(\mathbf{x}) = P^*(\mathbf{x})$

ÚTIA

Modifikace EM algoritmu - neúplné datové vektory

neúplná data: $\mathbf{x} = (x_1, -, x_3, x_4, -, -, x_7, \ldots, x_N) \in \mathcal{X}$

$\mathcal{N}(\mathbf{x}) = \{n \in \mathcal{N} : \text{souřadnice } x_n \text{ je definovaná v } \mathbf{x}\}, \quad \mathbf{x} \in \mathcal{X}$

$\mathcal{S}_n = \{\mathbf{x} \in \mathcal{S} : n \in \mathcal{N}(\mathbf{x})\}, \quad \approx$ vektory $\mathbf{x} \in \mathcal{S}$ s definovanou souřadnicí x_n

předpoklad: součinové komponenty $\Rightarrow$ Snadný výpočet marginál

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \log\left[\sum_{m=1}^{M} f(m) \bar{F}(\mathbf{x}|m)\right], \qquad \bar{F}(\mathbf{x}|m) = \prod_{n \in \mathcal{N}(x)} f_n(x_n|m)$$

iterační rovnice: $(m \in \mathcal{M}, n \in \mathcal{N}, \mathbf{x} \in \mathcal{S})$

$$q(m|\mathbf{x}) = \frac{f(m)\bar{F}(\mathbf{x}|m)}{\sum_{j=1}^{M} f(j)\bar{F}(\mathbf{x}|j)}, \qquad f^{'}(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x})$$

$$f_n^{'}(.|m) = \arg \max_{f_n(.|m)} \left\{ \sum_{\mathbf{x} \in \mathcal{S}_n} q(m|\mathbf{x}) \log f_n(x_n|m) \right\}$$

POZN. Nahrazování chybějících údajů pomocí odhadů ovlivňuje data.

Strukturní model směsi (Grim et al. 1986, 1999, 2002)

binární strukturní parametry: $\phi_m = (\phi_{m1}, \ldots, \phi_{mN}) \in \{0,1\}^N$

$$F(\mathbf{x}|m) = \prod_{n \in \mathcal{N}} f_n(x_n|m)^{\phi_{mn}} f_n(x_n|0)^{1-\phi_{mn}}, \qquad (\text{obvykle: } f_n(x_n|0) = P_n(x_n))$$

$\phi_{mn} = 0$: místo $f_n(x_n|m)$ se v součinu dosadí fixní distribuce $f_n(x_n|0)$

$$P(\mathbf{x}) = \sum_{m \in \mathcal{M}} F(\mathbf{x}|m) f(m) = \sum_{m \in \mathcal{M}} F(\mathbf{x}|0) G(\mathbf{x}|m, \phi_m) f(m)$$

$$G(\mathbf{x}|m, \phi_m) = \prod_{n \in \mathcal{N}} \left[\frac{f_n(x_n|m)}{f_n(x_n|0)} \right]^{\phi_{mn}}, \qquad F(\mathbf{x}|0) = \prod_{n \in \mathcal{N}} f_n(x_n|0)$$

motivace: **"distribuce pozadí "** $F(\mathbf{x}|0)$ **se vykrátí v Bayesově vzorci:**

$$p(\omega|\mathbf{x}) = \frac{P(\mathbf{x}|\omega) p(\omega)}{P(\mathbf{x})} = \frac{\sum_{m \in \mathcal{M}_\omega} G(\mathbf{x}|m, \phi_m) f(m)}{\sum_{j \in \mathcal{M}} G(\mathbf{x}|j, \phi_j) f(j)} \approx \sum_{m \in \mathcal{M}_\omega} G(\mathbf{x}|m, \phi_m) f(m)$$

Vlastnosti strukturního modelu

Strukturní modifikace EM algoritmu (diskrétní proměnné)

strukturní optimalizace je součástí odhadu parametrů směsi

$$L = \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} \log \Big[\sum_{m \in \mathcal{M}} F(\mathbf{x}|0) G(\mathbf{x}|m, \phi_m) f(m) \Big], \qquad F(\mathbf{x}|0) = \prod_{n \in \mathcal{N}} f_n(x_n|0)$$

iterační rovnice: $(m \in \mathcal{M}, n \in \mathcal{N}, \mathbf{x} \in \mathcal{S})$

$$q(m|\mathbf{x}) = \frac{G(\mathbf{x}|m, \phi_m) f(m)}{\sum_{j \in \mathcal{M}} G(\mathbf{x}|j, \phi_j) f(j)}, \qquad f'(m) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x})$$

$$f_n'(\xi|m) = \frac{1}{\sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x})} \sum_{\mathbf{x} \in \mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x})$$

strukturní optimalizace: $\phi'_{mn} = 1$ pro r nejvyšších hodnot γ'_{mn}

$$\gamma'_{mn} = \sum_{\mathbf{x} \in \mathcal{S}} \frac{q(m|\mathbf{x})}{|\mathcal{S}|} \log \Big[\frac{f_n'(x_n|m)}{f_n(x_n|0)} \Big] = f'(m) \sum_{x_n \in \mathcal{X}_n} f_n'(x_n|m) \log \frac{f_n'(x_n|m)}{f_n(x_n|0)}$$

Odvození kritéria

Celkové shrnutí

Výpočetní vlastnosti součinových distribučních směsí

- efektivní výpočet parametrů směsi pomocí EM algoritmu z velkých datových souborů v mnohorozměrném prostoru (!)
- snadný výpočet marginálních rozložení pravděpodobnosti (!)
- vhodné pro aproximaci multimodálních rozložení pravděpodobnosti
- při velkém počtu komponent se vlastnosti součinové směsi blíží obecnosti neparametrického jádrového odhadu
- směsi jsou jednodušší než jádrové odhady (méně komponent) není třeba řešit problém optimalizace vyhlazení
- možnost odhadu parametrů směsi z neúplných datových vektorů
- EM algoritmus lze použít na vážená data (specifické aplikace, možnost zrychlení výpočtu agregací dat)
- součinové směsi lze interpretovat jako neuronovou síť (až na úrovni funkčních vlastností neuronů)

Historie problému Literatura

Z historie problému identifikace směsí

- **Pearson (1894):** "Contributions to the mathematical theory of evolution. 1. Dissection of frequency curves." *Philosophical Transactions of the Royal Society of London* **185**, 71-110. (odhady směsi dvou jednorozměrných normálních hustot metodou momentů)
- **Kale (1962), Hasselblad (1966), Day (1969), Wolfe (1970):** jednoduché iterační schema (EM algoritmus) původně odvozené intuitivně algebraickou úpravou věrohodnostních rovnic pro normální směsi, metoda snadno použitelná v mnoharozměrných případech, v každé iteraci zvyšuje hodnotu věrohodnostní funkce
- **Hosmer (1973):** "Iterative m.-l. estimates were proposed by Hasselblad and subsequently have been looked at by Day, Hosmer and Wolfe"
- **Peters a Walker (1978):** "... we have observed in experiments that the convergence is monotone, i.e. that the likelihood function is actually increased in each iteration, but we have been unable to prove it."

