تمرین ۱ درس یادگیری ماشین

آریا جلالی
۹۸۱۰۵۶۶۵

۱ جبرخطی

۱.۱ Derivative

الف

از تعریف مشتق جزیی داریم

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(\mathbf{a} + h.e_i) - f(\mathbf{a})}{h}$$

$$\begin{aligned} f(x+h) = (x+h)^T A(x+h) &= (x^T A + h^T A)(x+h) \\ &= x^T Ah + x^T Ax + h^T Ax + h^T Ah \\ &= f(x) + x^T Ah + h^T Ax + h^T Ah \end{aligned}$$

$$\lim_{\|h\| \to 0} \frac{f(x+h) - f(x) - x^t(A+A^T)h}{\|h\|} = 0$$

دقت کنید عبارت $h^T Ah$ عبارتی از $\mathcal{O}(|h^2|)$ است و میتوانیم از آن در صورت کسر صرف نظر کنیم.
با توجه به تعریف مشتق در $\mathbf{R^n}$ نتیجه میشود مشتق $x^T Ax$ برابر با $x^T(A+A^T)$ است که تنها در صورت متقارن بودن ماتریس A برابر با $2x^T A$ میشود.

ب

۲ حالت را در نظر میگیریم

$x \in \mathcal{R}^n$

$$trace(x^T Ax) = x^T Ax \to \frac{\partial f}{\partial x} = x^T(A+A^T)$$

در واقع اگر x یک بردار باشد، بخش ب حالت کلی بخش الف است و فرض متقارن بودن A در آن برقرار نیست.

$x \in \mathcal{M}_{nm}$

$$\begin{aligned} [AX]_{ik} = \sum_j A_{ij} X_{jk} \Rightarrow [X^T AX]_{il} = \sum_j X^T_{ij} [AX]_{jl} &= \sum_j X^T_{ij} \sum_k A_{jk} X_{kl} \\ = \sum_j \sum_k X^T_{ij} A_{jk} X_{kl} &= \sum_j \sum_k X_{ji} A_{jk} X_{kl} \end{aligned}$$

$$trace(X^TAX) = \sum_c \sum_j \sum_k X_{jc}A_{jk}X_{kc}$$

$$[\frac{\partial trace(X^TAX)}{X}]_{ab} = \frac{\partial}{\partial x_{ab}} \sum_c \sum_j \sum_k X_{jc}A_{jk}X_{kc}$$

برای اینکه جملات بالا مشتق نباشد، باید $c = b$ پس c را فیکس می‌کنیم.

$$[\frac{\partial trace(X^TAX)}{X}]_{ab} = \frac{\partial}{\partial x_{ab}} \sum_j \sum_k X_{jb}A_{jk}X_{kb}$$

دو حالت $j = a$ و $k = a$ را در نظر می‌گیریم و فعلا حالت $k = j = a$ را بررسی نمی‌کنیم.

$$j = a \rightarrow \frac{\partial}{\partial x_{ab}} \sum_{k \neq a} X_{ab}A_{ak}X_{kb} = \sum_{k \neq a} A_{ak}X_{kb}$$

$$k = a \rightarrow \frac{\partial}{\partial x_{ab}} \sum_{j \neq a} X_{jb}A_{ja}X_{ab} = \sum_{j \neq a} X_{jb}A_{ja}$$

$$k = j = a \rightarrow \frac{\partial}{\partial x_{ab}} X_{ab}^2A_{ab} = 2X_{ab}A_{ab}$$

با جمع کردن داریم

$$\begin{aligned} 2X_{ab}A_{ab} + \sum_{j \neq a} X_{jb}A_{ja} + \sum_{k \neq a} A_{ak}X_{kb} &= \sum_j X_{jb}A_{ja} + \sum_k A_{ak}X_{kb} \\ &= \sum_j A_{aj}^TX_{jb} + \sum_k A_{ak}X_{kb} = [A^TX]_{ab} + [AX]_{ab} \end{aligned}$$

و چون شرطی روی a و b نداشته بودیم، می‌توانیم بنویسیم

$$[\frac{\partial trace(X^TAX)}{X}] = A^TX + AX = (A^T + A)X$$

دقت کنید نتیجه‌ی بالا طبق قرارداد Denominator نوشته شده است و اگر آن را طبق قرارداد Numerator بنویسیم، جواب برابر با $X^T(A + A^T)$ خواهد بود، که خواسته‌ی سوال است.

