یونا ... تکلیف سری پنجم

استنباط آماری 9810100294

(1)

a) $\beta(\alpha, \beta) = \frac{\Gamma(\alpha, \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)}$ (فراین)

طبق فرمول داریم: $p(\theta | x) = \frac{p(x|\theta) p(\theta)}{\int_\theta p(x|\theta) p(\theta) d\theta} \propto p(x|\theta) p(\theta)$ (تخمین مقدار دال)

$X_i \sim binom(N, \theta) \longrightarrow P_r\{X = x_i | N, \theta\} = \binom{N}{x_i} \theta^{x_i} (1-\theta)^{N-x_i}$

$p(\theta) = \frac{1}{\beta(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$

$\Longrightarrow p(\theta | x_i) = \frac{1}{\beta(\beta, \alpha)} \binom{N}{x_i} \theta^{\overbrace{x_i+\alpha}^{\alpha'} -1} (1-\theta)^{\overbrace{N+\beta-x_i}^{\beta'} -1} \Longrightarrow x \triangleq \{x_1, \dots x_+\} \to$

$p(\theta | x) = \binom{N}{x_i} \times \frac{1}{\beta(\alpha, \beta)} \times p^{\overbrace{\sum x_i - \alpha}^{\alpha''} -1} \cdot (1-p)^{\overbrace{nN+\beta-\sum x_i}^{\beta''} -1} \longrightarrow$ posterior dist

$p(\theta | x'') \sim Beta(\sum x_i - \alpha, nN + \beta - \sum x_i)$

b)

تابع توزیع تابع پیشین (prior) در فرم نوع توزیع پسین مشخص است اما با افزایش تعداد دفعات آزمایشی تعداد این تاثیر کاهش می‌یابد:

c)

تابع یکنواخت (uniform) نوع خاصی از توزیع (بتا) می‌باشد ($\alpha = \beta = 1$) در کد

$p(\theta | x) \sim beta(\sum x_i + 1, nN - \sum x_i + 1) \longrightarrow p(\theta | x) \propto \theta^{\sum x_i} \cdot (1-\theta)^{\sum N - x_i}$

در اینجا تابع پیشین یک binominal خواهد شد

d $\to$ code

(2)

a) good dice $\longrightarrow P_r(n \geq 4) = \frac{4}{12} = 0.75$,

bad dice $\longrightarrow P_r(n \geq 4) = 0.5$, $\longrightarrow$

$P_r(H_1 | n \geq 4) = \frac{P_r(H_1) \cdot P_r(n \geq 4 | H_1)}{P_r(n \geq 4)} = \frac{0.5 \times 0.75}{0.5 \times 0.75 + 0.5 \times .75} = \boxed{0.6}$

b) $\to$ code

$P_r(H_2 | n \geq 4) = \boxed{0.4}$

c) bayesian به باور ذهنی وابسته است به این صورت که ما با این باورهای ذهنی شروع کرده و فرضیات را اصلاح می‌کنیم و هر بار احتمال‌های پیشین را که برای ما اطلاعات ذهنی بودند برای ما update می‌کنیم و در واقع اگر باور ذهنی ما ناشناخته باشد

(1)

2

ج ادامه

(در روش ما هیچ فرضی نداریم همچنین در روش frequentist ما بر اساس مشاهدات به صورت کلی و یک جانبه از کل آنها نتیجه می‌گیریم

حال آنکه در bayesian چنان چه گفته شد با اطلاعات گذشته) باورهای خود را اصلاح می‌کنیم.

3

تابع پیشین

prior

$$f_X(x) \to \frac{1}{\sqrt{2\pi}\sigma_x} e^{\frac{-x^2}{2\sigma_x^2}} \longrightarrow \begin{matrix} X = x \\ Y = x + w \end{matrix} \longrightarrow Y|X=x \sim N(x, \sigma_w^2) \Rightarrow f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_w} e^{\frac{-(y-x)^2}{2\sigma_w^2}}$$

تابع پسین

posterior

$$\to f_{Y|X}(y|x)\, f_X(x) = \frac{1}{2\pi\sigma_x\sigma_w} e^{-\left[\frac{(y-x)^2}{2\sigma_w^2} + \frac{x^2}{2\sigma_x^2}\right]} \xrightarrow{max}$$

