بهینه‌سازی محدب

نیم‌سال دوم ۱۴۰۳-۱۴۰۲

مدرس: دکتر امیر نجفی

دانشکده‌ی مهندسی کامپیوتر

تمرین اول

مسئله‌ی ۱. مرور احتمال (۱۰ نمره)

الف) فرض کنید یک تاس منصفانه به شما داده شده است و باید تصمیم بگیرید که کدام وظیفه سخت‌تر است: (i) حدس زدن مقدار یک بار پرتاب تاس یا (ii) دو بار پرتاب تاس و به دست آوردن همان مقدار در هر دو بار. با توجه به اینکه تاس منصفانه است (هر طرف وزن ۱/۶ دارد)، آیا رویداد (i) شانس بیشتری برای موفقیت دارد یا رویداد (ii) یا هر دو احتمال موفقیت یکسانی دارند؟ توضیح دهید.

ب) اکنون این نتیجه را به تاس‌های n–وجهی با هر توزیع (ممکن است غیر یکنواخت) تعمیم می‌دهیم. ابتدا این حقیقت مفید را ثابت کنید: برای هر $\alpha_1, \alpha_2, \ldots, \alpha_n$ که $\sum_i \alpha_i = 1$، رابطه زیر برقرار است:

$$0 \leqslant \sum_{i=1}^{n} (\alpha_i - 1/n)^2 = \sum_{i=1}^{n} \alpha_i^2 - \frac{1}{n}.$$

ج) بگذارید X_1 مقدار اولین پرتاب و X_2 مقدار دومین پرتاب باشد. نشان دهید که $\Pr(X_1 = X_2) \geqslant 1/n$ (راهنمایی: از قسمت (ب) استفاده کنید). برای چه توزیعی این نابرابری به حداقل می‌رسد؟

مسئله‌ی ۲. آماره U (۱۰ نمره)

فرض کنید $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ تابعی متقارن و نامنفی از ورودی هایش باشد. با داشتن دنباله‌ای مستقل و هم‌توزیع (i.i.d) از متغیر های تصادفی X_k برای $k \geqslant 1$ عبارت زیر را در نظر بگیرید:

$$U := \frac{1}{\binom{n}{2}} \sum_{j<k} g(X_j, X_k)$$

الف) نشان دهید U تخمین‌گری نااریب برای $\mathbb{E}[g(X_1, X_2)]$ می‌باشد.

ب) فرض کنید برد تابع g همواره کمتر یا مساوی b باشد، نشان دهید:

$$\mathbb{P}[|U - \mathbb{E}[U]| \geqslant t] \leqslant 2e^{-\frac{nt^2}{2b^2}}$$

مسئله‌ی ۳. گراف تصادفی (۱۰ نمره)

برای گراف G عدد خوشه‌ای $\mathcal{C}(G)$ را برابر بزرگترین مجموعه از رئوس تعریف می‌کنیم که دوبه‌دو به هم یال داشته باشند. یک گراف تصادفی به این شکل می‌سازیم که بین هر دو راس u, v به طور مستقل از سایر جفت راس ها یک راس به احتمال p قرار می‌دهیم. نشان دهید اگر تعداد یال‌های گراف برابر با n باشد برای این گراف داریم:

$$\mathbb{P}[\frac{1}{n}|\mathcal{C}(G) - \mathbb{E}[\mathcal{C}(G)]| \geqslant \delta] \leqslant 2e^{-2n\delta^2}$$

مسئله‌ی ۴. تابع غیر قابل یادگیری (۱۰ نمره)

فرض کنید $\mathcal{S}$ کلاس همه‌ی زیرمجموعه‌هایی از $[0, 1]$ باشد که تعداد اعضای آن متناهی است. کلاس توابع $\mathcal{F}$ را به صورت زیر تعریف می‌کنیم:

$$\mathcal{F}_{\mathcal{S}} = \{\mathbb{1}_S(.)|S \in \mathcal{S}\}$$

فرض کنید که X_i ها از توزیعی روی $[0, 1]$ سمپل‌گیری شده‌باشند که تکینگی ندارد. (یعنی $\mathbb{P}(\{x\}) = 0$ برای هر $x \in [0, 1]$) نشان دهید که همگرایی یکنواخت برای خانواده توابع $\mathcal{F}$ و توزیع $\mathbb{P}$ برقرار نیست.

مسئله‌ی ۵. دایره‌های هم مرکز (۱۰ نمره)

فرض کنید $\mathcal{X} = \mathbb{R}^2$ و فرض کنید که نقاط کلاس مثبت‌مان توسط دایره‌ای به مرکز مبدا مختصات تعریف شده است. همچنین فرض کنید که فضای فرضیه ما نیز همین دایره های به مرکز مبدا مختصات باشند. نشان دهید این کلاس از مسائل را می‌توان با $\mathrm{PAC}\text{-}(\epsilon, \delta)$ با تعداد نمونه $n \geqslant (1/\epsilon)\log(1/\delta)$ یاد گرفت.

