بسمه تعالی

پاسخ سری پنجم تمرینات درس یادگیری ماشین

امیرحسین رمضانی بناب (۹۹۲۱۰۲۹۴)

۱ سوال ۱

۱.۱ آ

نشان می‌دهیم در نقطه‌ی بهینه خواهیم داشت

$$\forall n \;:\; \xi_n \geq 0$$

فرض کنیم یک داده داشته باشیم که $\xi_i < 0$. در آن صورت داریم:

$$\mathsf{y}^{(i)}(\mathbf{w}^\intercal \mathsf{x}^{(i)} + b) \geq 1 + |\xi_i|$$

حال اگر ξ_i را تبدیل به صفر کنیم، اولا تابع هدف کاهش می‌یابد. ثانیا شرایط مساله نقض نمی‌شود زیرا

$$\mathsf{y}^{(i)}(\mathbf{w}^\intercal \mathsf{x}^{(i)} + b) \geq 1 + |\xi_i| \geq 1$$

پس همچنان برای تمام نقاط شرایط مسئله برقرار است و تابع هدف نیز کاهش یافته است. که این با بهینه‌بودن ξ_n در تناقض است.

۲.۱ ب

با استفاده از ضرایب لاگرانژ به مسئله‌ی زیر خواهیم رسید

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2}\mathbf{w}^\intercal\mathbf{w} + \frac{C}{2}\sum_{n=1}^{N}\xi_n^2 - \sum_{n=1}^{N}\alpha_n(y_n(\mathbf{w}^\intercal\mathbf{x}_n + b) - 1 + \xi_n)$$

باید این عبارت را نسبت به $\mathbf{w}$ و b و ξ کمینه و نسبت به $\alpha \geq 0$ بیشینه کنیم. داریم:

$$\nabla_{\mathbf{w}}\mathcal{L} = \mathbf{w} - \sum_{n=1}^{N}\alpha_n y_n \mathbf{x}_n = 0 \qquad \Rightarrow \mathbf{w} = \sum_{n=1}^{N}\alpha_n y_n \mathbf{x}_n \qquad (*)$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{n=1}^{N}\alpha_n y_n = 0 \qquad \Rightarrow \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\frac{\partial \mathcal{L}}{\partial \xi_n} = C\xi_n - \alpha_n = 0 \qquad \Rightarrow \xi_n = \frac{\alpha_n}{C} \qquad (**)$$

۳.۱ پ

با جایگذاری عبارات بالا در $\mathcal{L}$ خواهیم داشت

$$\mathcal{L}(\alpha) = \sum_{n=1}^{N}\alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}y_n y_m \alpha_n \alpha_m \mathbf{x}_n^\intercal \mathbf{x}_m - \frac{1}{2C}\sum_{n=1}^{N}\alpha_n^2$$

با توجه به اینکه همیشه برای $y_n \in \{-1,+1\}$ داریم

$$y_n y_n = +1$$

دو عبارت جمع بالا را در یک عبارت خلاصه می‌کنیم

$$\mathcal{L}(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m (\mathbf{x}_n^\intercal \mathbf{x}_m + \frac{\delta_{ij}}{C})$$

که δ_{ij} تابعی است که در $i = j$ مقدار ۱ و در بقیه نقاط مقدار ۰ دارد.
به این ترتیب، هدف کمینه کردن $\mathcal{L}$ است به شکلی که

$$\forall n \ : \ \sum_{n=1}^{N} \alpha_n y_n = 0 \ , \ \alpha_i \geq 0$$

۴.۱ ت

وقتی $C \to \infty$ مسئله‌ی بهینه سازی زیر را خواهیم داشت:

$$minimize \ \mathcal{L}(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m \mathbf{x}_n^\intercal \mathbf{x}_m$$

که

$$\forall n \ : \ \sum_{n=1}^{N} \alpha_n y_n = 0 \ , \ \alpha_i \geq 0$$

این همان مسئله‌ی بهینه‌سازی Hard Margin SVM است.
وقتی $C \to 0$ از رابطه‌ی ** داریم

$$C\xi_n - \alpha_n = 0 \Rightarrow \alpha_n = 0$$

در نتیجه از رابطه‌ی * داریم

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n = 0$$

پس در این حالت همه‌ی وزن‌ها برابر صفر می‌شوند و عملا دسته‌بندی انجام نخواهد شد.

