



دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف

استاد درس: دکتر محمدحسین رهبان

بهار ۱۴۰۰

تمرین پنجم درس یادگیری ماشین

نام و نام خانوادگی: امیر پورمند

شماره دانشجویی: ۹۹۲۱۰۲۵۹

آدرس ایمیل pourmand1376@gmail.com

۱ سوال ۱

۱.۱ الف

خب باید نشان دهیم حذف این شرط تاثیری ندارد. خب یک مثال نقض بزنیم به ازای $\xi_i < 0$ چون در ترم اصلی تابع هدف عبارت ξ_i^2 را داریم میتونیم همیشه مقدار ξ_i را به صفر افزایش دهیم و تابع هدف کمتر میشود و شرط نیز به ازای صفر قابل ارضا خواهد بود پس این شرط برای وقتی که از نرم ۲ تابع استفاده میکنیم لازم نخواهد بود.

۲.۱ ب

ابتدا بازنویسی این تابع را با ضرایب لاگرانژ انجام دهیم:

$$\begin{aligned}\mathcal{L}(w, \xi, b) &= \frac{1}{2}w^T w + \frac{C}{2} \sum_{i=1}^n \xi_i^2 + \sum_{i=1}^n \alpha_i \left(1 - \xi_i - y^{(i)} (w^T x^{(i)} + b) \right) \\ &= \frac{1}{2}w^T w + \frac{C}{2} \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i \left(\xi_i + y^{(i)} (w^T x^{(i)} + b) - 1 \right)\end{aligned}$$

که این عبارت باید نسبت به a_i ها ماکزیمم شود و بعد از آن از کل عبارت مینیمم گرفته شود البته یک شرط داریم که $a_i \geq 0$ باشد.

حال باید minmax را به maxmin تبدیل کنیم که چون شرایط KKT برقرار است قضیه اوکی است. یک سری مشتق بگیریم ببینیم چه میشود!

$$\nabla_w \mathcal{L} = w - \sum_{i=1}^n \alpha_i y^{(i)} x^{(i)} = 0 \Rightarrow w = \sum_{i=1}^n \alpha_i y^{(i)} x^{(i)}$$

$$\nabla_{\xi} \mathcal{L} = C\xi_i - \alpha_i = 0 \Rightarrow C\xi_i = \alpha_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^n \alpha_i y^{(i)} = 0$$

۳.۱ ب

فرم دوگان از جایگذاری مشتق های بالای بدست آمده در عبارت لاگرانژین بدست می اید که قدری کار جبر است!

$$\begin{aligned}
 f(a) &= \frac{1}{2} \left[\sum_{i=1}^n \alpha_i y^{(i)} x^{(i)} \right]^T \left[\sum_{i=1}^n \alpha_i y^{(i)} x^{(i)} \right] + \\
 &\frac{C}{2} \sum_{i=1}^n \left(\frac{a_i}{C} \right)^2 - \sum_{i=1}^n \alpha_i \left(\frac{a_i}{C} + y^{(i)} \left(\left[\sum_{j=1}^n \alpha_j y^{(j)} x^{(j)} \right]^T x^{(i)} + b \right) - 1 \right) \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)} \right)^T x^{(j)} + \\
 &\frac{1}{2C} \sum_{i=1}^n (\alpha_i)^2 - \sum_{i=1}^n \alpha_i \frac{a_i}{C} - \sum_{i=1}^n \alpha_i y^{(i)} \left[\sum_{j=1}^n \alpha_j y^{(j)} x^{(j)} \right]^T x^{(i)} \\
 &\quad - \sum_{i=1}^n \alpha_i y^{(i)} b + \sum_{i=1}^n \alpha_i \\
 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)} \right)^T x^{(j)} \\
 &\quad - \sum_{i=1}^n \sum_{j=1}^n \alpha_i y^{(i)} \alpha_j y^{(j)} \left(x^{(j)} \right)^T x^{(i)} \\
 &\quad + \sum_{i=1}^n \alpha_i - \frac{1}{2C} \sum_{i=1}^n (\alpha_i)^2 \\
 &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)} \right)^T x^{(j)} - \frac{1}{2C} \sum_{i=1}^n (\alpha_i)^2
 \end{aligned}$$

پس فرم دوگان مسئله ما میشود:

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)} \right)^T x^{(j)} - \frac{1}{2C} \sum_{i=1}^n (\alpha_i)^2$$

