Základní pojmy z bayesovské statistiky

Obsah

1 Úvod

1.1 Veličiny

- Řízení u_t - veličina, kterou nastavujeme a tak ovlivňujeme řízený proces.
  - Výstup y_t - modelovaná veličina, kterou lze na konci periody změřit.
  - Vstup v_t - externí veličina, která ovlivňuje výstup, můžeme ji měřit ale nemůžeme ji ovlivňovat.
  - Šum e_t - neměřená porucha (nulová střední hodnota a konstantní rozptyl r)
  - Stav x_t - vnitřní dynamická veličina, kterou nelze měřit a která má vliv na výstup.
  - Ukazovátko c_t - vnitřní veličina, která v modelu směsi ukazuje na aktivní komponentu.

1.2 Kódování diskrétních veličin

Několik diskrétních veličin lze zakódovat do jedné, s více hodnotami.

Příklad: $x_1, x_2 \in \{1, 2\}$ lze kódovat do jedné y podle následujícího schematu

x_1	x_2		y
1	1	$\rightarrow$	1
1	2	$\rightarrow$	2
2	1	$\rightarrow$	3
2	2	$\rightarrow$	4

Potom $y \in \{1, 2, 3, 4\}$.

1.3 Náhodná veličina

Náhodná veličina je veličina, která je ovlivněna náhodou (poruchou, šumem apod.). Její charakteristiky (jako např. střední hodnota, rozptyl apod.) jsou konstantní.

Existují dva druhy náhodných veličin

1. diskrétní - má jen konečný nebo spočetný počet různých hodnot,
2. spojité - její hodnoty jsou reálná čísla (často nezáporná).

1.4 Náhodný proces

Náhodný proces je náhodná veličina indexovaná časem, tj. náhodná veličina, jejíž charakteristiky se mění v čase.

Jak náhodná veličina tak i čas mohou být buď diskrétní nebo spojité.

Nás budou zajímat jen procesy diskrétní v čase - tzv diskrétní posloupnosti s hodnotami jak diskrétními tak i spojitými.

1.5 Diskrétní čas

Pod pojmem vzorkování v diskrétním čase máme na mysli měření veličiny v okamžicích, následujících po sobě s nějakou pevnou periodou. Označíme-li τ spojitý čas T periodu vzorkování, pak platí

$$\tau = tT + \tau_0$$

kde τ_0 je počáteční čas měření (pro $t = 0$), který se pro jednoduchost klade roven nule a $t = 1,\, 2,\, \cdots$ je diskrétní čas. Lze tedy říci, že diskrétní čas počítá periody vzorkování.

2 Model

2.1 Podmíněná hustota pravděpodobnosti

Uvažujeme dvě náhodné veličiny x a y. Podmíněná hustota pravděpodobnosti náhodné veličiny x pro $x = x_0$ se značí $f(y|x_0)$. Je to hustota pravděpodobnosti veličiny y určená z dvojic $(y, x = x_0)$.

Například

Pro soubor tvořený hodnotami v následující tabulce

y	*1*	*2*	*1*	*1*	*2*	*2*	*1*	*2*	*1*	*1*	*2*	*1*	*2*	*2*	*1*	*2*
x	*1*	*1*	*1*	*2*	*1*	*1*	*2*	*1*	*1*	*2*	*1*	*2*	*2*	*1*	*2*	*2*

bude $f(y|1)$ tvořeno z dvojic

y	*1*	*2*	*1*	*2*	*2*	*2*	*1*	*2*	*2*
x	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*

a tedy

$$f(x|1) = \begin{array}{c|cc} y & 1 & 2 \\ \hline P(y) & \frac{3}{9} = \frac{1}{3} & \frac{6}{9} = \frac{2}{3} \end{array}$$

Tomu odpovídá i definice

$$f(y|x = x_0) = \frac{f(y, x = x_0)}{f(x = x_0)},$$

tedy $f(y, x = x_0)$ - je počet dvojic $(y, x = x_0)$ dělený počtem všech dvojic; a $f(x = x_0)$ je počet dvojic kde $x = x_0$ dělený opět počtem všech dvojic. Počet všech dvojic se vykrátí a počet dvojic, kde $x = x_0$ normalizuje podmíněno pravděpodobnost tak, aby součet přes y byl jedna.

