دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین بهار ۱۴۰۰

پاسخ تمرین سری پنجم

مدرس: دکتر محمّدحسین رهبان

پاسخ سوال ۱ *SVM*

آ) فرض کنید (w^*, b^*, ζ^*) جواب بهینه این مسئله بدون در نظر گرفتن $\forall i : \zeta_i \geq 0$ باشد. کافی است ثابت کنیم $\forall i : {\zeta_i}^* > 0$. فرض کنیم این‌طور نیست و $\zeta_i^* < 0$ است. در این صورت : $y_i({w^*}^T x_j + b^*) \geq 1 - \zeta_i^* > 1 - 0$ که از این نتیجه می‌شود $\zeta_i = 0$ یک نتیجه قابل قبول است. حال آن‌که $\zeta_i^2 = 0$ کوچکتر از ${\zeta_i^*}^2$ است و این در تناقض با این است که ζ_i^* بهینه است. بنابراین فرض خلف باطل و حکم اثبات می شود.

ب)

$$min_{w,b,\zeta} max_{\lambda} \frac{1}{2}||w||^2 + \sum_{i=1}^{N} \frac{C}{2}\zeta_i^2 - \sum_{i=1}^{N} \lambda_i((y_i(w^T x_i + b) - 1 + \zeta_i)$$

$$s.t. \ \forall i : \lambda_i \geq 0$$

به دلیل برقراری *strong duality* می‌توانیم جای *min* و *max* را عوض کنیم .

$$L(w,b,\zeta,\lambda) = max_{\lambda} min_{w,b,\zeta} \frac{1}{2}||w||^2 + \sum_{i=1}^{N} \frac{C}{2}\zeta_i^2 - \sum_{i=1}^{N} \lambda_i((y_i(w^T x_i + b) - 1 + \zeta_i)$$

حال مشتق‌گیری را انجام می‌دهیم.

$$\nabla_w L = 0 \Rightarrow w - \sum_{i=1}^{N} \lambda_i x_i y_i = 0 \Rightarrow w = \sum_{i=1}^{N} \lambda_i x_i y_i$$

$$\frac{\partial L}{\partial \zeta_i} = C\zeta_i - \lambda_i = 0 \Rightarrow C\zeta_i = \lambda_i$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} \lambda_i y_i = 0$$

حال مقادیر فوق را در عبارت L جای‌گذاری می‌کنیم.

$$L(w,b,\zeta,\lambda) = max_{\lambda} \frac{1}{2}(\sum_{i=1}^{N} \lambda_i x_i^T y_i)(\sum_{i=1}^{N} \lambda_i x_i y_i) + \frac{1}{2}\sum_{i=1}^{N} \frac{\lambda_i}{\zeta_i}\zeta_i^2 - \sum_{i=1}^{N} \lambda_i[y_i(\sum_{j=1}^{N} \lambda_j x_j^T y_j)x_i + b) - 1 + \zeta_i]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\lambda_j y_i y_j x_i^T x_j + \frac{1}{2}\sum_{i=1}^{N} \lambda_i\zeta_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i\lambda_j y_i y_j x_i^T x_j - \sum_{i=1}^{N} \lambda_i y_i b + \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \lambda_i\zeta_i$$

در عبارت فوق $\sum_{i=1}^{N}\lambda_i y_i b = 0$ است. در نهایت عبارت لاگرانژی بهینه به صورت زیر است:

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j - \frac{1}{2}\sum_{i=1}^{N}\lambda_i\zeta_i + \sum_{i=1}^{N}\lambda_i$$

پ) فرم دوگان :

$$max_\lambda - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j - \frac{1}{2}\sum_{i=1}^{N}\lambda_i\zeta_i + \sum_{i=1}^{N}\lambda_i$$

$$s.t.\ \ \forall i : \lambda_i \geq 0,\ \ \sum_{i=1}^{N}\lambda_i y_i = 0$$

ت) در حالت $C \Rightarrow \infty$، $margin$ ها اهمیت زیادی دارند و با تغییر کوچکی در ζ ها ممکن است نقطه بهینه‌ای که یافته‌ایم را از دست بدهیم؛ در نتیجه باید ζ ها را نزدیک صفر نگه داریم. بنابراین الگوریتم حالت $hard\ margin$ را در نظر می‌گیرد.

