به نام خدا

تمرین اول مقدمه‌ای بر یادگیری ماشین

حسین ابراهیمی ـ ۹۵۱۰۵۳۰۲

مدرس : دکتر جمال‌الدین گلستانی

سوال T۱

الف.

کافی است نشان دهیم ضرب داخلی دو بردار ω و $u-v$ برابر صفر است که معادل آن است با این که دو بردار بر هم عمود هستند.
طبق فرض سوال می‌دانیم که دو بردار u و v در ابرصفحه HP قرار دارند پس خواهیم داشت:

$$\begin{cases} u \in HP \implies <\omega, u> + b = 0 \\ v \in HP \implies <\omega, v> + b = 0 \end{cases} \implies <\omega, u> + b - <\omega, v> - b = 0$$
$$\implies <\omega, u> - <\omega, v> = 0 \implies <\omega, u-v> = 0$$

در نتیجه بردار ω بر $u-v$ از ابرصفحه HP عمود است پس بر خود ابر صفحه HP نیز عمود خواهد بود.

ب.

کافی است اثبات کنیم که $\omega^T u + b$ بزرگتر از صفر است که به این معنا است که بردار u در نیم فضای $\omega^T x + b > 0$ قرار دارد.

$$x \in HP \implies <\omega, x> + b = 0, \qquad u = x + \alpha\omega \quad \alpha > 0$$

$$\omega^T u + b = <\omega, u> + b = <\omega, x + \alpha\omega> + b = <\omega, \alpha\omega> + \underbrace{<\omega, x> + b}_{0} = \alpha <\omega, \omega> = \alpha ||\omega||^2 > 0$$

در نتیجه u در نیم‌فضای مذکور قرار دارد.

ج.

نشان می‌دهیم هر v که در نیم‌فضای $\omega^T x + b > 0$ قرار دارد برای وقتی که $\omega \to \alpha\omega$ و $b \to \alpha b$ تغییر کند خواهیم داشت:

$$\omega'^T v + b' = <\omega', v> + b' = <\alpha\omega, v> + \alpha b = \alpha <\omega, v> + \alpha\omega = \alpha(<\omega, v> + b) = \alpha(\omega^T v + b)$$

حال اگر $\alpha > 0$ باشد، با توجه به این که $\omega^T v + b > 0$ خواهیم داشت که $\omega'^T v + b' > 0$ است پس یعنی هر برداری که در نیم‌فضای مثبت قرار داشته در حالت جدید نیز در همان نیم‌فضا قرار خواهد داشت .

اما اگر $\alpha < 0$ باشد داریم که $\omega'^T v + b' < 0$ پس هر برداری که در نیم‌فضای مثبت قرار داشته در حالت جدید جای آن عوض شده و در نیم‌فضای منفی ابرصفحه جدید قرار دارد.
به حالت مشابه قبل اگر v در نیم‌فضای $\omega^T x + b < 0$ قرار داشته باشد در حالتی که $\alpha > 0$ است چون $\omega^T x + b < 0$ پس $\omega'^T v + b' < 0$ یعنی در همان نیم‌فضای قبلی خود خواهد ماند. اما اگر $\alpha < 0$ باشد، خواهیم داشت $\omega'^T v + b' > 0$ که بدین معناست که v در حالت جدید در نیم‌فضای مثبت قرار دارد و نسبت به حالت قبل خود تغییر جا داده است.
پس در حالت کلی وقتی که $\alpha > 0$ باشد، نقاط در هر دو نیم‌فضای HP تغییر وضعیت نمی‌دهند ولی اگر $\alpha < 0$ باشد جای دو نیم‌فضا باهم عوض می‌شود.

د.

کافی است مقدار مسافتی که لازم است در جهت $u + \alpha\omega$ طی کنیم تا به نقطه‌ای از HP به نام v برسیم را بدست آوریم. پس خواهیم داشت:

$$v = u + \alpha\omega$$

حال با توجه به این که نقطه v در HP قرار دارد خواهیم داشت:

$$v \in HP \implies \omega^T v + b =< \omega, v > +b = 0 \implies < \omega, u + \alpha\omega > +b = 0 \implies < \omega, u > + < \omega, \alpha\omega > +b = 0$$
$$\implies \alpha < \omega, \omega >= -(< \omega, u > +b) \implies \alpha = -\frac{< \omega, u > +b}{||w||^2} = -\frac{\omega^T u + b}{||w||^2}$$

با توجه به این که فاصله مقداری مثبت است، قدرمطلق مقدار α را در نظر می‌گیریم:

$$minimum\ distance = |\alpha| = \frac{|\omega^T u + b|}{||w||^2}$$

سوال T۲

الف.

