«بسم الله الرحمن الرحیم»

تمرین سری چهارم

حسین ابراهیمی - 95105302

سوال T15

$$\min_{w\in\mathbb{R}^d} L_S(w) = \frac{1}{m}\sum_{i=1}^{m} |w^T x_i - y_i| \qquad x_i\in\mathbb{R}^d,\ y_i\in\mathbb{R}$$

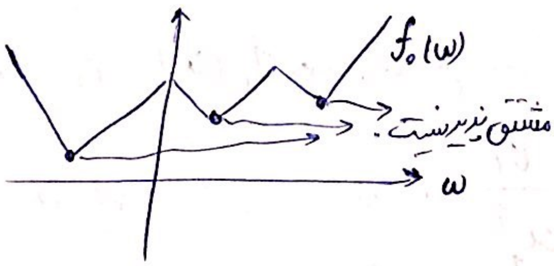
(الف) خیر، مسئله‌ی بهینه‌سازی خطی باید فرم زیر را داشته باشد:

$$\min f_0(w) = c^T w$$
$$\text{s.t.}\quad e_i^T w + r_i \leq 0$$
$$h_i^T w + g_i = 0$$

(به حالت شکسته)

اما در مسئله ما، تابع هدف ترکیبی از چندین Hyperplane است و یک خط برای بهینه‌سازی نیست همچنین مسئله غیرمقید است و نیازی به چک کردن این که قیود خطی هستند نیست.

در فضای 1 بعدی



(ب) بله، کافی است نشان دهیم:

$$\forall \alpha\in[0,1],\ \forall w_1,w_2\in\mathbb{R}^d:\ f(\alpha w_1+(1-\alpha)w_2) \leq \alpha f(w_1) + (1-\alpha) f(w_2)$$

برای این کار، مسئله را اینگونه می‌نویسیم:

$$\min \underbrace{\frac{1}{m}|\langle w,x_1\rangle - y_1|}_{f_1(w)} + \underbrace{\frac{1}{m}|\langle w,x_2\rangle - y_2|}_{f_2(w)} + \dots + \underbrace{\frac{1}{m}|\langle w,x_m\rangle - y_m|}_{f_m(w)}$$

چون ضرایب ضرب شده در توابع همگی $\frac{1}{m}$، و مثبت هستند کافی است نشان دهیم هر

یک از $f_i(w)$ ها محدب است:

$$f_i(\alpha w_1 + (1-\alpha) w_2) = |\langle \alpha w_1 + (1-\alpha) w_2 , x_i \rangle - y_i|$$

$$= |\langle \alpha w_1 , x_i \rangle + \langle (1-\alpha) w_2 , x_i \rangle - y_i|$$

$$= |\alpha \langle w_1 , x_i \rangle + (1-\alpha) \langle w_2 , x_i \rangle - y_i|$$

$$= |\alpha \langle w_1 , x_i \rangle + (1-\alpha) \langle w_2 , x_i \rangle - \alpha y_i - (1-\alpha) y_i|$$

$$= |\alpha (\langle w_1 , x_i \rangle - y_i) + (1-\alpha)(\langle w_2 , x_i \rangle - y_i)|$$

$$\leqslant \alpha |\langle w_1 , x_i \rangle - y_i| + (1-\alpha) |\langle w_2 , x_i \rangle - y_i| \quad (\alpha, 1-\alpha \geqslant 0)$$

$$\leqslant \alpha f_i(w_1) + (1-\alpha) f_i(w_2)$$

پس تمام $f_i(w)$ $i=1,\dots,m$ محدب هستند پس $\frac{1}{m}\sum_{i=1}^{m} f_i(w)$ نیز محدب است.

ج) خیر، در ریشه‌های $f_i(w)$ ها، تابع هدف مشتق‌پذیر نیست.

به طور مثال در نقطه $w = \left(\frac{y_1}{x_1^1}, 0, 0, \dots, 0\right)$ مشتق‌پذیر نیست.

(منظور از x_1^1، مؤلفه اول x_1 است.)

