« بسمه تعالی »

نظریه اطلاعات، آمار و یادگیری — امتحان پایان‌ترم_ نیمسال اول ۱۴۰۱_۱۴۰۲

۱. متغیر تصادفی لاپلاس (یا نمایی متقارن) با پارامتر $\tau > 0$ و میانگین μ متغیری پیوسته با تابع چگالی احتمال $f_X(x) = \frac{\tau}{2}\mathrm{e}^{-\tau|x-\mu|}$ است. حال آزمون فرضی را در نظر بگیرید که در فرض H_0 نمونه از توزیع لاپلاس P با پارامترهای (τ, μ) داریم و در فرض H_1 نمونه از توزیع لاپلاس Q با پارامتر های (τ, ν) داریم. نشان دهید کمترین احتمال خطا در آزمون فرض بیزی برابر است با

$$\mathrm{e}^{-\frac{\tau}{2}|\mu-\nu|}$$

ضمنا تست بهینه را هم تعیین نمایید.

۲. انحراف من‌درآوردی! با توجه به فرم وردشی انحراف KL، تعمیم زیر از این انحراف ارایه شده است:

$$V_{\alpha,\beta,r,s,t}(P_X||Q_X) = \sup_{f:\mathcal{X}\mapsto\mathbb{R}} \mathbb{E}_{P_X}[f(X)] - r\mathbb{E}_{Q_X}[f(X)] - s\log\mathbb{E}_{Q_X}[\exp(\alpha f(X))] - t\log\mathbb{E}_{Q_X}[\exp(\beta f(X))]$$

در اینجا (s,t)اعداد نامنفی و r, α, β اعداد حقیقی هستند.

- نشان دهید که چنانچه $r + \alpha s + \beta t \neq 1$ برقرار باشد، مقدار انحراف همواره بینهایت است و در نتیجه تعریف فوق به درد نخور است! در قسمتهای بعد فرض می‌کنیم که تساوی $r + \alpha s + \beta t = 1$ برقرار است.
- نشان دهید انحراف فوق همیشه نامنفی است.
- نشان دهید که انحراف فوق وقتی $P_X = Q_X$ است برابر با صفر است. آیا عکس آن درست است؟
- برای توزیعهای مشترک P_{XY} و Q_{XY} نشان دهید

$$V_{\alpha,\beta,r,s,t}(P_X||Q_X) \leq V_{\alpha,\beta,r,s,t}(P_{XY}||Q_{XY})$$

- نشان دهید پردازش یکسان انحراف را افزایش نمی‌دهد، یعنی

$$V_{\alpha,\beta,r,s,t}(P_X W_{Y|X}||Q_X W_{Y|X}) = V_{\alpha,\beta,r,s,t}(P_X||Q_X)$$

از اینجا نامساوی پردازش داده‌ها را برای انحراف V بیان کرده و ثابت نمایید. همچنین نشان دهید که انحراف فوق نسبت به زوج (P_X, Q_X) محدب است.

- خاصیت بالاجمعی (super-additive) زیر را ثابت نمایید:

$$V_{\alpha,\beta,r,s,t}(P_{XY}||Q_X Q_Y) \geq V_{\alpha,\beta,r,s,t}(P_X||Q_X) + V_{\alpha,\beta,r,s,t}(P_Y||Q_Y)$$

- قرار دهید $W_\alpha(P_X||Q_X) = V_{\alpha,0,1-\frac{1}{\alpha},\frac{1}{\alpha^2},0}(P_X, Q_X)$. حد زیر را بیابید:

$$\lim_{\alpha\to 0} W_\alpha(P_X||Q_X)$$

- مستقیما یا با استفاده از قسمتهای قبل نشان دهید:

$$\chi^2(P_{XY}||Q_X Q_Y) \geq \chi^2(P_X||Q_X) + \chi^2(P_Y||Q_Y)$$

- ⋆⋆ مقدار $W_\alpha(P_X||Q_X)$ را برای هر α_ی دلخواه بین صفر و یک محاسبه کنید.

۳. نامساوی زیر را بین انحراف KL و فاصله TV اثبات کنید،

$$D(P||Q) \geq \log\frac{1 + \mathsf{TV}(P,Q)}{1 - \mathsf{TV}(P,Q)} - \frac{2\mathsf{TV}(P,Q)}{1 + \mathsf{TV}(P,Q)} \qquad (1)$$

در حالتهای حدی(مجانبی) این نامساوی را با نامساوی پینسکر مقایسه نمایید. ضمنا نامساوی فوق را با کشیدن نمودار (نمودار کامپیوتری منظور است نه دستی!) با نامساوی پینسکر مقایسه نمایید.

