به نام خدا

دانشکده‌ی مهندسی کامپیوتر

سررسید: ۱۹ آبان ماه جمعه ۲۳:۵۹	پاییز ۱۴۰۲
یادگیری ماشین	
مدرس: مهدی جعفری سیاوشانی	تمرین ۲: تئوری یادگیری

- سررسید این تمرین جمعه ۱۹ آبان ماه ساعت ۵۹ : ۲۳ است.
- این تمرین نظری است و تاخیر مجاز ندارد
- در صورت کشف تقلب، بار اول برای افراد درگیر تقلب، نمره‌ی همان سوال(های) خاص صفر در نظر گرفته می‌شوند. در صورت تکرار، نمره کل تمرین صفر در نظر گرفته می‌شود و در صورت تکرار، درس برای افراد حذف خواهد شد.
- تمامی پاسخ‌های خود را در یک فایل با فرمت HW2-[SID]-[Fullname].zip (.pdf) روی کوئرا قرار دهید.

پرسش‌ها

۱ تئوری یادگیری

۱.۱ پرسش اول (۴۵ نمره)

به سوالات زیر در رابطه با بعد VC و نقطه شکست پاسخ دهید.

۱. کوچکترین نقطه شکست برای مدل Perceptron را در $\mathbb{R}^3$ بدست آورید. (۵ نمره)

۲. در هر یک از موارد زیر، بعد VC را در فضای فرضیه توصیف شده بدست آورید و ادعای خود را ثابت کنید.

(آ) توابع Parity:
$\chi = \{0,1\}^N, \quad \forall I \subseteq \{1,2,3,...,N\} \quad h_I(x) = (\sum_{i\in I} x_i) \mod 2. \quad \mathcal{H} = \{h_S, \forall S \subseteq \{1,2,...,N\}\}$
(۱۰ نمره)

(ب) مستطیل هایی در فضای $\mathbb{R}^d$ که اضلاع آن ها موازی محورهای مختصات است. (۱۰ نمره)

(ج) فرض کنید F یک فضای تابعی خطی از مقادیر حقیقی با بعد d باشد. به ازای هر g که یک تابع حقیقی است: $\mathcal{H} = \{sign(f+g) : f \in F\}$ (۱۰ نمره)

(د) مجموعه دایره های در فضای $\mathbb{R}^2$. (۱۰ نمره)

۲.۱ پرسش دوم (۲۵ نمره)

در نظر بگیرید بهترین فرض برای یک مسئله بر اساس true error (خطایی که با فرض دانستن توزیع حاکم بر داده‌ها تعریف می‌شود که برخی منابع به آن Generalization Error نیز می‌گویند) به صورت زیر به دست می‌آید:

$$h* = argmin_{h\in H}\epsilon(h)$$

همچنین generalization error فرض *h را η در نظر میگیریم. حال یک فرض $h_0 \in H$ داریم و میخواهیم تعیین کنیم که آیا فرض h_0 دارای generalization error کوچکتر مساوی $\eta > 0$ است یا خیر. به طور مشخص، وقتی $\epsilon(h) \leq \epsilon(h*) + \eta$ باشد می‌گوییم فرض h، $\eta - optimal$ است.
حال میخواهیم به این سوال پاسخ دهیم: با داشتن فرض h_0 آیا آن فرض $\eta - optimal$ است؟
ابتدا $\delta > 0$ را یک ثابت در نظر بگیرید و فرض کنید مجموعه‌ی متناهی H از فرض ها، با اندازه |H|=k داریم. برای هر $h \in H$، $\hat{\epsilon}(h)$ را به عنوان training error مربوط به h در نظر بگیرید که از تعدادی training set با m نمونه IID بدست آمده و $\hat{h} = argmin_{h\in H}\hat{\epsilon}(h)$ را به عنوان فرضی در نظر گیرید که training error را به حداقل می‌رساند.
حال الگوریتم زیر را در نظر بگیرید:

1.set $\gamma := \sqrt{\frac{1}{2m}log\frac{2k}{\delta}}$
2.If $\hat{\epsilon}(h_0) > \hat{\epsilon}(\hat{h}) + \eta + 2\gamma$ then return NO
3.If $\hat{\epsilon}(h_0) < \hat{\epsilon}(\hat{h}) + \eta - 2\gamma$ then return YES
4.Otherwise, return UNSURE

به طور شهودی الگوریتم با مقایسه کردن training error ها عمل میکند و فقط زمانی NO یا YES برمی‌گرداند که $\hat{\epsilon}(h_0)$ به طرز چشمگیری بزرگتر یا کوچکتر از $\hat{\epsilon}(\hat{h}) + \eta$ باشد.