د)

$$\text{①} \min_{w \in \mathbb{R}^d} |f(w)| \quad \overset{\text{معادل}}{\equiv} \quad \text{②} \min c \quad \text{s.t.} \quad c \geqslant f(w), \quad c \geqslant -f(w)$$

در ابتدا فرض می‌کنیم w^* نقطه بهینه مسئله بهینه‌سازی ① باشد در این صورت خواهیم

داشت:

ادامه سوال T15، قسمت د

$$\forall w \in \mathbb{R}^d : \quad |f(w^*)| \leq |f(w)|$$

در نتیجه برای w^* نیز برقرار است، پس داریم:

$$|f(w^*)| \leq |f(w^*)| \Rightarrow \quad -\underbrace{|f(w^*)|}_{c^*} \leq f(w^*) \leq \underbrace{|f(w^*)|}_{c^*}$$

$$\Rightarrow \begin{cases} f(w^*) \leq c^* \\ -f(w^*) \leq c^* \end{cases}$$

در نتیجه (w^*, c^*) در قیود مسئله 2 صدق می‌کند.

فرض خلف: فرض می‌کنیم (w^*, c^*) جواب بهینه مسئله 2 نباشد در نتیجه $(\hat{w}, \hat{c})$ ای وجود دارد که:

$$\begin{cases} \hat{c} \geq f(\hat{w}) \\ \hat{c} \geq -f(\hat{w}) \end{cases} \longrightarrow \quad \hat{c} < c^* = |f(w^*)|$$

$$|f(\hat{w})| \leq \hat{c}$$

$$\Rightarrow |f(\hat{w})| < |f(w^*)| \quad \times$$

با فرض اینکه w^* جواب بهینه مسئله 1 است در تناقض است. پس فرض خلف باطل و حکم صادق است.

حال برای تبدیل به طریق مشابه بالا اثبات می‌شود که

$$\min_{w} |f_1(w)| + |f_2(w)| \equiv \min c$$

$$c \geq f_1(w) + f_2(w)$$

$$c \geq f_1(w) - f_2(w)$$

$$c \geq -f_1(w) + f_2(w)$$

$$c \geq -f_1(w) - f_2(w)$$

در نتیجه برای m تابع به شکل مقابل خواهیم داشت:

$$\min_{w} \left(\left|\frac{f_1(w)}{m}\right| + \left|\frac{f_2(w)}{m}\right| + \dots + \left|\frac{f_m(w)}{m}\right| \right)$$

که معادل با مسئله‌ی بهینه‌سازی زیر خواهد بود:

$$\min \quad c$$

$$\text{s.t.} \quad c \geq \frac{f_1(w)}{m} + \frac{f_2(w)}{m} + \dots + \frac{f_m(w)}{m}$$

$$c \geq -\frac{f_1(w)}{m} + \frac{f_2(w)}{m} + \dots + \frac{f_m(w)}{m}$$

$$\vdots$$

$$c \geq -\frac{f_1(w)}{m} - \frac{f_2(w)}{m} - \dots - \frac{f_m(w)}{m}$$

در نتیجه مسئله‌ی بهینه‌سازی بالا شامل 2^m قید خواهد بود که هر کدام حالت مثبت و یا منفی یکی
از $\frac{f_i(w)}{m}$ ها خواهد بود که $f_i(w)$ برابر است با: $f_i(w) = \langle w, x_i \rangle - y_i$

سوال $T1\cancel{7}$

If $m \geq m_{\mathcal{H}}(\varepsilon) \xrightarrow[D]{\text{for every}} \underset{S\sim D^m}{\mathbb{E}}\left[L_D(A(S))\right] \leq \min_{h\in\mathcal{H}} L_D(h) + \varepsilon$

حال اگر m را بزرگتر از $m_{\mathcal{H}}(\varepsilon\delta)$ بگیریم، خواهیم داشت:

$$m \geq m_{\mathcal{H}}(\varepsilon\delta) \Rightarrow \underset{S\sim D^m}{\mathbb{E}}\left[L_D(A(S))\right] \leq \min_{h\in\mathcal{H}} L_D(h) + \epsilon\delta. \quad (*)$$

