Apriorní rozdělení

Jan Kracík
jan.kracik@vsb.cz

Apriorní rozdělení

Apriorní rozdělení (spolu s modelem) reprezentuje informaci o neznámém parametru θ, která je dostupná předem, tj. bez informace z dat.

Má zásadní význam v případě, že data jsou málo informativní.

Problém

Jak sestavit apriorní rozdělení při nedostatečném množství apriorní informace?

Obecné poznámky

- Apriorní rozdělení často volíme “přibližné”. Lepší než nic!
- Apriorní rozdělení musí nést co nejméně “špatné” informace.
- Preferujeme *kompatibilitu s apriorní informací* před jednoduchostí. V praxi volíme kompromis.

Apriorní rozdělení - příklad

Příklad

Viz segmentace obrazu, Gibbsovo apriorní rozdělení:

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

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

$$\sum_{(i,j)\in S} \left(\sum_{(k,l)\in N(i,j)} \delta(x_{ij}, x_{kl}) \right) = 2\left(n^2 - \text{délka hranice}\right)$$

Apriorní rozdělení - příklad

- Zvolené apriorní rozdělení neobsahuje informace o počtu objektů a jejich poloze, jen o délce hranice: Preferuje segmentaci s kratší délkou hranice mezi oblastmi ("správná" informace).
- Apriorní rozdělení je přibližné. Lze např. zvolit vhodnější potenciál (lineární členy) i parametry.
- Obsahuje "špatnou" informaci: nejpravděpodobnější je segmentace s délkou hranice 0.

$$x \in \operatorname{Argmax} f(x) \Leftrightarrow (\forall (i,j) \in S : x_{ij} = 0) \vee (\forall (i,j) \in S : x_{ij} = 1)$$

- Díky informaci z dat je "špatná" část informace potlačena, "správná" zůstává (je kompatibilní s daty). Celkově má apriorní hustota pozitivní vliv na dobrý výsledek.
- Apriorní rozdělení lze dále vylepšit (např. hierarchický model).

Aproximace apriorního rozdělení

V praxi často využíváme přibližné postupy.

- Θ konečná množina: subjektivní přiřazení pravděpodobností
- Θ nekonečná (nespočetná): konečné dělení množiny, dále jako pro diskrétní θ
  problém: určení pravděpodobností chvostů
- někdy je jednodušší zvolit apriorní rozdělení $f(x) = \int f(x|\theta)f(\theta)d\theta$; odtud lze (někdy) určit $f(\theta)$.

Apriorní rozdělení s maximální entropií

Informace ve formě momentů, pravděpodobností zajímavých jevů, ...

Obecně:

$$\int g_k(\theta) f(\theta) d\theta = \omega_k, \tag{1}$$

kde $g_k : \Theta \to \mathbb{R}, \omega_k \in \mathbb{R}, k = 1, \ldots n$. Apriorní rozdělení pak lze volit tak, aby mělo maximální entropii vzhledem k daným omezením.

Pro diskrétní parametr θ:

$$f(\theta = \theta_i) = \frac{\exp\left(-\sum_k \lambda_k g_k(\theta_i)\right)}{\sum_i \exp\left(-\sum_k \lambda_k g_k(\theta_i)\right)},$$

kde λ_k jsou Lagrangeovy multiplikátory; určíme tak, aby platila omezení (1).

Apriorní rozdělení s maximální entropií

Podobně pro spojitý parametr:

$$f(\theta) = \frac{\exp\left(-\sum_k \lambda_k g_k(\theta)\right)}{\int \exp\left(-\sum_k \lambda_k g_k(\theta)\right) d\theta}$$

Hierarchické apriorní rozdělení

Často lze zvolit vhodné apriorní rozdělení až na neznámý parametr. K tomuto parametru přistupujeme stejně jako ke všem ostatním: považujeme jej za náhodnou veličinu a hledáme pro něj vhodné apriorní rozdělení.

Ve složitějších úlohách velmi častý přístup.