زمانی‌که $C \Rightarrow 0$، یعنی خیلی کوچک باشد، الگوریتم می‌تواند $margin$ ها را به راحتی حرکت دهد؛ بنابراین می‌تواند $margin$ بزرگ‌تری را در نظر بگیرد.

پاسخ سوال ۲ *Kernel*

۱.

$$k_1(x_1, x_2) = \phi_1(x_1)^T\phi_1(x_2)$$

$$k_2(x_1, x_2) = \phi_2(x_1)^T\phi_2(x_2)$$

$$k_3(x_1, x_2) = k_1(x_1, x_2) + k_2(x_1, x_2)$$

$$= \phi_1(x_1)^T\phi_1(x_2) + \phi_2(x_1)^T\phi_2(x_2)$$

$$= (\phi_1(x_1), \phi_2(x_1))^T(\phi_1(x_2), \phi_2(x_2))$$

حال تعریف می‌کنیم: $\forall x : \phi_3(x) = (\phi_1(x), \phi_2(x))$. در واقع ϕ_3 را از *concat* کردن ϕ_1 و ϕ_2 بدست می‌آوریم.

بنابراین داریم: $\phi_3(x_1)^T\phi_3(x_2) = k_3(x_1, x_2)$ و طبق تعریف k_3 یک هسته معتبر است.

۲. مشابه قسمت قبل

$$k_4(x_1, x_2) = k_1(x_1, x_2)k_2(x_1, x_2)$$

$$k_1(x_1, x_2) = \phi_1(x_1)^T\phi_1(x_2) = \sum_i \phi_{1i}(x_1)\phi_{1i}(x_2)$$

$$k_2(x_1, x_2) = \phi_2(x_1)^T\phi_2(x_2) = \sum_j \phi_{2j}(x_1)\phi_{2j}(x_2)$$

حال دو عبارت فوق را در هم ضرب می‌کنیم:

$$\sum_i \sum_j \phi_{1i}(x_1)\phi_{1i}(x_2)\phi_{2j}(x_1)\phi_{2j}(x_2)$$

حال تعریف می‌کنیم:

$$\phi'_{ij}(x) = \phi_{1i}(x)\phi_{2j}(x)$$

در نتیجه داریم:

$$k_4(x_1, x_2) = \sum_{i,j} \phi'_{ij}(x_1)\phi'_{ij}(x_2) = \phi'(x_1)^T \phi'(x_2)$$

بنابراین k_4 طبق تعریف هسته‌ای معتبر است.

۳. اگر k یک هسته معتبر باشد به ازای یک اسکالر مثبت c ، ck نیز یک هسته معتبر است.

$$k'(x_1, x_2) = ck(x_1.x_2) = c\phi(x_1)^T \phi(x_2) = \sqrt{c}\phi(x_1)^T \sqrt{c}\phi(x_2) = \phi'(x_1)^T \phi'(x_2)$$

همچنین با استقرا بر روی بخش ب، اگر k هسته معتبر باشد k^n نیز هسته‌ای معتبر است.

می دانیم بسط تیلور e^x به صورت زیر است :

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ...$$

بنابراین می‌توان برای هسته k_1 اینطور نوشت:

$$e^{k_1(x_1,x_2)} = \sum_{n=0}^{\infty} \frac{k_1(x_1, x_2)^n}{n!}.$$

همان‌طور که گفتیم، $k_1(x_1, x_2)^n$ یک هسته معتبر است. ضرب یک عدد مثبت در هسته نیز خود یک هسته معتبر است. بنابراین $\frac{k_1(x_1,x_2)^n}{n!}$ هسته‌ای معتبر است. از طرفی حاصل جمع هسته‌های معتبر خود یک هسته معتبر است (بخش ۱). در نتیجه عبارت نهایی، هسته‌ای معتبر است.

۴. بسط تیلور $\frac{1}{1-\alpha}$ به صورت زیر است:

$$\frac{1}{1-\alpha} = 1 + \alpha + \alpha^2 + \alpha^3 + ... = \sum_{n=0}^{\infty} \alpha^n$$

همچنین $k(x_1, x_2) = x_1^T x_2$ یک هسته معتبر است. زیرا در صورتی که $\phi(x) = x$ آن‌گاه داریم:

$$k(x_1, x_2) = x_1^T x_2 = \phi(x_1)^T \phi(x_2) \Rightarrow valid\ kernel$$

بنابراین عبارت $\frac{1}{1-k(x_1,x_2)}$ مجموع عباراتی به فرم k^n است که طبق اثبات بخش قبل هسته‌ای معتبر است.