2.2 Bílý šum

Bílý šum e_t je posloupnost navzájem nezávislých identicky rozdělených náhodných veličin s nulovou dtření hodnotou a konstantním rozptylem. Používá se pro něj zkratka **i.i.d.** (Inedpendent and Identically Distributed).

V regresním modelu je to náhodná porucha. Tady je důležité, že tato náhodná veličina je nezávislá na hodnotách v regresní vektoru ψ_t - což bílý šum splňuje.

2.3 Spojitý model

Spojitý stochastický model je obecně vyjádřen jako podmíněná hustota pravděpodobnosti

$$f(y_t|\psi_t, \Theta)$$

kde y_t je výstup, ψ_t je vektor, obsahující hodnoty veličin, které v čase t ovlivňují výstup y_t a Θ jsou parametry modelu.

Poznámka

Veličiny y_t a ψ_t určují strukturu modelu, parametry Θ vyjadřují vztahy mezi veličinami modelu.

Nejčastěji používaným modelem je regresní model, který zmíněnou hustotu pravděpodobnosti definuje pomocí rovnice

$$y_t = \psi_t^{'}\theta + e_t =$$

$$= b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + \cdots + a_n y_{t-n} + b_n u_{t-n} + k + e_t$$

kde a_i, b_i jsou regresní koeficienty, k je absolutní člen a e_t je bílý šum.

2.4 Diskrétní model

Diskrétní stochastický model je definován jako pravděpodobnostní tabulka (uvedeme pro veličiny $y_t \mid u_t$ a y_{t-1}, což je řízený model s pamětí, a tedy v dostatečné obecnosti)

$$f(y_t|u_t, y_{t-1}) = \begin{array}{c|cccc} [u_t,\ y_{t-1}] & y_t = 1 & y_t = 2 & \cdots & y_t = n_y \\ \hline 1,\ 1 & \Theta_{1|11} & \Theta_{2|11} & \cdots & \Theta_{n_y|11} \\ 1,\ 2 & \Theta_{1|2} & \Theta_{2|12} & \cdots & \Theta_{n_y|12} \\ 2,\ 1 & \Theta_{1|21} & \Theta_{2|21} & \cdots & \Theta_{n_y|21} \\ 2,\ 2 & \Theta_{1|22} & \Theta_{2|22} & \cdots & \Theta_{n_y|22} \end{array}$$

Poznámky

- Tento model je napsán pro binární veličiny. Pro vícehodnotové veličiny bude model stejný, jen větší.
  - Pokud bude v modelu více veličin, nebo veličiny budou vícerozměrné, lze použít kódování veličin.

2.5 Stavový model

Stavový stochastický model je zadán dvěma podmíněnými hustotami pravděpodobností

$$f\left(x_{t+1}|x_t,u_t\right) \text{ a } f\left(y_t|x_t,u_t\right)$$

které, ve spojitém případě lze zadat rovnicemi

$$\begin{aligned} x_{t+1} &= Mx_t + Nu_t + w_t \\ y_t &= Ax_t + Bu_t + v_t \end{aligned}$$

kde M, N, A, B jsou maticové parametry modelu, x_t je stav, y_t, u_t jsou data a w_t, v_t jsou poruchy s nulovou střední hodnotou a konstantním kovariančními maticemi r_w a r_v.

2.6 Směsový model

Model směsi distribucí je zadán modely komponent (spojité nebo diskrétní modely běžného typu) a modelem ukazovátka (diskrétní model). Ukazovátko svou hodnotou v každém čase ukazuje na tzv. aktivní komponentu (tj. komponentu, která v daném čase nejlépe popisuje aktuální data).

Model má tvar

$$f\left(y_t, c_t|y\left(t-1\right), \Theta, \alpha\right) = f\left(y_t|c_t, \psi_t, \Theta_{c_t}\right) f\left(c_t|\varphi_t, \alpha\right)$$

kde c_t je ukazovátko v čase t, ψ_t, φ_t jsou regresní vektory a Θ, α jsou parametry. $f\left(y_t|c_t, \psi_t, \Theta_{c_t}\right)$ pro $c_t = 1, 2, \cdots, n_c$ jsou komponenty a $f\left(c_t|\varphi_t, \alpha\right)$ je model ukazovátka.

