نظریه یادگیری ماشین

نیم‌سال اول ۱۴۰۱ ـ ۱۴۰۰

مدرس: دکتر بیگی

دانشکده‌ی مهندسی کامپیوتر

تمرین سری سوم

موعد تحویل:۹ اردیبهشت

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

مسئله‌ی ۱. (۲۱ نمره)

(الف) در صفحه مختصات دو بعدی نقاط {(0,2),(0,-2),(-2,0),(1,0)} دارای برچسب 1- و نقاط {(3,1),(3,-1),(4,1),(4,-1)} دارای برچسب ۱+ هستند ابتدا نقاط را در صفحه مختصات رسم و سپس معادله صفحه انتخاب شده بعنوان مرز تصمیم گیری را توسط SVM بدست آورید.

(ب) در صفحه مختصات دو بعدی نقاط {(-1,-1),(+1,+1),(-1,+1),(+1,-1)} دارای برچسب 1- و نقاط {(-2,-2),(+2,+2),(-2,+2),(+2,-2)} دارای برچسب ۱+ هستند ابتدا نقاط را در صفحه مختصات رسم و سپس معادله صفحه انتخاب شده بعنوان مرز تصمیم گیری را توسط SVM بدست آورید. و تفاوت این حالت را با حالت قسمت (الف) بیان کنید

(ج) با در نظر گرفتن قسمت‌های الف و ب؛ راجع به تفاوت بین دسته‌بند پرسپترون[۱] و SVM تحقیق کنید و با ذکر دلیل به ۲ مورد از نقاط ضعف مدل پرسپترون اشاره کنید و توضیح دهید که SVM چگونه بر آنها غلبه می‌کند. (راهنمایی: در هر یک از قسمت‌های قبل در صورت استفاده از دسته‌بند پرسپترون یک ایراد وارد است. آنها را بررسی کنید)

مسئله‌ی ۲. (۲۴ نمره)

ابتدا درست یا غلط بودن هر یک از گزاره‌های زیر را مشخص کنید و سپس برای هر کدام دلیل خود را توضیح دهید.

(الف) مارجین[۲] فاصله بین مرز تصمیم‌گیری انتخاب شده توسط SVM و بردارهای پشتیبان[۳] می‌باشد.

(ب) ماشین بردار پشتیبان خطی ابرپارامتری برای تنظیم شدن توسط cross validation ندارد.

(ج) بصورت معمول (و نه همیشه) ماشین بردار پشتیبان دقیق‌تر از Logistic Regression عمل می‌کند.

(د) ماشین بردار پشتیبان سریع است زیرا از موارد زیر استفاده می‌کند:
۱.بهینه سازی درجه دوم[۴]

[۱] perceptron
[۲] margin
[۳] vectors support
[۴] optimization quadratic

۲.استفاده از کرنل
۳.استفاده از فرم دوگانه [۵] که در آن از تعداد نقاط کمتری استفاده می‌شود.

(ه) ثابت می‌شود مسئله SVM (در همه حالات) یک مسئله بهینه‌سازی محدب[۶] است.

(و) مقادیر مارجین بدست آمده از دو هسته متفاوت $k_1(\phi_1,\phi_2)$ و $k_2(\phi_1,\phi_2)$ که از آموزش بر روی یک مجموعه آموزشی بدست آمده اند، به ما اطلاعی از اینکه کدام هسته عملکرد بهتری بر روی مجموعه تست دارد نمی‌دهند.

مسئله‌ی ۳. (۵ نمره)

در مسئله بهینه سازی SVM در حالت soft margin در تابع متغیرهای slack به جای استفاده از $\xi \mapsto \sum_{i=1}^{m} \xi_i$ می‌توان از $\xi \mapsto \sum_{i=1}^{m} \xi_i^p$، with $p > 1$ استفاده کرد. برای این مسئله و این حالت یک فرم دوگانه ارائه دهید و آن را اثبات کنید.

مسئله‌ی ۴. (۲۰ نمره)

کلاس توابع زیر را درنظر بگیرید.

$$\mathcal{F} \triangleq \{\mathbf{x} \to \operatorname{sign}(\langle \boldsymbol{\theta}, \mathbf{x} \rangle) \mid \boldsymbol{\theta} \in \Theta\}$$

فرض کنید که $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ نمونه‌های ما در فضا باشند به طوری که $x_i \in \mathbb{R}^d$ و همچنین $\|\mathbf{x}_i\|_\infty \leqslant b$ برقرار باشد.

