

Bayesovské metody - příklady

Jan Kracík

jan.krakik@vsb.cz

Příklad - odhad střední hodnoty normálního rozdělení

Klasický přístup

Nechť $x_1, x_2, \dots, x_t$ jsou i.i.d. náhodné veličiny,
normální rozdělení, známý rozptyl ($\sigma^2 = 1$)

$$f_{\mu}(x_{\tau}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_{\tau} - \mu)^2}{2}\right),$$

$\mu \in \mathbb{R} \dots$ neznámý parametr (stř. hodnota)

Z nezávislosti plyne

$$f_{\mu}(x_1, x_2, \dots, x_t) = \prod_{\tau=1}^t f_{\mu}(x_{\tau}) = (2\pi)^{-\frac{t}{2}} \exp\left(-\frac{1}{2} \sum_{\tau=1}^t (x_{\tau} - \mu)^2\right).$$

Věrohodnostní funkce:

$$l(\mu) = f_{\mu}(x_1, \dots, x_t)$$

Příklad - odhad střední hodnoty normálního rozdělení

Klasický přístup - maximálně věrohodný odhad

maximálně věrohodný odhad $\hat{\mu}_{ML}$:

$$\hat{\mu}_{ML}(x_1, x_2, \dots, x_t) = \arg \max_{\mu \in \mathbb{R}} l(\mu)$$

Maximalizací věrohodnostní funkce dostaneme:

$$\hat{\mu}_{ML}(x_1, \dots, x_t) = \frac{1}{t} \sum_{\tau=1}^t x_{\tau}.$$

Příklad - odhad střední hodnoty normálního rozdělení

Klasický přístup - intervalový odhad

při známém rozptylu σ^2 :

$$\frac{\frac{1}{t} \sum_{\tau=1}^t x_{\tau} - \mu}{\sigma} \sqrt{t} \sim \mathcal{N}(0, 1)$$

Odtud dostáváme intervalový odhad

$$\left\langle \frac{1}{t} \sum_{\tau=1}^t x_{\tau} - \frac{1}{\sqrt{t}} z_{1-\frac{\alpha}{2}}, \frac{1}{t} \sum_{\tau=1}^t x_{\tau} + \frac{1}{\sqrt{t}} z_{1-\frac{\alpha}{2}} \right\rangle,$$

kde $z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}$ jsou kvantily $\mathcal{N}(0, 1)$.

Pro všechna $\mu \in \mathbb{R}$ tedy platí:

$$P \left(\frac{1}{t} \sum_{\tau=1}^t x_{\tau} - \frac{1}{\sqrt{t}} z_{1-\frac{\alpha}{2}} \leq \mu \leq \frac{1}{t} \sum_{\tau=1}^t x_{\tau} + \frac{1}{\sqrt{t}} z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - model

Statistický model: podmíněná hustota pravděpodobnosti

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2}\right)$$

Neznámý parametr: μ , náhodná veličina

Sdružená hustota náhodného vektoru $x_1, x_2, \dots, x_t$:

$$f(x_1, x_2, \dots, x_t|\mu) = \prod_{\tau=1}^t f(x_\tau|\mu) = (2\pi)^{-\frac{t}{2}} \exp\left(-\frac{1}{2} \sum_{\tau=1}^t (x_\tau - \mu)^2\right).$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - apriorní hustota

- z praktických důvodů normální rozdělení
- velký rozptyl - nemáme vhodnou apriorní informaci

Například:

$$f(\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{\mu^2}{100}\right)$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - aposteriorní hustota

$$\begin{aligned} f(\mu|x_1, \dots, x_t) &\propto f(x_1, \dots, x_t|\mu)f(\mu) \\ &\propto \exp\left(-\frac{1}{2}\sum_{\tau=1}^t(x_\tau - \mu)^2\right) \exp\left(-\frac{1}{2}\frac{\mu^2}{100}\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\left(t + \frac{1}{100}\right)\mu^2 - 2\left(\sum_{\tau=1}^t x_\tau\right)\mu\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\frac{\left(\mu - \frac{1}{t + \frac{1}{100}}\sum_{\tau=1}^t x_\tau\right)^2}{\frac{1}{t + \frac{1}{100}}}\right) \end{aligned}$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - aposteriorní hustota

$$f(\mu|x_1, \dots, x_t) \propto \exp \left(-\frac{1}{2} \frac{\left(\mu - \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau} \right)^2}{\frac{1}{t + \frac{1}{100}}} \right)$$

$$\mu|x_1, \dots, x_t \sim \mathcal{N} \left(\frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau}, \frac{1}{t + \frac{1}{100}} \right)$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - bodový odhad