2.7 Statický model

Statický model je model bez paměti, tj. v regresním vektoru se neobjevují hodnoty starší modelované veličiny.

Statický regresní model

$$y_t = b_0 u_t + b_1 u_{t-1} + \cdots + b_n u_{t-n} + k + e_t$$

Statický diskrétní model

u_t	1	2	$\cdots$	m
$f\left(y_t\|u_t\right)$	p_1	p_2	$\cdots$	p_m

kde $p_i \geq 0$ a $\sum p_i = 1$.

2.8 Dynamický model

Dynamický model popisuje dynamické jevy, tj. v regresní vektoru se objevují hodnoty starších modelovaných veličin.

Dynamický regresní model

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + \cdots + a_n y_{t-n} + b_n u_{t-n} + k + e_t$$

Dynamický diskrétní model

$$
f\left(y_t|u_t, y_{t-1}\right) = \begin{array}{c|cccc}
[u_t,\ y_{t-1}] & y_t = 1 & y_t = 2 & \cdots & y_t = n_y \\
\hline
1,\ 1 & \Theta_{1|11} & \Theta_{2|11} & \cdots & \Theta_{n_y|11} \\
1,\ 2 & \Theta_{1|2} & \Theta_{2|12} & \cdots & \Theta_{n_y|12} \\
2,\ 1 & \Theta_{1|21} & \Theta_{2|21} & \cdots & \Theta_{n_y|21} \\
2,\ 2 & \Theta_{1|22} & \Theta_{2|22} & \cdots & \Theta_{n_y|22}
\end{array}
$$

2.9 Řád modelu

Řád modelu je dán maximálním zpožděním modelované veličiny v regresním vektoru.

2.10 Parametry modelu

Parametry modelu určují vztahy mezi modelovanou veličinou a veličinami v regresním modelu. Hodnoty parametrů se většinou získají odhadem z měřených veličin.

U regresního vektoru jsou regresní koeficienty a rozptyl šumu; u diskrétního modelu jsou to pravděpodobnosti z tabulky modelu.

Obecným popisem parametrů je hustota pravděpodobnosti parametrů podmíněná změřenými daty

$$f\left(\Theta|d\left(t\right)\right)$$

kde $d\left(t\right)$ jsou data, změřená od začátku odhadu. Jsou zde započítána i tzv. apriorní data - změřená ještě před začátkem odhadování.

2.11 Konstanta modelu

Konstanta modelu k patří mezi parametry-modelu. Použije se v případě, kdy modelovaná veličina nemá nulovou střední hodnotu.

2.12 Regresní vektor

Regresní vektor modelu obsahuje hodnoty veličin, na kterých závisí modelovaná veličina v daném časovém okamžiku.

3 Odhad

3.1 Statistika odhadu

Odhad parametrů se provádí na základě měřených dat. Funkce, která uchovává informaci o parametrech z naměřených dat se nazývá statistika. Například statistikou pro odhad parametrů statického normálního regresního modelu

$$y_t = k + e_t$$

je součet výstupů $S_t = \sum_{\tau=1}^{t} y_t$ a počet měření $\kappa_t = t$.

3.2 Exponenciální třída rozdělení

Důležitou vlastností aposteriorního rozdělení je tzv. reprodukovatelnost. Tou myslíme skutečnost, že při přepočtu apriorní hustoty na aposteriorní podle Bayesova vzorce, dostaneme stejný tvar hustoty pravděpodobnosti, jen s přepočtenými statistikami.