Hierarchickým apriorním rozdělením rozumíme rozdělení s hustotou $f(\theta)$ ve tvaru

$$f(\theta) = \int_{\Theta_1 \times \cdots \times \Theta_n} f(\theta|\theta_1) f(\theta_1|\theta_2) \dots f(\theta_{n-1}|\theta_n) f(\theta_n) d\theta_1 \dots d\theta_n.$$

Hierarchické apriorní rozdělení

Příklad
Skrytý markovovský řetězec s neznámými pravděpodobnostmi přechodu

Pozorované veličiny: $y_1, \ldots, y_t$
Skryté veličiny (parametry): $x_1, \ldots, x_t, \eta$

$$\begin{aligned} f(y_1, \ldots, y_t | x_0, \ldots, x_t, \eta) &= \prod_{\tau=1}^{t} f(y_\tau | x_\tau) \\ f(x_1, \ldots, x_t | x_0, \eta) &= \prod_{\tau=1}^{t} f(x_\tau | x_{\tau-1}, \eta) \\ f(\eta) &= \ldots \end{aligned}$$

Konjugované rodiny apriorních rozdělení

Množina (rodina) $\mathcal{F}$ hustot pravděpodobnosti parametru θ se nazývá *konjugovaná* pro model $f(x|\theta)$, jestliže pro každé apriorní rozdělení $f(\theta)$ z $\mathcal{F}$ a každé pozorování x je aposteriorní rozdělení $f(\theta|x)$ jakožto hustota θ rovněž z $\mathcal{F}$.

Příklady

Viz odhad parametru

- alternativního rozdělení,
- normálního rozdělení se známým rozptylem.

Poznámky

- Množina všech hustot θ je konjugovaná, ale nezajímavá.
- Zajímavé jsou malé a parametrizované konjugované rodiny.
- Konjugovaná rodina hustot je určena modelem.
- Ne pro každý model existují konjugovaná apriorní rozdělení.
- Výpočet aposteriorního rozdělení spočívá v určení parametru.
- Velmi praktické, ale ne vždy vhodné.

Definice

Uvažujme statistický model $f(x|\theta)$ náhodné veličiny X s hodnotami v $\mathcal{X}$. Nechť pro $T : \mathcal{X} \to \mathbb{R}^n$, $n \in \mathbb{N}$ platí

$$f(x|T(X) = t, \theta) = f(x|T(X) = t).$$

Statistiku T pak nazýváme *postačující statistikou modelu* $f(x|\theta)$.

Postačující statistiky

Postačující statistika nese stejnou informaci o parametru θ jako náhodná veličina X.

Nechť $t = T(x)$. Pak

$$\begin{aligned}
f(\theta|x) &= f(\theta|x,t) \\
&= \frac{f(x,t|\theta)f(\theta)}{\int f(x,t|\theta)f(\theta)d\theta} \\
&= \frac{f(x|t,\theta)f(t|\theta)f(\theta)}{\int f(x|t,\theta)f(t|\theta)f(\theta)d\theta} \\
&= \frac{f(x|t)f(t|\theta)f(\theta)}{f(x|t)\int f(t|\theta)f(\theta)d\theta} \\
&= \frac{f(t,\theta)f(\theta)}{\int f(t,\theta)f(\theta)d\theta} = f(\theta|t)
\end{aligned}$$

Exponenciální rodina

Statistický model, který lze vyjádřit ve tvaru

$$f(x|\theta) = C(\theta)\exp\left(\sum_{i=1}^{n} B_i(\theta)S_i(x)\right),$$

kde $B_1, \ldots, B_n : \Theta \to \mathbb{R}$, $S_1, \ldots, S_n : \mathcal{X} \to \mathbb{R}$, $C : \Theta \to \mathbb{R}$, se nazývá *exponenciální rodina* s postačujícími statistikami $S_1, \ldots, S_n$.