$$L_s(h) = \frac{1}{3}\sum_{i=1}^{3}(h(x_i) - y_i)^2 = \frac{1}{3}\sum_{i=1}^{3}(a_0 x_i + a_1 x_i + a_2 x_i^2 - y_i)^2$$
$$\implies L_s(a_0, a_1, a_2) = \frac{1}{3}[(a_0 - 1)^2 + (a_0 + a_1 + a_2)^2 + (a_0 + 2a_1 + 4a_2 - 4)^2]$$

ب.

$$\frac{\partial L_s}{\partial a_0} = \frac{1}{3}[2(a_0 - 1) + 2(a_0 + a_1 + a_2) + 2(a_0 + 2a_1 + 4a_2 - 4)] = \frac{2}{3}[3a_0 + 3a_1 + 5a_2 - 5] = 0$$
$$\frac{\partial L_s}{\partial a_1} = \frac{1}{3}[2(a_0 + a_1 + a_2) + 4(a_0 + 2a_1 + 4a_2 - 4)] = \frac{2}{3}[3a_0 + 5a_1 + 9a_2 - 8] = 0$$
$$\frac{\partial L_s}{\partial a_2} = \frac{1}{3}[2(a_0 + a_1 + a_2) + 8(a_0 + 2a_1 + 4a_2 - 4)] = \frac{2}{3}[5a_0 + 9a_1 + 17a_2 - 16] = 0$$
$$\implies \begin{cases} 3a_0 + 3a_1 + 5a_2 = 5 \\ 3a_0 + 5a_1 + 9a_2 = 8 \\ 5a_0 + 9a_1 + 17a_2 = 16 \end{cases} \implies a_0 = 1,\ a_1 = -3.5,\ a_2 = 2.5$$

ج.

$$\psi(x) = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \implies \psi(x_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \psi(x_2) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \psi(x) = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$$

$$\Psi = \begin{pmatrix} \psi(x_1) & \psi(x_2) & \psi(x_3) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 4 \end{pmatrix} \implies \Psi^T = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}$$

$$A = \Psi\Psi^T = \begin{pmatrix} 3 & 3 & 5 \\ 3 & 5 & 9 \\ 5 & 9 & 17 \end{pmatrix},\ b = \Psi Y = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \\ 16 \end{pmatrix}$$

ستون‌های ماتریس A مجموعه‌ای مستقل خطی می‌باشد در نتیجه ماتریس A، $full\ rank$ است پس معکوس‌پذیر می‌باشد.

$$\omega^* = A^{-1}b = \begin{pmatrix} 1 & -1.5 & 0.5 \\ -1.5 & 6.5 & -3 \\ 0.5 & -3 & 1.5 \end{pmatrix} \begin{pmatrix} 5 \\ 8 \\ 16 \end{pmatrix} \implies \omega^* = \begin{pmatrix} 1 \\ -3.5 \\ 2.5 \end{pmatrix} \implies a_0 = 1,\ a_1 = -3.5,\ a_2 = 2.5$$

ضرایب دقیقا برابر با ضرایبی است که از قسمت "ب" سوال بدست آمده بود.

سوال T۳

ابتدا نشان می‌دهیم که با تغییر کردن شکل آپدیت کردن ω در هر مرحله از الگوریتم، برای (x_i, y_i)هایی که $y_i\omega^{(t)}x_i < 0$ است بدین معنا که این نقاط در طرف اشتباه predict شده‌اند، در هر بار اجرا کردن الگوریتم یک مقدار مثبت به آن‌ها می‌افزاید:

$$\omega^{(t+1)} = \omega^{(t)} + \eta x_i y_i \quad for\ some\ \eta > 0, \quad \omega^{(1)} = 0$$

$$\begin{aligned} y_i\omega^{(t+1)}x_i &= y_i < \omega^{(t+1)}, x_i >= y_i < \omega^{(t)} + \eta x_i y_i, x_i >= y_i(< \omega^{(t)}, \eta x_i y_i > + < \omega^{(t)}, x_i >) \\ &= y_i < \omega^{(t)}, x_i > + \eta \underbrace{y_i^2}_{1} < x_i, x_i >= y_i\omega^{(t)}x_i + \underbrace{\eta||x_i||^2}_{>0} > y_i\omega^{(t)}x_i \end{aligned}$$

حال برای اینکه نشان دهیم الگوریتم $perceptron$ تغییر یافته در همان تعداد $step$ به همان نقطه همگرا می‌شود، متغیرهای زیر را همانند خود الگوریتم تعریف می‌کنیم:

$$\omega^* = argmin\ ||\omega||^2 \quad s.t. \quad y_i\omega^{*T}x_i \geqslant 1 \quad i = 1, 2, 3, \cdots, m$$