Z historie problému identifikace směsí

- **Schlesinger M.I. (1968):** "Relation between learning and self learning in pattern recognition", (in Russian), *Kibernetika*, (Kiev), No. 2, 81-88. (první důkaz monotonie EM algoritmu)

достигается тогда, когда величины x_i пропорциональны величинам α_i.

Лемма легко может быть доказана для $s = 2$, а затем методом математической индукции обобщена для любого s.

Теорема 1. Пусть $A^{(t)}$, $A^{(t+1)}$ — значения неизвестных параметров, полученных соответственно после t-ой и $(t+1)$-ой итераций алгоритма самообучения. В таком случае, если $A^{(t)} \neq A^{(t+1)}$, то $L(A^{(t)}) < L(A^{(t+1)})$.

Доказательство. На основании того, что $\sum_{k=1}^{s} \alpha_{ik} = 1$ для всех i (см. формулу (7)), выражение для $L(A^{(t)})$ можно записать следующим образом:

$$L(A^{(t)}) = \sum_{i=1}^{m} \log \sum_{k=1}^{s} p_k^{(t)} \cdot p(v_i/a_k^{(t)}) =$$

$$= \sum_{k=1}^{s} \sum_{i=1}^{m} \alpha_{ik}(A^{(t)}) \log p_k^{(t)} +$$

$$\leqslant \sum_{k=1}^{s} \sum_{i=1}^{m} \alpha_{ik}(A^{(t)}) \log p(v_i/a_k^{(t+1)}); \qquad (10)$$

$$\sum_{i=1}^{m} \sum_{k=1}^{s} \alpha_{ik}(A^{(t)}) \log \frac{p_k^{(t)} \cdot p(v_i/a_k^{(t)})}{\sum_{k=1}^{s} p_k^{(t)} \cdot p(v_i/a_k^{(t)})} \geqslant$$

$$\geqslant \sum_{i=1}^{m} \sum_{k=1}^{s} \alpha_{ik}(A^{(t)}) \log \frac{p_k^{(t+1)} \cdot p(v_i/a_k^{(t+1)})}{\sum_{k=1}^{s} p_k^{(t+1)} \cdot p(v_i/a_k^{(t+1)})}, \qquad (11)$$

причем, по крайней мере, одно из первых двух неравенств выполняется строго.

Докажем неравенство (9).

По определению (этап 2 алгоритма) величина $p_k^{(t+1)}$ пропорциональна величине $\sum_{i=1}^{m} \alpha_{ik}(A^{(t)})$. К тому же очевидным является равенство $\sum^{s} p_k^{(t+1)} =$

Foto

- **Ajvazjan et al. (monografie, 1974):** cituje Schlesingerův výsledek
- **Isaenko a Urbach (1976):** cituje Schlesingerův výsledek

Z historie problému identifikace směsí

- **Dempster et al. (1977):** "Maximum likelihood from incomplete data via the EM algorithm." *J. Roy. Statist. Soc., B*, Vol. 39, pp.l-38.
  (Scholar Google, 2008: 12000 citací)
  navržen název EM algoritmus, směsi pouze jako speciální případ

Maximum Likelihood from Incomplete Data via the *EM* Algorithm

By A. P. Dempster, N. M. Laird and D. B. Rubin

Harvard University and Educational Testing Service

[Read before the Royal Statistical Society at a meeting organized by the Research Section on Wednesday, December 8th, 1976, Professor S. D. Silvey in the Chair]

for all $p \geq p(\varepsilon)$ and all $r \geq 1$, where each term in the sum is non-negative.

Applying assumption (2) in the theorem for $p, p+1, p+2, \ldots, p+r-1$ and summing, we obtain from (3.12)

$$\varepsilon > \lambda \sum_{j=1}^{r} (\phi^{(p+j)} - \phi^{(p+j-1)})(\phi^{(p+j)} - \phi^{(p+j-1)})^{\mathrm{T}}, \tag{3.13}$$

whence

$$\varepsilon > \lambda(\phi^{(p+r)} - \phi^{(p)})(\phi^{(p+r)} - \phi^{(p)})^{\mathrm{T}}, \tag{3.14}$$

as required to prove convergence of $\phi^{(p)}$ to some ϕ^*.

Theorem 1 implies that $L(\phi)$ is non-decreasing on each iteration of a GEM algorithm, and is strictly increasing on any iteration such that $Q(\phi^{(p+1)} | \phi^{(p)}) > Q(\phi^{(p)} | \phi^{(p)})$. The corollaries

Z historie problému identifikace směsí

- **Boyles R.A. (1983):** On the convergence of the EM algorithm. *J. Roy. Statist. Soc., B*, Vol. 45, pp. 47-50.
- **Wu C.F.J. (1983):** "On the convergence properties of the EM algorithm." *Ann. Statist.*, Vol. 11, pp. 95-103.

  problém konvergence posloupnosti parametrů směsi

> applications in statistics.
>
> However, the proof of convergence of EM sequences in DLR contains an error. The implication from (3.13) to (3.14) in their Theorem 2 fails due to an incorrect use of the triangle inequality. Additional comments on this proof are given in Section 2.2. Therefore the convergence of EM sequence as proved in their Theorems 2 and 3 is cast in doubt. Other results on the monotonicity of likelihood sequence and the convergence rate of EM sequence (Theorems 1 and 4 of DLR) remain valid.
>
> Despite its slow numerical convergence, the EM algorithm has become a very popular

- **Everitt, B.S. and D.J. Hand (1981)** - monografie
- **Titterington et al. (1985)** - monografie
- **McLachlan and Peel (2000)** - monografie
- **období 1895 - 1965:** - publikováno asi 80 prací o identifikaci směsí
- **Scholar Google (2008):** 62 000 článků s heslem "EM algorithm"

Vlastnosti neparametrického jádrového odhadu

Theorem (Parzen, 1962; Cacoullos, 1966)

Nechť $\mathcal{S}_K$ je posloupnost K nezávislých realizací N-rozměrného náhodného vektoru s nějakou neznámou hustotou pravděpodobnosti $P^*(\mathbf{x})$.
Neparametrický jádrový odhad hustoty $P^*(\mathbf{x})$

$$P(\mathbf{x}) = \frac{1}{K} \sum_{y \in \mathcal{S}_K} \prod_{n=1}^{N} \frac{1}{\sqrt{(2\pi)}\sigma_K} \exp\left\{ \frac{(x_n - y_n)^2}{2\sigma_K^2} \right\}$$

je v každém bodě spojitosti $P^*(\mathbf{x})$ asymptoticky nestranný, tj. platí

$$\lim_{K \to \infty} \mathrm{E}_{\mathcal{S}_K}\{P(\mathbf{x})\} = P^*(\mathbf{x}),$$

