به نام خدا

نظریه‌ی اطّلاعات، آمار و یادگیری (۲۵۱۱۰-۱)

تمرین سری اوّل

ترم بهار ۱۴۰۲-۰۳

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

استاد: دکتر محمّدحسین یاسائی میبدی

مهلت تحویل: جمعه ۱۵ فروردین ۱۴۰۴ ساعت ۲۳:۵۹

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

۱ خواص Total Variation

۱. ثابت کنید:

$$d_{\mathrm{TV}}\left(\prod_{i=1}^{n} p_i, \prod_{i=1}^{n} q_i\right) \leq \sum_{i=1}^{n} d_{\mathrm{TV}}\left(p_i, q_i\right).$$

۲. فرض کنید $g : \mathcal{X} \mapsto \mathcal{Y}$ تابعی یک‌به‌یک است و تعریف می‌کنیم $Y = g(X)$. ثابت کنید:

$$d_{\mathrm{TV}}\left(p_X, q_X\right) = d_{\mathrm{TV}}\left(p_Y, q_Y\right).$$

۳. ثابت کنید:

$$d_{\mathrm{TV}}\left(p_0, p_1\right) = d_{\mathrm{TV}}\left(p_0 \otimes q, p_1 \otimes q\right).$$

۴. ثابت کنید:

$$d_{\mathrm{TV}}\left(\mathcal{N}(0, \mathbf{C}), \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})\right) = 1 - 2\Phi\left(\frac{1}{2}\|\mathbf{C}^{\frac{-1}{2}}\boldsymbol{\mu}\|_2\right),$$

که در آن $\boldsymbol{\mu} \in \mathbb{R}^n$، $\mathbf{C} \in \mathbb{R}^{n\times n}$ و داریم:

$$\Phi(a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$

(راهنمایی: ابتدا مساله را برای یک بعد حل کنید و سپس با استفاده از سفید کردن و بخش‌های قبل، جوابِ یك بعد را به تعداد بعد دلخواه تعمیم دهید.)

۲ D_{KL} برای توزیع‌های گوسی

برای دو توزیع گوسی $\mathcal{N}\left(\boldsymbol{\mu}_1, \mathbf{C}_1\right)$ و $\mathcal{N}\left(\boldsymbol{\mu}_2, \mathbf{C}_2\right)$ که در آن $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^n$ و $\mathbf{C}_1, \mathbf{C}_2 \in \mathbb{R}^{n\times n}$ انحراف KL را محاسبه کنید.

۳ آماره‌ی بسنده

آماره‌ی $\mathbf{T}$ را یک آماره‌ی بسنده[1] از $\mathbf{X} \sim p_{\boldsymbol{\theta}}$ می گوییم اگر توزیع $p_{\mathbf{X}|\mathbf{T}}$ مستقل از $\boldsymbol{\theta}$ باشد. به عبارتی $\mathbf{T}$ همه‌ی اطّلاعات برای تخمین $\boldsymbol{\theta}$ را در بر دارد.

۱. تابع $\mathbf{t} : \mathcal{X} \mapsto \mathbb{R}^n$ را در نظر بگیرید. نشان دهید $\mathbf{T} = \mathbf{t}(\mathbf{X})$ یک آماره‌ی بسنده از $\mathbf{X}$ است، اگر و فقط اگر تابع درست نمایی را بتوان به صورت زیر نوشت:
$$p_{\boldsymbol{\theta}}(\mathbf{x}) = g(\mathbf{t}(\mathbf{x}), \boldsymbol{\theta})h(\mathbf{x}).$$

۲. آزمون فرض $H_۰ : \mathbf{X} \sim p_{\mathbf{X}}, H_۱ : \mathbf{X} \sim q_{\mathbf{X}}$ را در نظر بگیرید. نشان دهید $\mathbf{T}$ آماره‌ای بسنده از $\mathbf{X}$ است، اگر و فقط اگر نامساوی پردازش داده به تساوی ختم شود و داشته باشیم: $D_{\mathrm{KL}}(p_{\mathbf{X}}\|q_{\mathbf{X}}) = D_{\mathrm{KL}}(p_{\mathbf{T}}\|q_{\mathbf{T}})$.