با توجه به این که می‌دانیم داده‌ی آموزشی $separable$ است پس ωای وجود دارد که به ازای آن برای هر i خواهیم داشت که $y_i\omega^T x_i > 0$ است. حال کوچکترین این $y_i\omega^T x_i$ ها را برابر α در نظر می‌گیریم و داریم که $\frac{1}{\alpha}\omega$ برداری است که به ازای آن برای هر i، $y_i\omega^T x_i \geqslant 1$ برقرار است. پس بدین صورت ω^* با شرایطای که باید داشته باشد بدست می‌آوریم.

$$R \triangleq max_i\ ||x_i|| \qquad B \triangleq ||\omega^*||$$

اثبات می‌کنیم پس از T گام الگوریتم جدید خواهیم داشت:

$$\frac{< \omega^*, \omega^{(T+1)} >}{||\omega^*||\ ||\omega^{(T+1)}||} \geqslant \frac{\sqrt{T}}{RB}$$

برای صورت کسر داریم:

$$< \omega^*, \omega^{(t+1)} > - < \omega^*, \omega^{(t)} > = < \omega^*, \omega^{(t+1)} - \omega^* >= \eta y_i < \omega^*, x_i >= \eta \underbrace{y_i \omega^{*T} x_i}_{\geqslant 1} \geqslant \eta$$

حال با توجه به اینکه $\omega^{(1)} = 0$ است، بعد از T گام خواهیم داشت:

$$< \omega^*, \omega^{(T+1)} > = \sum_{t=1}^{T} (< \omega^*, \omega^{(t+1)} > - < \omega^*, \omega^{(t)} >) \geqslant \eta T$$

حال برای مخرج کسر داریم:

$$||\omega^{(t+1)}||^2 = < \omega^{(t)} + \eta x_i y_i, \omega^{(t)} + \eta x_i y_i > = ||\omega^{(t)}||^2 + \underbrace{2\eta y_i < \omega^{(t)}, x_i >}_{\leq 0} + \eta^2 y_i^2 ||x_i||^2 \leq ||\omega^{(t)}||^2 + \eta^2 R^2$$

و چون داریم که $||\omega^{(1)}||^2 = 0$ پس بعد از T گام از اجرای الگوریتم تغییر یافته داریم:

$$\implies ||\omega^{(T+1)}||^2 \leq \eta^2 T R^2 \implies ||\omega^{(T)}|| \leq \eta T R$$

پس خواهیم داشت :

$$\frac{< \omega^*, \omega^{(T+1)} >}{||\omega^*|| \, ||\omega^{(T+1)}||} \geqslant \frac{\eta T}{\eta B R \sqrt{T}} = \frac{\sqrt{T}}{BR}$$

در نتیجه رابطه بالا اثبات شد. پس الگوریتم تغییر یافته همانند الگوریتم $perceptron$ در حداکثر R^2B^2 گام خاتمه می‌یابد.
برای نشان دادن این که برداری که بدست می‌آید(ω') هم جهت با بردار $perceptron$ (ω) است از استقرا استفاده می‌کنیم:
پایه استقرا: $\omega'^{(1)} = \eta\omega^{(1)} = 0$
فرض استقرا: فرض می‌کنیم تا t مرتبه $iterate$، داریم $\omega'^{(t)} = \eta\omega^t$ برای t ,$i = 0, 1, 2, \cdots$ حال اثبات می‌کنیم $\omega'^{(t+1)} = \eta\omega^{(t+1)}$.
حال وقتی ما در یکی از الگوریتم‌ها $iterate$ کنیم حتما باید در الگوریتم دیگر نیز $iterate$ انجام دهیم چون فرض کنید i ای وجود داشته باشد که $y_i\omega^{(t)}x_i < 0$ باشد، در این صورت چون η مثبت است، خواهیم داشت که $y_i\eta\omega^{(t)}x_i = y_i\omega'^{(t)}x_i < 0$ نیز هست پس در الگوریتم دیگر نیز نیاز به $iterate$ هستیم. حالت برعکس نیز چون η مثبت است برقرار است.
طبق فرض استقرا داریم:

$$\omega^{(t+1)} = \omega^{(t)} + y_i x_i$$

$$\omega'^{(t+1)} = \omega'^{(t)} + \eta y_i x_i = \eta\omega^{(t)} + \eta y_i x_i = \eta \underbrace{(\omega^t + y_i x_i)}_{\omega^{(t+1)}} = \eta\omega^{(t+1)}$$

پس داریم $\omega'^{(t+1)} = \eta\omega^{(t+1)}$ که حکم استقرا را ثابت می‌کند.