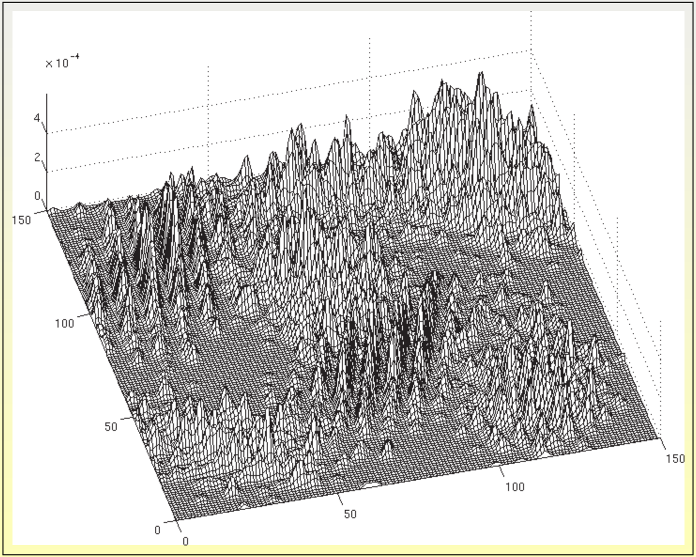
pokud $\lim_{K \to \infty} \sigma_K = 0$. *Platí-li navíc* $\lim_{K \to \infty} K\sigma_K^N = \infty$,
potom $P(\mathbf{x})$ je také asymptoticky konzistentní v kvadratickém průměru:

$$\lim_{K \to \infty} \mathrm{E}_{\mathcal{S}_K}\{[P^*(\mathbf{x}) - P(\mathbf{x})]^2\} = 0.$$

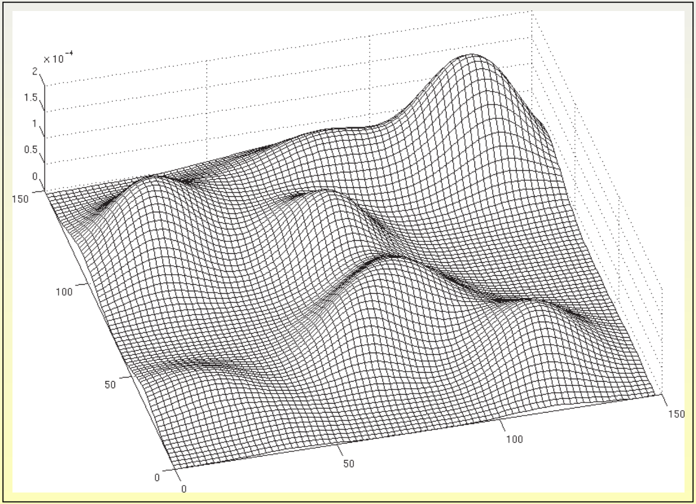
NEVÝHODY: výpočetní náročnost, problém vyhlazení

Zpět: Srovnání

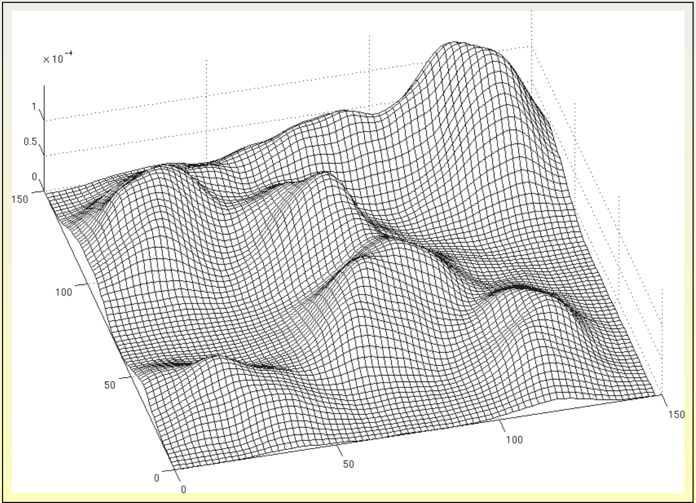
"Podhlazený" jádrový odhad



Příliš vyhlazený jádrový odhad



Optimálně vyhlazený jádrový odhad



(normální jádro s obecnou kov. maticí)

Zpět: Norm. směs

Odvození marginálních distribucí ze součinové směsi

$$P(\mathbf{x}) = \sum_{m=1}^{M} f(m)F(\mathbf{x}|m) = \sum_{m=1}^{M} f(m)\prod_{n=1}^{N} f_n(x_n|m), \qquad \mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}$$

$$\sum_{x_i\in\mathcal{X}_i} P(\mathbf{x}) = \sum_{m=1}^{M} f(m)(\sum_{x_i\in\mathcal{X}_i} f_i(x_i|m)) \prod_{n\in\mathcal{N}\setminus i} f_n(x_n|m) \sum_{m=1}^{M} f(m) \prod_{n\in\mathcal{N}\setminus i} f_n(x_n|m)$$

. . . .

$$\mathbf{x}_C = (x_{i_1}, x_{i_2}, \dots, x_{i_k}) \in \mathcal{X}_C, \quad \mathcal{X}_C = \mathcal{X}_{i_1} \times \cdots \times \mathcal{X}_{i_k}, \quad C = \{i_1, \dots, i_k\} \subset \mathcal{N}$$

$$P_C(\mathbf{x}_C) = \sum_{m=1}^{M} f(m)F_C(\mathbf{x}_C|m), \qquad F_C(\mathbf{x}_C|m) = \prod_{n\in C} f_n(x_n|m)$$

$$P_{n|C}(x_n|\mathbf{x}_C) = \frac{P_{nC}(x_n, \mathbf{x}_C)}{P_C(\mathbf{x}_C)} = \sum_{m=1}^{M} \frac{f(m)F_C(\mathbf{x}_C|m)}{P_C(\mathbf{x}_C)} f_n(x_n|m)$$

$$P_{n|C}(x_n|\mathbf{x}_C) = \sum_{m=1}^{M} W_m(\mathbf{x}_C) f_n(x_n|m), \qquad W_m(\mathbf{x}_C) = \frac{f(m)F_C(\mathbf{x}_C|m)}{P_C(\mathbf{x}_C)}$$

Explicitní řešení kroku M

Theorem (podrobněji viz Grim, 1982)

Nechť maximálně věrohodný odhad parametru $\mathbf{b}$ *hustoty pravděpodobnosti* $F(\mathbf{x}|\mathbf{b})$ *je aditivní funkcí dat* $\mathbf{x} \in \mathcal{S}$*:*

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \log F(\mathbf{x}|\mathbf{b}), \quad \mathbf{x} \in \mathcal{X}, \qquad \mathbf{b} \approx \textit{parametr}$$

$$\mathbf{b}^* = \arg\max_{\mathbf{b}} \left\{ \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \log F(\mathbf{x}|\mathbf{b}) \right\} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \mathbf{a}(\mathbf{x})$$