Apriorní hustoty ve tvaru

$$f(\theta) \propto (C(\theta))^v \exp\left(\sum_{i=1}^{n} B_i(\theta) V_i\right) \tag{2}$$

pro nějaké $(v, V_1, \dots, V_n) \in \mathbb{R}^{n+1}$ představují konjugované hustoty pro exponenciální rodinu s postačujícími statistikami $S_1, \dots, S_n$.

$$f(\theta|x_1, \dots, x_t) \propto (C(\theta))^{v+t} \exp\left(\sum_{i=1}^{n} B_i(\theta)\left(V_i + \sum_{\tau=1}^{t} S_i(x_\tau)\right)\right)$$

Aposteriorní rozdělení má stejný tvar jako apriorní pro hodnoty parametrů

$$\left(v + t, V_1 + \sum_{\tau=1}^{t} S_1(x_\tau), \dots, V_n + \sum_{\tau=1}^{t} S_n(x_\tau)\right).$$

Exponenciální rodina

Pro libovolnou posloupnost pozorování $x_1, \ldots, x_t$ je vektorová funkce

$$\left(\sum_{\tau=1}^{t} S_1(x_\tau), \ldots, \sum_{\tau=1}^{t} S_n(x_\tau) \right)$$

postačují statistikou modelu

$$f(x_1, \ldots, x_t | \theta) = (C(\theta))^t \exp\left(\sum_{i=1}^{n} B_i(\theta) \sum_{\tau=1}^{t} S_i(x_\tau) \right).$$

Dimenze této postačující statistiky *nezávisí* na počtu pozorování t.

Konjugovaná apriorní hustoty vybraných jednoparametrických modelů

Neznámý parametr: θ, ostatní parametry jsou známé

Model	Rozdělení x	Konjug. rozdělení θ
Normální	$\mathcal{N}(\theta, \sigma^2)$	$\mathcal{N}(\mu, \xi^2)$
Normální	$\mathcal{N}(\mu, \frac{1}{\theta})$	$\mathcal{G}(\alpha, \beta)$
Poissonovo	$\mathcal{P}(\theta)$	$\mathcal{G}(\alpha, \beta)$
Gama	$\mathcal{G}(\nu, \theta)$	$\mathcal{G}(\alpha, \beta)$
Binomické	$\mathcal{B}i(n, \theta)$	$\mathcal{B}e(\alpha, \beta)$
Multinomické rozdělení	$\mathcal{M}_k(\theta_1, \dots, \theta_k)$	$\mathcal{D}(\alpha_1, \dots, \alpha_k)$

Neinformativní apriorní hustoty

V případě, že není k dispozici žádná vhodná apriorní informace, lze využít tzv. neinformativních apriorních rozdělení. Nejčastěji používaná jsou Jeffreysovy a referenční hustoty.

Uvažujme statistický model $f(x|\theta)$. *Fisherova informace* pro $\Theta \subset \mathbb{R}$ je definována vztahem

$$I(\theta) = E\left(\left(\frac{\partial \ln f(x|\theta)}{\partial \theta}\right)^2 \middle| \theta\right)$$

Poznámka

$$E\left(\left(\frac{\partial \ln f(x|\theta)}{\partial \theta}\right) \middle| \theta\right) = 0$$

$I(\theta)$ charakterizuje množství informace, které o parametru θ přinese pozorování x. Tato informace je závislá na θ.

Jeffeysova apriorní hustota

Jeffreysova apriorní hustota pro model $f(x|\theta)$ je definována jako

$$f(\theta) \propto (I(\theta))^{\frac{1}{2}}.$$

Jeffreysovya apriorní hustota je často nevlastní (nemá konečný integrál, nejde o skutečnou hustoty).

Jeffeysova apriorní hustota

Pro $\Theta \subset \mathbb{R}^n$ definujeme *Fisherovu informační matici*:

$$(I(\Theta))_{ij} = E\left(\left(\frac{\partial \ln f(x|\theta)}{\partial \theta_i}\right)\left(\frac{\partial \ln f(x|\theta)}{\partial \theta_j}\right)\middle|\, \theta\right)$$

Jeffreysova apriorní hustota:

$$f(\theta) \propto (\det I(\theta))^{\frac{1}{2}}$$