« بسمه تعالی »

نظریه اطلاعات، آمار و یادگیری — امتحان پایان‌ترم_ نیمسال اول ۱۴۰۰_۱۴۰۱

• زمان : • سوالات را به دقت خوانده و به تمامی اجزای آن پاسخ دهید.

۱. آزمون فرض زیر را در نظر بگیرید.

$$\begin{aligned} H_0 &: X_1,\ldots,X_n \overset{\text{i.i.d.}}{\sim} P=\mathcal{N}(0,1) \\ H_1 &: X_1,\ldots,X_n \overset{\text{i.i.d.}}{\sim} Q=\mathcal{N}(\mu,1) \end{aligned} \tag{1}$$

(آ) برای آزمون فرض فوق نرخ بهینه کاهش خطا یعنی

$$-\lim_{n\to\infty}\frac{1}{n}\log\beta_{1-\epsilon}$$

را محاسبه کنید.

(ب) حال حالت بیزی با توزیع یکنواخت روی فرض ها در نظر بگیرید. فرض کنید I اندیس فرض انتخاب شده و $\hat{I}$ برداشت ما از I با توجه به نمونه‌هاست. نشان دهید در حالت کلی نرخ بهینه کاهش خطا از رابطه زیر بدست می‌آید (اثبات مطالبی که در کلاس بیان شده لازم نیست):

$$-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}[\hat{I}\neq I]=\max_{0\le s\le 1}-\log\mathbb{E}_Q\left[\left(\frac{P(X)}{Q(X)}\right)^s\right]$$

سمت راست این عبارت به انحراف چرنف معروف است که البته یک f_انحراف نیست. مقدار آن را برای آزمون فرض قسمت قبل را محاسبه نمایید.

۲. متغیر تصادفی یکنواخت U_θ با پارامتر θ، متغیری یکنواخت روی بازه $[\theta,\theta+1]$ است. هدف تخمین‌زدن پارامتر θ تحت تابع زیان مجذور فاصله از روی n مشاهده iid از توزیع است.

(آ) ابتدا اطلاعات فیشر را برای این مدل محاسبه کنید.

(ب) با استفاده از روش دونقطه‌ای Le-Cam نشان دهید

$$\inf_{\hat{\theta}}\sup_{\theta\in\mathbb{R}}\mathbb{E}[(\theta-\hat{\theta})^2]\ge\frac{C}{n^2}$$

ضمنا یک تخمینگر ساده برای این مسئله ارائه کنید که به کران فوق میرسد. به عبارت دیگر کران بالا tight است.

(ج) کران فوق آیا با کرانهای معمول برای مسائل مشابه همخوانی دارد؟

۳. f_انحراف و نامساوی فانو.

هدف از این مسئله ارائه تعمیم سرراستی از نامساوی فانو با استفاده از f_انحرافهاست. فرض کنید M متغیری با توزیع یکنواخت روی $[1:\mathsf{M}]$ باشد (مثلا در مسئله آزمون فرض بیزی با توزیع پیشین یکنواخت). X مشاهده ما از فرض ناپیدای M است و توزیع مشترک P_{MX} داده شده است. ضمنا یک قاعده تصمیم‌گیر تصادفی $T_{\hat{M}|X}$ داده شده است. فرض کنید که احتمال خطا برابر با p_e باشد.

(آ) Q_X را توزیع احتمال دلخواهی در نظر بگیرید. نشان دهید

$$D_f(P_{MX}T_{\hat{M}|X}||P_MQ_XT_{\hat{M}|X})=D_f(P_{MX}||P_MQ_X)$$

(ب) نشان دهید

$$D_f(P_{MX}||P_MQ_X)\ge D_f(P_{M\hat{M}}||P_MQ_{\hat{M}})$$

که در آن $P_{M\hat{M}}$ توزیع حاشیه‌ای القا شده از توزیع P_{MX} و قاعده تصمیم‌گیری است.