۳. نشان دهید $T = \log\left(\frac{p_{\mathbf{X}}(\mathbf{X})}{q_{\mathbf{X}}(\mathbf{X})}\right)$ یک آماره‌ی بسنده در آزمون فرض فوق است.

۴ $\mathcal{R}(P, Q)$ برای توزیع‌های ساده

۱. آزمون فرض دوتایی $H_۰ : X \sim \mathrm{Bernoulli}(p),\ H_۱ : X \sim \mathrm{Bernoulli}(q)$ را در نظر بگیرید. ناحیه‌ی Neyman-Pearson $\mathcal{R}(\mathrm{Bernoulli}(p), \mathrm{Bernoulli}(q))$ را محاسبه و رسم کنید.

۲. توزیع‌های پیوسته P و Q را بر روی $[۰, ۳]$ در نظر بگیرید که توابع چگالی احتمال آن‌ها به صورت زیر است:

$$p_X(x) = \begin{cases} \frac{x}{۳}, & ۰ \leq x \leq ۱, \\ \frac{۱}{۳}, & ۱ \leq x \leq ۲, \\ \frac{x-۱}{۳}, & ۲ \leq x \leq ۳, \\ ۰, & \text{otherwise} \end{cases} \qquad q_X(x) = \begin{cases} \frac{۱}{۳}, & ۰ \leq x \leq ۳, \\ ۰, & \text{otherwise} \end{cases}$$

تابع $\beta_\alpha(P, Q)$ را محاسبه کنید و ناحیه $\mathcal{R}(P, Q)$ را رسم کنید. همچنین تست فرضی که به $\beta_\alpha(P, Q)$ می‌رسد برای $\alpha = \frac{۵}{۶}$ و $\alpha = \frac{۱}{۲}$ مشخص کنید.

۳. آزمون فرض دوتایی $H_۰ : X \sim \mathcal{N}(۰, ۱),\ H_۱ : X \sim \mathcal{N}(\mu, ۱)$ را در نظر بگیرید. ناحیه‌ی Neyman-Pearson $\mathcal{R}(\mathcal{N}(۰, ۱), \mathcal{N}(\mu, ۱))$ را محاسبه و رسم کنید.

۴. این بار آزمون فرض با n نمونه به صورت $H_۰ : X_۱, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(۰, ۱),\ H_۱ : X_۱, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu, ۱)$ در نظر بگیرید و ناحیه‌ی Neyman-Pearson $\mathcal{R}(\mathcal{N}^{\otimes n}(۰, ۱), \mathcal{N}^{\otimes n}(\mu, ۱))$ را محاسبه کنید.با افزایش تعداد نمونه ها ناحیه چه تغییری می کند؟
(راهنمایی: ابتدا با استفاده از پرسش ۳ نشان دهید میانگین n نمونه آماره‌ی بسنده از $X_۱, ..., X_n$ است.)

۵ $\mathcal{R}(P, Q)$ و d_{TV}

۱. نشان دهید:
$$d_{\mathrm{TV}}(P, Q) = \sup_{۰ \leq \alpha \leq ۱} \{\alpha - \beta_\alpha(P, Q)\}.$$
بر این اساس، فاصله‌ی $d_{\mathrm{TV}}(P, Q)$ را چگونه می توان از روی $\mathcal{R}(P, Q)$ محاسبه کرد؟

۲. در درس دیدیم در آزمون فرض بیزی یکنواخت احتمال خطای بهینه به صورت زیر است:
$$\mathbb{P}_E = \frac{۱}{۲}(۱ - d_{\mathrm{TV}}(P, Q)).$$

[1] sufficient statistic

در حالت کلی برای توزیع اولیه‌ی $\pi = (\pi_0, \pi_1)$ احتمال خطا به صورت زیر تعریف می‌شود:

$$P_E = \inf_{p_{Z|X}} \pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}.$$

آزمون بهینه را برای حالت کلی به دست آورید. همچنین نشان دهید برای کمینه کردن احتمال خطای بیزی کافی است آزمون‌های یقینی را در نظر بگیریم.