Jestliže $\gamma(\mathbf{x})$ *je relativní četnost vektoru* $\mathbf{x}$ *v* $\mathcal{S}$*, platí ekvivalentně:*

$$L = \sum_{\mathbf{x} \in \bar{\mathcal{X}}} \gamma(\mathbf{x}) \log F(\mathbf{x}|\mathbf{b}), \qquad \bar{\mathcal{X}} = \{\mathbf{x} \in \mathcal{X} : \gamma(\mathbf{x}) > 0\}, \qquad \Big(\sum_{\mathbf{x} \in \bar{\mathcal{X}}} \gamma(\mathbf{x}) = 1\Big)$$

$$\mathbf{b}^* = \sum_{\mathbf{x} \in \bar{\mathcal{X}}} \gamma(\mathbf{x}) \; \mathbf{a}(\mathbf{x}) = \arg\max_{\mathbf{b}} \left\{ \sum_{\mathbf{x} \in \bar{\mathcal{X}}} \gamma(\mathbf{x}) \log F(\mathbf{x}|\mathbf{b}) \right\}$$

Explicitní řešení kroku M - normální směs

normální směs s obecnou kovariační maticí:

$$P(\mathbf{x}) = \sum_{m=1}^{M} f(m) F(\mathbf{x}|\mathbf{c}_m, A_m)$$

$$F(\mathbf{x}|\mathbf{c}_m, A_m) = \frac{1}{\sqrt{((2\pi)^N \det A_m)}} \exp\{-\frac{1}{2}(\mathbf{x} - \mathbf{c}_m)^T A_m^{-1} (\mathbf{x} - \mathbf{c}_m)\}$$

implicitní tvar kroku M:

$$(\mathbf{c}_m^{'}, A_m^{'}) = \arg \max_{(\mathbf{c}_m, A_m)} \left\{ \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{\sum_{\mathbf{y}\in\mathcal{S}} q(m|\mathbf{y})} \log F(\mathbf{x}|\mathbf{c}_m, A_m) \right\}$$

explicitní řešení:

$$\mathbf{c}_m^{'} = \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{\sum_{\mathbf{y}\in\mathcal{S}} q(m|\mathbf{y})} \mathbf{x}, \qquad \left(\gamma(\mathbf{x}) = \frac{q(m|\mathbf{x})}{\sum_{\mathbf{y}\in\mathcal{S}} q(m|\mathbf{y})}\right)$$

$$A_m^{'} = \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{\sum_{\mathbf{y}\in\mathcal{S}} q(m|\mathbf{y})} (\mathbf{x} - \mathbf{c}_m^{'})(\mathbf{x} - \mathbf{c}_m^{'})^T$$

Zpět: EM algoritmus

Odvození M-kroku pro diskrétní součinovou směs

$$f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \left\{ \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log f_n(x_n|m) \right\}, \quad n \in \mathcal{N}, \quad m \in \mathcal{M}$$

$$f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \left\{ \sum_{\mathbf{x}\in\mathcal{S}} (\sum_{\xi\in\mathcal{X}_n} \delta(\xi, x_n)) q(m|\mathbf{x}) \log f_n(x_n|m) \right\}$$

$$f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \left\{ \sum_{\xi\in\mathcal{X}_n} \sum_{\mathbf{x}\in\mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x}) \log f_n(\xi|m) \right\}$$

$$f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \left\{ \sum_{\xi\in\mathcal{X}_n} \left(\sum_{\mathbf{x}\in\mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x}) \right) \log f_n(\xi|m) \right\}$$

$$f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \left\{ \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \sum_{\xi\in\mathcal{X}_n} \left(\frac{\sum_{\mathbf{x}\in\mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x})}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})} \right) \log f_n(\xi|m) \right\}$$

$$\Rightarrow \qquad f_n^{'}(\xi|m) = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})} \sum_{\mathbf{x}\in\mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x})$$

Zpět: EM algoritmus

Důkaz

Důkaz neidentifikovatelnosti diskrétní součinové směsi

Definice identifikovatelnosti směsi

Třída směsí $\mathcal{P} = \{P(\mathbf{x}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ je identifikovatelná, jestliže parametry $\boldsymbol{\theta}, \boldsymbol{\theta}^{'} \in \Theta$ libovolných dvou ekvivalentních směsí

$$P(\mathbf{x}, \boldsymbol{\theta}) = P(\mathbf{x}, \boldsymbol{\theta}^{'}), \quad \forall\ \mathbf{x} \in \mathcal{X}$$

se mohou lišit pouze pořadím komponent.

Theorem (srv. Teicher, 1963, 1968; Gyllenberg et al., 1994; Grim, 2001)

Libovolná diskrétní součinová směs $(x_n \in \mathcal{X}_n, \quad |\mathcal{X}_n| < \infty)$

$$P(\mathbf{x}) = \sum_{m=1}^{M} f(m)F(\mathbf{x}|m)] = \sum_{m=1}^{M} f(m) \prod_{n=1}^{N} f_n(x_n|m)$$

může být ekvivalentně vyjádřena nekonečně mnoha různými způsoby (tj. pomocí různých množin parametrů), pokud alespoň jedna z podmíněných distribucí $f_i(x_i|m)$ *splňuje podmínku*

$$0 < f_i(x_i|m) < 1, \quad \textit{pro nějaké} \ \ x_i \in \mathcal{X}_i.$$

Důkaz neidentifikovatelnosti diskrétní součinové směsi

Důkaz:

Jestliže pro nějaké $i \in \mathcal{N}, x_i \in \mathcal{X}_i$ a $m \in \mathcal{M}$ platí $0 < f_i(x_i|m) < 1$ potom jednorozměrné rozložení pravděpodobnosti $f_i(\cdot|m)$ můžeme nekonečně mnoha způsoby vyjádřit jako konvexní kombinaci dvou různých distribucí, např. $(0 < \alpha < 1,\ \beta = 1 - \alpha)$:

$$f_i(\xi|m) = \alpha f_i^{'}(\xi|m) + \beta f_i^{''}(\xi|m), \quad \xi \in \mathcal{X}_i$$

S využitím předchozí substituce můžeme napsat

$$f(m)F(\mathbf{x}|m) = f^{'}(m)F^{'}(\mathbf{x}|m) + f^{''}(m)F^{''}(\mathbf{x}|m)$$

kde

$$f^{'}(m) = \alpha f(m), \quad f^{''}(m) = \beta f(m)$$

$$F^{'}(\mathbf{x}|m) = f^{'}(x_i|m) \prod_{n \in \mathcal{N}, n \neq i} f_n(x_n|m), \quad F^{''}(\mathbf{x}|m) = f^{''}(x_i|m) \prod_{n \in \mathcal{N}, n \neq i} f_n(x_n|m)$$

a po dosazení za komponentu $f(m)F(\mathbf{x}|m)$ můžeme původní směs $P(\mathbf{x})$ vyjádřit pomocí netriviálně odlišných parametrů, cbd.