حال بر اساس نامساوی مارکوف خواهیم داشت:

$$L_D(A(S)) - \min_{h\in\mathcal{H}} L_D(h) \geq 0 \Rightarrow \mathbb{P}\left[L_D(A(S)) - \min_{h\in\mathcal{H}} L_D(h) \geq \varepsilon\right]$$

$$\leq \frac{\mathbb{E}\left[L_D(A(S)) - \min_{h\in\mathcal{H}} L_D(h)\right]}{\varepsilon} = \frac{\mathbb{E}\left[L_D(A(S))\right] - \min_{h\in\mathcal{H}} L_D(h)}{\epsilon}.$$

$$\xrightarrow{(*)} \quad \leq \frac{1}{\epsilon}\cdot\epsilon\delta = \delta$$

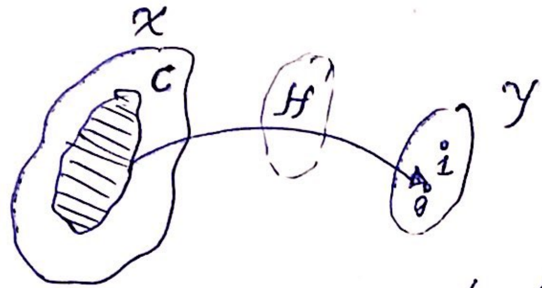
$$\Rightarrow \mathbb{P}\left[L_D(A(S)) - \min_{h\in\mathcal{H}} L_D(h) \geq \epsilon\right] \leq \delta$$

پس با احتمال $1-\delta$ خواهیم داشت که:

$$L_D(A(S)) \leq \min_{h\in\mathcal{H}} L_D(h) + \epsilon.$$

سوال T17

$$H' \subseteq H \rightarrow VCdim(H') \leq VCdim(H)$$



فرض کنید مجموعه C، مجموعه ای باشد که H' بتوان غالب است و حال چون $H' \subseteq H$ است
h های که در H' هستند که تمام حالت های مقداردهی C را پوشش می دهند، در H نیز قرار
دارند پس H نیز بتوان ها غالب است و چون H دارای توابع بیشتری نسبت به H'
است پس ممکن است C' ای با $|C'| > |C|$ وجود داشته باشد که H بتوان
ها غالب است اما H' نیست، در نتیجه خواهیم داشت:

$$VCdim(H') \leq VCdim(H)$$

سوال T18

① Hard SVM: $w^* = \underset{w}{\operatorname{argmin}} \|w\|^2$

$$\text{s.t. } y_i w^T x_i \geq 1, \quad i = 1, 2, \dots, m$$

② Soft SVM: $\hat{w} = \min_{w} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1}^{m} \max(0, 1 - y_i \cdot w^T x_i)$

$$i = 1, 2, 3, \dots, m$$

می خواهیم اثبات کنیم که برای وقتی که مجموعه آموزشی معنی که بصورت خطی قابل جداسازی
باشد جواب مسئله های بهینه سازی بالا باهم برابر اند.

بدین منظور فرض می کنیم که w^*، جواب بهینه مسئله Hard SVM باشد، حال

ادعا می‌کنیم درصورتی‌که S ویژگی مورد نظر را داشته باشد، w^* جواب مسئله soft-svm نیز می‌باشد. در مسئله 2 چون قیدی نداریم درنتیجه w^* جزء کاندیداهای جواب مسئله خواهد بود.

• برهان خلف: فرض می‌کنیم w^* جواب مسئله 1 نیست درنتیجه $\hat{w}$ وجود دارد که دارای کمترین $\frac{\lambda}{2}\|\hat{w}\|^2 + L_S^{hinge}(w)$ بین تمام w هاست، پس خواهیم داشت: (دو حالت)

I. حالت اول، حالتی است که $L_S^{hinge}(w)$ برای $\hat{w}$ برابر با صفر باشد:

$$L_S^{hinge}(\hat{w}) = 0 \Rightarrow \frac{1}{m}\sum_{i=1}^{m} \max(0, 1-y_i \hat{w}^T x_i) = 0 \Rightarrow \underbrace{y_i \hat{w}^T x_i \geqslant 1}_{\text{در مسئله 1 صدق می‌کند.}}.$$