۳. $(*)$ نشان دهید:

$$D_{\mathrm{KL}}(P\|Q) = -\int_0^1 \log\left(\frac{d}{d\alpha}\beta_\alpha(P,Q)\right) d\alpha.$$

۶ $(*)$ $\beta_{1-\epsilon}(P^{\otimes n}, Q^{\otimes n})$ با دقت بیشتر

نشان دهید در حالتی که $V(P\|Q) < \infty$ داریم:

$$\frac{1}{n}\log \beta_{1-\epsilon}(P^{\otimes n}, Q^{\otimes n}) = -D_{\mathrm{KL}}(P\|Q) + \sqrt{\frac{V(P\|Q)}{n}} Q^{-1}(\epsilon) + o\left(n^{-\frac{1}{2}}\right)$$

که در آن $V(P\|Q) := \mathrm{Var}_P\left[\log \frac{dP}{dQ}\right]$ و $Q^{-1}(.)$ معکوس تابع مکمل cdf گوسی استاندارد یعنی $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$ است. دقت کنید که در حد $n \to \infty$ این به همان نتیجه‌ای که در کلاس به آن رسیدید منجر می‌شود.

۷ نگاه بیزی و خواص $\mathcal{R}(P,Q)$

در درس، روش کاهش خطای β به ازای حداقل احتمال موفقیت α را به عنوان یك معیار برای به دست آوردن یك روش تصمیم‌گیری در مسئله‌ی آزمون فرض مشاهده کردیم. یک روش دیگر برای به دست آوردن یک روش تصمیم گیری در مسئله آزمون فرض به صورت زیر است:

$$\min_{P_{Z|X}} \pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}, \tag{1}$$

که در آن π_0 و π_1 به ترتیب احتمال اولیه‌ی فرض $H_0 : X \sim P$ و $H_1 : X \sim Q$ هستند. به این معیار معیار بیزی می‌گویند.

فرض کنید برای دو توزیع P, Q، مسئله‌ی آزمون فرض منحنی مرزی پایینی ناحیه‌ی $\mathcal{R}(P,Q)$ برابر $\beta_\alpha(P,Q) = \alpha^2$ باشد. اگر جواب مسئله بهینه سازی خطای کل در رابطه‌ی (1) یك روش تصمیم‌گیری به صورت Log-Likelihood Ratio (LLR) که در قضیه‌ی Neyman-Pearson مطرح شد، باشد،

۱. مقدار τ را برحسب π_0 و π_1 به دست آورید.

۲. برای روش تصمیم گیری LLR با معیار بیزی، مقدار α و β را بر حسب π_0 و π_1 به دست آورید.

۳. برای اینکه جواب بهینه‌ی معیار بیزی به فرمت LLR بتواند بیان شود، چه شرایطی روی احتمال‌های اوّلیه‌ی π_0, π_1 باید برقرار باشد؟

۸ $(*)$ به هم چسبیدگی!

اگر $(\mathcal{S}_n, \mathcal{F}_n)$ فضاهای اندازه باشند و $\mathbb{P} = \{P_n\}_{n=1}^\infty$ و $\mathbb{Q} = \{Q_n\}_{n=1}^\infty$ دنباله‌ای از اندازه‌های احتمال روی این فضاها باشند، می‌گوییم $\mathbb{P}$ به $\mathbb{Q}$ چسبیده است و با $\mathbb{P} \lhd \mathbb{Q}$ نشان می‌دهیم، اگر برای هر دنباله از مجموعه‌های $A_n \in \mathcal{F}_n$ که $Q_n(A_n) \xrightarrow{n\to\infty} 0$ داشته باشیم: $P_n(A_n) \xrightarrow{n\to\infty} 0$.

۱. ثابت کنید اگر $\mathbb{P} \lhd \mathbb{Q}$ و یا $\mathbb{Q} \lhd \mathbb{P}$ در یک تست برای تشخیص بین فرضیه‌ی اول با توزیع P_n و فرضیه‌ی دوم با توزیع Q_n (احتمال پیشین دو فرضیه را یکسان در نظر بگیرید) مجموع احتمال خطای نوع اول و دوم نمی‌تواند به صفر همگرا شود (وقتی $n \to \infty$).