چون در توان e^{-x}

x min شود تابع ما ماکزیمم (بیشینه) خواهد بود

$\xrightarrow{\text{داریم}}$

$$\hat{x}_{map} = \arg\min_x \frac{(y-x)^2}{2\sigma_w^2} + \frac{x^2}{2\sigma_x^2}$$

$$\xrightarrow{\frac{\partial(\cdot)}{\partial x} = 0} -\frac{(y-x)}{\sigma_w^2} + \frac{x}{\sigma_x^2} = 0 \longrightarrow \boxed{\hat{x}_{map} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_w^2} y}$$

b) $\hat{x}_{map} \to 0$ $\qquad \hat{x}_{map} \propto \frac{1}{\sigma_w^2}$ $\qquad \sigma_w^2 \to \infty$? (3

درحقیقت اگر σ_w^2 زیاد شود معادل این است که هر آنچه می‌آییم x مشارکتی در آن ندارد چون σ_w قدری بزرگ است که بتواند به Y خروجی را صفر کند.

(4)

$$\begin{cases} H_0 : x = -1 \\ H_1 : x = 1 \end{cases}$$

توابع پیشین (prior) $\to \begin{cases} P_r(H_0) = 1-p \longrightarrow Y = -2+w \to Y|H_0 \sim N(-2,\sigma^2) \\ P_r(H_1) = p \longrightarrow Y = 2+w \to Y|H_1 \sim N(2,\sigma^2) \end{cases}$

توابع پسین (posterior)

$$\begin{cases} P_r(H_0|Y=y) = f_Y(y|H_0)\,P_r(H_0) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y+2)^2}{2\sigma^2}} \cdot (1-p) \\ P_r(H_1|Y=y) = f_Y(y|H_1)\,P_r(H_1) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y-2)^2}{2\sigma^2}} \cdot p \end{cases}$$

فرض صفر (H_0) زمانی رد می‌شود که احتمال پسین (posterior) فرض جایگزین (H_1) از احتمال پسین فرض صفر (H_0) بزرگتر باشد:

داریم $\to \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y-2)^2}{2\sigma^2}} \cdot p > \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y+2)^2}{2\sigma^2}} \cdot (1-p) \to \boxed{y > \frac{\sigma^2}{2} \ln\frac{(1-p)}{p}}$

در این صورت می‌توان H_0 را رد کرد و در غیر این صورت H_0 رد نمی‌شود.

(5) الگوریتم Metropolis Hastings یک روش برای به دست آوردن دنباله‌ای از نمونه‌های تصادفی از توزیع احتمال معینی است که نمونه‌برداری به طور مستقیم از آن‌ها دشوار می‌باشد این روش یکی از روش‌های خانواده مونت کارلو است و در نمونه‌برداری از توزیع‌های پسین (posterior) کاربرد دارد.

شروط الگوریتم:

(1) وجود توزیع ثابت $\pi(x)$ به شکل detailed balance یعنی: هر $x' \to x$ قابل برگشت باشد (احتمال بودن در state، x و گذار به x' برابر با بودن در x'

و گذار به x باشد یعنی: $\pi(x)\,p(x'|x) = \pi(x')\,p(x|x')$

(2) یکتایی توزیع $\pi(x)$

داریم $\to p(x)\cdot p(x'|x) = p(x')\cdot p(x|x') \to \frac{p(x'|x)}{p(x|x')} = \frac{p(x')}{p(x)}$ طبق شرط detailed balance

هدف تقسیم گذار به دو مرحله proposal و acceptance-reject است: داریم

$$\underbrace{P(x',x)}_{\text{احتمال گذار}} = \underbrace{g(x'|x)}_{\text{proposal}}\,\underbrace{A(x',x)}_{\text{قبول کردن Acceptance}}$$

ادامه

a)

$\xrightarrow{(I)} \dfrac{A(x',x)}{A(x,x')} = \dfrac{p(x')\,g(x|x')}{p(x)\cdot g(x'|x)}$

acceptance/rejection

↓

$$A(x',x) = \min\left(1, \frac{P(x')}{P(x)} \cdot \frac{g(x|x')}{g(x'|x)}\right)$$