۲ سوال ۲

با توجه به اینکه در صورت سوال گفته شده است که هسته‌های k_1 و k_2 معتبر هستند. فرض کنیم ماتریس این هسته‌ها را با K_1 و K_2 نشان دهیم و همچنین تعداد داده ها را N در نظر بگیریم. باید دو شرط زیر را بررسی نماییم:

$$
\begin{aligned}
&k_i(x_1, x_2) = k_i(x_2, x_1) \quad for\ i = 1, 2 \\
&\forall a \in \mathbb{R}^N \ : a^\intercal K_i a \succeq 0 \quad for\ i = 1, 2
\end{aligned}
$$

۱.۲ ۱

برای هسته‌ی k_3 آن دو ویژگی را بررسی می‌کنیم.
شرط اول :

$$
\begin{aligned}
k_3(x_2, x_1) &= k_1(x_2, x_1) + k_2(x_2, x_1) \\
&= k_1(x_1, x_2) + k_2(x_1, x_2) \\
&= k_3(x_1, x_2)
\end{aligned}
$$

پس شرط اول برقرار است.
شرط دوم :

$$
\begin{aligned}
\forall a \in \mathbb{R}^N \ : a^\intercal K_3 a &= a^\intercal (K_1 + K_2) a \\
&= a^\intercal K_1 a + a^\intercal K_2 a \\
&\succeq 0
\end{aligned}
$$

پس شرط دوم برقرار است و هسته معتبر است.

۲.۲ ۲

برای هسته‌ی k_4 از معتبر بودن هسته‌های k_1 و k_2 استفاده می‌کنیم. از آنجایی که این دو هسته معتبر هستند، اگر فضای نمونه‌های مسئله را $\mathcal{X}$ بنامیم و فضاهای جدید را با $\mathcal{Z}_i$ نمایش دهیم دو تابع زیر را داریم:

$$
\begin{aligned}
\Phi_1 &: \mathcal{X} \to \mathcal{Z}_1 \\
\Phi_2 &: \mathcal{X} \to \mathcal{Z}_2
\end{aligned}
$$

که

$$
\begin{aligned}
k_1(x_1, x_2) &= \Phi_1(x_1)^\intercal \Phi_1(x_2) \\
k_2(x_1, x_2) &= \Phi_2(x_1)^\intercal \Phi_2(x_2)
\end{aligned}
$$

اگر فرض کنیم

$$
\begin{aligned}
\Phi_1(x_i) &=< \alpha_1^{(i)}, \alpha_2^{(i)}, ..., \alpha_N^{(i)} > \\
\Phi_2(x_i) &=< \beta_1^{(i)}, \beta_2^{(i)}, ..., \beta_M^{(i)} >
\end{aligned}
$$

خواهیم داشت

$$
\begin{aligned}
k_1(x_1, x_2) &= \sum_{i=1}^{N} \alpha_i^{(1)} \alpha_i^{(2)} \\
k_2(x_1, x_2) &= \sum_{i=1}^{M} \beta_i^{(1)} \beta_i^{(2)}
\end{aligned}
$$

پس

$$\begin{aligned}
k_4(x_1,x_2) &= k_1(x_1,x_2)k_2(x_1,x_2) \\
&= \Big(\sum_{i=1}^{N}\alpha_i^{(1)}\alpha_i^{(2)}\Big)\Big(\sum_{i=1}^{M}\beta_i^{(1)}\beta_i^{(2)}\Big) \\
&= \sum_{i=1}^{N}\sum_{i=j}^{M}\alpha_i^{(1)}\alpha_i^{(2)}\beta_j^{(1)}\beta_j^{(2)} \\
&= \sum_{i=1}^{N}\sum_{i=j}^{M}[\alpha_i^{(1)}\beta_j^{(1)}][\alpha_i^{(2)}\beta_j^{(2)}] \\
&= \sum_{i=1}^{N}\sum_{i=j}^{M}[\gamma_{ij}^{(1)}][\gamma_{ij}^{(2)}] \\
&= \gamma(x_1)^\intercal\gamma(x_2)
\end{aligned}$$