Například pro odhad parametrů statického normálního regresního modelu

$$y_t = k + e_t$$

je statistikou součet výstupů $S_t = \sum_{\tau=1}^{t} y_t$ a počet měření $\kappa_t = t$. Aposteriorní hustotou pravděpodobnosti pro odhad parametru k je

$$f\left(k|d\left(t\right)\right) \sim r^{-0.5\kappa_t} \exp\left\{-0.5\right\}$$

3.3 Věrohodnostní funkce

Věrohodnostní funkce je definována jako součin hustot pravděpodobností modelu s dosazenými změřenými daty. Tedy:

Máme model

$$f\left(y_t|\psi_t, \Theta\right)$$

a měřená data $d\left(N\right) = \left\{y_t, u_t\right\}_{t=1}^{N}$; $\psi_t = \left[u_t, y_{t-1}, u_{t-1}, \cdots, y_{t-n}, u_{t-n}\right]$ je regresní vektor. Potom věrohodnostní funkce je

$$L_N\left(\Theta\right) = \prod_{t=1}^{N} f\left(y_t|\psi_t, \Theta\right)$$

Poznámka

Věrohodnostní funkce je podstatnou části aposteriorní hustoty pravděpodobnosti

$$f\left(\Theta|d\left(N\right)\right) \propto L_N\left(\Theta\right) f\left(\Theta|d\left(0\right)\right)$$

kde $f\left(\Theta|d\left(0\right)\right)$ je počáteční (apriorní) rozdělení na parametrech konstruované z expertní (apriorní) znalosti.

3.4 Bayesův vzorec

Bayesův vzorec dává vztah mezi apriorní a aposteriorní hustotou pravděpodobnosti. Popisuje vývoj hustoty pravděpodobnosti na parametrech pro průběžně měřená data. Je základem pro bayesovské odhadování.

Pro měřená data $d\left(N\right) = \left\{y_t, u_t\right\}_{t=1}^{N}$ a model $f\left(y_t|\psi_t, \Theta\right)$ platí

$$f\left(\Theta|d\left(t\right)\right) \propto f\left(y_t|\psi_t, \Theta\right) f\left(\Theta|d\left(t-1\right)\right)$$

a $f\left(\Theta|d\left(0\right)\right)$ je počáteční rozdělení na parametrech.

Poznámka

Pro správné fungování Bayesova vzorce je třeba předpokládat platnost přirozených podmínek řízení, tedy

$$f\left(\Theta|u_t, d\left(t-1\right)\right) = f\left(\Theta|d\left(t-1\right)\right)$$

nebo, což je totéž

$$f\left(u_t|\Theta, d\left(t-1\right)\right) = f\left(\Theta|d\left(t-1\right)\right)$$

3.5 Bodový odhad

Bodový odhad je číslo (vektor), které v určitém smyslu optimálně vyjadřuje hodnotu parametru.

Podle definice optimality to může být

- střední hodnota $\hat{\Theta}_t = E\left[\Theta|d\left(t\right)\right] = \int_{\Theta^*} \Theta f\left(\Theta|d\left(t\right)\right)$
- maximum-likelihood (ML) $\hat{\Theta}_t = \arg\max_{\Theta\in\Theta^*} L_t\left(\Theta\right)$

Poznámka

Pokud v druhém případě místo věrohodnostní funkce použijeme aposteriorní hustotu pravděpodobnosti, hovoříme o MAP odhadu (maximum aposteriori probability)

- *maximum aposterori probability (MAP)* $\hat{\Theta}_t = \arg\max_{\Theta\in\Theta^*} f\left(\Theta|d\left(t\right)\right)$

3.6 Apriorní hustota pravděpodobnosti

Apriorní hustotu pravděpodobnosti nazýváme tu hustotu, ze které rekurzivně počítáme aposteriorní (podle Bayesova vztahu). V čase t je to tedy

$$f\left(\Theta|d\left(t-1\right)\right).$$

3.7 Aposteriorní hustota pravděpodobnosti

Aposteriorní hustota pravděpodobnosti dává kompletní stochastický odhad parametrů, založený na počáteční znalosti a průběžně měřených datech. V čase t s daty $d\left(t\right)$ je to

$$f\left(\Theta|d\left(t\right)\right)$$

3.8 Počáteční rozdělení parametrů

Počátečním (apriorním na začátku odhadování) rozdělení nazýváme rozdělení na parametrech vycházející pouze z expertní znalosti nebo dat, měřených ještě před začátkem odhadování - tzv. apriorních dat.

4 Predikce

4.1 Prediktivní hustota pravděpodobnosti

Prediktivní hustota pravděpodobnosti je

$$f\left(y_{t+k}|d\left(t\right)\right)$$

která odhaduje výstup y_{t+k} při znalosti dat do času t - tedy k kroků dopředu.