(I) مقدار دهی اولیه

state $\theta = \theta_0 (= 0)$

$t = 0$

(II) Iteration (for —)

تولید state x'

↳ Acceptance/rejection → $A(x', x_t) = \min\left(1, \dfrac{P_r(x')}{P_r(x_t)} \cdot \dfrac{g(x_t|x')}{g(x'|x_t)}\right)$

(III) Accept / reject

- $u \in [0,1]$ ← random
- $u \leq A(x', x_t)$ → accept x' as new state $\quad x_{t+1} = x'$
- $u > A(x', x_t)$ → reject x' as new state $\quad x_{t+1} = x_t$

(IV) repeat the iteration $t = t+1$ (untill the states are out.)

---

(6)

a) در اینجا نکته‌ای که مهم است این است که ~~فراوانی (تعداد)~~ بیماری در جامعه آماری نکته‌ای بسیار مهم است و دانستن آن می‌تواند در تحلیل تست و در داشتن استنتاج مفید باشد. حالا آنکه ما در تفسیر specificity و sensitivity را بدون توجه به تعداد فراوانی بررسی می‌کنیم که باعث خطای شناختی می‌شود. به این پارادوکس، veridicial paradox می‌گویند یعنی زمانی که یک راه حل به نظر احمقانه می‌آید اما با این وجود درست است (غلط نیست).

---

b) odds: number of {sick : healthy} → $P(S)$, $p(h)$

positive مثبت شدن آزمایش → $p(p)$

روش odds تا نش

طبق video: update: $\{\text{sick} : \text{healthy}\} \times \dfrac{P_r\{P|s\}}{P_r\{P|h\}}$

داریم:

prior odds: 1:99

bayes' factor: $\dfrac{Pr\{P|s\}}{Pr\{P|h\}} = \dfrac{90\%}{9\%} = \boxed{10}$

posterior odds: $10:99 \rightarrow \dfrac{10}{109} \simeq \boxed{0.092}$

c)

$$\text{bayes' factor} = \frac{\text{sensitivity}}{FPR} = \frac{Pr\{p|s\}}{Pr\{p|h\}} \xrightarrow[\text{کم}]{\text{برای prior های}} \text{posterior} \simeq \text{prior}\,(\text{bayes' factor})$$

به جای prior ها اگر از odd استفاده کنیم چندان دقیق نیست.

داریم

راه معمولی $\longrightarrow P(s|p) = \dfrac{P(p|s)\cdot P(s)}{P(p|s) + P(p|h)} = \dfrac{\text{Prior}\cdot(\text{sensitivity})}{\text{Prior}\,(\text{sensitivity}) + (1-\text{Prior})\,(FPR)}$

odds $\longrightarrow O(s|p) = O(s)\cdot \dfrac{Pr\{p|s\}}{Pr\{p|h\}} = \boxed{\text{prior}\cdot\dfrac{\text{sensitivity}}{FPR}}$

در متد odds چونکه ترم‌های (FPR، sensitivity) و همچنین احتمال پیشین (prior) از هم و تولید جدا هستند اولاً باعث کاهش ابهام می‌شود و ثانیاً چون از هم جدا هستند می‌توان تاثیر prior های مختلف را مستقیماً مشاهده کرد حال آن که در روش معمول چون این ترم‌ها با هم مخلوط شده اند کسر پیچیده‌تر است بنابراین با ابهام بیشتری داریم و تاثیر prior ها را نیز به طور مستقل نمی‌توان مشاهده نمود. همچنین روش odds برای انجام iteration های زیاد محاسبات (تست‌های متوالی) و پیاده‌سازی کاربرد آن‌ها در نوشتن برنامه‌های کامپیوتری بیشتر از راه معمولی کمک می‌کند

---

d)

prior odds $O(\text{sick}) = 8:92$

sensitivity $= 0.8$

odds : $1:9 \longrightarrow FPR = 0.1$
(p|h)

$\Longrightarrow$ $\text{bayes' factor} = \dfrac{\text{sensitivity}}{FPR} = \boxed{8}$

$O(s|p) = \underset{\text{prior}}{O(s)} \times \text{bayes' factor} = 64:92$

$O = \dfrac{64}{156} \simeq \boxed{0.41}$ احتمال

CS CamScanner