به نام خدا

نظریه‌ی اطّلاعات، آمار و یادگیری (۱-۲۵۱۱۰)

تمرین سری دوم

ترم بهار ۰۴-۱۴۰۳

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

استاد: دکتر محمّدحسین یاسائی میبدی

مهلت تحویل: جمعه ۱۲ اردی‌بهشت ۱۴۰۴ ساعت ۲۳:۵۹

(*) مسائلی که با ستاره مشخّص شده‌اند امتیازی هستند و حل‌کردن آن‌ها نمره‌ی امتیازی خواهد داشت!

۱ انحراف بزرگ برای Log-Likelihood

فرض کنید P و Q دو توزیع احتمال باشند که $P \ll Q$. همین‌طور X_iها متغیّرهای تصادفی i.i.d از توزیع P و Y_iها متغیّرهای تصادفی i.i.d. از توزیع Q هستند. تعریف می‌کنیم: $Z_i = \log\left(\frac{p(X_i)}{q(X_i)}\right)$ و $W_i = \log\left(\frac{p(Y_i)}{q(Y_i)}\right)$.
در این سوال می‌خواهیم رابطه‌ی زیر را به ازای هر $t \geq 0, n \in \mathbb{N}$ ثابت کنیم:

$$\mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \leq \exp\left(-n\left(\alpha + \frac{t}{2}\right)\right)$$

که در آن $\alpha = -2\log\mathcal{B}(P,Q)$ و $\mathcal{B}(P,Q) = \mathbb{E}_{Y\sim Q}\left[\sqrt{\frac{p(Y)}{q(Y)}}\right]$.

۱. ثابت کنید:

$$\mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \leq \exp\left(-n\cdot F(t)\right),$$

که در آن: $F(t) = \sup_{\lambda\geq 0}\{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\}$، $\psi_P(\lambda) = \log\mathbb{E}\left[e^{\lambda Z_i}\right]$ و $\psi_Q(\lambda) = \log\mathbb{E}\left[e^{\lambda W_i}\right]$.

۲. ثابت کنید: $F(0) = -\psi_P(-\frac{1}{2}) - \psi_Q(\frac{1}{2}) = \alpha$

۳. ثابت کنید: $F(t) \geq F(0) + \frac{t}{2}$ سپس حکم را نتیجه بگیرید.

۲ زوج نرخ‌های قابل دسترس

در درس، دیدیم یک روش بررسی رفتار حدی خطاهای مسئله‌ی آزمون فرض آن است که خطای $\pi_{1|0}$ را کوچک نگه داریم و نرخ‌های همگرایی قابل دسترس برای خطای $\pi_{0|1}$ را به دست آوریم. حال در این مسئله می‌خواهیم برای هر دو عبارت خطا نرخ همگرایی به دست آوریم. منحنی مرزی ناحیه‌ی زوج نرخ‌های همگرایی قابل دسترس یعنی زوج نرخ‌هایی مانند E_0 و E_1 که برای آن‌ها روش تصمیم‌گیری‌ای وجود دارد که در آن داریم:

$$\pi_{1|0} \leq e^{-n\cdot E_0}, \qquad \pi_{0|1} \leq e^{-n\cdot E_1}$$

۱. استدلال کنید که چرا ناحیه‌ی (e^{-nE_0}, e^{-nE_1})، باید یک ناحیه‌ی محدّب باشد؟

۲. با استفاده از قضیه‌ی Neyman-Pearson و قرار دادن $\tau = n \cdot t$، که τ پارامتر روش تصمیم گیری LLR است، نشان دهید به شرط $-D_{\mathrm{KL}}(Q\|P) \leq t \leq D_{\mathrm{KL}}(P\|Q)$ داریم:

$$\pi_{۱|۰}^{(n)} \leq e^{-n\cdot\psi_P^*(t)}, \qquad \pi_{۰|۱}^{(n)} \leq e^{-n\psi_Q^*(t)}.$$

که: $\psi_P^*(t) = \sup_{\lambda\in\mathbb{R}} \{-\lambda t - \psi_P(\lambda)\}$ و $\psi_P(\lambda) = \log \mathbb{E}_{X\sim P}\left[\exp\left(\lambda \log \frac{q(X)}{p(X)}\right)\right]$.
همچنین: $\psi_Q^*(t) = \sup_{\lambda\in\mathbb{R}} \{\lambda t - \psi_Q(\lambda)\}$ و $\psi_Q(\lambda) = \log \mathbb{E}_{X\sim Q}\left[\exp\left(\lambda \log \frac{p(X)}{q(X)}\right)\right]$.