۲. برای فضای توابع $f: \mathcal{S}_n \mapsto \mathbb{R}$ ضرب داخلی زیر را تعریف می‌کنیم:

$$\langle f, g\rangle = \mathbb{E}_{X\sim Q_n}[f(X)g(X)]$$

ثابت کنید اگر تحت این ضرب داخلی داشته باشیم: $\|L_n\|^2 < \infty$ (که $L_n = \frac{dP_n}{dQ_n}$ همان نسبت likelihood است.) در این صورت داریم: $\mathbb{P} \triangleleft \mathbb{Q}$.

۹ ظرفیت کانال نقطه به نقطه

یک کانال مخابراتی بی‌حافظه که با توزیع شرطی $p_{Y|X}$ توصیف می‌شود را در نظر بگیرید. در کلاس دیده‌ایم که ظرفیت این کانال برابر است با $\max_{p_X} I(X;Y)$. می‌خواهیم یک بار روشی که در کلاس برای اثبات قابلیت حصول این نرخ به کار برده شد را مجدّداً بررسی کنیم. به تعبیر دیگر، می‌خواهیم کدی را معرّفی کنیم که با به کار بستنِ آن کد، با n بار استفاده از کانال، 2^{nR} پیام مختلف را بتوانیم ارسال کنیم و برای آن‌که با افزایش n، احتمال خطا به سمت صفر برود، شرط $R < \max_{p_X} I(X;Y)$ کافی باشد.

یک کد (n, R) از اجزای زیر تشکیل می‌شود:

- مجموعه‌ی پیام‌ها $\{1, 2, \ldots, 2^{nR}\} = [1 : 2^{nR}]$.
- یک تابع کدگذار $X^n : \{1, 2, \ldots, 2^{nR}\} \mapsto \mathcal{X}^n$ که کلمه‌کدهای متناظر با پیام‌ها را مشخّص می‌کند. این تابع مشخّص می‌کند که برای ارسال پیامِ $m \in [1 : 2^{nR}]$ در هربار استفاده از کانال، باید چه کلمه‌کدی ارسال شود.
- یک تابع کدگشا $\mathcal{D} : \mathcal{Y}^n \mapsto \{1, 2, \ldots, 2^{nR}\}$ که با توجّه به خروجی‌های دریافتی از n بار استفاده از کانال، پیام ارسالی را حدس می‌زند.

کدی که ما در نظر می‌گیریم به این صورت است: توزیع دلخواه p_X را در نظر می‌گیریم. کلمه‌کد متناظر با هر پیام $m \in [1 : 2^{nR}]$ عبارت است از n نمونه‌ی مستقل و هم‌توزیع از توزیع p_X. همچنین کدگشای انتخابی ما چنین عمل می‌کند که با مشاهده‌ی y^n، مقادیر $p^{\otimes n}_{Y|X}\big(y^n|X^n(1)\big), p^{\otimes n}_{Y|X}\big(y^n|X^n(2)\big), \ldots, p^{\otimes n}_{Y|X}\big(y^n|X^n(2^{nR})\big)$ را در نظر می‌گیرد و پیام خروجی را به صورت تصادفی با مکانیسم زیر انتخاب می‌کند:

$$\mathbb{P}[\hat{M} = \hat{m}] = \frac{p^{\otimes n}_{Y|X}\big(y^n|X^n(\hat{m})\big)}{\sum\limits_{m=1}^{2^{nR}} p^{\otimes n}_{Y|X}\big(y^n|X^n(m)\big)}.$$

پیام ارسالی از کانال را با متغیّر تصادفی M نشان می‌دهیم و فرض می‌کنیم توزیع یکنواختی روی مجموعه‌ی $[1 : 2^{nR}]$ دارد. احتمال صحّت مخابره برای کد توصیف‌شده را می‌توان به صورت زیر نوشت:

$$\mathbb{E}\left[\mathbb{P}[\hat{M} = M]\right] = \mathbb{E}\left[\sum_{m,y^n} p_M(m) p^{\otimes n}_{Y|X}(y^n|X^n(m)) p_{\hat{M}|Y^n}(m|y^n)\right]$$

که امید ریاضی روی کتاب‌های کد تصادفی و در حقیقت روی $X^n(1), X^n(2), \ldots, X^n(2^{nR})$ گرفته می‌شود. همچنین وابسته به نوع کانال، مجموع روی y^n ممکن است با انتگرال nگانه جایگزین شود. در این مسئله این حسّاسیت را کنار می‌گذاریم.

