SOUČINOVÉ DISTRIBUČNÍ SMĚSI

I. část: EM algoritmus

Jiří Grim

Ústav teorie informace a automatizace AV ČR

Oddělení rozpoznávání obrazů

Listopad 2014

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

ÚTIA

Outline

Metoda distribučních směsí

Výchozí informace:

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

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

Princip metody distribučních směsí:

aproximace mnohorozměrného multimodálního rozložení pravděpodobnosti $P^*(\mathbf{x})$ ve tvaru váženého součtu parametrických komponent $F(\mathbf{x}|m)$

$$P(\mathbf{x}) = \sum_{m=1}^{M} w_m F(\mathbf{x}|m), \quad \sum_{m=1}^{M} w_m = 1, \quad \sum_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x}|m) = 1 \; \left(= \int_{\mathcal{X}} F(\mathbf{x}|m) d\mathbf{x} \right)$$

Příklady aplikací:

rozpoznávání obrazů, klasifikace textových dokumentů, problém predikce, modelování textur, analýza obrazů, statistické modely dat, ...

Směsi jako "semiparametrický" model

parametrický přístup: např. odhad normální hustoty pravděpodobnosti

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

průměr: $\mathbf{c} = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} \mathbf{x}$, kovarianční 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)

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

problém: ▸ Optimální vyhlazení (volba parametrů σ_n)

Distribuční směsi jako semiparametrický multimodální model

- kompromis mezi omezujícím předpokladem parametrického modelu a obecností neparametrických odhadů
- **efektivní odhad parametrů směsi pomocí EM algoritmu**

Příklad - EM algoritmus pro směs normálních hustot

výpočet odhadu parametrů směsi z dat: $\mathcal{S} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(K)}\}$

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

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

Iterační rovnice: $\approx$ maximalizace věrohodnostní funkce

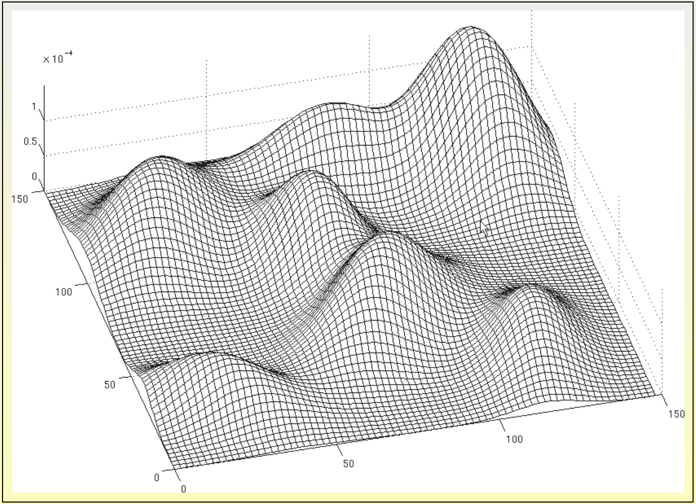
E-krok: $q(m|\mathbf{x}) = \frac{w_m F(\mathbf{x}|\mathbf{c}_m, A_m)}{\sum_{j=1}^{M} w_j F(\mathbf{x}|\mathbf{c}_j, A_j)}, \ \mathbf{x} \in \mathcal{S}, \quad m = 1, 2, \dots, M$

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

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

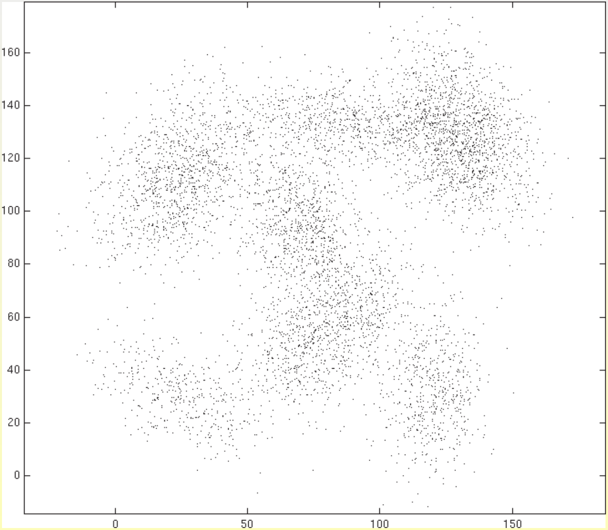
POZN. **Počet komponent směsi je nutné zvolit předem.**

Příklad: rekonstrukce normální směsi z dat



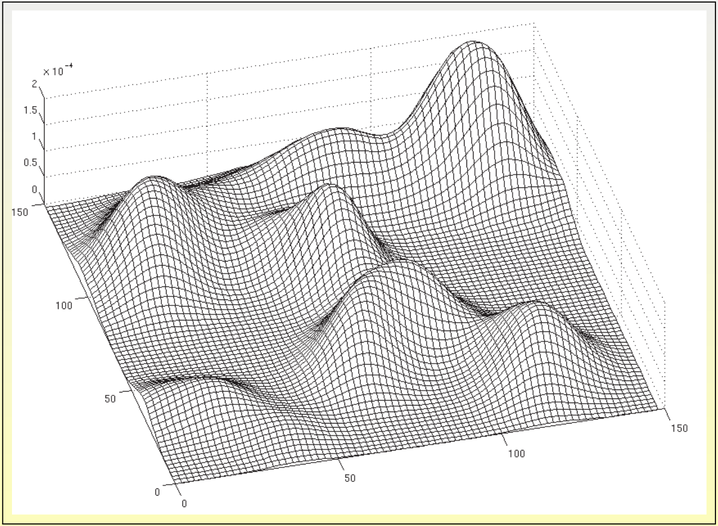
dimenze distribuční směsi $N = 2$, počet komponent směsi $M = 7$

Data náhodně generovaná podle normální směsi (M=7)



6000 bodů: pro ověření správnosti implementace EM algoritmu

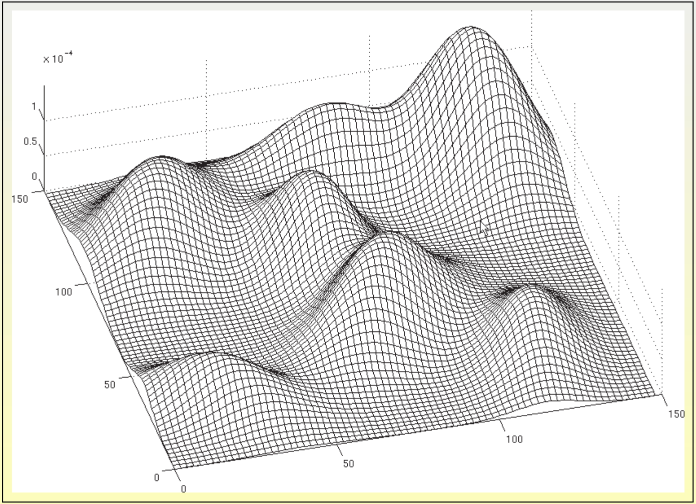
Příklad odhadu normální směsi (M=28)



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

(SROVNÁNÍ: jádrový odhad)

Původní směs normálních hustot (M=7)



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

Obecná verze EM algoritmu

EM algoritmus: maximalizuje věrohodnostní funkci (kritérium)

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

Iterační rovnice: $(m = 1, 2, \ldots, M, \quad \mathbf{x}\in\mathcal{S}, \quad \mathcal{S} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}\})$

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

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

---

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

$$\Rightarrow \quad f_n^{'}(.|m) = \arg\max_{f_n(.|m)}\left\{\frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log f_n(x_n|m)\right\}, \quad n = 1, \ldots, N$$

POZN. Explicitní řešení rovnic pro výpočet maxima: Lemma

Monotónní vlastnost EM algoritmu (Schlesinger, 1968)

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

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

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

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

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

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

ÚTIA

Důkaz monotónní vlastnosti

Lemma

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

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

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

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

přičemž první člen na levé straně je roven přírůstku kritéria L:

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

Důkaz monotónní vlastnosti

Z upravené nerovnosti dále plyne po dosazení z předchozí rovnice:

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

S využitím substituce podle kroku M

$$(**) \quad w_m^{'} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(m|\mathbf{x}), \quad m = 1, 2, \ldots, M$$

můžeme napsat nerovnost:

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

tzn. první suma na pravé straně výchozí nerovnosti (*) je nezáporná.

POZN. Definice (**) vah $w_m^{'}$ maximalizuje první sumu.

