دانشکده‌ی مهندسی برق

آنالیز احتمالاتی در ابعاد بالا		بهار ۱۴۰۲
	تمرین سری چهارم	
مدرس: دکتر محمد حسین یاسائی		موعد تحویل: ۱۲ خرداد

۱ محاسبه‌ی بعد VC

- نشان دهید بعد VC خانواده همه دایره‌ها در یک صفحه دو بعدی برابر با ۳ است.
- نشان دهید بعد VC خانواده همه مثلث‌ها در یک صفحه دو بعدی برابر با ۷ است.

۲ زیبایی تقارن!

n متغیر تصادفی مستقل $X_1, X_2, \ldots, X_n$ داده شده است، به طوری که همه آنها حول مبدا متقارن هستند. نامساوی زیر را نشان دهید:

$$\mathbb{P}\left[\sum_{k=1}^{n} X_k > t\sqrt{\sum_{k=1}^{n} X_k^2}\right] \leq \exp\left(-ct^2\right)$$

یک مقدار مناسب برای ضریب c بیان کنید.

۳ یادگیری آماری توابع حقیقی‌مقدار

مجموعه‌ی کران‌دار $\mathcal{Y} = [-1, +1]$ و یک فضای دلخواه $\mathcal{X}$ را در نظر بگیرید. فرض کنید $\mathcal{F}$ مجموعه‌ای از توابع $\mathcal{X} \to \mathcal{Y}$ باشد. نمونه‌های مستقل $\{x_t, y_t\}_{t=1}^{n}$ از یک توزیع ناشناخته‌ی P_{XY} روی $\mathcal{X} \times \mathcal{Y}$ داده شده‌اند. از نظریه‌ی یادگیری آماری برای توابع باینری، می‌دانیم که در آن‌جا، سایز مجموعه‌ی

$$\mathcal{F}|_{x_1,\ldots,x_n} = \{(f(x_1), \ldots, f(x_n)) : f \in \mathcal{F}\}$$

نقشی محوری در تحلیل مسائل طبقه‌بندی دارد. اما به وضوح، این تئوری در مسئله‌ی فعلی کاربردی نخواهد داشت زیرا این مجموعه، عموماً ناشمارا خواهد بود. همچنین امکان دارد که دو تابع، با وجود اینکه روی نمونه‌ها مقادیری اندکی متفاوت ایجاد می‌کنند، بسیار به یکدیگر شبیه باشند. در این تمرین، به بررسی مسئله‌ی یادگیری آماری توابع حقیقی- مقدار می‌پردازیم. در ابتدا، برای یکسان‌کردن نمادگذاری، چند تعریف را ارائه می‌کنیم:

- یک مجموعه‌ی $V \subseteq \mathbb{R}^n$ را یک α-cover روی $x_1, \ldots, x_n$ نسبت به نرم l_p، $p \in [1, \infty)$ می‌گوییم، اگر

$$\forall f \in \mathcal{F}, \exists v \in V \text{ s.t. } \left(\frac{1}{n}\sum_{t=1}^{n} |f(x_t) - v_t|^p\right)^{1/p} \leq \alpha.$$

- عدد پوششی $\mathcal{N}_p(\mathcal{F}, \alpha, x^n)$ نیز به عنوان سایز کوچکترین α-cover نسبت به نرم l_p تعریف می‌شود.

با توجه به این تعاریف، به سوالات زیر پاسخ دهید:

الف. نشان دهید که برای هر $1 \leq p \leq q \leq \infty$، داریم $\mathcal{N}_p(\mathcal{F}, \alpha, x^n) \leq \mathcal{N}_q(\mathcal{F}, \alpha, x^n)$.

ب. نشان دهید که برای هر $x^n = \{x_1, \dots, x_n\}$، پیچیدگی راداماخر تجربی یک مجموعه توابع داده‌شده، در نامساوی زیر صدق می‌کند:

$$\widehat{\mathcal{R}}(\mathcal{F}) \leq \inf_{\alpha \geq 0} \left\{ \alpha + \sqrt{\frac{2 \log \mathcal{N}_1 (\mathcal{F}, \alpha, x^n)}{n}} \right\}$$

ج. فرض کنید $\mathcal{X}$ برابر با گوی l_1 در $\mathbb{R}^d$ بوده و تعریف کنید

$$\mathcal{F} = \{ f(x) = \langle f, x \rangle : ||f||_\infty \leq 1 \}.$$

نشان دهید که پیچیدگی راداماخر، کران بالایی از مرتبه‌ی $O\left(\sqrt{\frac{d \log(n)}{n}}\right)$ پیدا می‌کند.

