# به نام خدا

# تمرین دوم مقدمه‌ای بر یادگیری ماشین


حسین ابراهیمی ـ ۹۵۱۰۵۳۰۲

مدرس : دکتر جمال‌الدین گلستانی


---

## سوال T۴

### الف.

با توجه به PriorKnowledge خود که یک منحنی درجه ۲ بر روی داده به خوبی fit می‌شود، hypothesis set خود را به شکل زیر در نظر می‌گیریم:

$$x = \begin{pmatrix} 1 & x^1 & x^2 \end{pmatrix}^T, \quad \mathcal{H} = \{h_w(x) = sign(w_0 + w_1(x^1) + w_2(x^2) + w_3(x^1)^2 + w_4(x^2)^2 + w_5(x^1x^2)); w_i \in \mathbb{R}\}$$

$$\implies \psi(x) = \begin{pmatrix} 1 \\ x^1 \\ x^2 \\ (x^1)^2 \\ (x^2)^2 \\ x^1x^2 \end{pmatrix}, w = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{pmatrix} \implies \mathcal{H} = \{h_w(x) = sign(w^T\psi(x)); w \in \mathbb{R}^6\} : \ linear\ model$$

### ب.

باید برای هر عضو از مجموعه $S$ داشته باشیم که :

$$y_i(w^T\psi(x_i)) > 0, \quad i = 1, 2, \cdots, 6$$

### ج.

$$w^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

حال برای $i = 1$ داریم که $y_1(w^T\psi(x_1)) = 0$ پس گام بعدی را بر اساس این عضو انجام می‌دهیم:

$$w^{(2)} = w^{(1)} + \psi(x_1)y_1$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T + \begin{pmatrix} -1 & 0 & -1 & 0 & -1 & 0 \end{pmatrix}^T = \begin{pmatrix} -1 & 0 & -1 & 0 & -1 & 0 \end{pmatrix}^T$$

حال برای $i = 1$ داریم که $y_1(w^{(2)^T}\psi(x_1)) = 3 > 0$ است.
باتوجه به بردار $w$ در گام دوم برای داده‌ی چهارم داریم که $y_4(w^{(2)^T}\psi(x_4)) = -3 < 0$. پس گام بعدی الگوریتم را بر روی این داده تکرار می‌کنیم:

$$w^{(3)} = w^{(2)} + y_4\psi(x_4)$$
$$= \begin{pmatrix}-1 & 0 & -1 & 0 & -1 & 0\end{pmatrix}^T + \begin{pmatrix}1 & 2 & 1 & 4 & 1 & 2\end{pmatrix}^T = \begin{pmatrix}0 & 2 & 0 & 4 & 0 & 2\end{pmatrix}^T$$

# سوال T5

با توجه به مسئله بهینه‌سازی Hard-SVM داریم :

$$input : S = \{(x_1, y_1), \cdots, (x_m, y_m)\}$$
$$slove : (w^*, b^*) = arg\min_{(w,b)} ||w^2|| \quad s.t. \quad y_i(< w, x_i > +b) \geqslant 1$$

$$output : \hat{w} = \frac{w^*}{||w^*||}, \hat{b} = \frac{b^*}{||w^*||}$$

همچنین می‌دانیم که الگوریتم Perceptron در صورتی که داده کاملا خطی قابل جداسازی باشد در حداکثر $R^2B^2$ که در آن $R = \max_i \; ||x_i||$ و $B = ||w^*||$ هستند خاتمه می‌یابد.
حال در این مسئله داریم که تمامی داده‌ها درون توپی با شعاع $\rho$ قرار دارند پس برای مقدار $R$ خواهیم داشت:

$$R = \max_i \; ||x_i|| \leq \rho$$

و همچنین با توجه به این که margin داده برابر با $\gamma$ است داریم:

$$\min_i \underbrace{\frac{1}{||\hat{w}||}}_{=1} y_i(< \hat{w}, x_i > +\hat{b}) = \min_i y_i(< \hat{w}, x_i > +\hat{b}) = \gamma \qquad \gamma > 0, \quad i = 1, 2, \cdots, m$$

با توجه به قضیه‌ای که در ۶ــالف اثبات خواهیم کرد، در مسئله‌ی بهینه‌سازی‌ای که حل می‌کنیم $x_j$ ای وجود دارد که یکی از قیدها را به حالت تساوی تبدیل می‌کند یعنی $y_j(w^{*T}x_j + b^*) = 1$ که کم‌ترین مقدار را نیز داراست. و همچنین می‌دانیم رابط بالا برای هر $i = 1, 2, \cdots, m$ برقرار است.

$$\min_i y_i(< \frac{w^*}{||w^*||}, x_i > + \frac{b^*}{||w^*||}) = \gamma \implies \frac{1}{||w^*||} \underbrace{\min_i y_j(< w^*, x_j > +b^*)}_{=1} = \gamma \implies \frac{1}{||w^*||} = \gamma$$

$$\implies B = ||w^*|| = \frac{1}{\gamma}$$

با توجه به نتایج بالا برای حداکثر گام‌های Perceptron خواهیم داشت:

$$T \leq R^2B^2 \leq \rho^2 \frac{1}{\gamma^2} = (\frac{\rho}{\gamma})^2$$

# سوال T6

$$d = \max_{(w,b)} \; \min_i \; \frac{1}{||w||}|w^Tx_i + b| \quad s.t. \quad y_i(w^Tx_i + b) > 0, \; i = 1, 2, \cdots, m \qquad (1)$$
$$\min \; ||w||^2 \quad s.t. \quad y_i(w^Tx_i + b) \geqslant 1 \qquad (5)$$