Zpět: EM algoritmus

ÚTIA

Problém omezené přesnosti reprezentace čísel

- výpočet komponent v logaritmickém tvaru (pro $N >> 1$):

$$\log[F(\mathbf{x}|m)f(m)] = \log f(m) + \sum_{n\in\mathcal{N}} \log f_n(x_n|m), \quad (\leq \log C_0, \ \Rightarrow C_0)$$

- řešení problému omezené přesnosti velmi malých čísel:

$$\exp\{-\log C_0 + \log f(m) + \sum_{n\in\mathcal{N}} \log f_n(x_n|m)\} = C_0^{-1} F(\mathbf{x}|m)f(m)$$

$$q(m|\mathbf{x}) = \frac{C_0^{-1} F(\mathbf{x}|m)f(m)}{\sum_{j=1}^{M} C_0^{-1} F(\mathbf{x}|j)f(j)} = \frac{F(\mathbf{x}|m)f(m)}{\sum_{j=1}^{M} F(\mathbf{x}|j)f(j)}$$

- možnost zanedbání (přerušení výpočtu) komponenty:

$$-\log C_0 + \log f(m) + \sum_{n\in\mathcal{N}} \log f_n(x_n|m) < \log R_{min} \Rightarrow q(m|\mathbf{x}) = 0$$

- kontrola správnosti programu: $L^{'} \geq L$

Zpět-Implementace

Důkaz monotónní konvergence EM algoritmu: $L' \geq L$

Kullback-Leiblerova informační divergence je nezáporná, tj. platí:

$$I(q(\cdot|\mathbf{x}), q'(\cdot|\mathbf{x})) = \sum_{m=1}^{M} q(m|\mathbf{x}) \log \frac{q(m|\mathbf{x})}{q'(m|\mathbf{x})} \geq 0,$$

Důkaz

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \log[\sum_{m=1}^{M} f(m) F(\mathbf{x}|m)], \qquad q(m|\mathbf{x}) = \frac{f(m) F(\mathbf{x}|m)}{\sum_{j=1}^{M} f(j) F(\mathbf{x}|j)}$$

Věrohodnostní funkci L resp. L' lze zapsat ekvivalentně pomocí $q(m|\mathbf{x})$:

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} q(m|\mathbf{x}) \log[f(m) F(\mathbf{x}|m)] - \sum_{m=1}^{M} q(m|\mathbf{x}) \log q(m|\mathbf{x}) \right\}$$

$$L' = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} q(m|\mathbf{x}) \log[f'(m) F'(\mathbf{x}|m)] - \sum_{m=1}^{M} q(m|\mathbf{x}) \log q'(m|\mathbf{x}) \right\}$$

Důkaz monotónní konvergence EM algoritmu: $L' \geq L$

Předchozí vzorce použijeme k vyjádření přírůstku věrohodnostní funkce:

$$L' - L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} q(m|\mathbf{x}) \log \left[\frac{f'(m) F'(\mathbf{x}|m)}{f(m) F(\mathbf{x}|m)} \right] + \sum_{m=1}^{M} q(m|\mathbf{x}) \log \frac{q(m|\mathbf{x})}{q'(m|\mathbf{x})} \right\}$$

Druhý člen na pravé straně představuje Kullback-Leiblerovu divergenci:

$$L' - L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} q(m|\mathbf{x}) \log \left[\frac{f'(m) F'(\mathbf{x}|m)}{f(m) F(\mathbf{x}|m)} \right] + I(q(\cdot|\mathbf{x}), q'(\cdot|\mathbf{x})) \right\}$$

Po vynechání nezáporné informační divergence dostaneme nerovnost

$$L' - L \geq \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \left\{ \sum_{m=1}^{M} q(m|\mathbf{x}) \log \left[\frac{f'(m) F'(\mathbf{x}|m)}{f(m) F(\mathbf{x}|m)} \right] \right\}$$

$$L' - L \geq \sum_{m=1}^{M} \left[\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \right] \log \frac{f'(m)}{f(m)} + \frac{1}{|\mathcal{S}|} \sum_{m=1}^{M} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \log \frac{F'(\mathbf{x}|m)}{F(\mathbf{x}|m)}$$

Důkaz monotónní konvergence EM algoritmu: $L^{'} \geq L$

S využitím substituce za $f^{'}(m)$ podle kroku M obdržíme nerovnost

$$\sum_{m=1}^{M}\left[\frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\right]\log\frac{f^{'}(m)}{f(m)} = \sum_{m=1}^{M} f^{'}(m)\log\frac{f^{'}(m)}{f(m)} \geq 0$$

Podle kroku M platí: $$F^{'}(.|m) = \arg\max_{F(.|m)}\left\{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\log F(\mathbf{x}|m)\right\}$$

a z implicitní definice v kroku M plyne:

$$(*) \qquad \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\log F^{'}(\mathbf{x}|m) \geq \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\log F(\mathbf{x}|m), \quad m \in \mathcal{M}$$

Z uvedených nerovností plyne monotónní vlastnost EM algoritmu:

$$L^{'} - L \geq \sum_{m=1}^{M} f^{'}(m)\log\frac{f^{'}(m)}{f(m)} + \frac{1}{|\mathcal{S}|}\sum_{m=1}^{M}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\log\frac{F^{'}(\mathbf{x}|m)}{F(\mathbf{x}|m)} \geq 0$$

POZN. Definice M-kroku je zbytečně silná, stačí aby nové parametry splňovaly nerovnost (*) $\Rightarrow$ GEM algoritmus

Zpět

Alternativní důkaz monotónní konvergence

Kullback-Leiblerova informační divergence $I(q(\cdot|\mathbf{x})||q^{'}(\cdot|\mathbf{x}))$ je nezáporná pro libovolné dvě podmíněné distribuce $q(\cdot|\mathbf{x}), q^{'}(\cdot|\mathbf{x})$ a rovná se nule právě když jsou obě distribuce identické. ▸ Důkaz

$$\Rightarrow \qquad \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} I(q(\cdot|\mathbf{x})||q^{'}(\cdot|\mathbf{x})) = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}[\sum_{m=1}^{M} q(m|\mathbf{x})\log\frac{q(m|\mathbf{x})}{q^{'}(m|\mathbf{x})}] \geq 0$$

Dosazením za $q(m|\mathbf{x}), q^{'}(m|\mathbf{x})$ podle kroku E dostaneme:

$$\frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}\sum_{m=1}^{M} q(m|\mathbf{x})\log\frac{P^{'}(\mathbf{x})}{P(\mathbf{x})} - \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}\sum_{m=1}^{M} q(m|\mathbf{x})\log[\frac{f^{'}(m)F^{'}(\mathbf{x}|m)}{f(m)F(\mathbf{x}|m)}] \geq 0.$$