Důkaz monotónní vlastnosti

Definici funkce $F^{'}(\cdot|m)$ v kroku M můžeme zapsat ve tvaru:

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

tzn. pro libovolné funkce $F(\cdot|m)$ platí nerovnost:

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

Předchozí nerovnost lze upravit na tvar:

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

tj. přírůstek věrohodnostní funkce je nezáporný:

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

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

Alternativní důkaz

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 Definice
- POTÍŽ: věrohodnostní funkce směsi má téměř vždy lokální maxima (zvláště při velké dimenzi a malém počtu dat)
- $\Rightarrow$ nalezené lokální maximum závisí na volbě počátečních parametrů
- POTÍŽ: kvalita odhadu směsi závisí na zvoleném počtu komponent a počátečních parametrech
- $\Rightarrow$ POČET KOMPONENT? POČÁTEČNÍ HODNOTY?

Problém aproximace neznámé pravděpodobnostní distribuce

- cílem je co nejpřesnější aproximace neznámého rozložení pravděpodobnosti pomocí distribuční směsi Problém aproximace × MLE
- směs nemusí být identifikovatelná (výhoda při aproximaci)
- konkrétní počet komponent směsi není důležitý
- počáteční parametry směsi je možné generovat náhodně

Výpočetní vlastnosti EM algoritmu

Typická aproximační úloha: velký počet dat + velký počet komponent

- efektivní aproximace multimodálních distribucí s velkou dimenzí
- při velkém počtu komponent ($M \approx 10^1 - 10^2$) lze zanedbat komponenty s řádově nižší váhou ($\Rightarrow$ konkrétní počet není důležitý)
- 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
- 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
- EM algoritmus lze použít na "vážená" data

POZN. **Některé z uvedených vlastností závisí na konkrétních datech, tj. neplatí obecně a nelze je dokázat.**

Z historie problému odhadování směsí

v období 1895 - 1965 bylo publikováno asi 80 prací o identifikaci 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)*

efektivní řešení problému odhadu směsí až s příchodem počítačů:

- **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 odhadování směsí

první důkaz monotonie EM algoritmu:

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

достигается тогда, когда величины x_i пропорцио-нальны 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 $\sum\limits_{k=1}^{s} \alpha_{ik} = 1$ для всех i (см. формулу (7)), выражение для $L(A^{(t)})$ можно записать следующим образом:

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

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

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

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

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

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 $\sum\limits_{i=1}^{m} \alpha_{ik}(A^{(t)})$. К тому же очевидным является равенство $\sum\limits^{s} p_k^{(t+1)} =$

Foto

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

ÚTIA

Z historie problému odhadování směsí

název EM algoritmus: z často citované práce ("all time top 10"):

- **Dempster et al. (1977):** "Maximum likelihood from incomplete data via the EM algorithm." *J. Roy. Statist. Soc., B, Vol. 39, pp.l-38.*

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 $p \geqslant p(\varepsilon)$ and all $r \geqslant 1$, 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)

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

whence

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

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

Google Scholar, listopad 2014: **39 000 citací článku Dempster et al.**

ÚTIA

Z historie problému odhadování směsí

chyba v důkazu konvergence posloupnosti parametrů:

- **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.*

> 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

Monografie: Titterington et al. (1985), McLachlan and Peel (2000)
Google Scholar (2013):

- termín "EM algorithm"- 2 930 000 publikací
- "EM algorithm & mixture"- 576 000 publikací

SOUČINOVÉ SMĚSI

směs součinových komponent (model podmíněné nezávislosti):

$$P(\mathbf{x}) = \sum_{m=1}^{M} w_m \prod_{n=1}^{N} f_n(x_n|m), \quad \mathbf{x} \in \mathcal{X}$$

Základní výhody použití součinových komponent

- zjednodušení výpočtu (odpadá např. inverze kovariančních matic)
- stabilita výpočtu (odpadá riziko špatně podmíněných matic)
- snadný výpočet marginálních rozložení pravděpodobnosti ▸ Odvození
- možnost odhadu parametrů směsi z neúplných datových vektorů

Nevýhody součinových směsí (+ mýty a pověry)

- skrytý předpoklad nezávislosti proměnných (?!) (platí pouze pro M=1)
- předpoklad součinových komponent je omezující (?) ▸ Jádrový (Parzenův) odhad
- vhodná normalizace dat před použitím součinové směsi (?)

Příklad EM algoritmu - součinová normální směs

KOMPONENTY: **normální hustoty s diagonální kovarianční maticí:**

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

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

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

Invariance vůči normování proměnných

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

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

$$(\sigma_{mn}^{'})^2=\frac{1}{w_m^{'}|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}(x_n-\mu_{mn}^{'})^2q(m|\mathbf{x})=\frac{1}{w_m^{'}|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}}x_n^2q(m|\mathbf{x})\ -\ (\mu_{mn}^{'})^2$$

odpadá inverze kovariančních matic $\Rightarrow$ stabilnější výpočet

Příklad EM algoritmu - diskrétní součinová směs

KOMPONENTY: **součiny jednorozměrných diskrétních rozložení psti**

$$F(\mathbf{x}|m) = \prod_{n=1}^{N} f_n(x_n|m), \quad \mathbf{x} = (x_1, \ldots, x_N) \in \mathcal{X}, \; x_n \in \mathcal{X}_n, \; |\mathcal{X}_n| < \infty$$

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

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

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

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

Podrobnější odvození

POZN. 1 Diskrétní součinová směs není identifikovatelná. Důkaz
(problém při shlukové analýze × výhoda při aproximaci)

POZN. 2 Každé diskrétní rozložení psti lze zapsat jako součinovou směs.

Příklad EM algoritmu - směs Bernoulliho rozložení

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

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

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

iterační rovnice:

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

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

POZN. Problém přesnosti součinů velkého počtu parametrů θ_{mn}.

Poznámky k implementaci EM algoritmu

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

$$\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \rightarrow w_m^{'}, \qquad \sum_{\mathbf{x}\in\mathcal{S}} x_n \; q(m|\mathbf{x}) \rightarrow \mu_{mn}^{'}, \theta_{mn}^{'}$$

- nutná podmínka správnosti programu: $L^{'} \geq L$
- 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
- užitečný údaj:

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

- pro data s velkou dimenzí ($N >> 1$) je typický malý "překryv" komponent, tj. poměrně vysoké hodnoty $\bar{q}_{max} \approx 0.85 - 0.99$
- $\Rightarrow$ špatně podmíněné kovarianční matice

Ověření algoritmu: generování umělých dat + reidentifikace parametrů

Implementace EM algoritmu v prostoru s velkou dimenzí

PROBLÉM: **numerická nestabilita v prostoru s velkou dimenzí**

- komponenty $F(\mathbf{x}|m)$ "podtékají" již při dimenzi $N \approx 30 - 40$
  $\Rightarrow$ ztráta přesnosti výpočtu (někdy obtížně zjistitelná)
  $\Rightarrow$ výpočet komponent je třeba provést v logaritmickém tvaru:

$$\log[F(\mathbf{x}|m)w_m] = \log w_m + \sum_{n \in \mathcal{N}} \log f_n(x_n|m)$$

maximum: $\log C(\mathbf{x}) = \max_m \{ \log[F(\mathbf{x}|m)w_m] \} \Rightarrow C(\mathbf{x})$

- **NORMALIZACE** hodnot $F(\mathbf{x}|m)$ a $P(\mathbf{x})$ pro výpočet $q(m|\mathbf{x})$:

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

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

Příklady zdrojového kódu: ▸ Bernoulliovská směs ▸ Normální součinová směs

ÚTIA

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

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

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

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

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

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

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

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

POZN. Lokální výběr příznaků, libovolná dimenze dat.

Strukturní modifikace EM algoritmu - obecně

strukturní optimalizace je součástí EM algoritmu

$$L = \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}} \log\left[\sum_{m\in\mathcal{M}} F(\mathbf{x}|0)G(\mathbf{x}|m,\phi_m)w_m\right]$$

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

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

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

strukturní optimalizace: $\phi_{mn}^{'} = 1$ pro r nejvyšších hodnot $\gamma_{mn}^{'}$ Důkaz

$$\gamma_{mn}^{'} = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\log\left[\frac{f_n^{'}(x_n|m)}{f_n(x_n|0)}\right]$$

POZN. Optimalizace může zahrnovat i "pozadí" $F(\mathbf{x}|0)$ (Grim, 1999)

ÚTIA

Strukturní modifikace EM algoritmu - diskrétní distribuce

$f_n(x_n|m),\ x_n \in \mathcal{X}_n,\ n \in \mathcal{N}\ \approx$ **diskrétní rozložení pravděpodobnosti**

$$L = \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}} \log\left[\sum_{m\in\mathcal{M}} G(\mathbf{x}|m,\phi_m)w_m\right], \qquad G(\mathbf{x}|m) = \prod_{n\in\mathcal{N}}\left[\frac{f_n(x_n|m)}{f_n(x_n|0)}\right]^{\phi_{mn}}$$