به طوری که دو شرط زیر برقرار باشد:

$$a_i \geq 0, \forall 1 \leq i \leq n$$

$$\sum_{i=1}^n a_i y^{(i)}$$

۴.۱ ت

در مسئله svm دو هدف وجود دارند که گاهی متناقض هستند و پارامتر C برای نوعی تعادل بین این دو بوجود آمده است. هدف اول این است که یک مارجین تا حد زیاد خیلی بزرگ فراهم شود و هدف دوم اینه که نقاط با الگوریتم به خوبی جدا و تفکیک شوند. C در واقع برای تنظیم هدف دوم است. اگر مقدار C خیلی زیاد باشد یعنی هدف دوم به شدت مهم است و این به این معناست که تفکیک پذیری نقاط برای ما خیلی مهم تر از اندازه مارجین است که نتیجتاً به حالتی میرسیم که همه یا اکثر نقاط به خوبی تفکیک شده اند اما مارجین خیلی زیادی نداریم. به صورت بالعکس اگر مقدار C خیلی کم باشد یعنی که اندازه مارجین برای ما خیلی مهم تر از تفکیک پذیری نقاط است و نهایتاً به چیزی میرسیم که ممکن است خیلی از نقاط حتی اشتباه دسته بندی شده باشند اما مارجین خوبی داریم. اگر بخواهم مسئله را بهتر توضیح دهم وقتی پارامتر C خیلی زیاد باشد ترم اول عبارت در سوال مجبور است خیلی کم شود که همه نقاط خوب دسته بندی شوند و برعکس هم مشخص است.

۲ سوال ۲

۱.۲ ۱

خب اگر بتوانیم ثابت کنیم که ماتریس معادل این کرنل همواره PSD است مشخص میشود که کرنل معتبر است. اگر کرنل K_1 ماتریس کرنل A را داشته باشد و کرنل K_2 ماتریس B را داشته باشد کرنل جمع آنها معادل جمع دو کرنل ماتریس این دو هست. پس داریم

$$a^T(A+B)a = a^T Aa + a^T Ba \geq 0, \forall a \in \mathbb{R}^N \quad (۱)$$

و میدانیم نامساوی آخر به این علت درست است که برای هر دو کرنل K_1 و K_2 داریم:

$$a^T Aa \geq 0, \forall a \in \mathbb{R}^N \quad (۲)$$

$$a^T Ba \geq 0, \forall a \in \mathbb{R}^N \quad (۳)$$

پس ثابت کردیم همواره جمع دو کرنل یک کرنل خواهد بود.

۲ ۲.۲

خب برای این سوال یک روش دیگه در پیش میگیریم. اگر بتونیم ثابت کنیم یه فضایی وجود داره که این ضرب داخلی داره در اون فضا انجام میشه یعنی که کرنل ما، یک کرنل معتبره. ابتدا طبق فرضمون باید دو تا فضا وجود داشته باشه که کرنل های K_1 و K_2 در اون فضاها ضرب داخلی بشن. پس داریم:

$$K_1(x, x') = q(x)^T q(x') = \sum_{i=1}^Q q_i q'_i \quad (۴)$$

$$K_2(x, x') = w(x)^T w(x') = \sum_{i=1}^W w_i w'_i \quad (۵)$$

حال داریم:

$$\begin{aligned} K_4(x, x') &= \sum_{i=1}^Q q_i q'_i \times \sum_{i=1}^W w_i w'_i \\ &= \sum_{i=1}^Q \sum_{j=1}^W q_i q'_i w_j w'_j \\ &= \sum_{i=1}^Q \sum_{j=1}^W (q_i w_j)(q'_i w'_j) \\ &= \sum_{i=1}^Q \sum_{j=1}^W r_{ij}(x) r_{ij}(x') \\ &= r(x)^T r(x') \end{aligned} \quad (۶)$$

پس انگار داریم در ماتریس r که یک ماتریس $Q \times W$ بعدی است تبدیل انجام میدهیم.

۳ ۳.۲

راهی جز بسط تیلور برای ما باقی نگذاشته اید! میدانیم:

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots \quad (۷)$$

حال جایگذاری کنیم:

$$e^x = 1 + k(x, x') + \frac{k(x, x')^2}{2} + \frac{k(x, x')^3}{3!} + \dots \quad (۸)$$

حال مشخص است که این کرنل معادل جمع بی نهایت ترم و ضرب کرنل های معتبر در خودشان است که در دو مثال قبل ثابت کردیم هر دو درست هستند.

تنها چیزی که قبلا ثابت نشده است ضرب یک عدد ثابت در کرنل است که باز نیز کرنل را معتبر نگه میدارد زیرا اگر طبق ۱.۲ ماتریس PSD معادل آن را بنویسیم صرفا در یک عدد مثبت ضرب شده است و باعث میشود که عبارت مثبت باقی بماند البته شرط مثبت بودن ضرایب پایرجاست و در اینجا هم رعایت شده است.