4.2 Bodová predikce

Bodová predikce je analogie bodového odhadu parametrů. Podle zvoleného kriteria počítáme hodnotu která nejlépe vyjadřuje odhad budoucí hodnoty výstupu. Počítáme jako

- střední hodnotu $\hat{y}_{t+k} = \int_{y^*_{t+k}} y_{t+k} f\left(y_{t+k}|d\left(t\right)\right)$ nebo
- argument maxima prediktivní hustoty pravděpodobnosti $\hat{y}_{t+k} = \arg\max_{y_{t+k}} f\left(y_{t+k}|d\left(t\right)\right)$.

Poznámka

Pro $k = 1$ dostáváme předpověď příštího výstupu, a to přímo z modelu

$$f\left(y_t|u_t, d\left(t-1\right)\right) = f\left(y_t|\psi_t\right)$$

Po změření výstupu y_t a výpočtu bodové jednokrokové predikce $\hat{y}_t$ můžeme spočítat chybu predikce $\hat{e}_t = y_t - \hat{y}_t$. Ta je důležitou charakteristikou např. pro kvalitu odhadu.

4.3 n-kroková bodová predikce

Spočteme ji s pomocí n-krokové prediktivní hustoty pravděpodobnosti $f\left(y_{t+n}|d\left(t\right)\right)$ například jako střední hodnotu

$$\hat{y}_{t+n} = \int_{y^*_{t+n}} y_{t+n} f\left(y_{t+n}|d\left(t\right)\right) dy_{t+n}$$

Příklad

Ukážeme tří-krokovou predikci s regresním modelem $y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} = e_t$.

Apriorní data jsou y_0 a předpokládáme znalost všech u_t, pro $t = 0, 1, \cdots, 3$.

Počítáme

$$\hat{y}_1 = b_0 u_1 + a_1 y_0 + b_1 u_1$$
$$\hat{y}_2 = b_0 u_2 + a_1 \hat{y}_1 + b_1 u_2$$
$$\hat{y}_3 = b_0 u_3 + a_1 \hat{y}_2 + b_1 u_3$$

Zde postupně neznámá y_i nahrazujeme jejich predikcemi $\hat{y}_i$, spočítanými v minulém kroku.

5 Filtrace

5.1 Kovariance

Správný odhad kovariančních matic šumů stavového modelu jsou velmi důležité pro správný odhad stavu systému. Tyto kovarianční matice by měly správně odrážet poruchy stavové a výstupní veličiny. Tyto poruchy jsou

$$w_t = x_{t+1} - Mx_t - Nu_t$$
$$v_t = y_t - Ax_t - Bu_t$$

Protože ale stav neznáme, není jednoduché tyto matice určit.

Možná interpretace je tato: kovariance w_t říká, jak moc dovolíme stavu aby se měnil a kovariance v_t říká totéž o měřeném výstupu. Tedy, poruchy v y_t, který měříme zahrneme do stavu, pokud jsou menší, než dovoluje kovariance w_t, zbytek jde do šumu výstupu.

5.2 Kalmanův filtr

5.3 Nelineární filtrace

6 Klasifikace

6.1 Klastrování

Klastrování je detekce shluků bodů v datovém prostoru. Každý datový záznam je číselný vektor představující bod v datovém prostoru.

Pokud je systém multimodální, to znamená, že systém pracuje v několika různých stavech, pak data z každého pracovního stavu tvoří shluk, který nazýváme klastr.

6.2 Klasifikace

Klasifikace je třídění přicházejících (měřených) dat do tříd. Třídy mohou být dány nějakou společnou vlastností (barva, tvar apod.) nebo nalezenými klastry.

7 Řízení

7.1 Kriterium

Zabýváme se optimálním řízením, proto musíme zadat nějaké kriterium optimality. Nejčastěji to bývá kvadratické kriterium

$$J = \sum_{t=1}^{N} \left(y_t^2 + \omega u_t^2 \right)$$

kde N je interval řízení, ω je penalizace řízení.