(ج) نشان دهید:

$$D_f(P_{M\hat{M}}||P_MQ_{\hat{M}})\ge D_f(\mathsf{Ber}(p_e)||\mathsf{Ber}(1-\frac{1}{M}))$$

(د) در نهایت نتیجه بگیرید:

$$\inf_{Q_X}\max_m D_f(P_{X|M=m}||Q_X)\ge\inf_{Q_X}D_f(P_{MX}||P_MQ_X)\ge D_f(\mathsf{Ber}(p_e)||\mathsf{Ber}(1-\frac{1}{M}))$$

۴. نقطه در برابر مخلوط!! در روش le-cam دو نقطه از فضا به دیدار هم می‌رفتند و مسئله به یک آزمون فرض باینری تبدیل می‌شد. در این مسئله تعمیم ساده‌ای از این روش را بیان می‌کنیم که در آن یک نقطه به دیدار مجموعه‌ای از نقاط دیگر می‌رود.

(آ) به صورت دقیقتر، فرض کنید برای تابع زیان $\ell : \Theta \times \widehat{\Theta} \mapsto \mathbb{R}$ نقطه θ_0 و مجموعه $\Theta_1 \subseteq \Theta$ موجود باشند به طوری‌که شرط زیر برقرار باشد:

$$\ell(\theta_0, a) + \ell(\theta_1, a) \geq \Delta, \quad \forall a \in \widehat{\Theta}, \theta_1 \in \Theta_1$$

نشان دهید برای هر توزیع احتمال دلخواه μ روی Θ_1، کران زیر برقرار است:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta[\ell(\theta, \hat{\theta})] \geq \frac{\Delta}{2}(1 - TV(P_{\theta_0}, \mathbb{E}_{\mu_{\Theta_1}} P_{\theta_1}))$$

دقت کنید که $\mathbb{E}_{\mu_{\Theta_1}} P_{\theta_1}$ یک توزیع احتمال مخلوط است.

(ب) در حالت کلی محاسبه کران بالا به خصوص محاسبه TV بین توزیع مخلوط ضربی و یک توزیع دیگر چالش برانگیز است. به همین منظور، از کرانی مبتنی بر فاصله χ^2 استفاده می‌شود. ابتدا نشان دهید:

$$TV(P, Q)^2 \leq c\chi^2(P, Q)$$

مقدار c چقدر است؟

(ج) نشان دهید:

$$1 + \chi^2(\mathbb{E}_\mu[P_\theta], Q) = \mathbb{E}\left[\int \frac{P_\theta(x) P_{\theta'}(x)}{Q(x)} dx\right]$$

که در آن θ, θ' مستقل از یکدیگر و از روی توزیع μ تولید شده‌اند. سپس رابطه زیر را نتیجه بگیرید:

$$1 + \chi^2(\mathbb{E}_\mu[P_\theta^{\otimes n}], Q^{\otimes n}) = \mathbb{E}\left[\left(\int \frac{P_\theta(x) P_{\theta'}(x)}{Q(x)} dx\right)^n\right]$$

(د) تست یکنواختی.

به عنوان مثالی از روش بالا، مسئله تست کردن یکنواختی را بررسی می‌کنیم. در این مسئله n نمونه از یک توزیع P روی یک مجموعه k عضوی داده شده است. می‌دانیم که توزیع فوق یا یکنواخت است یا حداقل به اندازه ϵ از توزیع یکنواخت تحت فاصله TV فاصله دارد. هدف از این مسئله بررسی آزمون فرض زیر است:

$$H_0 : P \sim U[1:k], \quad \text{vs} \quad H_1 : TV(P, U[1:k]) \geq \epsilon$$

می‌خواهیم حداقل تعداد نمونه‌های لازم n را برای رسیدن به خطای حداکثر یک دهم بدست آوریم (یک دهم یک عدد دلخواه است و نکته خاصی ندارد!). منظور از خطای یک دهم اینست که روابط زیر برقرار باشند:

$$\mathbb{P}_{U[1:k]}[\hat{H} = 1] \leq \frac{1}{10},$$
$$\mathbb{P}_Q[\hat{H} = 0] \leq \frac{1}{10}, \quad \forall Q : TV(U[1:k], Q) \geq \epsilon$$

نشان دهید که حداقل n از مرتبه $\Omega\left(\frac{\sqrt{k}}{\epsilon^2}\right)$ است.