که در عبارت بالا

$$\gamma_{ij}^{(k)} = \alpha_i^{(k)}\beta_j^{(k)}$$

و $\gamma(x_k)$ یک ماتریس $N \times M$ (معادل یک بردار NM بعدی)است که:

$$\gamma(x_k)[i,j] = \gamma_{ij}^{(k)}$$

پس چون توانستیم هسته را به شکل ضرب داخلی این دو بردار بنویسیم، این هسته معتبر می‌باشد.

۳ ۳.۲

برای هسته‌ی k_5 از سری تیلور استفاده می‌کنیم. طبق بسط تیلور داریم

$$\forall i,j \;:\; e^{k_5(x_i,x_j)} = \sum_{n=0}^{n=\infty}\frac{k_5(x_i,x_j)^n}{n!}$$

پس اگر ماتریس کرنل k_5 را K_5 بنامیم داریم:

$$e^{K_5} = \sum_{n=0}^{n=\infty}\frac{K_5^n}{n!}$$

که تمام عملیات به صورت element-wise انجام می‌شود. از آنجایی که در قسمت های قبلی ثابت کردیم جمع و ضرب هسته‌ها به شکل element-wise هسته‌ی معتبری را می‌سازد، چون سری تیلور از مجموعه‌ای از ضرب و جمع element-wise ساخته شده است نیز هسته‌ی معتبری می‌سازد.

۴ ۴.۲

در این سوال از خاصیت "همگنی" که جزو خواص هسته‌های معتبر در سوال ۳ ذکر شده است، استفاده می‌کنیم. اگر k_6 یک هسته‌ی معتبر باشد پس باید داشته باشیم:

$$\forall\alpha\in\mathbb{R}_{++} \;:\; k_6(\alpha x_1,x_2) = \alpha k_6(x_1,x_2)$$

(البته این تعریف همگنی در کوئرا اصلاح شد و به اعداد منفی تعمیم یافت. ولی در این نقطه ما به اعداد منفی نیازی نداریم و فقط روی اعداد مثبت بررسی می‌کنیم)

پس

$$\frac{1}{1-\alpha x_1^\intercal x_2} = \frac{\alpha}{1-x_1^\intercal x_2}$$

یعنی

$$\alpha - \alpha^2 x_1^\intercal x_2 = 1 - x_1^\intercal x_2$$

پس

$$x_1^\intercal x_2 = \frac{1}{1+\alpha}$$

از آنجایی که $x_1^\intercal x_2$ مقدار مشخصی دارد ولی تساوی به ازای هر α مثبت اتفاق می‌افتد به تناقض می‌رسیم.زیرا در آن صورت نتیجه می‌شود همه‌ی αهای مثبت با هم برابرند! پس این هسته معتبر نیست.

۳ سوال ۳

ابتدا خاصیت همگنی را استفاده می‌کنیم تا یک نتیجه بگیریم:

$$x = 0 \Rightarrow f(0, y) = \alpha f(0, y) \Rightarrow f(0, y) = 0$$

حال از خاصیت افزایشی استفاده می‌کنیم. داریم:

$$f(x + (-x), y) = f(x, y) + f(-x, y)$$

پس

$$f(x, y) + f(-x, y) = 0 \Rightarrow f(-x, y) = -f(x, y)$$

از اینجا به بعد منظور از خاصیت همگنی، همین خاصیت است.
از خواص تقارنی، افزایش و همگنی استفاده می‌کنیم تا h را ساده‌تر کنیم. داریم:

$$\begin{aligned}
g(x+y, x+y) &= g(x, x+y) + g(y, x+y) && (1)\\
&= g(x+y, x) + g(x+y, y) && (2)\\
&= g(x, x) + g(y, x) + g(x, y) + g(y, y) && (3)\\
&= g(x, x) + g(y, y) + 2g(x, y) && (4)
\end{aligned}$$