۱. نشان دهید:

$$\mathbb{E}\left[\mathbb{P}[\hat{M} = M]\right] = \mathbb{E}_{X^n(1),\ldots,X^n(2^{nR})}\left[\sum_{y^n} p^{\otimes n}_{Y|X}(y^n|X^n(1)) p_{\hat{M}|Y^n}(1|y^n)\right].$$

۲. نشان دهید:

$$\mathbb{E}\left[\mathbb{P}[\hat{M} = M]\right] \geq \mathbb{E}_{X^n,Y^n}\left[\frac{1}{1 + 2^{nR} \cdot 2^{-\log_2\left(\frac{p^{\otimes n}_{Y|X}(Y^n|X^n)}{p^{\otimes n}_{Y}(Y^n)}\right)}}\right].$$

۳. نشان دهید اگر $R < I(X;Y)$ باشد، با افزایش n، مقدار $\mathbb{E}\left[\mathbb{P}[\hat{M} = M]\right]$ به سمت ۱ میل می‌کند.

۱۰ ظرفیت کانال دسترسی چندگانه[۲]

یک کانال دسترسی چندگانه‌ی بی‌حافظه با توزیع شرطی $p_{Z|X,Y}$ توصیف می‌شود. می‌خواهیم با استفاده از کدی مشابه کدِ مسئله‌ی قبلی، درباره‌ی ظرفیت این کانال بحث کنیم. یک کد (n, R_X, R_Y) برای کانال دسترسی چندگانه از اجزای زیر تشکیل می‌شود:

- دو مجموعه‌ی پیام‌ها $[۱ : ۲^{nR_X}] = \{۱, ۲, \ldots, ۲^{nR_X}\}, [۱ : ۲^{nR_Y}] = \{۱, ۲, \ldots, ۲^{nR_Y}\}$.
- دو تابع کدگذار $X^n : \{۱, ۲, \ldots, ۲^{nR_X}\} \mapsto \mathcal{X}^n$ و $Y^n : \{۱, ۲, \ldots, ۲^{nR_Y}\} \mapsto \mathcal{Y}^n$ که کلمه‌کدهای متناظر با پیام‌ها را مشخّص می‌کند. این تابع مشخّص می‌کند که برای ارسال پیام‌های $m_X \in [۱ : ۲^{nR_X}], m_Y \in [۱ : ۲^{nR_Y}]$ در هربار استفاده از کانال، باید چه کلمه‌کدهایی ارسال شوند.
- یک تابع کدگشا $\mathcal{D} : \mathcal{Z}^n \mapsto \{۱, ۲, \ldots, ۲^{nR_X}\} \times \{۱, ۲, \ldots, ۲^{nR_Y}\}$ که با توجّه به خروجی‌های دریافتی از n بار استفاده از کانال، پیام‌های ارسالی را حدس می‌زند.

۱. از کد معرّفی‌شده در پرسش قبل الگو بگیرید و یک کد تصادفی برای کانال دسترسی چندگانه پیشنهاد بدهید.

۲. نشان دهید متوسّط احتمال صحّت کد تصادفی با افزایش n به سمت ۱ میل می‌کند، اگر:

$$\begin{aligned} &R_X \leq I(X; Z|Y) \\ &R_Y \leq I(Y; Z|X) \\ &R_X + R_Y \leq I(X, Y; Z). \end{aligned}$$

توجّه کنید که اطّلاعات متقابل شرطی به صورت زیر تعریف می شود:

$$\begin{aligned} I(X; Z|Y) &= \mathbb{E}_{X,Y,Z \sim p_{X,Y,Z}} \left[\log_۲ \left(\frac{p_{X,Y|Z}(X, Y|Z)}{p_{X|Z}(X|Z) p_{Y|Z}(Y|Z)} \right) \right] \\ &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} p_{X,Y,Z}(x, y, z) \log_۲ \left(\frac{p_{X,Y|Z}(x, y|z)}{p_{X|Z}(x|z) p_{Y|Z}(y|z)} \right). \end{aligned}$$

[2] Multiple Access Channel (MAC)