مسئله‌ی ۶. قضیه‌ی گلیونکو_ کانتلی (۱۰ نمره)

فرض کنید $X_1, \ldots, X_n$ نسخه‌هایی مستقل و هم‌توزیع از X باشند که تابع توزیع تجمعی (CDF) آن $F(t) = \mathbb{P}(X \leqslant t)$ است. تابع توزیع تجمعی تجربی X به صورت زیر تعریف می‌شود:

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(X_i \leqslant t)$$

۱. میانگین و واریانس $\hat{F}_n(t)$ را محاسبه کنید و نتیجه بگیرید که $\hat{F}_n(t) \to F(t)$ به صورت تقریباً حتمی وقتی $n \to \infty$. (راهنمایی: از لم بورل_ کانتلی استفاده کنید.)

۲. نشان دهید که برای $n \geqslant 2$:

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}}\left|\hat{F}_n(t) - F(t)\right| \leqslant C\sqrt{\frac{\log(n/\epsilon)}{n}}\right) = 1 - \epsilon$$

مسئله‌ی ۷. یادگیری اتحاد بازه‌ها (۱۰ نمره)

یک الگوریتم یادگیری PAC برای کلاس مفاهیم C_2 که از اجتماع دو بازه بسته به صورت $[a,b] \cup [c,d]$ به طوری که $a,b,c,d \in \mathbb{R}$ تشکیل می‌شود ارائه دهید. نتیجه خود را برای به‌دست‌آوردن یک الگوریتم یادگیری PAC برای کلاس مفهومی C_p گسترش دهید که از اتحاد $p \geqslant 1$ بازه بسته تشکیل شده است، یعنی $[a_1,b_1] \cup \cdots \cup [a_p,b_p]$، با $a_k, b_k \in \mathbb{R}$ برای $k \in [p]$. پیچیدگی زمانی و نمونه‌ای الگوریتم شما به عنوان تابعی از p چیست؟

مسئله‌ی ۸. کران‌های تمرکزی (۱۰ نمره)

الف) یک مجموعه از نمونه‌ها به صورت $X = (x_1, x_2, \ldots, x_m)$ به طوری که $\forall i, |x_i| \leqslant M$ را در نظر بگیرید. تابع زیر را تعریف کنید:

$$f(X) = \frac{1}{m}\sum_{i=1}^{m} x_i.$$

آیا می‌توانید کرانی برای احتمال $\Pr[|f(X) - E[f(X)] \geqslant \epsilon]$ ارائه دهید؟

ب) فرض کنید X و X' دو مجموعه با اندازه m باشند که دقیقاً در یک نقطه متفاوت‌اند. یعنی $|X \cap X'| = m - 1$. می‌گوییم یک تابع h پایدار است اگر برای همه‌ی X, X'ها، داشته باشیم $|h(X) - h(X')| \leqslant g(m)$ که g یک تابع نطولی است. سرعت نزول g به عنوان تابعی از m چقدر باید باشد تا نابرابری McDiarmid یک کران برای پیشامد $\Pr[|h(X) - E[h(X)] \geqslant \epsilon]$ ارائه دهد که با افزایش m به صفر همگرا شود؟

ج) آیا تابع f از قسمت (الف) پایدار است (همچنان فرض می‌کنیم $|x_i| \leqslant M$، $\forall i$)؟ آیا نابرابری McDiarmid یک کران همگرا ارائه می‌دهد؟ اگر بله، کران را بدست آورید. اکنون تابع $f'(X) = \max(X)$ را در نظر بگیرید. آیا f' پایدار است؟ آیا می‌توانید با نامساوی McDiarmid یک کران ارائه دهید؟

مسئله‌ی ۹. مدل دو_اوراکل (۱۰ نمره)

در اینجا ما یک سناریوی جایگزین برای یادگیری PAC را در نظر میگیریم که مدل دو اوراکل نامیده می‌شود. فرض کنید که به شما این توانایی داده شده است که به طور صریح درخواست نمونه‌های مثبت یا منفی بدهید، که به ترتیب از توزیع‌های D^+ و D^- کشیده می‌شوند. یک مفهوم به طور کارآمد قابل یادگیری PAC است اگر الگوریتم $\mathcal{A}$ وجود داشته باشد که با داشتن $m = \mathrm{poly}(1/\epsilon, 1/\delta)$ نمونه بتواند یک فرضیه h بدست آورد، که با اطمینان $(1-\delta)$ داشته باشیم $\Pr_{x\sim D^+}[h(x) = 0] \leqslant \epsilon$ و $\Pr_{x\sim D^-}[h(x) = 1] \leqslant \epsilon$. نشان دهید که اگر یک مسئله در معنای کلاسیک به طور کارآمد قابل یادگیر PAC باشد، همیشه در مدل دو اوراکل نیز به طور کارآمد قابل یادگیری PAC است.

مسئله‌ی ۱۰. یادگیری PAC برای مستطیل‌های n_بعدی (۱۰ نمره)

یک الگوریتم یادگیری-PAC برای C، مجموعه مستطیل‌های n_بعدی هم‌محور در $\mathbb{R}^n$ ارائه دهید، به طوری که $C = \{[a_1,b_1] \times \cdots \times [a_n,b_n] : a_i, b_i \in \mathbb{R}\}$. پیچیدگی نمونه به عنوان تابعی از n چگونه تغییر می‌کند؟ (اثبات شما باید دقیق باشد)