حال با توجه به اینکه قیود مسئله 1 برای $\hat{w}$ برابر است با: $1 \leqslant y_i w^{*T} x_i$ خواهیم داشت:

$$\frac{\lambda}{2}\|\hat{w}\|^2 + \underbrace{\frac{1}{m}\sum \max(0, 1-y_i \hat{w}^T x_i)}_{0} < \frac{\lambda}{2}\|w^*\|^2 + \underbrace{\frac{1}{m}\sum l^{hinge}}_{0}$$

$$\Rightarrow \|\hat{w}\|^2 < \|w^*\|^2 \Rightarrow \|\hat{w}\| < \|w^*\| \;\; \text{X}$$

با فرض اینکه w^* جواب بهینه مسئله I است در تناقض است. پس حکم برقرار است.

II. حالت دوم، حالتی است که در آن $L_S^{hinge}(\hat{w}) \neq 0$ باشد:

$$\frac{\lambda}{2}\|\hat{w}\|^2 + \frac{1}{m}\sum_{i=1}^{m} \max(0, 1-y_i \hat{w}^T x_i)$$

$$< \frac{\lambda}{2}\|w^*\|^2 + \underbrace{\frac{1}{m}\sum_{i=1}^{m} \max(0, 1-y_i w^{*T} x_i)}_{0}.$$

$$\frac{\lambda}{2}\|\hat{w}\|^2 + \frac{1}{m}\sum_{i=1}^{m}\max(0, 1-y_i\hat{w}^T x_i) < \frac{\lambda}{2}\|w^*\|^2$$

با توجه به اینکه دو عبارت سمت چپ هر دو مثبت هستند پس خواهیم داشت :

$$\frac{\lambda}{2}\|\hat{w}\|^2 < \frac{\lambda}{2}\|w^*\|^2 \Rightarrow \|\hat{w}\| < \|w^*\| \quad (*)$$

حال λ خود را با توجه به رابطه بالا به صورت زیر تعریف می‌کنیم :

$$\lambda = \frac{2}{m\|w^*\|^2} > 0$$

حال در نتیجه خواهیم داشت :

$$\frac{1}{\cancel{m}}\sum_{i=1}^{m}\max(0, 1-y_i\hat{w}^T x_i) < \frac{\lambda}{\cancel{2}}\left(\|w^*\|^2 - \|\hat{w}\|^2\right) = \frac{\cancel{2}}{\cancel{m}}\cdot\frac{1}{\|w^*\|^2}$$

$$\Rightarrow \sum_{i=1}^{m}\max(0, 1-y_i\hat{w}^T x_i) < \frac{\|w^*\|^2 - \|\hat{w}\|^2}{\|w^*\|^2} < 1$$

$$\Rightarrow \forall i \in \{1, 2, \ldots, m\} : \max(0, 1-y_i\hat{w}x_i) < 1$$

$$\Rightarrow \forall i : 1-y_i\hat{w}^T x_i < 1 \Rightarrow y_i\hat{w}^T x_i < 1$$

$$\Rightarrow \forall i : \boxed{0 < y_i\hat{w}^T x_i < 1}$$

حال در نتیجه $\hat{w}$ در فرم بهینه‌سازی Hard-svm زیر صدق می‌کند که :

$$\max_{w}\min_{i}\frac{1}{\|w\|}|w^T x_i| *$$

$$\text{s.t.} \quad y_i\acute{w}^T x_i > 0. \quad i = 1, 2, \ldots, m$$

که فرم بهینه‌سازی بالا معادل فرم معادل مسئله بهینه‌سازی زیر است که همان Hard-SVM اصلی:

$$\min_{w} \|w\|^2 \quad s.t. \quad y_i w^T x_i \geq 1$$

پس $\hat{w}$ کاندیدای جواب مسئله بهینه‌سازی بالا است که جواب مسئله w^* می‌باشد پس با توجه به تابع هدف خواهیم داشت:

$$\|w^*\| < \|\hat{w}\|$$

که با (*) در تناقض است پس فرض خلف باطل و حکم صادق است.