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

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

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

Podrobněji

strukturní optimalizace: $\phi_{mn}' = 1$ pro r nejvyšších hodnot γ_{mn}'

$$\gamma_{mn}' = \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{w_m'|\mathcal{S}|}\log\left[\frac{f_n'(x_n|m)}{f_n(x_n|0)}\right] = w_m'\sum_{\xi_n\in\mathcal{X}_n} f_n'(\xi_n|m)\log\frac{f_n'(\xi_n|m)}{f_n(\xi_n|0)}$$

Odvození

Strukturní modifikace EM algoritmu - normální směs

normální hustoty: $f_n(x_n|\mu_{mn},\sigma_{mn}) = \frac{1}{\sqrt{2\pi}\sigma_{mn}} \exp\left\{ -\frac{(x_n-\mu_{mn})^2}{2\sigma_{mn}^2}\right\}$

$$L = \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}} \log\left[\sum_{m\in\mathcal{M}} w_m \prod_{n\in\mathcal{N}} \left(\frac{f_n(x_n|\mu_{mn},\sigma_{mn})}{f_n(x_n|\mu_{0n},\sigma_{0n})}\right)^{\phi_{mn}}\right],$$

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

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

$$\mu_{mn}' = \frac{1}{w_m'|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} x_n q(m|\mathbf{x}), \quad (\sigma_{mn}')^2 = \frac{1}{w_m'|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} x_n^2 q(m|\mathbf{x}) - (\mu_{mn}')^2,$$

strukturní optimalizace: $\phi_{mn}' = 1$ pro r nejvyšších hodnot γ_{mn}'

$$\gamma_n'^{(m)} = \frac{w_m'}{2}\left[\frac{(\mu_{mn}'-\mu_{0n})^2}{(\sigma_{0n})^2} + \frac{(\sigma_{mn}')^2}{(\sigma_{0n})^2} - \log\frac{(\sigma_{mn}')^2}{(\sigma_{0n})^2} - 1\right]$$

Podrobněji

Vlastnosti strukturního modelu směsi

strukturní model směsi realizuje "podprostorový" přístup:

- **mechanizmus:** málo informativní jednorozměrné distribuce $f_n(x_n|m)$ nahrazuje model příslušným fixním "pozadím" $f_n(x_n|0)$
- 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)
- potlačuje vliv nespolehlivě odhadnutých parametrů
- 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: $L^{'} - L \geq 0$

Modifikace EM algoritmu - neúplné datové vektory

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

$\mathcal{N}(\mathbf{x}) = \{n \in \mathcal{N} :$ souřadnice x_n 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

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

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

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

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

POZN. Nahrazování chybějících údajů pomocí odhadů ovlivňuje data.

Modifikace EM algoritmu - vážená data

$\gamma(\mathbf{x}) > 0$: relativní četnost vektoru $\mathbf{x}$ v posloupnosti $\mathcal{S}$, $\quad (\sum_{\mathbf{x}\in\mathcal{X}} \gamma(\mathbf{x}) = 1)$

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

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

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

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

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

$$F'(.|m) = \arg\max_{F(.|m)} \left\{ \sum_{\mathbf{x}\in\bar{\mathcal{X}}} \frac{\gamma(\mathbf{x}) q(m|\mathbf{x})}{w_m'} \log F(\mathbf{x}|m) \right\}$$

POUŽITÍ: **agregace dat, virtuálně "nekonečná" data:** $\gamma(\mathbf{x}) = P^*(\mathbf{x})$

Sekvenční rozhodovací schema

Problém: postupné doplňování příznaků (př. lékařská diagnostika)

dané hodnoty: $\mathbf{x}_D = (x_{j_1}, \dots, x_{j_l}) \in \mathcal{X}_D, \quad \mathcal{D} = \{j_1, \dots, j_l\} \subset \mathcal{N}$

$$P(\mathbf{x}_D|\omega) = \sum_{m\in\mathcal{M}_\omega} w_m \prod_{n\in D} f_n(x_n|m,\omega), \qquad P(\mathbf{x}_D) = \sum_{\omega\in\Omega} P(\mathbf{x}_D|\omega)p(\omega)$$

Optimální sekvenční rozhodování: $p(\omega|x_n, \mathbf{x}_D), \ \omega \in \Omega$

Volba nejinformativnější proměnné $x_n, \ n \notin D$ při dané podmnožině známých vstupních údajů $\mathbf{x}_D = (x_{j_1}, \dots, x_{j_l}) \in \mathcal{X}_D$ podle kriteria **maximální podmíněné informace:** $I_{\mathbf{x}_D}(\mathcal{X}_n, \Omega)$

$$I_{\mathbf{x}_D}(\mathcal{X}_n, \Omega) = H_{\mathbf{x}_D}(\mathcal{X}_n) - H_{\mathbf{x}_D}(\mathcal{X}_n|\Omega), \qquad n^* = \arg\max_{n\notin D} \{I_{\mathbf{x}_D}(\mathcal{X}_n, \Omega)\}$$

$$H_{\mathbf{x}_D}(\mathcal{X}_n) = \sum_{x_n\in\mathcal{X}_n} -P_{n|D}(x_n|\mathbf{x}_D)\log P_{n|D}(x_n|\mathbf{x}_D), \quad p(\omega|\mathbf{x}_D) = \frac{P(\mathbf{x}_D|\omega)p(\omega)}{P(\mathbf{x}_D)},$$

$$H_{\mathbf{x}_D}(\mathcal{X}_n|\Omega) = \sum_{\omega\in\Omega} p(\omega|\mathbf{x}_D) \sum_{x_n\in\mathcal{X}_n} -P_{n|D\omega}(x_n|\mathbf{x}_D,\omega)\log P_{n|D\omega}(x_n|\mathbf{x}_D,\omega)$$

$$P_{n|D\omega}(x_n|\mathbf{x}_D,\omega) = P_{nD|\omega}(x_n, \mathbf{x}_D|\omega)/P_{D|\omega}(\mathbf{x}_D|\omega)$$

ÚTIA

Výběr nejinformativnějšího podprostoru

odhad podmíněných distribucí $P(\mathbf{x}|\omega), \omega \in \Omega$ ve tvaru součinové směsi

$\Rightarrow$ **informační kriterium pro výběr příznaků:**

$$I(\mathcal{X}_D, \Omega) = H(\mathcal{X}_D) - H(\mathcal{X}_D|\Omega)\}, \qquad \mathcal{D}^* = \arg\max_{\mathcal{D}\subset\mathcal{N}} \{I(\mathcal{X}_D, \Omega)\}$$

$$H(\mathcal{X}_D) = \sum_{\mathbf{x}_D \in \mathcal{X}_D} -P_D(\mathbf{x}_D) \log P_D(\mathbf{x}_D), \qquad \mathcal{D} = \{j_1, \ldots, j_k\} \subset \mathcal{N}, \qquad |\mathcal{D}| = k$$

$$H(\mathcal{X}_D|\Omega) = \sum_{\omega\in\Omega} p(\omega) \sum_{\mathbf{x}_D \in \mathcal{X}_D} -P_{D|\omega}(\mathbf{x}_D|\omega) \log P_{D|\omega}(\mathbf{x}_D|\omega)$$

$$P_{D|\omega}(\mathbf{x}_D|\omega) = \sum_{m\in\mathcal{M}_\omega} w_m \prod_{n\in\mathcal{D}} f_n(x_n|m)$$

optimální podmnožina $\mathcal{D} \subset \mathcal{N}$: úplné prohledání, přibližné metody

Motivace:

- výběr nejinformativnější podmnožiny příznaků pro rozpoznávání
- rychlá lokalizace 2D grafických objektů (číslice, písmena)

VLASTNOSTI SOUČINOVÝCH SMĚSÍ

SOUHRN: výpočetní vlastnosti součinových distribučních směsí

- efektivní odhad parametrů směsi v mnohorozměrném prostoru (!)
- snadný výpočet marginálních 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
- vhodné pro aproximaci multimodálních rozložení pravděpodobnosti
- možnost odhadu parametrů směsi z neúplných datových vektorů
- umožňují sekvenční rozhodování s postupným doplňováním nejinformativnějších příznaků
- existuje strukturní modifikace součinové směsi, která umožňuje "lokální" výběr příznaků a rozhodování v prostorech s velkou dimenzi
- součinové směsi lze interpretovat jako neuronovou síť

Příklady aplikací Literatura

A1: 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 $P(\mathbf{x})$ s vyhlazovacím parametrem σ_K

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

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:*

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

Výpočetní náročnost + problém vyhlazení

A2: Optimalizace vyhlazení Parzenova odhadu

Parzenův odhad s normálním jádrem:

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

optimalizace vyhlazení metodou cross-validace:

$\approx$ maximalizace upravené věrohodnostní funkce pomocí EM algoritmu:

$$L(\boldsymbol{\sigma}) = \sum_{\mathbf{x} \in \mathcal{S}} \log \left[\frac{1}{(|\mathcal{S}| - 1)} \sum_{\mathbf{y} \in \mathcal{S}, \mathbf{y} \neq \mathbf{x}} \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{ \frac{(x_n - y_n)^2}{2\sigma_n^2} \right\} \right]$$

$$q(\mathbf{y}|\mathbf{x}) = \frac{f(\mathbf{x}|\mathbf{y}, \boldsymbol{\sigma})}{\sum_{\mathbf{u} \in \mathcal{S}, \mathbf{u} \neq \mathbf{x}} f(\mathbf{x}|\mathbf{u}, \boldsymbol{\sigma})}, \quad \mathbf{y} \in \mathcal{S}$$

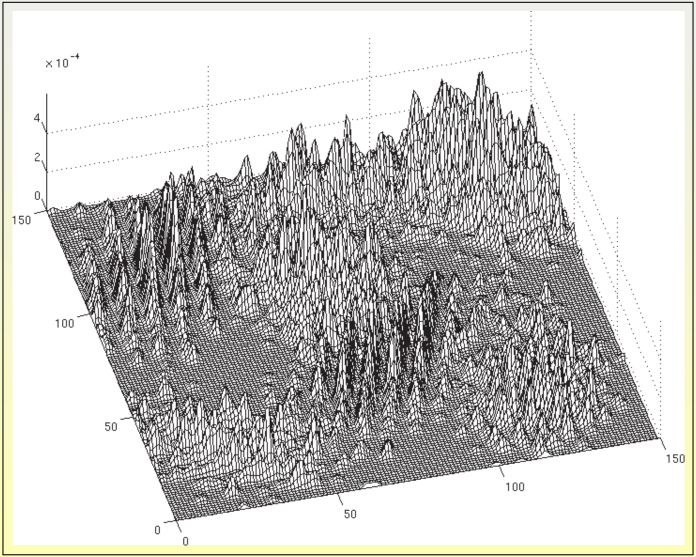
$$(\sigma_n^{'})^2 = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \sum_{\mathbf{y} \in \mathcal{S}, \mathbf{y} \neq \mathbf{x}} (x_n - y_n)^2 q(\mathbf{y}|\mathbf{x})$$

POZN. **Časově náročný výpočet!**

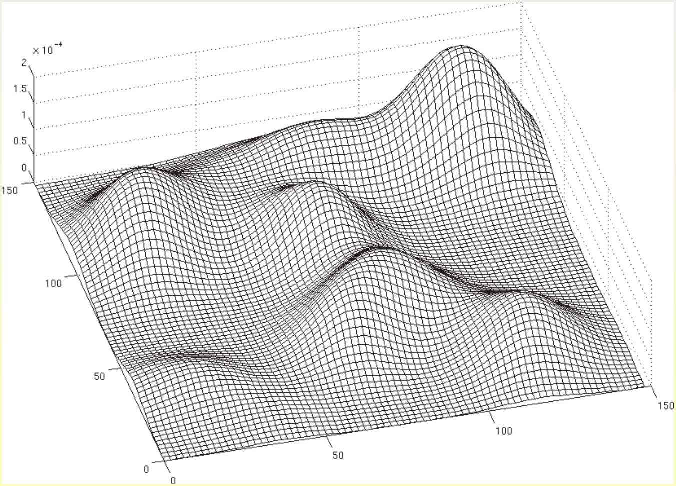
Zpět: Kompromis
Součinové směsi

ÚTIA

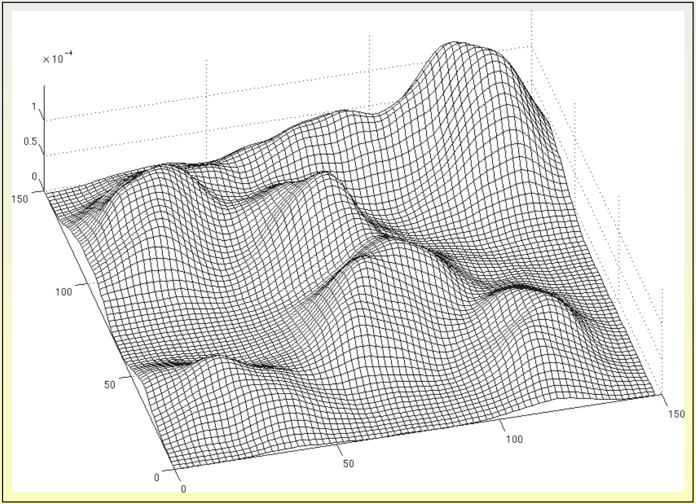
"Podhlazený "jádrový odhad



Příliš vyhlazený jádrový odhad



Optimálně vyhlazený jádrový odhad



(normální jádro s obecnou kov. maticí)

Zpět: Norm. směs

A3: Odvození marginálních distribucí ze součinové směsi

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

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

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

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

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

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

◂ Zpět - Nekompletní data ◂ Zpět - Výhody součinových směsí

A4: Explicitní řešení kroku M

Lemma (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 $\gamma(\mathbf{x}) = N(\mathbf{x})/|\mathcal{S}|$ *je relativní četnost vektoru* $\mathbf{x}$ *v* $\mathcal{S}$, *platí ekvivalentně:*

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

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

Důsledek: Maximum vážené věrohodnostní funkce lze vyjádřit jako vážený maximálně věrohodný odhad.

Zpět: EM algoritmus

ÚTIA

A4: Explicitní řešení kroku M - normální směs

normální směs s obecnou kovarianční maticí:

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

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

implicitní tvar kroku M:

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

explicitní řešení:

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

$$A_m^{'} = \sum_{\mathbf{x}\in\mathcal{S}} \gamma(\mathbf{x})\ (\mathbf{x} - \mathbf{c}_m^{'})(\mathbf{x} - \mathbf{c}_m^{'})^T = \sum_{\mathbf{x}\in\mathcal{S}} \gamma(\mathbf{x})\mathbf{x}\mathbf{x}^T - \mathbf{c}_m^{'}(\mathbf{c}_m^{'})^T$$

Zpět: EM algoritmus

A5: Odvození M-kroku pro diskrétní součinovou směs

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

$$\sum_{\xi\in\mathcal{X}_n} \delta(\xi, x_n) = 1, \quad x_n \in \mathcal{X}_n,$$

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

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

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

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

◂ Zpět: EM algoritmus ◂ Zpět: Strukturní EM

A5: Invariance vůči lineární transformaci

Invariance součinové normální směsi vůči lineární transformaci

Nechť parametry normální součinové směsi $\{w_m, \mu_{mn}, \sigma_{mn}, m \in \mathcal{M}, n \in \mathcal{N}\}$ jsou stacionárním bodem EM algoritmu, tj. splňují iterační rovnice. Dále nechť $\mathbf{y} = T(\mathbf{x})$ je lineární transformace dat $\mathbf{x} \in \mathcal{X}$ a parametrů směsi:

$$y_n = a_n x_n + b_n,\ \mathbf{x} \in \mathcal{S}, \quad \tilde{w}_m = w_m, \quad \tilde{\mu}_{mn} = a_n \mu_{mn} + b_n, \quad \tilde{\sigma}_{mn} = a_n \sigma_{mn}.$$

Potom transformované parametry $\{\tilde{w}_m, \tilde{\mu}_{mn}, \tilde{\sigma}_{mn}, m \in \mathcal{M}, n \in \mathcal{N}\}$ jsou rovněž stacionárním bodem EM algoritmu v transformovaném prostoru $\mathcal{Y}$.