پاسخ سوال ۳ *Properties of Kernels*

ابتدا $g(x+y,x+y)$ را ساده می‌کنیم.

$$g(x+y,x+y)$$

$$= g(x,x+y) + g(y,x+y)$$

$$= g(x+y,x) + g(x+y,y)$$

$$= g(x,x) + g(y,x) + g(x,y) + g(y,y)$$

$$= g(x,x) + 2g(x,y) + g(y,y)$$

مشابه روند بالا $g(x-y,x-y)$ را هم ساده می‌کنیم.

$$g(x-y,x-y)$$

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

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

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

$$= g(x,x) - 2g(x,y) + g(y,y)$$

در نهایت نتایج فوق را در عبارت اصلی جایگزین می‌کنیم.

$$h(x,y) = \frac{1}{4}(g(x,x) + 2g(x,y) + g(y,y) - g(x,x) + 2g(x,y) - g(y,y)) = \frac{1}{4}(4g(x,y)) = g(x,y)$$

از معتبر بودن هسته $g(x,y)$ نتیجه می‌گیریم $h(x,y)$ هم یک هسته معتبر است.

پاسخ سوال ۴ *Kernels Over Sets*

باید نشان دهیم هسته k یک هسته معتبر است. یعنی داریم:

$$k(A,B) = \phi(A)^T \phi(B)$$

$$k : U \times U \Rightarrow \mathbb{R}$$

که در آن U مجموعه جهانی است. تعریف می‌کنیم:

$$M = \cup_{S \in U} P(S)$$

که در آن $P(S)$ مجموعه توانی مجموعه S است.

بنابراین M را اجتماع همه ی زیرمجموعه های همه‌ی مجموعه‌ها در نظر می‌گیریم. M را مجموعه‌ای مرتب در نظر می‌گیریم. یعنی ترتیب اعضا در آن مهم است و به هر عضو آن یک عدد طبیعی نسبت می‌دهیم.

$$M = m_1, m_2, \ldots$$

که در آن هر کدام از m ها زیرمجموعه‌ی یکی از مجموعه‌ها است.

حال تبدیل ϕ را این‌طور تعریف می‌کنیم:

$$\phi : U \Rightarrow \mathbb{R}^{|M|}$$

$$\phi(A)_i = \begin{cases} 0, & m_i \notin A. \\ 1, & m_i \in A. \end{cases}$$

یعنی درآیه متناظر m_i را یک می‌گذاریم در صورتی که m_i یکی از زیرمجموعه‌های A باشد و در غیر این صورت صفر می‌گذاریم. طبیعی است که دقیقا $2^{|A|}$ درآیه از درآیه‌های $\phi(A)$ برابر ۱ است. حال می‌توانیم بنویسیم:

$$\phi(A)^T\phi(B) = \sum_i \phi(A)_i\phi(B)_i = \sum_{i:\ m_i \notin A,B} 0*0 + \sum_{\substack{i:\ m_i \in A \\ m_i \notin B}} 1*0 + \sum_{\substack{i:\ m_i \notin A \\ m_i \in B}} 0*1 + \sum_{\substack{i:\ m_i \in A \\ m_i \in B}} 1*1$$

$$= \sum_{\substack{i:\ m_i \in A \\ m_i \in B}} 1 = 1 * 2^{|A\cap B|} = 2^{|A\cap B|}$$

بنابراین فقط در اندیس‌هایی که زیرمجموعه متناظر آن، هم در A و هم در B باشد نتیجه ۱ است و این یعنی این هسته نشان دهنده تعداد زیرمجموعه‌های مشترک دو مجموعه است.

بنابراین k یک هسته معتبر است:

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

دقت شود تعداد زیرمجموعه‌های هر کدام از مجموعه‌ها شمارا است. بنابراین مجموعه‌ی تمام زیرمجموعه‌ها، یک مجموعه شمارا است و می‌توانیم به هر کدام از زیرمجموعه‌ها یک عدد طبیعی اختصاص دهیم.

پاینده باشید