۳. با استفاده از نامساوی‌های فوق نشان دهید که به ازای هر t که در شرط $-D_{\mathrm{KL}}(Q\|P) \leq t \leq D_{\mathrm{KL}}(P\|Q)$ صدق کند، زوج نرخ زیر قابل‌حصول هستند:

$$E_۰(t) = \psi_P^*(t), \qquad E_۱(t) = \psi_P^*(t) + t.$$

۴. حال نشان دهید که منحنی پارامتری $\begin{cases} E_۰(t) = \psi_P^*(t) \\ E_۱(t) = \psi_P^*(t) + t \end{cases}$، همان منحنیِ مرزیِ زوج‌های قابل‌حصول است.

۵. هدف آنست که نرخ بهینه‌ی همگرایی عبارت زیر را محاسبه کنیم:

$$\min_{P(Z|X^n)} \left\{\pi_۰\pi_{۱|۰} + \pi_۱\pi_{۰|۱}\right\} \tag{۱}$$

با استفاده از مرزی که برای ناحیه‌ی زوج نرخ‌های همگرایی قابل‌حصول به دست آوردیم، مسئله‌ی محاسبه‌ی نرخ بهینه‌ی همگرایی عبارت (۱) به ازای مقادیر ثابت احتمال‌های اولّیه‌ی $\pi_۰, \pi_۱$ را به صورت یک مسئله‌ی max-min درآورید و نشان دهید نرخ بهینه برابر است با $\psi_P^*(۰)$.

۳ اطّلاعات چرنف

فرض کنید نمونه‌های $X_۱, X_۲, \ldots, X_n$ به صورت i.i.d. از توزیع Q به ما داده شده باشد. حالت بیزی را در نظر بگیرید که می‌دانیم با احتمال اولّیه‌ی $\pi_۱$، داریم $Q = P_۱$ و همچنین با احتمال اولّیه‌ی $\pi_۲$ داریم $Q = P_۲$. همچنین فرض کنید $\mathcal{A}^{(n)} \subseteq \mathcal{X}^n$ ناحیه‌ی پذیرش فرضیه‌ی $H_۱$ باشد. احتمالات خطا را نیز به صورت زیر تعریف می‌کنیم:

$$\alpha_n = P_۱^{(n)}(\mathcal{X}^n \backslash \mathcal{A}^{(n)}), \qquad \beta_n = P_۲^{(n)}(\mathcal{A}^{(n)}).$$

در این صورت، احتمال خطای کل برابر خواهد شد با:

$$\mathbb{P}_{\mathsf{E}}^{(n)} = \pi_۱\alpha_n + \pi_۲\beta_n.$$

D^* را به صورت زیر تعریف می‌کنیم:

$$D^* = \lim_{n\to\infty} \frac{-۱}{n} \log \min_{\mathcal{A}^{(n)}\subseteq\mathcal{X}^n} \left\{\mathbb{P}_{\mathsf{E}}^{(n)}\right\}.$$

۱. نشان دهید D^* (بهترین نمای قابل دستیابی در احتمال خطای بیزی) برابر است با:

$$D^* = D_{\mathrm{KL}}(P_{\lambda^*}\|P_۱) = D_{\mathrm{KL}}(P_{\lambda^*}||P_۲),$$

که در آن داریم:

$$p_\lambda(x) = \frac{p_۱^\lambda(x) p_۲^{۱-\lambda}(x)}{\sum_{y\in\mathcal{X}} p_۱^\lambda(y) p_۲^{۱-\lambda}(y)},$$

و λ^* نیز مقداری از λ است که برای آن داریم

$$D_{\mathrm{KL}}(P_{\lambda^*}\|P_۱) = D_{\mathrm{KL}}(P_{\lambda^*}\|P_۲).$$

۲. اطّلاعات چرنف بین دو توزیع P_1 و P_2 به صورت زیر تعریف می‌شود:

$$\mathcal{C}(P_1, P_2) \triangleq - \min_{0 \leq \lambda \leq 1} \log \left(\mathbb{E}_{X \sim P_2} \left[\left(\frac{p_1(X)}{p_2(X)} \right)^{\lambda} \right] \right).$$

نشان دهید:

$$D^* = \mathcal{C}(P_1, P_2).$$

۴ $f-$ انحراف و آزمون فرض

در درس دیدیم که در مسئله‌ی آزمون فرض می‌توان کران ضعیف‌تر از سرعت کاهش نمای $\beta_{1-\epsilon}(P,Q)$ را با استفاده از نامساوی پردازش داده‌ها برای انحراف KL به دست آورد:

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{1-\epsilon}(P,Q)} \leq \frac{1}{1-\epsilon} D_{\mathrm{KL}}(P\|Q).$$

هدف این مسئله به دست آوردن کران چفت زیر با استفاده از $f-$ انحراف ها و نامساوی پردازش داده ها است:

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{1-\epsilon}(P,Q)} \leq D_{KL}(P\|Q).$$

۱. برای $\alpha > 0$ نشان دهید تابع $f_\alpha(x) = \frac{x^{\alpha}-1}{\alpha-1}$ تابعی محدّب از x روی $\mathbb{R}^+$ است.

۲. فرض کنید محمل[1] توزیع های P و Q برابر با $\{1, 2, \ldots, n\}$ باشد. رابطه‌ی $D_{f_\alpha}(P\|Q)$ را برحسب توابع جرم احتمال ساده کنید.

۳. نامساوی پردازش داده را برای انحراف داده‌شده در قسمت قبل نشان دهید.

۴. انحراف $\alpha-$رینی از روی انحراف معرّفی‌شده در بالا به صورت زیر تعریف می‌شود:

$$R_\alpha(P\|Q) = \frac{1}{1-\alpha} \log \left(1 + (\alpha - 1) D_{f_\alpha}(P\|Q)\right).$$

نشان دهید نامساوی پردازش داده‌ها برای انحراف رینی نیز برقرار است.

۵. نشان دهید:

$$\lim_{\alpha \downarrow 1} D_{f_\alpha}(P\|Q) = \lim_{\alpha\downarrow 1} R_\alpha(P\|Q) = D_{\mathrm{KL}}(P\|Q).$$

۶. حال به مسئله‌ی آزمون فرض برمی‌گردیم. فرض کنید n نمونه‌ی $X^n = (X_1, \ldots, X_n)$ از یکی از دو توزیع $\mathsf{H}_0 : X \sim P_X$ و $\mathsf{H}_1 : X \sim Q_X$ در اختیار ما گذاشته شده است. همچنین فرض کنید که آزمون تصادفی $T_{Z|X^n}$ دارای خطای نوع اوّل $P\left[Z = 1\right] < \epsilon$ و خطای نوع دوم $Q\left[Z = 0\right] = e^{-nE_n}$ است. نشان دهید:

$$nR_\alpha(P_X\|Q_X) \geq R_\alpha\left(\mathrm{Ber}(\epsilon)\|\mathrm{Ber}(1-e^{-nE_n})\right).$$

۷. برای سادگی فرض کنید $\epsilon = \frac{1}{2}$. با بررسی نامساوی قسمت قبل برای $\alpha = 1 + \frac{1}{\sqrt{n}}$ نشان دهید که برای nهای به اندازه‌ی کافی بزرگ داریم:

$$E_n \leq R_{1+\frac{1}{\sqrt{n}}}(P_X\|Q_X).$$

۸. از قسمتهای قبل رابطه‌ی زیر را نتیجه بگیرید:

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{1-\epsilon}(P,Q)} \leq D_{\mathrm{KL}}(P\|Q).$$

[1] Support

۵ انحراف من‌درآوردی!

با توجه به فرم وردشی انحراف KL، تعمیم زیر از این انحراف ارائه شده است:

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

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

۱. نشان دهید که چنانچه $r+\alpha s+\beta t \neq ۱$ برقرار باشد، مقدار انحراف همواره بینهایت است و در نتیجه تعریف فوق به درد نخور است!
در قسمت‌های بعدی فرض می‌کنیم که تساوی $r+\alpha s+\beta t = ۱$ برقرار است.

۲. نشان دهید که انحراف فوق همیشه نامنفی است.

۳. نشان دهید که انحراف فوق وقتی $P_X = Q_X$ باشد برابر با صفر است. آیا عکس آن درست است؟

۴. برای توزیع‌های مشترک P_{XY} و Q_{XY} نشان دهید

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

۵. 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 فوق بیان کرده و ثابت نمایید. همچنین نشان دهید که انحراف فوق نسبت به زوج (P_X, Q_X) محدّب است.