۴. تخمین توزیع نقاط داده $X_1, \cdots, X_n$ به طور iid از توزیع ناشناخته گسسته P روی $\{1, 2, \cdots, k\}$ در اختیار ما قرار داده شده است. هدف تخمین توزیع فوق از روی این مشاهدات است. هدف از این مسئله نشان دادن رابطه زیر است:

$$R^*_{\mathsf{TV}}(k,n) := \inf_{\hat{P}} \sup_{P\in\mathcal{P}_k} \mathbb{E}_P[\mathsf{TV}(\hat{P}, P)] \asymp \min\left\{\sqrt{\frac{k}{n}}, 1\right\} \qquad (2)$$

(آ) ابتدا نشان دهید که تخمینگر ML برای این مسئله چیزی جز تخمینگر تجربی نیست.

(ب) نشان دهید که خطای تخمینگر تجربی از مضربی از سمت راست (۲)، کمتر است.

(ج) با استفاده از یکی از روشهای فانو (محلی یا سرتاسری) یا Assouad و یا هر روشی که به ذهنتان می‌رسد!، کران پایینی از مرتبه سمت راست (۲) روی مینیمکس خطا ارائه دهید.

(د) برای تابع زیان انحراف KL به شکل $D(P||\hat{P})$، نشان دهید که تخمینگر تجربی در بدترین وضعیت به زیان بینهایت می‌رسد!

۵. تشخیص علامت در رگرسیون خطی تنک در رگرسیون خطی تنک، nـ اندازه‌گیری نویزی $Y_i = \langle X_i, \theta^\star \rangle + \xi_i$ از بردار تنک $\theta^\star \in \mathbb{R}^d$ داریم که دارای حداکثر $k \ll d$ درایه ناصفر است. همچنین در اینجا بردارهای X_i بردارهای ثابتی در $\mathbb{R}^d$ هستند که مربوط به وسیله اندازه‌گیری هستند. ضمنا فرض می‌کنیم که $\xi_i \sim \mathcal{N}(0, \sigma^2)$. هدف در اینجا بازیابی علامت درایه‌های بردار تنک $\theta^\star$ است که در آن علامت صفر را صفر در نظر می‌گیریم. بدین منظور فرض می‌کنیم که اندازه سطح درایه‌های ناصفر $\theta^\star$ از مقدار $\theta_{\min}$ بیشتر است (در غیر این صورت می‌توان راحت دید که مسئله بازیابی علامت در بدترین وضعیت ناممکن است!). هدف ما بدست‌آوردن کران پایینی روی تعداد اندازه‌گیری ها است. فرآیند زیر را در نظر بگیرید:

متناظر با $s \in \{-1, 0, 1\}^d$، بردار $\theta^{(s)}$ قرار می‌دهیم به طوریکه درایه‌های آن از رابطه $\theta_j^{(s)} = \theta_{\min} s_j$ تعیین می‌شوند. حال یک بردار تصادفی $S \in \{-1, 0, 1\}^d$ به شکل تصادفی و یکنواخت از $\mathcal{S}_k := \{s \in \{-1, 0, 1\}^d : \|s\|_1 = k\}$ (یعنی بردارهایی با دقیقا kـ عنصر ناصفر) انتخاب می‌کنیم. به شرط $S = s$، مشاهده $(Y_1, \cdots, Y_n)$ را از بردار متناظر $\theta^{(s)}$ خواهیم داشت.

(آ) با استفاده از نامساوی فانو، نشان دهید برای هر تخمینگر $\hat{S}$ از علامت بردار $\theta^{(S)}$، یعنی S، احتمال خطای $\mathbb{P}[\widehat{S} \neq S]$ حداقل $\frac{1}{2}$ است، مگر اینکه

$$n \geq c \frac{\frac{d}{k} \log \binom{d}{k}}{\|n^{-\frac{1}{2}} \mathbf{X}\|_{\mathsf{Fr}}^2} \frac{\sigma^2}{\theta_{\min}^2}$$

در اینجا ماتریس $\mathbf{X}$ ماتریسی $n \times d$ است که سطر iـ ام آن برابر با بردار اندازه‌گیر X_i است.

(ب) حال فرض کنید که درایه‌های ماتریس $\mathbf{X}$ برابر با ± 1 باشند. کران پایینی برای n به منظور بازیابی علامت ارائه دهید. ضمنا اهمیت جمله $\frac{\sigma^2}{\theta_{\min}^2}$ در کران قسمت قبل را شرح دهید.