Jiné varianty kriteria

- $J=\sum_{t=1}^{N}\left(y_t^2+\omega\left(u_t-u_{t-1}\right)^2\right)$ je kriterium s penalizací přírůstků řízení
- $J=T\left(y_t,u_t\right)$, kde

$$T=\begin{array}{c|cc} & y_t=1 & y_t=2 \\ \hline u_t=1 & J_{1|1} & J_{2|1} \\ u_t=2 & J_{1|2} & J_{2|2} \end{array}$$

  je kriterium pro řízení diskrétního systému.

7.2 Interval řízení

Interval řízení je časový úsek, pro který provádíme řízení.

Jednokrokové řízení (řízení na jeden krok dopředu) je často nestabilní, proto se většinou volí delší interval.

7.3 Dynamické programování

Pro delší interval řízení je úloha výpočtu řídící veličiny složitá. Počítá se od konce intervalu řízení dopředu (proti směru času). Jedná se o postupnou minimalizaci kritéria na celém intervalu řízení. Teprve když se dostaneme na začátek intervalu, můžeme pro konkrétní data postupně počítat hodnoty řídící veličiny.

7.4 Bellmanova rovnice

Bellmanova rovnice slouží k minimalizaci kriteria řízení na intervalu řízení. Rovnice se počítá rekurzivně od konce intervalu a má tvar

$$\varphi_t=E\left[J_t+\varphi_{t+1}^{*}|u\left(t\right),y\left(t-1\right)\right]$$

$$\varphi_t^{*}=\min_{u_t}\varphi_t$$

φ je tzv. Bellmanova funkce a φ^* dostaneme, když dosadíme optimální řízení u^* - tj. to řízení, které minimalizuje φ. Tedy

$$u_t^{*}=\arg\min_{u_t}\varphi_t.$$

7.5 Statické řízení

Statické řízení je řízení se statickým modelem. Syntéza řízení je v tomto případě velmi jednoduchá, ale kvalita řízení není moc dobrá. Optimální řízení spočteme přímo z modelu - pro řízení na nulu musí platit

$$0=b_0u_t+k\quad\rightarrow\quad u_t=-\frac{k}{b_0}$$

7.6 Dynamické řízení

Dynamické řízení je řízení s dynamickým modelem. Syntéza řízení se provádí pomocí dynamického programování na určitém intervalu řízení.

7.7 Metoda ustupujícího horizontu

Dynamické řízení se počítá pro model se známými parametry (současné řízení a odhadování není prakticky realizovatelné). Pokud jsou parametry modelu neznámé, musíme použít nějakou sub-optimální metodu. Nejčastěji se používá tzv. metoda ustupujícího horizontu. Při této metodě

1. Zvolíme interval řízení
2. Provedeme syntézu řízení na intervalu s odhady parametrů, které jsou k dispozici.
3. Z navrženého řízení realizujeme jen první krok (vypočteme jedno optimální řízení, realizujeme jej a změříme výstup soustavy).
4. Nová data přidáme k ostatním a provedeme nový odhad parametrů.
5. Interval řízení posuneme o jeden krok dopředu a celou proceduru opakujeme.

8 Směsi

8.1 Model směsi

Model směsi je tvořen několika komponentami a modelem ukazovátka.

Komponenty jsou obyčejné modely (spojité nebo diskrétní), které modelují chování systému v určitém pracovním módu.

Ukazovátko má diskrétní kategorický model a jeho (diskrétní) hodnoty ukazují v každém čase na aktivní komponentu (tj. tu komponentu, která nejlépe modeluji aktuální pracovní režim). Model má tvar

$$y_t = \sum_{c_t=1}^{n_c} f\left(y_t|c_t, \psi_t, \Theta_{c_t}\right) f\left(c_t|\alpha\right)$$

kde n_c je počet komponent, c_t je ukazovátko, Θ a α jsou parametry.

8.2 Komponenta

Komponenta modelu směsi je obyčejný model, označený číslem (hodnotou ukazovátka). Komponenty mají většinou stejnou strukturu a liší se jen parametry.

8.3 Model ukazovátka

Model ukazovátka je diskrétní kategorický model s pravděpodobnostní funkcí

c	1	2	$\cdots$	n_c
$f\left(c\|\alpha\right)$	α_1	α_2	$\cdots$	α_{n_c}

Rejstřík