دانشکده‌ی مهندسی کامپیوتر

نظریه یادگیری ماشین

نیم‌سال اول ۱۴۰۳-۱۴۰۴

مدرس: دکتر امیر نجفی

تمرین چهارم

مسئله‌ی ۱. رگرسیون محدب (۱۰ نمره)

فرض کنید می‌خواهیم مسئله‌ی رگرسیون به شکل زیر را حل کنیم:

$$\min_{f\in\mathcal{F}} \sum_{i=1}^{n}(f(x_i)-y_i)^2$$

که در اینجا $\mathcal{F}$ مجموعه همه توابع محدب است. روشی برای حل این مسئله‌ی بهینه‌سازی نامتنهای بعدی ارائه دهید. (نکته جالب: برخلاف حالت classification این مسئله قابل یادگیری است)

مسئله‌ی ۲. به‌روزرسانی وزن‌ها (۱۵ نمره)

فرض کنید که فرض main weak learner در الگوریتم AdaBoost برقرار باشد. فرض کنید h_t یادگیرنده پایه‌ای باشد که در مرحله t انتخاب شده است. نشان دهید که یادگیرنده پایه h_{t+1} که در مرحله $t+1$ انتخاب شده است باید متفاوت از h_t باشد.

مسئله‌ی ۳. لاسو گروهی با گروه های همپوشان (۱۵ نمره)

نشان دهید اگر از لاسو گروهی برای گروه های همپوشان استفاده کنیم در این که پرامتر های غیر صفر ما در اجتماعی از این گروه ها قرار بگیرند کمکی نمی‌کند. به طور دقیق تر استفاده از این هموارساز باعث می‌شود که پارامتر ها در مکمل اجتماعی از این گروه ها قرار بگیرند. (نیازی به اثبات این مورد نیست، برای حل این مشکل Overlapping group lasso) طراحی شده است که می‌توانید در مورد آن بیشتر بخوانید.

مسئله‌ی ۴. در جستجوی هموارساز خوب (۱۰ نمره)

در دو کرانی که برای پیچیدگی راداماخر مدل های خطی به دست آوردیم دو عبارت ظاهر می‌شد:

الف) برای نرم ۱ عبارتی که ظاهر می‌شد برابر $||w||_1||x||_\infty$ بود (در واقع کران هایی روی این ها ولی برای سادگی فعلا این عبارت ها را در نظر می‌گیریم.)

ب) برای نرم ۲ عبارت $||w||_2||x||_2$ ظاهر می‌شد.

فرض کنید w, x متغیر های تصادفی نزدیک مقادیر $\{-1, 1\}$ باشند. در حالت های زیر این دو کران را بررسی کنید و نتیجه بگیرید هموارسازی با نرم ۱ و ۲ چه زمان هایی مناسب تر هستند.

الف) بدون هیچ فرض اضافه‌ای.

ب) w تنک با حداکثر k مولفه ناصفر باشد.

ج) w چگال باشد مثلا با فرض $||w||_1 \approx \sqrt{d}||w||_2$

مسئله‌ی ۵. خطای نمایی Bayes (۱۵ نمره)

فرض کنید مجموعه ورودی $\mathcal{X}$ و فضای برچسب $Y = \{-1, +1\}$ باشد. در الگوریتم AdaBoost از تابع خطای نمایی زیر استفاده می‌شود:

$$\ell(h(x), y) = \exp(-yh(x)).$$

برای یک توزیع $\mathcal{D}$ روی $\mathcal{X} \times Y$، خطای نمایی به صورت زیر تعریف می‌شود:

$$R_l(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(h(x), y)].$$

خطای نمایی بیز برای یک توزیع $\mathcal{D}$ به صورت خطای کمینه روی توابع قابل اندازه‌گیری تعریف می‌شود:

$$R_l^* = \inf_{h:\mathcal{X}\to\mathbb{R}\text{measurable}} R_l(h).$$

فرضیه‌ای $h_{\exp}$ با $R_l(h_{\exp}) = R_l^*$ به‌عنوان راه‌حل بهینه بیز نامیده می‌شود. $\eta(x)$ را به صورت $\eta(x) = P[y = +1|x]$ تعریف کنید.

(a) عبارت رابطه راه‌حل بهینه بیز $h_{\exp}$ را برای خطای نمایی بر حسب $\eta(x)$ بدهید.

(b) خطای generalization و خطای بیز برای طبقه‌بندی باینری را به صورت زیر تعریف کنید:

$$R(h) = \mathbb{E}_{(x,y)\sim\mathcal{D}}[\mathbb{I}[sign(h(x)) \neq y]], \quad R^* = \inf_{h:\mathcal{X}\to\mathcal{Y}\text{ measurable}} R(h).$$

که $\text{sign}(t) = 1_{t\geqslant 0} - 1_{t<0}$.

نشان دهید $R(h_{\exp}) = R^*$.

مسئله‌ی ۶. خطای طبقه‌بند ضعیف (۱۰ نمره)

نشان دهید خطای h_t با توجه به توزیع $D^{(t+1)}$ دقیقاً برابر با ۱/۲ است. به عبارت دیگر، نشان دهید که برای هر $t \in [T]$ داریم:

$$\sum_{i=1}^{m} D_i^{(t+1)}\mathbb{I}[y_i \neq h_t(x_i)] = \frac{1}{2}.$$