د. به کمک ایده‌ی چینینگ و انتگرال دادلی، نشان دهید

$$\widehat{\mathcal{R}}(\mathcal{F}) \leq \inf_{\alpha \geq 0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_\alpha^1 \sqrt{\log \mathcal{N}_2 (\mathcal{F}, \delta, x^n)} d\delta \right\}.$$

ه. $\mathcal{F}$ را یک مجموعه از توابع روی $\mathcal{R}$ در نظر بگیرید. فرض کنید این توابع همگی نانزولی هستند. نشان دهید

$$\mathcal{N}(\mathcal{F}, \alpha, x^n)_\infty \leq n^{\frac{2}{\alpha}}.$$

برای این مجموعه از توابع، کران‌های بخش‌های قبل را مقایسه کنید.

یک سوال بسیار طبیعی، این است که آیا یک پارامتر ترکیبیاتی از مجموعه‌ی توابع وجود دارد که رفتار اعداد پوششی آن مجموعه از توابع را کنترل کند یا نه.

- می‌گوییم مجموعه‌ی توابع $\mathcal{F}$، نمونه‌های $(x_1, \dots, x_n)$ را شقّه می‌کند، اگر $(y_1, \dots, y_n) \in \mathcal{R}^n$ با شرایط زیر وجود داشته باشند:

$$\forall (b_1, \dots, b_n) \in \{0,1\}^n, \exists f \in \mathcal{F} \text{ s.t.} \quad f(x_t) > y_t + \frac{\alpha}{2} \text{ if } b_t = 1 \text{ and } f(x_t) < y_t - \frac{\alpha}{2} \text{ if } b_t = 0. \quad (۱)$$

بُعد «شقّه‌ی چاق» مجموعه‌ی $\mathcal{F}$ در سطح α، به عنوان سایز بزرگترین مجموعه‌ی شقّه‌شونده در سطح α تعریف می‌شود. این کمیت را با $\mathrm{fat}(\mathcal{F}, \alpha)$ نمایش می‌دهیم.

به کمک تعریف فوق، به سوالات زیر پاسخ دهید:

و. به صورت شهودی، بُعد شقه‌ی چاق را توضیح دهید.

ز. مجدداً مجموعه‌ی توابع نانزولی روی اعداد حقیقی را در نظر بگیرید. نشان دهید $\mathrm{fat}(\mathcal{F}, \alpha) \leq 4/\alpha$.

ح. **(امتیازی)** نشان دهید برای هر مجموعه‌ی $\mathcal{F}$ از توابع $\mathcal{X} \to [-1, +1]$، داریم

$$\mathcal{N}_2 (\mathcal{F}, \alpha, x^n) \leq \left(\frac{2}{\alpha} \right)^{K \cdot \mathrm{fat}(\mathcal{F}, c\alpha)}$$

که در آن K و c اعداد ثابت هستند.

ط. به کمک بخش قبل، نشان دهید

$$\widehat{\mathcal{R}}(\mathcal{F}) \leq \inf_{\alpha \geq 0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_\alpha^1 \sqrt{K \, \mathrm{fat}(\mathcal{F}, c\delta) \log \left(\frac{2}{\delta} \right)} d\delta \right\}.$$

۴ بُعد VC اجتماع

فرض کنید $\mathcal{H}_1, \dots, \mathcal{H}_r$، کلاس‌هایی از توابع باینری روی $\mathcal{X}$ باشند. بگیرید $d = \max_i \left(\mathrm{VCdim}(\mathcal{H}_i) \right)$. با فرض $d \geq 3$ نشان دهید

$$\mathrm{VCdim}\left(\bigcup_{i=1}^r \mathcal{H}_i \right) \leq 4d \log(2d) + 2 \log(r). \quad (۲)$$

همچنین، نشان دهید که برای $d = 2$ داریم $\mathrm{VCdim}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2d + 1$.

۵ بُعد ناتاراخان

فرض کنید $\mathcal{H}$ کلاسی از توابع $h : \mathcal{X} \to [k]$ باشد. می‌گوییم یک مجموعه‌ی $C \subset \mathcal{X}$ توسط $\mathcal{H}$ شقّه شده است، اگر دو تابع دلخواه $f_0, f_1 : C \to [k]$ وجود داشته باشند، به نحوی که:

- برای هر $x \in C$ داشته باشیم $f_0(x) \neq f_1(x)$.
- برای هر $B \subset C$، یک تابع $h \in \mathcal{H}$ وجود داشته باشد، به قسمی که

$$\forall x \in B: \quad h(x) = f_0(x), \text{ and } \forall x \in C - B, h(x) = f_1(x). \tag{۳}$$

بُعد ناتاراخان مجموعه‌ی $\mathcal{H}$ را سایز بزرگ‌ترین C شقّه‌شونده توسط $\mathcal{H}$ می‌گیریم و آن را با $\mathrm{Ndim}(\mathcal{H})$ نمایش می‌دهیم.