Ztrátová funkce: kvadratická, praktická volba

$$L(a, \mu) = (a - \mu)^2. \quad (1)$$

Optimálním rozhodnutí (odhad) $\hat{\mu}$:

$$\hat{\mu} \in \operatorname{Argmin}_{a \in \mathbb{R}} \int (a - \mu)^2 f(\mu | x_1, \dots, x_t) d\mu$$

$$\begin{aligned} & \int (a - \mu)^2 f(\mu | x_1, \dots, x_t) d\mu \\ &= a^2 - 2a \int \mu f(\mu | x_1, \dots, x_t) d\mu + \int \mu^2 f(\mu | x_1, \dots, x_t) d\mu \end{aligned}$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - bodový odhad

globální minimum v bodě $\int \mu f(\mu | x_1, \dots, x_t) d\mu$

Bodový odhad:

$$\hat{\mu} = \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau}.$$

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - bodový odhad MAP

Jiná možnost bodového odhadu: maximální a posteriori odhad, obdoba MLE

$$\hat{\mu}_{MAP}(x_1, x_2, \dots, x_t) = \arg \max_{a \in \mathcal{A}} f(\mu | x_1, x_2, \dots, x_t) \quad (2)$$

Pro normální a posteriori rozdělení:

$$\hat{\mu}_{MAP} = \hat{\mu}$$

Nebayesovské: Nelze vyjádřit jako minimalizaci ztrátové funkce.

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - intervalový odhad

Intervalový odhad: množina C , pro niž platí

$$P(\mu \in C | x_1, \dots, x_t) = 1 - \alpha.$$

Požadujeme, aby C byla co nejmenší, tj. C má tvar

$$C = \{\mu \in \mathbb{R} | f(\mu | x_1, \dots, x_t) \geq k\}$$

pro nějaké $k > 0$.

Příklad - odhad střední hodnoty normálního rozdělení

Bayesovský přístup - intervalový odhad

Aposteriorní rozdělení je normální, proto C je interval

$$C = \left\langle \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau} - \frac{1}{\sqrt{t + \frac{1}{100}}} z_{1-\frac{\alpha}{2}}, \right.$$

$$\left. \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau} + \frac{1}{\sqrt{t + \frac{1}{100}}} z_{1-\frac{\alpha}{2}} \right\rangle$$

Příklad - odhad střední hodnoty normálního rozdělení

Srovnání klasického a bayesovského přístupu

- Neznámý parametr μ :
 - klasický přístup: bod z $\mathbb{R}$
 - bayesovský přístup: náhodná veličina s hodnotami $\mathbb{R}$
- Statistický model:
 - klasický přístup: rozdělení pravděpodobnosti závislé na parametru μ
 - bayesovský přístup: podmíněnou hustotou pravděpodobnosti
 - pro konkrétní $\mu_0 \in \mathbb{R}$ jsou $f_{\mu_0}(x_\tau)$ i $f(x_\tau | \mu = \mu_0)$ stejná rozdělení pravděpodobnosti

Příklad - odhad střední hodnoty normálního rozdělení

Srovnání klasického a bayesovského přístupu

- Intervalový odhad:

- klasický: dvojice náhodných veličin $T_l(x)$, $T_u(x)$,

$$\forall \mu \in \mathbb{R} : P(T_l(x) \leq \mu \leq T_u(x)) = 1 - \alpha,$$

rozdělení $T_l(x)$, $T_u(x)$ je určeno parametrem μ .