Důkaz: Substitucí lze ověřit platnost rovnic:

$$F(\mathbf{y}|\tilde{\boldsymbol{\mu}}_m, \tilde{\boldsymbol{\sigma}}_m) = \frac{1}{\prod_{n=1}^{N} a_n} F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m), \quad \tilde{P}(\mathbf{y}) = \frac{1}{\prod_{n=1}^{N} a_n} P(\mathbf{x})$$

$$q(m|\mathbf{y}) = q(m|\mathbf{x}), \quad \mathbf{y} = T(\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}, \quad m \in \mathcal{M}$$

$$\tilde{\mu}_{mn} = \frac{1}{\tilde{w}_m |\mathcal{S}|} \sum_{\mathbf{y} \in \tilde{\mathcal{S}}} y_n q(m|\mathbf{y}), \quad (\tilde{\sigma}_{mn})^2 = \frac{1}{\tilde{w}_m |\mathcal{S}|} \sum_{\mathbf{y} \in \tilde{\mathcal{S}}} (y_n - \tilde{\mu}_{mn})^2 q(m|\mathbf{y})$$

Zpět: součinová normální směs

A6: Odvození kritéria strukturní optimalizace

strukturní směs je speciálním případem součinového modelu, tzn.

platí: $w_m^{'} = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}), \quad f_n^{'}(.|m) = \arg\max_{f_n(.|m)} \left\{ \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{w_m^{'}|\mathcal{S}|} \log f_n(x_n|m) \right\}$

je třeba dokázat, že optimalizace ϕ_{mn} zachovává monotonii, tj. :

$$L^{'} - L \geq \sum_{m=1}^{M} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \left\{ \sum_{m=1}^{M} q(m|\mathbf{x}) \log\left[\frac{F^{'}(\mathbf{x}|m)}{F(\mathbf{x}|m)}\right]\right\} \geq 0 \qquad \text{(viz důkaz)}$$

substitucí za $F^{'}(\mathbf{x}|m), F(\mathbf{x}|m)$ dostaneme:

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

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

Odvození kritéria strukturní optimalizace

předchozí nerovnost můžeme upravit na tvar:

$$(*) \quad L' - L \geq \sum_{m=1}^{M} \sum_{n=1}^{N} (\phi'_{mn} - \phi_{mn})\gamma'_{mn} + \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{\phi_{mn}}{|\mathcal{S}|} q(m|\mathbf{x}) \log \frac{f'_n(x_n|m)}{f_n(x_n|m)}$$

kde γ'_{mn} je kriterium strukturní optimalizace:

$$\gamma'_{mn} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log \frac{f'_n(x_n|m)}{f_n(x_n|0)}, \quad n \in \mathcal{N}, m \in \mathcal{M}$$

vzhledem k předchozí definici $f'_n(.|m)$ platí pro libovolné $f_n(\cdot|m)$ nerovnost:

$$\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log f'_n(x_n|m) \geq \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log f_n(x_n|m)$$

tj. poslední suma na pravé straně nerovnosti $(*)$ je nezáporná a ze stejného důvodu platí $\gamma'_{mn} \geq 0$ pro všechna $n \in \mathcal{N}, m \in \mathcal{M}$;

zvolíme-li $\phi'_{mn} = 1$ pouze pro r největších sčítanců γ'_{mn}, bude platit

$$L' - L \geq \sum_{m=1}^{M} \sum_{n=1}^{N} (\phi'_{mn} - \phi_{mn})\, \gamma'_{mn} \geq 0 \ \ \textbf{cbd.}$$

Zpět: Strukturní EM

Úpravy strukturního kritéria - diskrétní směs

$f_n(x_n|m),\; x_n \in \mathcal{X}_n,\; n \in \mathcal{N}$ $\approx$ **diskrétní rozložení pravděpodobnosti**

$$\gamma'_{mn} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log \frac{f'_n(x_n|m)}{f_n(x_n|0)}, \quad n \in \mathcal{N}, m \in \mathcal{M}$$

$$\sum_{\xi\in\mathcal{X}_n} \delta(\xi, x_n) = 1, \quad x_n \in \mathcal{X}_n,$$

$$\gamma'_{mn} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \Big[\sum_{\xi\in\mathcal{X}_n} \delta(\xi, x_n) \Big] \log \frac{f'_n(x_n|m)}{f_n(x_n|0)},$$

$$\gamma'_{mn} = \frac{1}{|\mathcal{S}|} \sum_{\xi\in\mathcal{X}_n} \Big[\sum_{\mathbf{x}\in\mathcal{S}} \delta(\xi, x_n) q(m|\mathbf{x}) \Big] \log \frac{f'_n(\xi|m)}{f_n(\xi|0)},$$

$$\gamma'_{mn} = w'_m \sum_{\xi\in\mathcal{X}_n} f'_n(\xi|m) \log \frac{f'_n(\xi|m)}{f_n(\xi|0)} = w'_m I(f'_n(\cdot|m), f_n(\cdot|0)),$$

γ'_{mn} $\approx$ **Kullback-Leiblerova informační divergence**

Zpět: Strukturní EM

Úpravy strukturního kritéria - normální směs

normální hustoty: $f_n(x_n|\mu_{mn},\sigma_{mn}) = \frac{1}{\sqrt{2\pi}\sigma_{mn}} \exp\left\{-\frac{(x_n-\mu_{mn})^2}{2\sigma_{mn}^2}\right\}$

$$\gamma'_{mn} = \frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x}) \log \frac{f_n(x_n|\mu'_{mn},\sigma'_{mn})}{f_n(x_n|\mu_{0mn},\sigma_{0n})}, \quad n\in\mathcal{N}, m\in\mathcal{M},$$

$$\gamma'_{mn} = w'_m \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{w'_m|\mathcal{S}|}\left[-\log\frac{\sigma'_{mn}}{\sigma_{0n}} - \frac{(x_n-\mu'_{mn})^2}{2(\sigma'_{mn})^2} + \frac{(x_n-\mu_{0n})^2}{2(\sigma_{0n})^2}\right],$$

$$\gamma'_{mn} = \frac{w'_m}{2}\left[\frac{(\mu'_{mn}-\mu_{0n})^2}{(\sigma_{0n})^2} + \frac{(\sigma'_{mn})^2}{(\sigma_{0n})^2} - 1 - \log\frac{(\sigma'_{mn})^2}{(\sigma_{0n})^2}\right] =$$

snadno lze ověřit:

$$= w'_m \int_{\mathcal{X}_n} f_n(x_n|\mu'_{mn},\sigma'_{mn}) \log\frac{f_n(x_n|\mu'_{mn},\sigma'_{mn})}{f_n(x_n|\mu_{0n},\sigma_{0n})} dx_n$$

Zpět: Strukturní EM

$\Rightarrow \quad \gamma'_{mn} \approx$ **"spojitá" Kullback-Leiblerova informační divergence**

A7: Důkaz neidentifikovatelnosti diskrétní součinové směsi

Definice identifikovatelnosti směsi (Teicher, 1963)

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. Zpět: identifikace x aproximace

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

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

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

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* Zpět: Diskrétní směs

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

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 rozložení $f_i^{'}(\cdot|m), f_i^{''}(\cdot|m)$, např. $(0 < \alpha < 1,\ \beta = 1 - \alpha)$:

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

(tj. $f_i(\cdot|m)$ je vnitřním bodem úsečky $\langle f_i^{'}(\cdot|m), f_i^{''}(\cdot|m)\rangle$)

S využitím předchozí substituce můžeme napsat

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

kde

$$w_m^{'} = \alpha w_m, \quad w_m^{''} = \beta w_m$$

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

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

Zpět: EM algoritmus

ÚTIA

A8: Alternativní důkaz monotónie 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

---

následující postup vychází z původního Schlesingerova důkazu

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

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

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

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

Alternativní důkaz monotónie EM algoritmu: $L^{'} \geq L$

Předchozí vzorce použijeme k vyjádření přírůstku věrohodnostní funkce:

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

Druhý člen na pravé straně představuje Kullback-Leiblerovu divergenci:

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

Další postup se shoduje s původním důkazem:

Po vynechání nezáporné informační divergence dostaneme nerovnost

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

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

Alternativní důkaz monotónie EM algoritmu: $L' \geq L$

S využitím substituce za w'_m podle kroku M obdržíme nerovnost

$$\sum_{m=1}^{M} \Big[\frac{1}{|\mathcal{S}|}\sum_{\mathbf{x}\in\mathcal{S}} q(m|\mathbf{x})\Big] \log\frac{w'_m}{w_m} = \sum_{m=1}^{M} w'_m \log\frac{w'_m}{w_m} \geq 0$$

Podle definice v kroku M: $\quad F'(.|m) = \arg\max_{F(.|m)} \Big\{ \sum_{\mathbf{x}\in\mathcal{S}} \frac{q(m|\mathbf{x})}{w'_m|\mathcal{S}|} \log F(\mathbf{x}|m) \Big\}$

tzn. pro libovolnou funkci $F(\mathbf{x}|m)$ platí nerovnost:

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

Z uvedených nerovností plyne monotónní vlastnost EM algoritmu:

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

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

Zpět

A9: 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:

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

která platí i pro posloupnosti $\{w^{(t)}(m)\}_{t=0}^{\infty}$, $\{q^{(t)}(\cdot|\mathbf{x})\}_{t=0}^{\infty}, m \in \mathcal{M}$:

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

Předchozí limity plynou z nerovnosti

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

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

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

A10: Maximálně věrohodné odhady a problém aproximace

Lemma

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

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

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):

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

A11: Důkaz nezápornosti Kullback-Leiblerovy divergence

Theorem (viz např. Vajda, 1992)

Pro libovolná dvě diskrétní rozložení pravděpodobnosti $\{q_1, q_2, \dots, q_M\}$, $\{q_1', q_2', \dots, q_M'\}$ *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 $q_1'/q_1 = \cdots = q_M'/q_M$, 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 - Důkaz Zpět (alternativní důkaz) Zpět (M-krok)

A12: Diskrétní součinová směs univerzálně aproximuje

Lemma (viz např. Grim, 2006)