První člen je roven přírůstku kritéria L:

$$\frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}\sum_{m=1}^{M} q(m|\mathbf{x})\log\frac{P^{'}(\mathbf{x})}{P(\mathbf{x})} = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}\log\frac{P^{'}(\mathbf{x})}{P(\mathbf{x})} = L^{'} - L.$$

Alternativní důkaz monotónní konvergence

Z upravené nerovnosti dále plyne s využitím předchozí rovnice:

$$L^{'} - L \geq \sum_{m=1}^{M} \left[\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \right] \log \frac{f^{'}(m)}{f(m)} +$$

$$+ \sum_{m=1}^{M} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \log \frac{F^{'}(\mathbf{x}|m)}{F(\mathbf{x}|m)}$$

Po substituci $f^{'}(m)$ podle kroku M je zřejmé, že první suma na pravé straně nerovnosti je nezáporná:

$$\sum_{m=1}^{M} \left[\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \right] \log \frac{f^{'}(m)}{f(m)} = \sum_{m=1}^{M} f^{'}(m) \log \frac{f^{'}(m)}{f(m)} \geq 0.$$

Alternativní důkaz monotónní konvergence

Vzhledem k definici v kroku M platí nerovnost:

$$\sum_{m=1}^{M} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \log F^{'}(\mathbf{x}|m) \geq \sum_{m=1}^{M} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \log F(\mathbf{x}|m),$$

tzn. následující suma je nezáporná

$$\sum_{m=1}^{M} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \log \frac{F^{'}(\mathbf{x}|m)}{F(\mathbf{x}|m)} \geq 0$$

a platí požadovaná nerovnost:

$$L^{'} - L \geq \sum_{m=1}^{M} f^{'}(m) \log \frac{f^{'}(m)}{f(m)} + \sum_{m=1}^{M} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}) \log \frac{F^{'}(\mathbf{x}|m)}{F(\mathbf{x}|m)} \geq 0$$

$$\Rightarrow \quad L^{'} \geq L$$

Zpět

Důsledky monotónní vlastnosti EM algoritmu

Neklesající shora omezená posloupnost hodnot kritéria $\{L^{(t)}\}_{t=0}^{\infty}$ má konečnou limitu $L^* < \infty$ a proto splňuje nutnou podmínku konvergence:

$$\lim_{t\to\infty} L^{(t)} = L^* < \infty \quad \Rightarrow \quad \lim_{t\to\infty} (L^{(t+1)} - L^{(t)}) = 0$$

Stejnou podmínku splňují i posloupnosti $\{f^{(t)}(\cdot)\}_{t=0}^{\infty}$, $\{q^{(t)}(\cdot|\mathbf{x})\}_{t=0}^{\infty}$:

$$\lim_{t\to\infty} ||f^{(t+1)}(\cdot) - f^{(t)}(\cdot)|| = 0, \qquad \lim_{t\to\infty} ||q^{(t+1)}(\cdot|\mathbf{x}) - q^{(t)}(\cdot|\mathbf{x})|| = 0.$$

Předchozí limity plynou z nerovnosti

$$L^{(t+1)} - L^{(t)} \geq I(f^{(t+1)}(\cdot)||f^{(t)}(\cdot)) + \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} I(q^{(t)}(\cdot|\mathbf{x})||q^{(t+1)}(\cdot|\mathbf{x}))$$

s použitím následující obecné nerovnosti (viz Kullback (1966)):

$$\sum_{\mathbf{x}\in\mathcal{X}} P^*(\mathbf{x}) \log\frac{P^*(\mathbf{x})}{P(\mathbf{x})} \geq \frac{1}{4}\Big(\sum_{\mathbf{x}\in\mathcal{X}} |P^*(\mathbf{x}) - P(\mathbf{x})|\Big)^2 \geq \frac{1}{4}\|P^*(\cdot) - P(\cdot)\|^2$$

Maximálně věrohodné odhady aproblém aproximace

Theorem

Maximalizace věrohodnostní funkce je asymptoticky ekvivalentní minimalizaci horní meze euklidovské vzdálenosti mezi skutečnou diskrétní distribucí P^* a její aproximací P.

Důkaz: Asymptoticky, pro $|\mathcal{S}| \to \infty$, platí

$$\lim_{|\mathcal{S}|\to\infty} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \log P(\mathbf{x}) = \lim_{|\mathcal{S}|\to\infty} \sum_{\mathbf{x}\in\mathcal{X}} \gamma(\mathbf{x}) \log P(\mathbf{x}) = \sum_{\mathbf{x}\in\mathcal{X}} P^*(\mathbf{x}) \log P(\mathbf{x})$$

kde $\gamma(\mathbf{x}) \geq 0$ je relativní četnost výskytu diskrétního vektoru $\mathbf{x}$ v posloupnosti $\mathcal{S}$ a P^* je skutečné rozložení pravděpodobnosti. Tvrzení věty plyne z nerovnosti (viz Kullback, 1966):

$$\sum_{\mathbf{x}\in\mathcal{X}} P^*(\mathbf{x}) \log \frac{P^*(\mathbf{x})}{P(\mathbf{x})} \geq \frac{1}{4} \Big(\sum_{\mathbf{x}\in\mathcal{X}} |P^*(\mathbf{x}) - P(\mathbf{x})| \Big)^2 \geq \frac{1}{4} \|P^*(\cdot) - P(\cdot)\|^2$$

Důkaz nezápornosti Kullback-Leiblerovy divergence

Theorem (viz např. Vajda, 1992)

Pro libovolná dvě diskrétní rozložení pravděpodobnosti $\{q_1, q_2, \ldots, q_M\}$, $\{q_1', q_2', \ldots, q_M'\}$ *platí nerovnost*

$$I(\mathbf{q} \| \, \mathbf{q}') = \sum_{m=1}^{M} q_m \log \frac{q_m}{q_m'} \geq 0$$

přičemž rovnost nastane právě tehdy, je-li $q_m' = q_m$, *pro všechna* $m \in \mathcal{M}$.

Důkaz: Bez ztráty obecnosti můžeme předpokládat $q_m > 0$ pro všechna $m \in \mathcal{M}$ (protože $0 \log 0 = 0$). Podle Jensenovy nerovnosti platí:

$$\sum_{m=1}^{M} q_m \log \frac{q_m'}{q_m} \leq \log \left(\sum_{m=1}^{M} q_m \frac{q_m'}{q_m} \right) = \log \left(\sum_{m=1}^{M} q_m' \right) = \log 1 = 0,$$

přičemž rovnost nastane právě tehdy, je-li $\zeta_1 = \cdots = \zeta_M$, cbd.