میتوان ثابت کرد این کرنل معتبر نیست. هر تابع کرنل معتبر تمام $R^d \times R^d$ را به R مپ میکند اما این تابع تعداد زیادی از بردارها که حاصل ضرب نقطه ای آنها ۱ یک را نمی تواند مپ کند و در نتیجه کرنل معتبری نیست. برای مثال اگر فرض کنیم x_1 و x_2 بعدی باشند و مثلاً هر دو مقدار ۱ را داشته باشند عبارت داخل پرانتز صفر میشود و معکوس آن تعریف نشده است پس به همین علت نمیتواند کرنل معتبری باشد. تا اینجا فرض خاصی نکرده بودم اما اگر فرض کنم که به نوعی $|K(x_1, x_2)| < 1$ است آنوقت این کرنل معتبر خواهد بود علت آن نیز با توجه به سری هندسی است. میدانیم:

$$1 + q + q^2 + q^3 + \dots = \frac{1}{1 - q}$$

پس داریم:

$$\frac{1}{1 - K(x_1, x_2)} = 1 + K(x_1, x_2) + K(x_1, x_2)^2 + \dots$$

که این کرنل یک کرنل معتبر است زیرا جمع یک سری کرنل معتبر می باشد و مشکلی ندارد.

۳ سوال ۳

ابتدا با استفاده از خاصیت متقارن بودن و افزایشی بر حسب ارگومان اول یعنی دو خاصیت اول میتوان نتیجه گرفت که برحسب ارگومان دوم نیز خاصیت دوم را داریم. یعنی:

$$f(x, z + y) = f(z + y, x) = f(z, x) + f(y, x) = f(x, z) + f(x, y) \quad (۹)$$

و همین نتیجه را برای سومی نیز داریم:

$$f(x, ay) = f(ay, x) = af(y, x) = af(x, y) \quad (۱۰)$$

ابتدا معادل صفر رو با توجه به خواص بدست بیارم که تکلیف منفی مشخص بشه!

$$0 = f(0, y) = f(-x + x, y) = f(-x, y) + f(x, y) \Rightarrow f(-x, y) = -f(x, y) \quad (۱۱)$$

$$0 = f(x, 0) = f(x, -y + y) = f(x, y) + f(x, -y) \Rightarrow f(x, -y) = -f(x, y) \quad (۱۲)$$

خب پیام که کرنل رو ساده کنم ببینم چی میشه:

$$\begin{aligned} h(x, y) &= \frac{1}{4} [g(x + y, x + y) - g(x - y, x - y)] \\ &= \frac{1}{4} [g(x, x) + 2g(x, y) + g(y, y) - [g(x, x) + 2g(x, -y) + g(-y, -y)]] \\ &= \frac{1}{4} (g(y, y) + 2g(x, y) - [g(-y, -y) + 2g(x, -y)]) \\ &= \frac{1}{4} [g(y, y) + 2g(x, y) - g(-y, -y) - 2g(x, -y)] \\ &= \frac{1}{4} [g(y, y) + 2g(x, y) - g(y, y) + 2g(x, y)] \\ &= \frac{1}{4} (4g(x, y)) \\ &= g(x, y) \end{aligned} \quad (۱۳)$$

پس اثبات کردیم این کرنل h همان کرنل g هست و حتما یک کرنل معتبر است.

۴ سوال ۴

خب باید ثابت کنیم این کرنل یک ضرب داخلی در یک فضایی را دارد انجام می‌دهد که باید فضا را تعریف کنیم.
 خب فضای Z را زیر مجموعه ای از فضای مرجع M در نظر می‌گیریم که مشخصاً تعداد ابعاد در این فضا برابر $2^{|M|}$ است.
 حال به ازای $W \subseteq Z$ میتوانیم فضا به این صورت تعریف کنیم:

$$\phi_W(Z) = \begin{cases} 1 & W \subseteq Z \\ 0 & otherwise \end{cases}$$

حال در این فضا یک اتفاق جالب توجه داریم. اگر $\phi(M)$ را در نظر بگیریم اولاً به تعداد $2^{|M|}$ عضو دارد ثانیاً به ازای هر زیرمجموعه ای مانند Z به ازای ابعادی که زیرمجموعه Z یک عضو در آن دارد این زیرفضا عدد یک را نشان می‌دهد و در غیراینصورت عدد صفر را.

حال به مسئله اصلی برگردیم دو مجموعه A و B مشخصاً هر دو زیرمجموعه مجموعه مرجع M خواهند بود به ازای آن عضوهایی که در دو مجموعه مشترک باشند عدد ۱ را در هر دو فضا داریم و به ازای عضوهایی که حداقل در یکی از دو مجموعه نباشد در یکی از فضاها عدد ۰ را داریم که ضرب آن مشخصاً صفر میشود. حال با جمع مقادیر ۱ در آنجاها که اشتراکی وجود دارد در واقع ضرب داخلی دو فضا بدست می‌آید که این دقیقاً معادل تعداد زیرمجموعه مشترک مجموعه A و B است.
 علت آن امر نیز آن است که تعداد یک ها در فضایی که ضرب داخلی انجام میشود برابر تعداد زیرمجموعه های مشترک این دو است که تعداد زیرمجموعه های مشترک را میتوان از دو راه بدست آورد: یکی اشتراک گرفتن دو مجموعه و سپس حساب کردن تعداد اعضای مجموعه توانی که این کار در کرنل انجام شده است و دیگری در همان اول کل مجموعه توانی A و B را در نظر گرفته و سپس اشتراک های آن ها را جمع زده است که جواب آن مشخصاً یکی است.