Nechť $p^{(k)}, k = 1, \ldots, K,\ K = |\mathcal{X}|$ *jsou tabulkou definované hodnoty libovolného diskrétního rozložení pravděpodobnosti* $P(\mathbf{x})$ *na prostoru* $\mathcal{X}$*:*

$$P(\mathbf{x}^{(k)}) = p^{(k)}, \quad \mathbf{x}^{(k)} \in \mathcal{X}, \quad k = 1, \ldots, K, \quad \mathcal{X} = \cup_{k=1}^{K}\{\mathbf{x}^{(k)}\}$$

Potom diskrétní rozložení pravděpodobnosti $P(\mathbf{x})$ *může být vyjádřeno ve tvaru součinové distribuční směsi pomocí delta-funkcí*

$$P(\mathbf{x}) = \sum_{k=1}^{K} w_k F(\mathbf{x}|k) = \sum_{k=1}^{K} p^{(k)} \prod_{n \in \mathcal{N}} \delta(x_n, x_n^{(k)}), \quad \mathbf{x} \in \mathcal{X}.$$

Důkaz: Je zřejmý z uvedeného vzorce. Komponenty směsi definované pomocí delta-funkcí jsou přiřazeny jednotlivým bodům prostoru $\mathbf{x}^{(k)} \in \mathcal{X}$ a váha komponenty je rovna příslušné tabulkové hodnotě $p^{(k)}$:

$$F(\mathbf{x}|k) = \prod_{n \in \mathcal{N}} \delta(x_n, x_n^{(k)}), \quad w_k = p^{(k)}, \quad k = 1, \ldots, K.$$

POZN. Uvedený konstruktivní důkaz má pouze formální význam, aproximace s využitím EM algoritmu je numericky výhodnější.

Zpět - ("lze zapsat")

A13: EM algoritmus pro Bernoulliovskou směs

základní schema EM algoritmu v C++: **směs Bernoulliho rozložení**

```
//    Odhad parametru smesi Bernoulliho rozlozeni pomoci EM algoritmu
//=================================================================
//int       NN;                         //  dimenze binarniho vektoru (DNN=NN+1)
//int       MM;                         //  pocet komponent smesi (DMM=MM+1)
//short     X[DNN];                     //  binarni datovy vektor
//double    P[DMM][DNN],SP[DMM][DNN];//  parametry smesi (theta) a scitaci promenne
//double    W[DMM],SW[DMM];             //  vahy komponent a prislusne scitaci promenne
//double    FX[DMM];                    //  hodnoty komponent pro dany vektor X[DNN]
//double    FXM,SWM,Q,SUM,SWM;          //  pomocne promenne
//int       N,M,IT,ITERMAX;             //  pomocne promenne

for(IT=1; IT<=ITERMAX; IT++)
//*****************************************************************************
{  for(M=1; M<=MM; M++) {SW[M]=0.0; for(N=1; N<=NN; N++) SP[M][N]=0.0;}
   Q=0.0;
   for(J=1;J<=JJ;J++)                   //  cyklus pres vsechny datove vektory X
   { READ(X);  SUM=0.0;                 //  nacteni X ze vstupniho souboru
     for(M=1; M<=MM; M++)
     {  FXM=W[M];
        for(N=1; N<=NN; N++) if(X[N]==1) FXM*=P[M][N]; else FXM*=(1-P[M][N]);
        FX[M]=FXM;  SUM+=FXM;
     }  // end of M-loop
     Q=Q+log(SUM);
     for((M=1; M<=MM; M++)
     {  G=FX[M]/SUM; SW[M]+=G;  for(N=1; N<=NN; N++) if(X[N]==1) SP[M][N]+=G;
     }  //  end of M-loop
   } //  end of J-loop
   Q=Q/JJ;
   for(M=1; M<=MM; M++)                 //  vypocet novych parametru komponent
   {  SWM=SW[M];   W[M]=SWM/JJ;  for((N=1; N<=NN; N++) P[M][N]=SP[M][N]/SWM;
   }  // end of M-loop
   print(IT,Q);
}  //  end of IT-loop
//*****************************************************************************
printf("\nKonec EM algoritmu\n\n");
```

Zpět

A14: EM algoritmus pro součinovou normální směs

EM algoritmus v C++: **součinová normální směs s velkou dimenzí**

```
//    Odhad parametru normalni soucinove smesi pomoci EM algoritmu
//===================================================================================
//int IT,N,M;  long K;  double F,G,FXM,SWM,SUM,FMAX,Q0;    // globalni promenne:
//short     X[DNN];                          // datovy vektor (DNN=NN+1)
//double    FX[DMM],W[DMM],SW[DMM];          // komponenty, vahy a odhady vah komponent
//double    C[DMM][DNN], A[DMM][DNN];        // vektory prumeru a rozptylu (DMM=MM+1)
//double    SC[DMM][DNN],SA[DMM][DNN];       // nove odhady vektory prumeru a rozptylu
for(IT=1; IT<=ITMAX; IT++)
//***********************************************************************************
{ Q=0.0
  for(M=1; M<=MM; M++)                       // logaritmicke parametry a nulovani stradacu
  {  SW[M]=RMIN;          F=log(W[M]+RMIN)-NN2LN2PI;
     for(N=1; N<=NN; N++) {F-=log(A[M][N]); SC[M][N]=RMIN; SA[M][N]=RMIN;}
     W[M]=2*F;                               // kvuli deleni pri vypoctu exponentu
  }  // end of M-loop
  for(I=1;I<=K;I++)                          // cyklus pres vsechny datove vektory X
  {  READ(X);   FMAX=-RMAX;
     for(M=1; M<=MM; M++)                    // vypocet logaritmu komponent
     {  FXM=W[M]; for(N=1; N<=NN; N++) {F=(X[N]-C[M][N])/A[M][N]; FXM-=F*F;}
        FXM/=2.0f;     FX[M]=FXM;  if(FXM>FMAX) FMAX=FXM;
     } // end of M-loop
     SUM=0.0;
     for(M=1; M<=MM; M++)                    // odlogaritmovani komponent a vypocet P(X)
     {  FXM=FX[M]-FMAX;  if(FXM>MINLOG) {FXM=exp(FXM); SUM+=FXM;} else FXM=0.0;
        FX[M]=FXM;
     }  // end of M-loop
     Q=Q+log(SUM)+FMAX;                      // vypocet hodnoty verohodnostni funkce
     for(M=1; M<=MM; M++)
     {  G=FX[M]/SUM;   SW[M]+=G;
        for(N=1; N<=NN; N++) {F=X[N]; SC[M][N]+=G*F; SA[M][N]+=G*F*F;}
     }  //  end of M-loop
  }  //  end of K-loop
  Q/=K;
  for(M=1; M<=MM; M++)                        //  vypocet novych parametru komponent
  {  SWM=SW[M];    W[M]=SWM/K;
     for(N=1; N<=NN; N++)
     {  F=SC[M][N]/SWM;   C[M][N]=F;   A[M][N]=sqrt(SA[M][N]/SWM-F*F);
     }  //  end of N-loop
  }   //  end of M-loop
  printf("\nIT=%2d   Q=%15.7lf  \n",IT,Q);
  //*********************************************************************************
} //  end of IT-loop
```

Prof. M.I. Schlesinger se svou ženou

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

Zpět

Literatura 1/12

- 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.

Zpět

Literatura 2/12

- 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.18, No.3, pp.173-190. *http://dml.cz/dmlcz/124132*
- 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. (1984): On structural approximating multivariate discrete probability distributions. *Kybernetika*, Vol. 20, No. 1, pp. 1-17, 1984. *http://dml.cz/dmlcz/125676*
- Grim J. (1986): Multivariate statistical pattern recognition with nonreduced dimensionality, *Kybernetika*, Vol. 22, pp. 142-157. *http://dml.cz/dmlcz/125022*

Zpět

ÚTIA

Literatura 3/12

Grim, J. (1986): Sequential decision-making in pattern recognition based on the method of independent subspaces. In: *Proceedings of the DIANA II Conference on Discriminant Analysis*, (Ed. F. Zitek), Mathematical Institute of the AS CR, Prague 1986, pp. 139-149.

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. (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*, (Ed. R.Trappl), Vienna, April 1992, World Scientific, Singapore 1992, pp. 997-1005. ▸ Paper Award

Grim J. and 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.

◂ Zpět

Literatura 4/12

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.

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. and 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 et al. (Eds.), (pp: 527-540), Valencia: Sociedad Espanola de Sistemas Generales, 1999.

Grim J. (2000): Self-organizing maps and probabilistic neural networks. Neural Network World, 3(10): 407-415.