پس $0 < \lambda = \frac{2}{m\|w^*\|^2}$ وجود دارد که به ازای آن جواب مسئله‌ها بهینه‌سازی soft-SVM و hard-SVM برای حالتی که به صورت خطی قابل جداسازی باشد وجود دارد.

T19 سوال

1.

$$f(\omega) = \lambda \|\omega\|^2 + \frac{1}{2m}\sum_{i=1}^{m}\left(\langle w, \psi(x_i)\rangle - y_i\right)^2$$

با توجه به ← Representer theorm داریم که جواب بهینه‌سازی تابع هدف بالا

برابر است با: $\sum_{i=1}^{m} \alpha_i \psi(x_i)$: $\alpha \in \mathbb{R}^m$

$$\langle w, \psi(x_i)\rangle = \left\langle \sum_{j=1}^{m} \alpha_j \psi(x_j), \psi(x_i)\right\rangle = \sum_{j=1}^{m} \alpha_j \langle \psi(x_j), \psi(x_i)\rangle$$

$$= \sum_{j=1}^{m} \alpha_j K(x_j, x_i) = \sum_{j=1}^{m} \alpha_j G_{ij} = (G\alpha)_i = \langle \alpha, G_{o,i}\rangle$$

$$\|\omega\|^2 = \langle w, w\rangle = \left\langle \sum_{j=1}^{m} \alpha_j \psi(x_j), \sum_{k=1}^{m} \alpha_k \psi(x_k)\right\rangle$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_j \alpha_k \langle \psi(x_j), \psi(x_k)\rangle = \sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_j \alpha_k K(x_j, x_k)$$

$$= \sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_j \alpha_k G_{jk} = \alpha^T G \alpha.$$

$$\min_{\omega} \lambda\|\omega\|^2 + \frac{1}{2m}\sum_{i=1}^{m}\left(\langle w, x_i\rangle - y_i\right)^2 \equiv \min_{\alpha} \lambda \cdot \alpha^T G\alpha + \frac{1}{2m}\sum_{i=1}^{m}\left(\langle \alpha, G_{o,i}\rangle - y_i\right)^2$$

حال اگر: $$\alpha^* = \min_{\alpha} \lambda \cdot \alpha^T G \alpha + \frac{1}{2m}\sum_{i=1}^{m}\left(\langle \alpha, G_{o,i}\rangle - y_i\right)^2$$

$$\Rightarrow w^* = \sum_{i=1}^{m} \alpha_i^* \psi(x_i)$$

$$f(\omega) = \lambda \|\omega\|^2 + \frac{1}{2m} \sum_{i=1}^{m} (\langle w, \psi(x_i) \rangle - y_i)^2 \quad .2$$

$$= \lambda w^T w + \frac{1}{2m} \sum_{i=1}^{m} (w^T \psi(x_i) - y_i)^2$$

$$\nabla_w f(\omega) = 0 \Rightarrow \nabla_w f(w) = 2\lambda W + \frac{1}{2m} \sum_{i=1}^{m} 2 \psi(x_i) (w \psi^T(x_i) - y_i)$$

$$\Rightarrow 2\lambda W^* + \frac{1}{m} (W^* \sum_{i=1}^{m} \psi(x_i) \psi^T(x_i) - \sum_{i=1}^{m} \psi(x_i) y_i) = 0$$

$$\Rightarrow \cdot \; X = (\psi(x_1) \quad \psi(x_2) \; \cdots \; \psi(x_m))$$

$$\Rightarrow 2\lambda W^* + \frac{1}{m} (w^* X X^T - Y X) = 0$$

$$\Rightarrow 2m\lambda W^* + X X^T w^* = Y X$$

$$\Rightarrow (2m\lambda I + X X^T) w^* = Y X$$

$$w^* = \underbrace{(2m\lambda I + X X^T)^{-1} Y}_{\alpha^*} X$$

$$\Rightarrow \boxed{\alpha^* = (2m\lambda I + X X^T)^{-1} Y}$$