در ۱ از خاصیت افزایشی، برای رسیدن به ۲ از خاصیت تقارنی، برای رسیدن به ۳ از خاصیت افزایشی و برای رسیدن به ۴ از خاصیت تقارنی استفاده شد.
همچنین داریم:

$$\begin{aligned}
g(x-y, x-y) &= g(x, x-y) + g(-y, x-y) && (5)\\
&= g(x-y, x) - g(x-y, y) && (6)\\
&= g(x, x) + g(-y, x) - g(x, y) - g(-y, y) && (7)\\
&= g(x, x) - g(y, x) - g(x, y) + g(y, y) && (8)\\
&= g(x, x) + g(y, y) - 2g(x, y) && (9)
\end{aligned}$$

در ۵ از خاصیت افزایشی، برای رسیدن به ۶ از خاصیت تقارنی و همگنی، برای رسیدن به ۷ از خاصیت افزایشی، برای رسیدن به ۸ از خاصیت همگنی و برای رسیدن به ۹ از خاصیت تقارنی استفاده شد.

حال داریم:

$$\begin{aligned}
h(x, y) &= \frac{1}{4}(g(x+y, x+y) - g(x-y, x-y))\\
&= \frac{1}{4}[(g(x, x) + g(y, y) + 2g(x, y)) - (g(x, x) + g(y, y) - 2g(x, y))]\\
&= \frac{1}{4}[4g(x, y)]\\
&= g(x, y)
\end{aligned}$$

پس چون g یک کرنل معتبر است، h نیز چنین است.

۴ سوال ۴

فرض کنیم مجموعه‌ی متناهی M مجموعه‌ی مرجع ما باشد که n عضو دارد. قرار می‌دهیم $P = 2^M$. پس P مجموعه‌ی توانی M است. از آنجایی که P متناهی است عناصرش را به شکل زیر در نظر می‌گیریم

$$P = \{\zeta_1, \zeta_2, ..., \zeta_{2^n}\}$$

حال برای مجموعه‌ی A تابع Φ را به شکل زیر تعریف می‌کنیم:

$$\Phi : P \to \{0,1\}^{2^n}$$

که برای هر $A \in P$ داریم:

$$\Phi(A) = (\alpha_1, \alpha_2, ..., \alpha_{2^n})$$

که

$$\alpha_i = \begin{cases} 1 & \zeta_i \subseteq A \\ 0 & \zeta_i \not\subseteq A \end{cases}$$

حال اگر $B \in P$ یک مجموعه باشد، و داشته باشیم

$$\Phi(B) = (\beta_1, \beta_2, ..., \beta_{2^n})$$

که

$$\beta_i = \begin{cases} 1 & \zeta_i \subseteq B \\ 0 & \zeta_i \not\subseteq B \end{cases}$$

خواهیم داشت

$$\Phi(A)^\intercal \Phi(B) = \sum_{i=1}^{2^n} \alpha_i \beta_i = |\{i | 1 \leq i \leq 2^n \ : \ \alpha_i = 1 \land \beta_i = 1\}|$$

از آن‌جایی که

$$\begin{aligned} |\{i | 1 \leq i \leq 2^n \ : \ \alpha_i = 1 \land \beta_i = 1\}| &= |\{i | 1 \leq i \leq 2^n \ : \ \zeta_i \subseteq A \land \zeta_i \subseteq B\}| \\ &= |\{i | 1 \leq i \leq 2^n \ : \ \zeta_i \subseteq A \cap B\}| \\ &= 2^{|A \cap B|} \end{aligned}$$

پس

$$\Phi(A)^\intercal \Phi(B) = 2^{|A \cap B|}$$

حال چون

$$k(A, B) = 2^{|A \cap B|}$$

پس

$$k(A, B) = \Phi(A)^\intercal \Phi(B)$$

که نشان می‌دهد این هسته معتبر است.

۵ سوال ۵

در این سوال ابتدا از طریق هسته‌ی RBF به دقت حدود 90.3 درصدی رسیدم و سپس با تغییر C به دقت 93.18 درصد دست یافتم. دلیل بهبود دقت افزایش پارامتر C بود که منجر شد کمی از Regularization کاسته شود و به اهمیت Violation افزوده شود.