Paper Award Zpět

Literatura 5/12

Grim J., Kittler J., Pudil P. and Somol P. (2000): Combining multiple classifiers in probabilistic neural networks, In *Multiple Classifier Systems*, Eds. Kittler J., Roli F., Springer, 2000, pp. 157 - 166.

Grim J., Pudil P. and 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. Paper Award

Grim J., Kittler J., Pudil P. and 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., Boček P. and 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.

Grim J. (2001): Latent Structure Analysis for Categorical Data. Research Report No. 2019. ÚTIA AV ČR, Praha 2001, 13 pp. 23

Literatura 6/12

- Grim J., Kittler J., Pudil P. and Somol P. (2002): Multiple classifier fusion in probabilistic neural networks. *Pattern Analysis & Applications* Vol. 5, No. 7, pp. 221-233.
- Grim J. and Haindl M. (2003): Texture Modelling by Discrete Distribution Mixtures. Computational Statistics and Data Analysis, 3-4 **41**, pp. 603-615.
- Grim J., Just P. and Pudil P. (2003): Strictly modular probabilistic neural networks for pattern recognition. *Neural Network World*, Vol. 13 , No. 6, pp. 599-615.
- Grim J., Somol P., Pudil P. and 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., Hora J. and Pudil P. (2004): Interaktivní reprodukce výsledků sčítání lidu se zaručenou ochranou anonymity dat. *Statistika*, Vol. 84, No. 5, pp. 400-414.

Zpět

Literatura 7/12

- Grim J., Haindl M., Somol P., Pudil P. and Kudo M. (2004): A Gaussian mixture-based colour texture model. In: *Proc. of the 17th International Conference on Pattern Recognition.* IEEE, Los Alamitos 2004, pp. 177-180.

- Grim J., Somol P., Haindl M. and 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., Haindl M., Pudil P. and 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.

- Grim J. (2006): EM cluster analysis for categorical data. In: *Structural, Syntactic and Statistical Pattern Recognition.* (Yeung D. Y., Kwok J. T., Fred A. eds.), (LNCS 4109). Springer, Berlin 2006, pp. 640-648.

Zpět

Literatura 8/12

J. Grim (2007): Neuromorphic features of probabilistic neural networks. *Kybernetika*, Vol. 43, No. 5, pp.697-712. *http://dml.cz/dmlcz/135807*

Grim J. and Hora, J. (2008): Iterative principles of recognition in probabilistic neural networks. *Neural Networks*, Special Issue, 6 **21**, 838–846 Paper Award

Grim J., Somol P., Haindl M. and J. Daneš (2009): Computer-Aided Evaluation of Screening Mammograms Based on Local Texture Models. *IEEE Trans. on Image Processing*, Vol. 18, No. 4, pp. 765-773. Paper Award

Grim J., Hora J., Boček P., Somol P. and P. Pudil (2010): Statistical Model of the 2001 Czech Census for Interactive Presentation. *Journal of Official Statistics*. Vol. 26, No. 4, pp. 673–694.

Grim J., Somol P. and Pudil P. (2010): Digital Image Forgery Detection by Local Statistical Models. *Proc. 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Los Alamitos, IEEE computer society, Echizen, I. et al., eds., pp. 579-582. Paper Award

Zpět

Literatura 9/12

- J. Grim (2011): Preprocessing of Screening Mammograms Based on Local Statistical Models. *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL 2011*, Barcelona, ACM, pp. 1-5

- Grim, J. (2014). Sequential pattern recognition by maximum conditional informativity. *Pattern Recognition Letters*, Vol. 45C, pp. 39-45. *http:// dx.doi.org/10.1016/j.patrec.2014.02.024* Paper Award

- Gyllenberg M., Koski T., Reilink E. and M. Verlaan (1994): Non-uniqueness in probabilistic numerical identification of bacteria. *Journal of Applied Probability*, Vol. 31, pp. 542–548.

- 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.

Zpět

Literatura 10/12

- Isaenko O.K. and 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.
- McLachlan, G.J. and Krishnan, T. (1997): The EM algorithm and extensions, John Wiley & Sons, New York.
- McLachlan G.J. and Peel D. (2000): *Finite Mixture Models*, John Wiley & Sons, New York, Toronto, (2000)
- Meng X.L. and Van Dyk D. (1997): The EM Algorithm—an Old Folk-song Sung to a Fast New Tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 59, No. 3, pp. 511-567.

Zpět

Literatura 11/12

- 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. and 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.
- Teicher H. (1963): Identifiability of finite mixtures. *Ann. Math. Statist.*, Vol. 34, pp. 1265-1269.
- Teicher H. (1968): Identifiability of mixtures of product measures. *Ann. Math. Statist.*, Vol. 39, pp. 1300-1302.

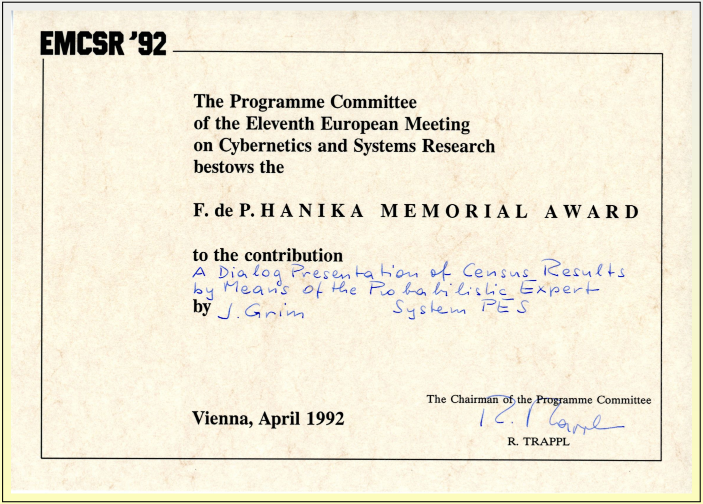
Zpět

Literatura 12/12

- Schlesinger M.I. (1968): Relation between learning and self learning in pattern recognition (in Russian), *Kibernetika*, (Kiev), No. 2, pp. 81-88.
- 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. and 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.

Zpět

Paper Award



EMCSR '92

The Programme Committee
of the Eleventh European Meeting
on Cybernetics and Systems Research
bestows the

F. de P. HANIKA MEMORIAL AWARD

to the contribution
A Dialog Presentation of Census Results
by Means of the Probabilistic Expert
System PES
by J. Grim

Vienna, April 1992

The Chairman of the Programme Committee
R. TRAPPL

Eleventh European Meeting on Cybernetics and Systems Research,
Vienna, April 1992

Zpět

Paper Award


Certificate

Jiri Grim

has been recognized with a
SPRINGER BEST PRESENTATION AWARD
at the International ICSC Symposium on
NEURAL COMPUTATION (NC'2000)

NC 2000

Prof. Hans-Heinrich Bothe
Symposium Chair

Hans R. Ryffel
ICSC President

Second ICSC Symposium on Neural Computation, Berlin, 2000

Paper Award

ÚTIA

Ústav teorie informace a automatizace
Akademie věd České republiky v.v.i.

Uděluje v soutěži o Nejlepší publikaci 2009

2. cenu

Jiřímu Grimovi

za práci

Computer-Aided Evaluation of Screening Mammograms Based on Local Texture Models

IEEE Trans. on Image Processing, vol. 18, No. 4, pp. 765–773, ISSN 1057–7149, April 2009.

(spoluautoři Somol P., Haindl M., Daneš J.)

ředitel ústavu

IEEE Transactions on Image Processing 18(4): 765-773, 2009

Paper Award

Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing
IIH-MSP-2010
October 15-17, 2010, Darmstadt, Germany

Best Paper Award Certificate

in recognition of the professional excellence of the paper

Digital Image Forgery Detection by Local Statistical Models

by

Jiri Grim, Petr Somol

&

Pavel Pudil

Xiamu Niu
General Chair
Harbin Institute of Technology
China

Alexander Nouak ,
Conference executive Chair
Fraunhofer IGD
Germany

Jeng-Shyang Pan
Program Committee Chair
National Kaohsiung University of Applied Sciences
Taiwan

Isao Echizen,
Program Committee Chair
National Institute of Informatics
Japan

Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP Darmstadt, 2010

Paper Award


ÚTIA

Ústav teorie informace a automatizace
Akademie věd České republiky v.v.i.

Uděluje v soutěži o Nejlepší publikaci 2014
3. cenu

Jiřímu Grimovi
za práci

Sequential pattern recognition by maximum conditional informativity

Pattern Recognition Letters, Elsevier (2014), pp. 39-45

ředitel ústavu

Pattern Recognition Letters, Vol. 45C, pp. 39-45, 2014

Zpět

Paper Award

ÚTIA

Ústav teorie informace a automatizace
Akademie věd České republiky

Uděluje v soutěži o Nejlepší publikaci 2008

2. cenu

Jiřímu Grimovi

za práci

Iterative principles of recognition in probabilistic neural networks

Neural Networks 21 (2008), 838-846.
(spoluautor Hora J.)