- bayesovský: množina C , $P(\mu \in C | x_1, \dots, x_t) = 1 - \alpha$,
přímočará interpretace v porovnání s klasickým intervalovým odhadem

- Nezávislost pozorování:

- klasický: $x_1, x_2, \dots, x_t$ nezávislé
- bayesovský: $x_1, x_2, \dots, x_t$ podmíněně nezávislé při daném μ

$$f(x_1, x_2, \dots, x_t) = \int f(\mu) \prod_{\tau=1}^t f(x_\tau | \mu) d\mu,$$

$$\prod_{\tau=1}^t f(x_\tau) = \prod_{\tau=1}^t \int f(\mu) f(x_\tau | \mu) d\mu.$$

Příklad - odhad střední hodnoty normálního rozdělení

Srovnání klasického a bayesovského přístupu

- Apriorní informace:

Všechna rozdělení veličiny $x_{\mathcal{T}}$ tvoří nekonečně rozměrný prostor.

Předpoklad, že data pochází z rozdělení $f_{\mu}(x)$ pro nějaké $\mu \in \mathbb{R}$ přináší informaci, že skutečné rozdělení $x_{\mathcal{T}}$ leží v konkrétním konečně rozměrném podprostoru.

Předpoklad *konkrétního statistického modelu* $f_{\mu}(x)$ je tedy silnou *apriorní informací* o neznámém rozdělení veličin $x_{\mathcal{T}}$.

Příklad - odhad střední hodnoty normálního rozdělení

Srovnání klasického a bayesovského přístupu

- Apriorní informace (pokračování):
 - klasický přístup: Veškerá apriorní informace vyjádřena statistickým modelem. Nemáme možnost do úlohy jednoduše zanést další informaci.
 - bayesovský přístup: Kromě informace reprezentované modelem do úlohy vstupuje detailnější informace reprezentována apriorní hustotou.

Příklad - Bernoulliovy pokusy

Model

$x_1, x_2, \dots, x_t \dots$ posloupnost Bernoulliových pokusů

Statistický model:

$$f(x_\tau | p) = p^{\delta(x_\tau, 1)} (1 - p)^{\delta(x_\tau, 0)},$$

$$f(x_1, x_2, \dots, x_t | p) = \prod_{\tau=1}^t f(x_\tau | p),$$

$p \in \langle 0, 1 \rangle$ je neznámý parametr.

Příklad - Bernoulliovy pokusy

Apriorní hustota

Apriorní rozdělení: beta rozdělení s parametry $\nu_0, \nu_1 \in \mathbb{R}^+$:

$$f(p) = \frac{(1-p)^{\nu_0-1} p^{\nu_1-1}}{B(\nu_0, \nu_1)},$$

kde $B : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ je beta funkce

$$\begin{aligned} B(a, b) &= \int_0^1 t^{a-1} (1-t)^{b-1} dt \\ &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \end{aligned}$$

Gamma : $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ je gamma funkce,

$$\Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} dt.$$

Pro $a > 0$ platí

$$\Gamma(a) = (a-1)\Gamma(a-1).$$

Příklad - Bernoulliovy pokusy

Aposteriorní hustota

Aposteriorní hustota:

$$\begin{aligned}f(p|x_1, \dots, x_t) &\propto f(p)f(x_1, \dots, x_t|p) \\&\propto f(p) \prod_{\tau=1}^t f(x_\tau|p) \\&\propto (1-p)^{\nu_0-1} p^{\nu_1-1} \prod_{\tau=1}^t p^{\delta(x_\tau,1)} (1-p)^{\delta(x_\tau,0)}.\end{aligned}$$

Příklad - Bernoulliovy pokusy

Aposteriorní hustota

Označme $V = \sum_{\tau=1}^t \delta(x_\tau, 1)$. Pak

$$\begin{aligned} f(p|x_1, \dots, x_t) &\propto (1-p)^{\nu_0-1} p^{\nu_1-1} p^V (1-p)^{t-V} \\ &\propto (1-p)^{\nu_0+t-V-1} p^{\nu_1+V-1}. \end{aligned}$$