Důsledek: následující suma na levé straně je maximální pro $\mathbf{q} = \mathbf{q}'$

$$\sum_{m=1}^{M} q_m \log q_m' \leq \sum_{m=1}^{M} q_m \log q_m$$

◂ Zpět ◂ Zpět (alternativní důkaz) ◂ Zpět (M-krok)

Vlastnosti strukturního modelu směsi

strukturní model směsi realizuje "podprostorový" přístup:

- řeší obecný problém výběru příznaků individuálně pro každou komponentu - jako součást EM algoritmu
- umožňuje bayesovské rozhodování nezávisle na dimenzi prostoru (řešení rozhodovacích problémů bez redukce dimenze prostoru)
- snižuje počet parametrů modelu (i komponent) a tím omezuje riziko "nadměrného" přizpůsobení modelu datům (overpeaking)
- kritériem strukturní optimalizace odvozeným z EM algoritmu je Kullback-Leiblerova informační divergence (pro diskrétní směs)
- umožňuje statisticky korektní řešení problému neúplného propojení vstupních proměnných a neuronové vrstvy
- umožňuje strukturní optimalizaci neuronové sítě při zachování monotónní vlastnosti EM algoritmu: $L^{'} - L \geq 0$

POZN. Distribuci pozadí obvykle zvolenou jako součin marginálních rozložení $P_n(x_n)$ lze rovněž optimalizovat (problém: $f_n(x_n|0) \to 0$).

Zpět - Strukturní model

Odvození kritéria strukturní optimalizace

v implicitní rovnici kroku M můžeme vynechat konstantní $F(\mathbf{x}|0)$:

$$G^{'}(.|m,\phi_m^{'}) = \arg \max_{G(.|m,\phi_m)} \left\{ \frac{1}{|\mathcal{S}|} \sum_{x\in\mathcal{S}} q(m|\mathbf{x}) \log G(\mathbf{x}|m,\phi_m) \right\}$$

po dosazení za $G(\mathbf{x}|m,\phi_m)$ upravíme výraz v závorce:

$$\sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{|\mathcal{S}|} \log \prod_{n\in\mathcal{N}} \left[\frac{f_n(x_n|m)}{f_n(x_n|0)}\right]^{\phi_{mn}} = \sum_{n\in\mathcal{N}} \phi_{mn} \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{|\mathcal{S}|} \log \left[\frac{f_n(x_n|m)}{f_n(x_n|0)}\right]$$

$\Rightarrow$ výchozí implicitní vztah lze rozepsat odděleně pro jednorozměrné distribuce $f_n^{'}(x_n|m)$ a strukturní parametry $\phi_{mn}^{'}$:

$$f_n^{'}(.|m) = \arg \max_{f_n(.|m)} \left\{ \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{|\mathcal{S}|} \log f_n(x_n|m) \right\}$$

$$\phi_m^{'} = \arg\max_{\phi_m} \left\{ \phi_{mn} \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{|\mathcal{S}|} \log \left[\frac{f_n^{'}(x_n|m)}{f_n(x_n|0)}\right] \right\} = \arg\max_{\phi_m} \left\{ \phi_{mn}\gamma_{mn}^{'} \right\}$$

Prof. M.I. Schlesinger se svou ženou

Při výletu na Karlštejn během pobytu v Praze v roce 1995.

Zpět

Literatura 1/10

- Ajvazjan S.A., Bezhaeva Z.I., Staroverov O.V. (1974):*Classification of Multivariate Observations*, (in Russian). Moscow: Statistika.
- Boyles R.A. (1983): On the convergence of the EM algorithm. *J. Roy. Statist. Soc., B*, Vol. 45, pp. 47-50.
- Cacoullos I. (1966): Estimation of a multivariate density. *Ann. Inst. Stat. Math.*, Vol. 18, pp. 179-190.
- Carreira-Perpignan M.A., Renals S. (2000): Practical identifiability of finite mixtures of multivariate Bernoulli distributions. *Neural Computation*, Vol. 12, pp. 141-152.
- Day N.E. (1969): Estimating the components of a mixture of normal distributions. *Biometrika*, Vol. 56, pp. 463-474.
- Dempster A.P., Laird N.M. and Rubin D.B. (1977): Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc., B*, Vol. 39, pp.l-38.

Literatura 2/10

- Duda R.O., Hart P.E. (1973): Pattern Classification and Scene Analysis. New York: Wiley-Interscience.
- Everitt, B.S. and D.J. Hand (1981): *Finite Mixture Distributions*. Chapman & Hall: London, 1981.
- Grim J. (1982): On numerical evaluation of maximum - likelihood estimates for finite mixtures of distributions. *Kybernetika*, Vol.l8, No.3, pp.173-190.
- Grim J. (1982): Design and optimization of multilevel homogeneous structures for multivariate pattern recognition. In *Fourth FORMATOR Symposium 1982*, Academia, Prague 1982, pp. 233-240.
- Grim J. (1986): Multivariate statistical pattern recognition with nonreduced dimensionality, *Kybernetika*, Vol. 22, pp. 142-157.

Literatura 3/10

- Grim J. (1992): A dialog presentation of census results by means of the probabilistic expert system PES, in *Proceedings of the Eleventh Europeccn Meeting on Cybernetics and Systems Research*, Vienna, April 1992, (Ed. R.Trappl), pp. 997-1005, World Scientific, Singapore 1992.

- Grim J. (1994): Knowledge representation and uncertainty processing in the probabilistic expert system PES, *International Journal of General Systems*, Vol. 22, No. 2, p. 103 - 111.

- Grim J., Boček P. (1995): Statistical Model of Prague Households for Interactive Presentation of Census Data, In *SoftStat'95. Advances in Statistical Software 5*, pp. 271 - 278, Lucius & Lucius: Stuttgart, 1996.

- Grim J., (1996): Maximum Likelihood Design of Layered Neural Networks. In: *Proceedings of the 13th International Conference on Pattern Recognition* **IV** (pp. 85-89), Los Alamitos: IEEE Computer Society Press.

- Grim J., (1996a): Design of multilayer neural networks by information preserving transforms. In: E. Pessa, M.P. Penna, A. Montesanto (Eds.), *Proceedings of the Third European Congress on System Science* (pp. 977-982), Roma: Edizzioni Kappa.

Literatura 4/10

Grim J. (1998): A sequential modification of EM algorithm. In *Studies in Classification, Data Analysis and Knowledge Organization*, Gaul W., Locarek-Junge H., (Eds.), pp. 163 - 170, Springer, 1999.

Grim J., Somol P., Novovičová J., Pudil P., Ferri F., (1998b): Initializing normal mixture of densities. In *Proc. 14th Int. Conf. on Pattern Recognition ICPR'98*, A.K. Jain, S. Venkatesh, B.C. Lovell (Eds.), pp. 886-890, IEEE Computer Society: Los Alamitos, California, 1998

Grim J. (1999): Information approach to structural optimization of probabilistic neural networks. In proceedings of: 4th System Science European Congress, L. Ferrer, A. Caselles, R. Beneyto, R. Pla, I. Martinez, V. Rossi, J. Martinez, J.R. Hernandez (Eds.), (pp: 527-540), Valencia: Sociedad Espanola de Sistemas Generales, 1999.