الف. نشان دهید که اگر $k = 2$ آن‌گاه $\mathrm{VCdim}(\mathcal{H}) = \mathrm{Ndim}(\mathcal{H})$.

ب. نشان دهید

$$|\mathcal{H}| \leq |\mathcal{X}|^{\mathrm{Ndim}(\mathcal{H})} . k^{2\mathrm{Ndim}(\mathcal{H})}.$$

برای این اثبات، سعی کنید استدلال مربوط به قضیه‌ی ساور_شلاح که در کلاس مطرح شد را اندکی تغییر دهید.

۶ برنشتاین ماتریسی با وابستگی به بعد ذاتی

ماتریس‌های تصادفی و مستقل A_k با میانگین صفر را در نظر بگیرید به طوری که $\|A_k\|_{op} \leq 1$. هدف این مسئله پیداکردن کرانی برای نرم مجموع ماتریسهای فوق است. در اینجا به جای استفاده از تابع مولد گشتاور به طور مستقیم (آنچنان که در درس گفت شد)، از تابع صعودی $\varphi(t) := e^t - t - 1$ استفاده خواهیم کرد. قرار می‌دهیم $S_n := \sum_{k=1}^n A_k$.

۱. رابطه زیر را نشان دهید.

$$\mathbb{P}\left[\lambda_{\max}\left(S_n\right) > \delta\right] \leq \frac{\mathrm{Tr}\left(\mathbb{E}\left[\varphi\left(tS_n\right)\right]\right)}{\varphi(t\delta)} \quad \forall t > ۰$$

۲. بررسی کنید که برای هر ماتریس تصادفی A با نرم همواره محدود یک و میانگین صفر رابطه ماتریسی زیر برقرار است:

$$\log \mathbb{E}[\exp(tA)] \preceq \varphi(t) \underbrace{\mathrm{var}(A)}_{=\mathbb{E}[A^2]}$$

۳. با استفاده از دو قسمت قبل و مباحث درس، نشان دهید:

$$\mathbb{P}\left[\lambda_{\max}\left(S_n\right) > \delta\right] \leq \frac{\mathrm{Tr}(V)}{\|V\|_{\mathrm{op}}} \inf_{t>۰} \frac{e^{\varphi(t)\|V\|_{\mathrm{op}}}}{\varphi(t\delta)}$$

که در آن $V := \sum_{k=1}^n \mathrm{var}[A_k]$

۴. n بردار تصادفی مستقل و با توزیع یکسان، میانگین صفر $X_1, X_2, \ldots, X_n$ را در نظر بگیرید. ماتریس کوواریانس این بردارها را با C نمایش می دهیم. فرض کنید X_kها dبعدی هستند و اندازه آن ها همواره از یک کمتر است. با استفاده از قسمت قبل، یک کران بالا برای احتمال زیر بیابید:

$$\mathbb{P}\left[\left\|\sum_{k=1}^n X_k\right\|_2 > \delta\right]$$

همچنین یک کران بالا با استفاده از کران برنشتاین ماتریسی بیان شده در درس بیابید و نتیجه ها را مقایسه کنید. مزیت بنیادین کران حاصل از قسمت قبل بر برنشتاین ماتریسی چیست؟

راهنمایی: ماتریس $A_k := \begin{bmatrix} 0 & X_k^T \\ X_k & \mathbf{0}_{d\times d} \end{bmatrix}$ را تعریف کنید و از قسمت‌های قبل کمک بگیرید.

۷ بهبود کران‌های طبقه‌بندی (امتیازی)

فرض کنید $\mathcal{F}$ یک مجموعه از توابع از $\mathcal{X}$ به $\mathcal{Y}$ بوده و $S = \{x_1, \ldots, x_n\}$ نمونه‌هایی مشخص از $\mathcal{X}$ باشند. مجموعه‌ی $\mathcal{W} \subseteq \mathcal{F}$ را یک مجموعه‌ی ϵ-separated نسبت به نرم l_p و نمونه‌های $x_1, \ldots, x_n$ می‌گوییم، اگر برای هر $w_1, w_2 \in \mathcal{W}$، داشته باشیم:

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\big|w_1(x_i) - w_2(x_i)\big|^p\Big)^{\frac{1}{p}} \geq \epsilon.$$

عدد گنجایشی با فاصله‌ی ϵ، نمونه‌های $\{x_1, \ldots, x_n\}$ و فاصله‌ی l_p را به عنوان سایز بزرگترین زیرمجموعه‌ی ϵ-separated از $\mathcal{F}$ می‌گیریم و با $\mathcal{M}_p(\mathcal{F}, \epsilon, \{x_1, \ldots, x_n\})$ نمایش می‌دهیم. همچنین در کلاس دیدیم که عدد پوششی، یعنی $\mathcal{N}_p(\mathcal{F}, \epsilon, \{x_1, \ldots, x_n\})$، به عنوان سایز کوچکترین $\mathcal{V} \subseteq \mathcal{F}$ در نظر گرفته می‌شود که به ازای هر $f \in \mathcal{F}$، وجود داشته باشد $v_f \in \mathcal{V}$ به قسمی که

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\big|f(x_i) - v_f(x_i)\big|^p\Big)^{\frac{1}{p}} < \epsilon.$$

الف. نشان دهید

$$\mathcal{N}_p(\mathcal{F}, \epsilon, \{x_1, \ldots, x_n\}) \leq \mathcal{M}_p(\mathcal{F}, \epsilon, \{x_1, \ldots, x_n\}).$$

از این به بعد، فرض کنید $\mathcal{Y} = \{0, 1\}$ و $d = \mathrm{VCdim}(\mathcal{F})$.

ب. نشان دهید که

$$\mathcal{N}_p(\mathcal{F}, \epsilon, \{x_1, \ldots, x_n\}) \leq \big(\frac{en}{d}\big)^d.$$

توجه کنید که کران بخش قبل، ربطی به شعاع ϵ نداشت. حال قصد داریم به بهتر کردن کران بخش (ب) بپردازیم. یک زیرمجموعه‌ی ϵ-separated از $\mathcal{F}$ با نرم l_1، نمونه‌های $\{x_1, \ldots, x_n\}$ و فاصله‌ی ϵ در نظر گرفته و آن را $\mathcal{W}_\epsilon$ بنامید. تعریف کنید

$$V = \{f_i - f_j | f_i \neq f_j \in \mathcal{W}_\epsilon\}.$$

ج-۱. متغیر های تصادفی $X_1, \ldots, X_t$ را مستقل و هر کدام را یکنواخت روی $\{x_1, \ldots, x_n\}$ در نظر بگیرید. نشان دهید

$$\forall v \in V: \quad \mathbb{P}\big[\forall i \in [t] : v(X_i) = 0\big] \leq (1-\epsilon)^t.$$

ج-۲. نشان دهید

$$\mathbb{P}\big[\forall\, v \in V\ \exists i \in [t] : |v(X_i)| = 1\big] \geq 1 - D^2(1-\epsilon)^t.$$

که در آن D سایز $\mathcal{W}_\epsilon$ است.

ج-۳. به کمک بخش قبل، اثبات کنید که $\sigma \subset \{1, \ldots, n\}$ با $|\sigma| = \frac{2\log(D)}{\epsilon}$ وجود دارد به شکلی که

$$\Big|\big\{\big(f(x_i)\big)_{i \in \sigma} | f \in \mathcal{W}_\epsilon\big\}\Big| = D.$$

ج-۴. نشان دهید $D \leq \Big(\frac{2e\log(D)}{d\epsilon}\Big)^d$ و به کمک آن $\log(D) \leq d\log\Big(\frac{2e^2}{\epsilon}\ln\Big(\frac{2e}{\epsilon}\Big)\Big)$.

ج-۵. نشان دهید که برای هر p،

$$\mathcal{N}_p(\mathcal{F}, \epsilon, \{x_1, \ldots, x_n\}) \leq \Big((2pe^2)\log\big(\frac{2e^2}{\epsilon}\big)\Big)^d \big(\frac{1}{\epsilon}\big)^{pd}.$$

راهنمایی: اگر $\alpha \geq 1$ و $\alpha\log^{-1}(\alpha) \leq \beta$، آن‌گاه $\alpha \leq \beta\log(e\beta\log(\beta))$.

د. با استفاده از بخش (د) تمرین قبل، کران $O\Big(\sqrt{\frac{d}{n}}\Big)$ را برای پیچیدگی راداماخر در مسئله‌ی طبقه‌بندی به دست آورید.

۸ ون‌هاندل (امتیازی)

سوال ۷.۱ کتاب ون‌هاندل (صفحه ۲۰۳) را حل کنید.