ředitel ústavu

Neural Networks, 21(6): 838–846, 2008

Zpět

Paper Award

ÚTIA

Ústav teorie informace a automatizace
Akademie věd České republiky

Cena za nejlepší článek

udělena v roce 2000

Jiřímu Grimovi

za práci

*Self-Organizing Maps
and Probabilistic Neural Networks*

Neural Network World No. 3, 2000, 407-415

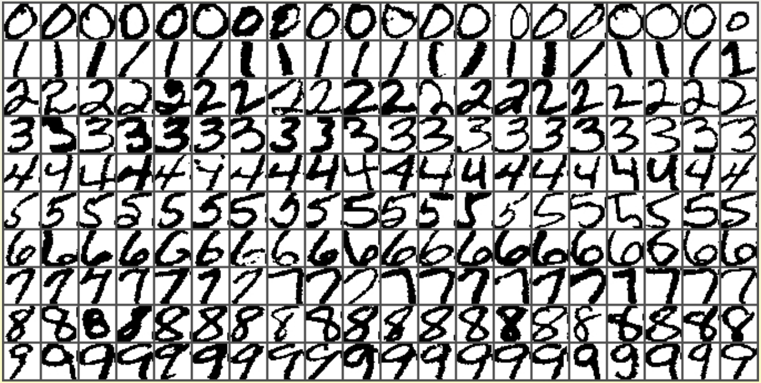
ředitel ústavu

Neural Network World, 3(10): 407-415, 2000

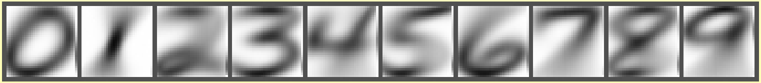
Zpět

Příklad 1: Databáze rukou psaných číslic NIST SD19

příklady číslic NIST SD19 normalizovaných na velikost rastru 32x32



"průměrné číslice" (marginální pravděpodobnosti trénovacích dat)



Příklad 1: Rozpoznávání číslic na binárním rastru

řádky: četnost jednotlivých rozhodnutí pro číslice z dané třídy
poslední sloupec: procento chybně klasifikovaných číslic z dané třídy
poslední řádek: četnost číslic chybně zařazených do dané třídy

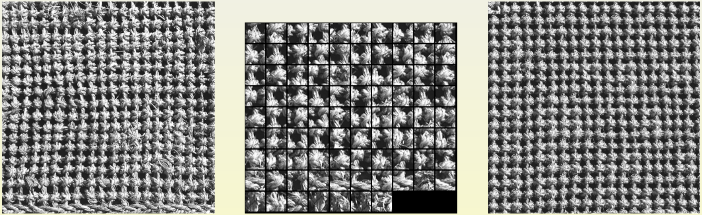
CLASS	0	1	2	3	4	5	6	7	8	9	chybně
testovací:	20182	22352	20038	20556	19577	18303	19969	20947	19790	19767	neg.:
0	19950	8	43	19	39	32	36	0	38	17	1.1%
1	2	22162	30	4	35	7	18	56	32	6	0.9%
2	32	37	19742	43	30	9	8	29	90	16	1.5%
3	20	17	62	20021	4	137	2	28	210	55	2.6%
4	11	6	19	1	19170	11	31	51	30	247	2.1%
5	25	11	9	154	4	17925	39	6	96	34	2.1%
6	63	10	17	6	23	140	19652	1	54	3	1.6%
7	7	12	73	10	73	4	0	20497	22	249	2.1%
8	22	25	53	97	30	100	11	11	19369	72	2.1%
9	15	13	25	62	114	22	3	146	93	19274	2.5%
chybně poz.:	197	139	537	396	352	462	148	328	665	699	1.84%

Celková chyba v procentech: 1.84%

Příklad 2: Modelování textur pomocí normální směsi

textura "ratan": predikce pomocí optimálních "dlaždic" μ_m^*

původní textura | optimální "dlaždice" | syntéza vzorkováním



"realistická" syntéza: komponenty $\boldsymbol{\mu}_m$ nahrazeny podobnými částmi původní textury $\boldsymbol{\mu}_m^*$ optimálně vyhledanými podle kriteria:

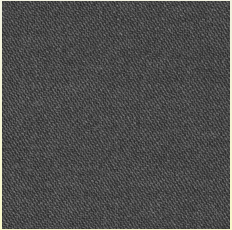
$$\boldsymbol{\mu}_m^* = \arg\min_{\mathbf{x}\in\mathcal{S}} \{\|\mathbf{x} - \boldsymbol{\mu}_m\|^2\}$$

POZN. Metoda "stochastického vzorkování" je blízká modelování textur kombinováním konečného počtu spojitě navazujících "dlaždic".

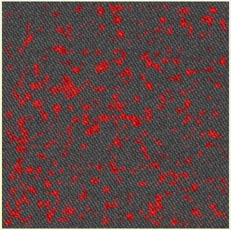
Příklad 3: Vyhledávání poruch a odchylek v textuře

Analýza nepravidelnosti textury “obklad”:

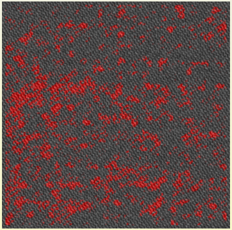
původní obrázek



L-nevěrohodnost



LR-nevěrohodnost

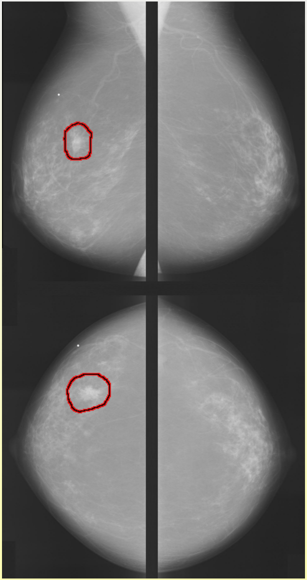


červené zabarvení: $\approx$ neobvyklá (atypická) místa (poruchy) v textuře

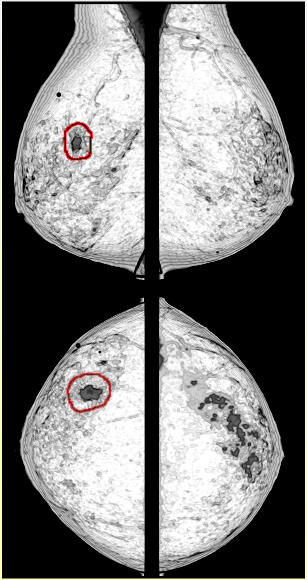
- střední obrázek: $\approx$ nízké hodnoty věrohodnosti: $\log P(\mathbf{x})$
- pravý obrázek: $\approx$ nízké hodnoty věrohodnostního poměru: $\log \frac{P(\mathbf{x})}{P_0(\mathbf{x})}$

Příklad 4: Vyhodnocování screeningových mamogramů

ověřený nález



věrohodnostní analýza



Příklad 5: Forenzní analýza obrazových dat

Původní obrázek s vloženou oválnou částí v levém horním rohu

Příklad 5: Forenzní analýza obrazových dat

Oválná část v levém horním rohu je zřetelně světlejší v důsledku odlišných lokálních vlastností textury

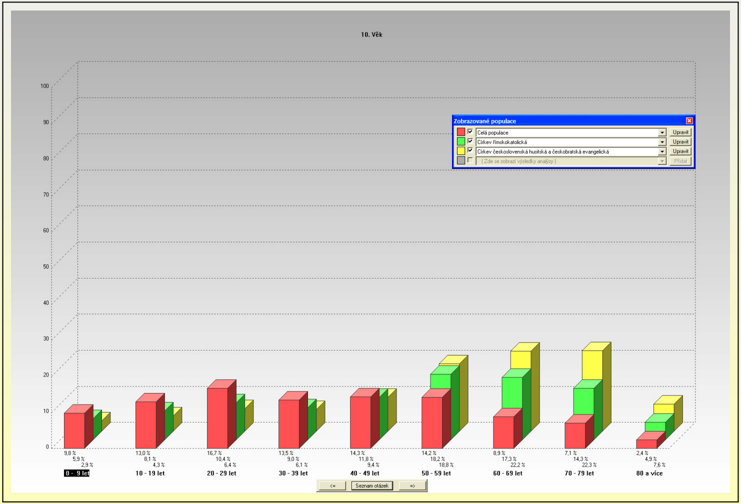
Příklad 6: Predikce chybějících částí obrázku

původní poškozený obrázek

Příklad 6: Predikce chybějících částí obrázku

opravený obrázek

Census: srovnání věkového složení různých subpopulací



POZN. Možnost analýzy subpopulací je omezena pouze přesností modelu.