۶. خاصیت بالاجمعی زیر را ثابت نمایید:

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

۷. $(*)$ قرار دهید:

$$W_\alpha(P_X\|Q_X) = V_{\alpha,۰,۱-\frac{۱}{\alpha},\frac{۱}{\alpha^۲},۰}(P_X\|Q_X).$$

حدّ زیر را بیابید:

$$\lim_{\alpha\to ۰} W_\alpha(P_X\|Q_X).$$

۸. مستقیماً یا با استفاده از قسمت‌های قبل نشان دهید:

$$D_{\chi^۲}(P_{XY}\|Q_X Q_Y) \geq D_{\chi^۲}(P_X\|Q_X) + D_{\chi^۲}(P_Y\|Q_Y).$$

۹. $(*)$ مقدار $W_\alpha(P_X\|Q_X)$ را برای هر مقدار α بین صفر و یک محاسبه کنید.

۱۰. $(*)$ آیا می توان تابعی محدب مانند f پیدا کرد که $f(۱) = ۰$ و داشته باشیم:

$$V_{\alpha,\beta,r,s,t}(P_X\|Q_X) = D_f(P_X\|Q_X).$$

۶ نامساوی‌های اطّلاعاتی

در درس با خاصیت بالاجمعی انحراف‌ها آشنا شدید.

۱. با استفاده از فرم وردشی رابطه‌ی زیر را برای انحراف $\chi^۲$ اثبات کنید:

$$D_{\chi^۲}(P_{XY}\|Q_X Q_Y) \geq D_{\chi^۲}(P_X\|Q_X) + D_{\chi^۲}(P_Y\|Q_Y).$$

در ادامه‌ی این پرسش قصد داریم کمی با خاصیت زیرجمعی انحراف‌ها نیز آشنا شویم. فرض کنید $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$. یک f–انحراف را زیرجمعی می‌نامیم، اگر:

$$D_f(P_{XYZ} \| P_X P_{YZ}) \leq D_f(P_{XY} \| P_X P_Y) + D_f(P_{XZ} \| P_X P_Z). \tag{۲}$$

۲. برای دست‌گرمی ثابت کنید انحراف KL دارای خاصیت زیرجمعی است.

۳. تابع f_{SKL} را به صورت $f_{\text{SKL}}(x) = (x - ۱)\log x$ در نظر می‌گیریم و انحراف Kullback–Leibler متقارن را از روی آن تعریف می‌کنیم:

$$D_{\text{SKL}}(P \| Q) = D_{f_{\text{SKL}}}(P \| Q).$$

ثابت کنید برای انحراف Kullback–Leibler متقارن، نابرابری (۲) همواره به صورت تساوی برقرار است. به عبارت دیگر نشان دهید:

$$D_{\text{SKL}}(P_{XYZ} \| P_X P_{YZ}) = D_{\text{SKL}}(P_{XY} \| P_X P_Y) + D_{\text{SKL}}(P_{XZ} \| P_X P_Z).$$

۴. $(*)$ فرض کنید $P_{XYZ} = P_X P_{Y|X} P_{Z|X}$ و $Q_{XYZ} = Q_X Q_{Y|X} Q_{Z|X}$. در این حالت خاص، نابرابری زیر را برای انحراف Hellinger ثابت کنید:

$$(P_{XYZ} \| Q_{XYZ}) \leq (P_{XY} \| Q_{XY}) + (P_{XZ} \| Q_{XZ}).$$

۵. با استفاده از قسمت های قبل آیا می توان گفت انحراف Hellinger یک انحراف زیرجمعی است؟

۶. درباره زیر جمعی بودن انحراف‌های TV و $\chi^۲$ بحث کنید.

۷. $(*)$ به جز توابعی که در بالا بررسی کردید (و ترکیبات خطی آن ها) آیا می‌توان تابعی (یا دسته‌ای از توابع) محدّب مانند f پیدا کرد که $f(۱) = ۰$ و دارای خاصیت زیرجمعی باشند؟
می توانید از شبیه‌سازی کامپیوتری برای بررسی ادعای خود استفاده کنید.

۷ $(*)$ **TV در برابر KL**

نامساوی زیر را بین انحراف KL و انحراف TV اثبات کنید:

$$D_{\text{KL}}(P \| Q) \geq \log\left(\frac{۱ + d_{\text{TV}}(P, Q)}{۱ - d_{\text{TV}}(P, Q)}\right) - \frac{۲ d_{\text{TV}}(P, Q)}{۱ + d_{\text{TV}}(P, Q).}$$

در حالات حدی (مجانبی) این نامساوی را با نامساوی Pinsker مقایسه کنید. ضمناً نامساوی فوق را با کشیدن نمودار (نمودار کامپیوتری منظور است نه دستی!) با نامساوی Pinsker مقایسه کنید.