Odtud plyne: Aposteriorní rozdělení je také beta rozdělení,

$$f(p|x_1, \dots, x_t) = \frac{(1-p)^{\nu_0+t-V-1} p^{\nu_1+V-1}}{B(\nu_0+t-V, \nu_1+V)}.$$

Příklad - Bernoulliovy pokusy

Predikce

Predikce x_{t+1} na základě pozorování $x_1, \dots, x_t$:

$$\begin{aligned} f(x_{t+1} | x_1, \dots, x_t) &= \int f(x_{t+1}, p | x_1, \dots, x_t) dp \\ &= \int f(x_{t+1} | p, x_1, \dots, x_t) f(p | x_1, \dots, x_t) dp \\ &= \int f(x_{t+1} | p) f(p | x_1, \dots, x_t) dp \\ &= \frac{(1-p)^{\delta(x_{t+1}, 0) + \nu_0 + t - V - 1} p^{\delta(x_{t+1}, 1) + \nu_1 + V - 1}}{B(\nu_0 + t - V, \nu_1 + V)} \\ &= \frac{B(\delta(x_{t+1}, 0) + \nu_0 + t - V, \delta(x_{t+1}, 1) + \nu_1 + V)}{B(\nu_0 + t - V, \nu_1 + V)} \end{aligned}$$

Příklad - Bernoulliovy pokusy

Predikce

Pro $x_{t+1} = 0$ dostaneme:

$$\begin{aligned} f(x_{t+1} = 0 | x_1, \dots, x_t) &= \frac{B(\nu_0 + t - V + 1, \nu_1 + V)}{B(\nu_0 + t - V, \nu_1 + V)} \\ &= \frac{\Gamma(\nu_0 + t - V + 1)\Gamma(\nu_1 + V)}{\Gamma(\nu_0 + \nu_1 + t + 1)} \frac{\Gamma(\nu_0 + \nu_1 + t)}{\Gamma(\nu_0 + t - V)\Gamma(\nu_1 + V)} \\ &= \frac{\nu_0 + t - V}{\nu_0 + \nu_1 + t} \end{aligned}$$

Obdobně pro $x_{t+1} = 1$:

$$f(x_{t+1} = 1 | x_1, \dots, x_t) = \frac{\nu_1 + V}{\nu_0 + \nu_1 + t}.$$

Příklad - segmentace digitálního obrazu

Veličiny

Digitální obraz:

- $n \times n$ bodů
- souřadnice: $(i, j) \in S = \{1, \dots, n\} \times \{1, \dots, n\}$
- dva druhy náhodné výplně: náhodný vektor

$$x = (x_{ij})_{(i,j) \in S}$$

s hodnotami v $\{0, 1\}^{(n^2)}$

- obrazová informace: náhodný vektor

$$y = (y_{ij})_{(i,j) \in S}$$

s hodnotami v $\mathbb{R}^{(n^2)}$.

Příklad - segmentace digitálního obrazu

Model

Pozorované veličiny y_{ij} v bodě (i, j)

- závisí pouze na druhu textury v tomto bodě (hodnotě x_{ij})
- mají normální rozdělení se známými parametry $(\mu_0, \sigma_0^2), (\mu_1, \sigma_1^2) \in \mathbb{R} \times \mathbb{R}^+$

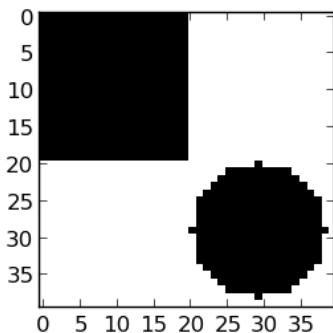
$$f(y|x) = \prod_{(i,j) \in S} f(y_{ij}|x_{ij})$$

$$f(y_{ij}|x_{ij}) = \prod_{c \in \{0,1\}} \left(\frac{1}{\sqrt{2\pi}\sigma_c} \exp \left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2} \right) \right)^{\delta(x_{ij}, c)}$$

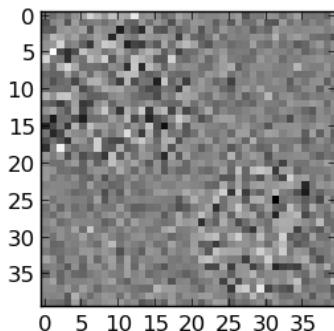
Příklad - segmentace digitálního obrazu

Data

Příklad segmentace a obrazu o rozměru 40×40 px
Textury se tedy liší pouze rozptylem.