## الف.

فرض می‌کنیم برای ابرصفحه با پارامترهای $(w^*, b^*)$، مقدار آن بروی $m$ قید بزرگتر از ۱ باشد یعنی : (فرض خلف)

$$y_i(w^{*T}x_i + b^*) > 1$$

حال کمترین مقدار ابرصفحه روی $m$ قید را برابر $\alpha$ در نظر می‌گیریم و $(w', b')$ را به شکل زیر تعریف می‌کنیم:

$$\alpha = \min_i \ y_i(w^{*T}x_i + b^*) > 1, \qquad w' = \frac{1}{\alpha}w^*, \ b' = \frac{1}{\alpha}b^*$$

در این صورت داریم:

$$\min_i \ y_i(w'^T x_i + b') = \min_i \ y_i(\frac{1}{\alpha}w^T x_i + \frac{1}{\alpha}b) = \frac{1}{\alpha}\underbrace{\min_i \ y_i(w^{*T}x_i + b^*)}_{\alpha} = 1$$

بدین معناست که کمترین مقدار قیود روی نقاط برای ابرصفحه $(w', b')$ برابر ۱ است . حال با توجه به مسئله بهینه‌سازی داریم که $w^*$ دارای کمترین اندازه بین $w$هایی است که در مسئله (5) صدق می‌کنند که $(w', b')$ با توجه به فرض عضو این مجموعه می‌باشد اما داریم:

$$||w'|| = ||\frac{1}{\alpha}w^*|| = \frac{1}{\alpha}||w^*|| \ \overset{\alpha>1}{\Longrightarrow} \ ||w'|| < ||w^*||$$

که با فرض در تناقض است. پس فرض خلف باطل است و وجود دارد $j$ای که $y_j(w^{*T}x_j + b^*) = 1$ است.

## ب.

$(w^*, b^*)$ جواب مسئله (5) است پس در قید مسئله صدق می‌کند که در این صورت برای هر $i = 1, 2, \cdots, m$ داریم:

$$y_i(w^{*T}x_i + b^*) \geqslant 1 \implies y_i(w^{*T}x_i + b^*) > 0$$

## ج.

$$d = \max_{(w,b)} \ \min_i \ \frac{1}{||w||}|w^T x_i + b| = \max_{(w,b)} \ \min_i \ \frac{1}{||w||}y_i(w^T x_i + b)$$
$$s.t. \quad y_i(w^T x_i + b) > 0 \qquad i = 1, 2, \cdots, m$$

با توجه به قسمت الف می‌دانیم که کمترین مقدار قیود روی نقاط برای ابرصفحه $(w^*, b^*)$ برابر با ۱ است:

$$d^* = \min_i \ \frac{1}{||w^*||}y_i(w^{*T}x_i + b^*) = \frac{1}{||w^*||}\underbrace{\min_i \ y_i(w^{*T}x_i + b^*)}_{=1} = \frac{1}{||w^*||}$$

## د.

$$\tilde{d} = \min_i \ \frac{1}{||\tilde{w}||}|\tilde{w}^T x_i + \tilde{b}| \implies \tilde{d}||\tilde{w}|| = \min_i \ y_i(\tilde{w}^T x_i + \tilde{b}) \quad (*), \qquad w' = \alpha w, \quad b' = \alpha b, \quad \alpha = \frac{1}{||\tilde{w}||\tilde{d}}$$
$$\min_i \ y_i(w'^T x_i + b') = \min_i \ y_i(\alpha\tilde{w}^T x_i + \alpha\tilde{b}) = \alpha \min_i y_i(\tilde{w}^T x_i + \tilde{b}) = \alpha\tilde{d}||\tilde{w}|| = 1$$

کم‌ترین مقدار قیود برای ابرصفحه $(w', b')$ برابر با یک است در نتیجه در $m$ تا قید مسئله بهینه‌سازی (5) صدق می‌کند.

حال چون $(w', b')$ در قیود مسئله $(5)$ صدق می‌کند پس در مسئله $(1)$ نیز صادق است. هم‌چنین می‌دانیم که با $\alpha$ برابر کردن پارامترهای ابرصفحه یعنی $w$ و $b$، ابر صفحه تغییر نمی‌کند پس خواهیم داشت:

$$\tilde{d} = \min_i \frac{1}{||w'||} y_i(w'^T x_i + b') = \frac{1}{||w'||} \underbrace{\min_i y_i(w'^T x_i + b')}_{=1} = \frac{1}{||w'||}$$

حال با توجه فرض خلف‌ای که گرفتیم داریم و هم‌چنین نتایج قسمت ج و بالا خواهیم داشت:

$$\tilde{d} > d^* \implies \frac{1}{||w'||} > \frac{1}{||w^*||} \implies ||w^*||^2 > ||w'||^2$$

نتیجه با فرض این که $w^*$ جواب بهینه مسئله بهینه‌سازی $(5)$ است بدین معنا که کمترین اندازه را در بین $w$هایی که در قیود صدق می‌کنند دارد در تناقض است پس فرض خلف باطل است.