۱. ابتدا نشان دهید که اگر $\epsilon(h_0) \leq \epsilon(h*) + \eta$ (که h_0 ، $\eta - optimal$ است) آنگاه احتمال آن که الگوریتم NO برگرداند حداکثر δ است. (۷ نمره)

۲. نشان دهید که اگر $\epsilon(h_0) > \epsilon(h*) + \eta$ (که h_0 ، $\eta - optimal$ نیست) آنگاه احتمال آن که الگوریتم YES برگرداند حداکثر δ است. (۷ نمره)

۳. حال فرض کنید $h_0 = h*$ باشد و $\eta > 0$ و $\delta > 0$ را ثابت در نظر بگیرید. نشان دهید اگر m به اندازه کافی بزرگ باشد، آنگاه احتمال آن که الگوریتم YES برگرداند حداقل $1 - \delta$ است. (۱۱ نمره)
راهنمایی: برای η و δ ثابت، وقتی m به بی نهایت میل میکند داریم:

$$\gamma = \sqrt{\frac{1}{2m} log \frac{2k}{\delta}} \to 0$$

این به این معناست که مقادیری از m وجود دارد که به ازای آن ها $2\gamma < \eta - 2\gamma$ می‌باشد.

۳.۱ پرسش سوم (۷ + ۱۸ نمره)

فرض کنید متغیرهای $Z_۱, Z_۲, ..., Z_m$ مستقل هستند و توزیع برنولی یکسانی دارند که $P(Z_i = ۱) = \phi$ و $P(Z_i = ۰) = ۱ - \phi$ است. با فرض $\hat{\phi} = \frac{۱}{m}\sum_{i=۱}^{m} Z_i$ و $\gamma > ۰$ مطابق نابرابری هوفدینگ داریم:

$$P(|\phi - \hat{\phi}| > \gamma) \leq ۲\exp(-۲\gamma^۲ m)$$

این نابرابری بر این فرض استوار است که متغیرهای تصادفی Z_i استقلال توأم دارند . در این مسئله قصد داریم این فرض را به فرض استقلال دوتایی Z_i، کاهش دهیم. در این مورد نمی‌توانیم نابرابری هوفدینگ را اعمال کنیم، اما نابرابری زیر (نابرابری چبیشف) صادق است:

$$P(|\phi - \hat{\phi}| > \gamma) < \frac{\mathrm{Var}(Z_i)}{m\gamma^۲}$$

که $\mathrm{Var}(Z_i)$ نشان دهنده واریانس متغیر تصادفی Z_i است و برای $Z_i \sim \mathrm{Bernoulli}(\phi)$ داریم $\mathrm{Var}(Z_i) = \phi(۱ - \phi)$.
با توجه به مجموعه فرضیه $\mathcal{H} = \{h_۱, h_۲, ..., h_k\}$ می‌توان تضمین کرد خطا تعمیم بهترین فرضیه به صورت زیر است:

$$\hat{h} = \arg\min_{h \in \mathcal{H}} \hat{\epsilon}(h)$$

تعریف می‌کنیم: $\hat{\epsilon}(h) = \frac{۱}{m}\sum_{i=۱}^{m} ۱\{h(x^{(i)}) \neq y^{(i)}\}$ که $(x^{(i)}, y^{(i)})$ نشان دهنده مجموعه آموزش است.

الف) حداکثر مقدار ممکن $Var(Z_i) = \phi(۱ - \phi)$ چقدر است؟ (۶ نمره)

ب) (۱۲ نمره) با فرض $\gamma > ۰$.

۱. برای احتمال $|\hat{\epsilon}(\hat{h}) - \epsilon(\hat{h})| > \gamma$ یک کران بالا غیر بدیهی(نه ثابت ۱) بدست آورید. (۶ نمره)

۲. با در نظر گرفتن اینکه $\delta \in (۰, ۱)$ ، با استفاده از کران بالا بدست آمده، اندازه نمونه m چقدر باید باشد تا بتوان تضمین کرد: $\mathbb{P}(|\hat{\epsilon}(\hat{h}) - \epsilon(\hat{h})| > \gamma) \leq \delta$ (۳ نمره)

۳. چگونه اندازه نمونه با آنچه که با استفاده از نابرابری هوفدینگ قابل دستیابی است مقایسه می شود؟ (۳ نمره)

پ) (پرسش امتیازی - ۷ نمره) نشان دهید که با حداقل احتمال $1-\delta$ ، تفاوت بین خطای تعمیم و خطای تعمیم (Generalization Error) بهترین فرضیه در H به مقدار $\sqrt{\frac{k}{m\delta}}$ محدود می‌شود.

موفق باشید