(a) Segmentace (hodnota x)



(b) Pozorovaný obraz (hodnota y)

Příklad - segmentace digitálního obrazu

Formulace problému

Odhad segmentace $\leftrightarrow$ nalezení odhadu $\hat{x}$

Potřebujeme aposteriorní hustotu $f(x|y)$!

Náhodný vektor x dimenze n^2 je neznámým parametrem.

Potřebujeme vhodnou apriorní hustotu $f(x)$.

Příklad - segmentace digitálního obrazu

Uniformní apriorní hustota

Úvaha: “Nemáme žádnou informaci o tom, jak jsou jednotlivé typy výplně zastoupeny ani informaci o tom, kde se jednotlivé objekty v obraze nachází a proto žádné hodnoty segmentace nemůžeme preferovat před jinými.”

→ Volíme rovnoměrnou apriorní hustotu.

$$f(x) = \frac{1}{2^{(n^2)}}, \quad (3)$$

Pro každé $(i, j) \in S$ je $f(x_{ij}) = \frac{1}{2}$, a veličiny x_{ij} jsou navzájem nezávislé, tj.

$$f(x) = \prod_{(i,j) \in S} f(x_{ij}). \quad (4)$$

Příklad - segmentace digitálního obrazu

Uniformní apriorní hustota

$$\begin{aligned}f(x|y) &= \dots \\&= \prod_{(i,j) \in S} f(x_{ij}|y) \\&= \prod_{(i,j) \in S} f(x_{ij}|y_{ij}).\end{aligned}$$

$$f(x_{ij} = c|y_{ij}) = \frac{\frac{1}{\sigma_c} \exp\left(-\frac{(y_{ij}-\mu_c)^2}{2\sigma_c^2}\right)}{\sum_{\tilde{c} \in \{0,1\}} \frac{1}{\sigma_{\tilde{c}}} \exp\left(-\frac{(y_{ij}-\mu_{\tilde{c}})^2}{2\sigma_{\tilde{c}}^2}\right)}$$

Příklad - segmentace digitálního obrazu

Ztrátová funkce

Zvolme tzv. 0 – 1 ztrátovou funkci

$$L(a, x) = 1 - \delta(a, x)$$

Střední hodnotu ztr. funkce vzhledem k aposteriornímu rozdělení:

$$E[L(a, x)|y] = 1 - f(x = a|y)$$

Odtud plyne

$$\hat{x} \in \underset{a \in \{0,1\}^{(n^2)}}{\text{Argmax}} f(a|y).$$

Příklad - segmentace digitálního obrazu

Optimální odhad

x_{ij} podmíněně nezávislé při daném y

$$E[L(a, x)|y] = 1 - \prod_{(i,j) \in S} f(x_{ij} = a_{ij}|y_{ij}).$$

Odtud plyne:

$$\hat{x} \in \underset{a \in \{0,1\}^{(n^2)}}{\operatorname{Argmin}} E[L(a, x)|y]$$

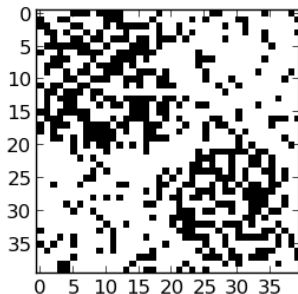
$$\Leftrightarrow$$

$$\hat{x}_{ij} \in \underset{x_{ij} \in \{0,1\}}{\operatorname{Argmax}} f(x_{ij}|y_{ij}).$$

Příklad - segmentace digitálního obrazu

Optimální odhad pro uniformní apriorní hustotu

Výsledný odhad segmentace (pro data z předchozího obrázku):



Obrázek: Odhad segmentace pro uniformní apriorní hustotu

Nevyužili jsme důsledně apriorní informaci!