راهنمایی ۱: قیود مسئله آزمون فرض را به فرم minmax با تابع زیان مناسب تبدیل کنید و سپس از قسمتهای قبل استفاده نمایید.

راهنمایی ۲: هر توزیع Q شامل $\frac{k}{2}$ نقطه با احتمال $\frac{1+2\epsilon}{k}$ و $\frac{k}{2}$ نقطه با احتمال $\frac{1-2\epsilon}{k}$ در فاصله ϵ از توزیع یکنواخت قرار دارد.

تذکر: حل این قسمت مستقل از قسمتهای قبل و با استفاده از روشهای دیگر بلامانع است! و می‌تواند نمره اضافه در بر داشته باشد.

۵. در روش فانو سرتاسری یک روش برای بدست‌آوردن کران بالا روی اطلاعات متقابل با استفاده از مفهوم عدد پوششی KL بدست آوردیم. در این مسئله، مشابه آن را برای کران زدن فاصله TV بیان می‌کنیم.

خانواده $\mathcal{P}$ از توزیع‌های احتمال را در نظر بگیرید و فرض کنید که $(X_1, \cdots, X_n) \overset{iid}{\sim} P \in \mathcal{P}$.

نشان خواهیم داد که تخمینگر $\widehat{P}(X_1, \cdots, X_n)$ وجود دارد به طوری که

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P[\text{TV}(\hat{P}, P)] \leq C \inf_{\epsilon > 0} \left(\epsilon + \sqrt{\frac{\log N(\mathcal{P}, \text{TV}, \epsilon)}{n}}\right) \quad (2)$$

(آ) فرض کنید که P_n توزیع تجربی حاصل از نمونه‌های $(X_1, \cdots, X_n)$ باشد. حال تخمینگر زیر را در نظر بگیرید:

$$\tilde{P} = \arg\min_{P \in \mathcal{P}} TV(P, P_n)$$

استدلال کنید که تخمینگر فوق برای حالتی که خانواده $\mathcal{P}$ از توزیعهای پیوسته تشکیل شده باشد، به سرانجامی نمی‌رسد!

(ب) به یاد بیاورید که

$$TV(P,Q) = \sup_{A} P(A) - Q(A)$$

مشکل اصلی با TV اینست که دو توزیع را روی مجموعه همه پیشامدها مقایسه می‌کند (ارتباط این جمله با قسمت قبل چیست؟). به منظور رفع این مشکل، از یک زیرمجموعه از پیشامدهای $\mathcal{A}$ استفاده می‌کنیم و شبه‌ـ فاصله زیر را تعریف می‌کنیم:

$$\mathrm{dist}(P,Q) = \sup_{A\in\mathcal{A}} |P(A) - Q(A)| \qquad (۳)$$

نشان دهید که نامساوی مثلث برقرار است.

(ج) قرار دهید

$$\hat{P} = \arg\min_{P\in\mathcal{P}} \mathrm{dist}\,(P, P_n)$$

نشان دهید: $\mathrm{dist}(\hat{P}, P) \le ۲\,\mathrm{dist}\,(P, P_n)$.

(د) (امتیازی.)
نشان دهید

$$\mathbb{E}\left[\mathrm{dist}\,(P,P_n)\right] \le \sqrt{\frac{C\log|\mathcal{A}|}{n}}$$

(ه) حال خانواده $\mathcal{A}$ را تعریف می‌نماییم. ϵـ پوشش $Q_۱,\cdots,Q_N$ از $\mathcal{P}$ را در فاصله TV در نظر بگیرید. حال برای هر $۱ \le i < j \le N$ تعریف کنید: $A_{ij} := \{x : Q_i(x) \ge Q_j(x)\}$. تقریب زیر از TV را بر حسب شبه‌ـ فاصله dist را اثبات کنید.

$$\mathrm{TV}(P,Q) \le \mathrm{dist}(P,Q) + ۴\epsilon, \quad \forall P,Q\in\mathcal{P}$$

(و) در نهایت با در نظرگرفتن مراحل طی شده رابطه (۲) را بدست آورید.