Grim J., Boček P., Pudil P. (2001): Safe dissemination of census results by means of interactive probabilistic models. In: *Proceedings of the ETK-NTTS 2001 Conference*, (Hersonissos (Crete), June 18-22, 2001), Vol.2, pp. 849-856, European Communities 2001.

Literatura 5/10

- Grim J. (2001): Latent Structure Analysis for Categorical Data. Research Report No. 2019. ÚTIA AV ČR, Praha 2001, 13 pp.
- Grim J., Haindl M.: Texture Modelling by Discrete Distribution Mixtures. Computational Statistics and Data Analysis, 3-4 **41** (2003) 603-615
- Grim J., Hora J., Pudil P. (2004): Interaktivní reprodukce výsledků sčítání lidu se zaručenou ochranou anonymity dat. *Statistika*, Vol. 84, No. 5, pp. 400-414.
- Grim J., Just P., Pudil P. (2003): Strictly modular probabilistic neural networks for pattern recognition. *Neural Network World*, Vol. 13 , No. 6, pp. 599-615.
- Grim J., Kittler J., Pudil P., Somol P. (2000): Combining multiple classifiers in probabilistic neural networks, In *Multiple Classifier Systems*, Eds. Kittler J., Roli F., Springer, 2000, pp. 157 - 166.

Literatura 6/10

- Grim J., Kittler J., Pudil P., Somol P. (2001): Information analysis of multiple classifier fusion. In: *Multiple Classifier Systems 2001*, Kittler J., Roli F., (Eds.), Lecture Notes in computer Science, Vol. 2096, Springer-Verlag, Berlin, Heidelberg, New York 2001, pp. 168 - 177.
- Grim J., Kittler J., Pudil P., Somol P., (2002): Multiple classifier fusion in probabilistic neural networks. *Pattern Analysis & Applications* Vol. 5, No. 7, pp. 221-233.
- Grim J., Pudil P., Somol P. (2000): Recognition of handwritten numerals by structural probabilistic neural networks. In: Proceedings of the Second ICSC Symposium on Neural Computation, Berlin, 2000. (Bothe H., Rojas R. eds.). ICSC, Wetaskiwin, 2000, pp 528-534.
- Grim J., Somol P., Novovičová J., Pudil P., Ferri F. (1998): Initializing normal mixtures of densities. In *Proceedings of the 14th International Conference on Pattern Recognition ICPR'98*, Brisbane, August 16 - 20, 1998, Eds. A.K. Jain, S. Venkatesh, B.C. Lovell, pp. 886-890, IEEE Computer Society: Los Alamitos, California, 1998

Literatura 7/10

- Grim J., Somol P., Pudil P., Just P. (2003): Probabilistic neural network playing a simple game. In *Artificial Neural Networks in Pattern Recognition.* (Marinai S., Gori M. Eds.). University of Florence, Florence 2003, pp. 132-138.
- Grim J., Somol P., Haindl M., Pudil P. (2005): A statistical approach to local evaluation of a single texture image. In: Proceedings of the 16-th Annual Symposium PRASA 2005. (Nicolls F. ed.). University of Cape Town, 2005, pp. 171-176.
- Grim J.: EM Cluster Analysis for Categorical Data. In: *Proceedings of the SSPR'06*, Hong-Kong (2006), (to appear).
- Gyllenberg M., T. Koski, E. Reilink, M. Verlaan (1994): Non-uniqueness in probabilistic numerical identification of bacteria. *Journal of Applied Probability*, Vol. 31, pp. 542–548.
- Haindl M., Grim J., Somol P., Pudil P., Kudo M. (2004): A Gaussian mixture-based colour texture model. In: Proceedings of the 17th IAPR International Conference on Pattern Recognition. IEEE, Los Alamitos 2004, pp. 177-180.

Literatura 8/10

Haindl M., Grim J., Pudil P., Kudo M. (2005): A Hybrid BTF Model Based on Gaussian Mixtures. In: Texture 2005. Proceedings of the 4th International Workshop on Texture Analysis. (Chantler M., Drbohlav O. eds.). IEEE, Los Alamitos 2005, pp. 95-100.

Hasselblad V. (1966): Estimation of prameters for a mixture of normal distributions. *Technometrics*, Vol. 8, pp. 431-444.

Hasselblad V. (1969): Estimation of finite mixtures of distributions from the exponential family. *Journal of Amer. Statist. Assoc.*, Vol. 58, pp. 1459-1471.

Isaenko O.K., Urbakh K.I. (1976): Decomposition of probability distribution mixtures into their components (in Russian). In: *Theory of probability, mathematical statistics and theoretical cybernetics*, Vol. 13, Moscow: VINITI.

Kullback S. (1966): An information-theoretic derivation of certain limit relations for a stationary Markov Chain. *SIAM J. Control*, Vol. 4, No. 3, pp. 454-459.

Literatura 9/10

- McLachlan G.J. and Peel D. (2000): *Finite Mixture Models*, John Wiley & Sons, New York, Toronto, (2000)
- Parzen E. (1962): On estimation of a probability density function and its mode. *Annals of Mathematical Statistics*, Vol. 33., pp. 1065-1076.
- Pearson C. (1894): Contributions to the mathematical theory of evolution. 1. Dissection of frequency curves." *Philosophical Transactions of the Royal Society of London* **185**, 71-110.
- Peters B.C., Walker H.F. (1978): An iterative procedure for obtaining maximumlikelihood estimates of the parameters for a mixture of normal distributions. *SIAM Journal Appl. Math.*, Vol. 35, No. 2, pp. 362-378.
- Schlesinger M.I. (1968): Relation between learning and self learning in pattern recognition (in Russian), *Kibernetika*, (Kiev), No. 2, 81-88.
- Teicher H. (1963): Identifiability of finite mixtures. *Ann. Math. Statist.*, Vol. 34, pp. 1265-1269.

Literatura 10/10

- Teicher H. (1968): Identifiability of mixtures of product measures. *Ann. Math. Statist.*, Vol. 39, pp. 1300-1302.
- Titterington D.M., Smith A.F.M. and Makov U.E. (1985): *Statistical analysis of finite mixture distributions*, John Wiley & Sons: Chichester, New York.
- Vajda I., Grim J. (1998): About the maximum information and maximum likelihood principles in neural networks, *Kybernetika*, Vol. 34, No. 4, pp. 485-494.
- Wolfe J.H. (1970): Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, Vol. 5, pp. 329-350.
- Wu C.F.J. (1983): On the convergence properties of the EM algorithm. *Ann. Statist.*, Vol. 11, pp. 95-103.
- Xu L. and Jordan M.I. (1996): On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation*, Vol. 8. pp. 129-151.