(الف) با فرض اینکه $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^d \mid \|\boldsymbol{\theta}\|_2 \leqslant r\}$ ثابت کنید

$$\mathcal{R}_n(\mathcal{F}) \leqslant \frac{rb\sqrt{d}}{\sqrt{n}}$$

(ب) حال با فرض اینکه $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^d \mid \|\boldsymbol{\theta}\|_1 \leqslant r\}$ ثابت کنید که

$$\mathcal{R}_n(\mathcal{F}) \leqslant \frac{rb\sqrt{r\log(rd)}}{\sqrt{n}}$$

(پ) حال فرض کنید که $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ یک کرنل و H فضای RKHS متناظر آن باشد. همچنین تعریف می‌کنیم $\mathcal{H}_\lambda = \{f \in \mathcal{H} : \|f\|_K \leqslant \lambda\}$. با فرض اینکه $C_K^2 = \sup_{\mathbf{x}} K(\mathbf{x}, \mathbf{x})$ نشان دهید برای هر $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$ خواهیم داشت

$$\mathcal{R}_n(\mathcal{H}_\lambda) \leqslant \frac{C_K \lambda}{\sqrt{n}}$$

مسئله‌ی ۵. (۱۸ نمره)

فرض کنید $\Phi : \mathcal{X} \to \mathcal{H}$ یک نگاشت به فضای بعد بالای H با بعد N باشد. کرنل PSD زیر را درنظر بگیرید.

$$\begin{cases} k : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \\ k(x, x') = \mathbb{E}_{i \sim D}\left[[\Phi(x)]_i \, [\Phi(x')]_i\right] \end{cases}$$

---

[۵] dual
[۶] convex

منظور از $[\Phi(x)]_i$ درایه‌ی iام $\Phi(x)$ است و D نیز یک توزیع احتمال روی i است. فرض کنید

$$\forall x \in \mathcal{X}\,,\ i \in [N], \quad |[\Phi(x)]_i| \leqslant R$$

برای محاسبه تقریبی این کرنل از تابع $k'\left(x, x'\right)$ استفاده می‌کنیم. ابتدا یک زیرمجموعه‌ی $I \subseteq [N]$ مطابق با توزیع D انتخاب کرده و سپس عبارت

$$k'\left(x, x'\right)=\frac{1}{n} \sum_{i \in I} D(i)[\Phi(x)]_i\left[\Phi\left(x'\right)\right]_i$$

را محاسبه می‌کنیم که در آن می‌دانیم $|I|=n$ است.

(الف) به ازای دو نقطه‌ی ثابت $x, x' \in \mathcal{X}$ ثابت کنید که

$$\mathbb{P}_{I \sim D^n}\left[\left|k\left(x, x'\right)-k'\left(x, x'\right)\right|>\epsilon\right] \leqslant 2 \exp \left(\frac{-n \epsilon^2}{2 R^4}\right)$$

(ب) فرض کنید که K و K' ماتریس کرنل‌های k و k' باشند. نشان دهید برای هر $\epsilon, \delta>0$ و سپس $n > \frac{2R^4}{\epsilon^2} \log \left(\frac{m(m+1)}{\delta}\right)$ با احتمال حداقل $1-\delta$ داریم

$$\forall i, j \in [m], \quad \left|\mathbf{K}_{ij}'-\mathbf{K}_{ij}\right| \leqslant \epsilon$$

مسئله‌ی ۶. (۱۲ نمره)

(الف) فضای احتمال $(\Omega, \mathcal{F}, \mathbb{P})$ را درنظر بگیرید. نشان دهید که تابع

$$K(A, B)=\mathbb{P}(A \cap B)-\mathbb{P}(A) \mathbb{P}(B)$$

که روی $\mathcal{F} \times \mathcal{F}$ تعریف شده است، یک کرنل PSD است.

(ب) با فرض اینکه S یک مجموعه‌ی متناهی است نشان دهید تابع

$$\left\{\begin{array}{l}K: P(S) \times P(S) \rightarrow \mathbb{R} \\ K(A, B)=2^{|A \cap B|}\end{array}\right.$$

که در آن $P(A)$ مجموعه‌ی توانی A است، یک کرنل PSD است.

موفق باشید :)