Příklad - segmentace digitálního obrazu

Znovu a lépe - Gibbsova apriorní hustota

Apriorní informace: scéna zachycuje “větší jednoduché objekty”!

Oblasti se stejnou hodnotou segmentace budou s větší pravděpodobností tvořit větší souvislé plochy.

Gibbsova distribuce:

$$f(x) = \frac{1}{Z(\beta)} \exp(-\beta H(x))$$
$$Z(\beta) = \int \exp(-\beta H(x)) dx$$

$H(x)$... “energie”

Příklad - segmentace digitálního obrazu

Energetická funkce

Energii $H(x)$ lze vhodně zvolit úměrnou počtu sousedních pixelů s odlišnými hodnotami segmentace.

Množina bodů, které přímo sousedí s bodem $(i, j) \in S$:

$$N(i, j) = \{(k, l) \in S : |i - k| + |j - l| = 1\}$$

Energii $H(x)$ lze vyjádřit ve tvaru

$$H(x) = - \sum_{(i,j) \in S} \left(\sum_{(k,l) \in N(i,j)} \delta(x_{ij}, x_{kl}) \right).$$

$H(x)$ je přímo úměrná celkové délce hranice mezi jednotlivými oblastmi.

Příklad - segmentace digitálního obrazu

Aposteriorní hustota

$$f(x|y) \propto$$

$$\left(\prod_{(i,j) \in S} \prod_{c \in \{0,1\}} \left(\frac{1}{\sigma_c} \exp \left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2} \right) \right)^{\delta(x_{ij}, c)} \right) \exp(-\beta H(x)).$$

Nelze faktorizovat $\Rightarrow$ obtížná maximalizace!

Příklad - segmentace digitálního obrazu

Aposteriorní hustota - normalizační konstanta

Normalizační konstanta pro aposteriorní hustotu:

$$\left(\sum_{x \in \{0,1\}^{(n^2)}} \left(\prod_{(i,j) \in S} \prod_{c \in \{0,1\}} \left(\frac{1}{\sigma_c} \exp \left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2} \right) \right)^{\delta(x_{ij}, c)} \right) \right)^{-1}$$

Pro obrázek 40×40 pixelů řádově 10^{481} operací!

Příklad - segmentace digitálního obrazu

Aposteriorní hustota - praktické řešení

Lze řešit pomocí Monte Carlo (MC) metod - generujeme nezávislé vzorky - odhad stř. hodnoty z aposteriorního rozdělení atd.

Problémy:

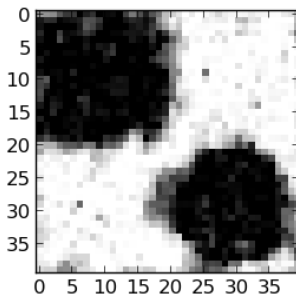
- obrovská dimenze, neefektivní vzorkování, prakticky nepoužitelné
- neumíme najít normalizační konstantu, nelze použít běžné MC metody

Řešení: Markov Chain Monte Carlo - generujeme závislé vzorky z Markovského řetězce se stacionárním rozdělením $f(x|y)$.

Příklad - segmentace digitálního obrazu

Aposteriorní hustota - vizualizace výsledku pomocí MCMC

Aproximace aposteriorních marginálních hustot pro Gibbsovu apriorní hustotu s $\beta = 0.4$



Obrázek: Vizualizace marginálních aposteriorních hustot veličin x_{ij} . Bod na souřadnici (i, j) je zobrazen s odstínem šedi úměrným $f(x_{ij} = 1 | y)$ (pro $f(x_{ij} = 1 | y) = 0$ černě, $f(x_{ij} = 1 | y) = 1$ bíle)

Příklad - segmentace digitálního obrazu

Závěr

- Využití apriorní informace může být rozhodující pro úspěšné řešení.
- S pomocí vhodných metod (MCMC) lze odhadovat i modely s obrovským množstvím parametrů.