۲.۱ Eigenvalue and Eigenvectors

طبق تعریف میدانیم رابطه مشخصه برای ماتریس A از رابطه $det(tI-A)$ بدست می‌آید.
حال داریم

$$det(tI-A)=\begin{vmatrix} t-a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & t-a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & t-a_{nn} \end{vmatrix}$$

از جبرخطی میدانیم معادله‌ی بالا چندجمله‌ای با درجه‌ی n است و مقادیر ویژه‌ی ماتریس A ریشه‌های آن هستند. پس می‌توانیم طبق قضیه‌ی فاکتور کردن چندجمله‌ای آن را به صورت مقابل نیز بازنویسی کنیم.

$$p_A(t)=C(t-\lambda_1)\dots(t-\lambda_n)$$

دقت کنید ضریب جمله‌ی توان n باید ۱ باشد، زیرا اگر از قانون لاپلاس برای محاسبه دترمینان استفاده کنیم داریم

$$det(tI-A)=(t-a_{11})\times M_{11}+\mathcal{O}(t^{n-2})$$

$$M_{11}=(t-a_{22})\times M_{11_{M_{22}}}+\mathcal{O}(t^{n-3})$$

یعنی ضرایب جملات t^{n-1} و t^n از عبارت $(t-a_{11})*(t-a_{22})*\dots*(t-a_{nn})$ بدست می‌آید. که $C=1$ را نشان می‌دهد.
حال با برابر قرار دادن ضرایب جمله‌ی t^{n-1} در ۲ رابطه به نتیجه‌ی مقابل می‌رسیم

$$\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}=a_{ii}=trace(A)$$

با ۰ قرار دادن t داریم

$$det(-A)=(-1)^n\prod_{i=1}^{n}\lambda_i\rightarrow(-1)^n det(A)=(-1)^n\prod_{i=1}^{n}\lambda_i\rightarrow det(A)=\prod_{i=1}^{n}\lambda_i$$

۳.۱ Rank

فرض کنید v بردار ویژه‌ی ماتریس M متناظر با مقدار ویژه‌ی λ_1 باشد داریم

$$B=P^{-1}MP\rightarrow PBP^{-1}=M\rightarrow PBP^{-1}v=Mv=\lambda_1 v\rightarrow BP^{-1}v=\lambda_1 P^{-1}v$$

یعنی تمام مقادیر ویژه‌های ماتریس M مقادیر ویژه‌ی ماتریس $B=P^{-1}MP$ هستند و از طرفی چون می‌توانیم بنویسیم $PBP^{-1}=S^{-1}BS=M$ نتیجه می‌گیریم تمام مقادیر ویژه‌های ماتریس $B=P^{-1}MP$ مقادیر ویژه‌ی ماتریس M نیز هستند که برابری مقادیر ویژه را نشان می‌دهد.

۴.۱ Norms

نرم‌های $l1$ و $l\infty$ در کلاس توضیح داده شدند و برابر هستند با ماکسیمم جمع قدرمطلق اعضای هر ستون و سطر (به ترتیب) داریم

$$l1(A) = max_i \sum_{j=1}^{n} a_{ij} = |5| + |-1| + |-2| = 8$$

$$l\infty(A) = max_i \sum_{j=1}^{n} a_{ji4} = |5| + |-4| + |2| = 11$$

حال برای $l2$ داریم

$$l2(A)^2 = max_v \frac{\|Av\|_2}{\|v\|_2} = max_v \frac{v^T A^T A v}{v^T v} = \frac{v^T M v}{v^T v} \quad M = A^T A$$

کسر نهایی همان Rayleigh quotient است که می‌دانیم مقدار ماکسیمم آن برابر با بزرگترین مقدار ویژه ماتریس متقارن $M = A^T A$ است. حال برای بدست آوردن مقادیر ویژه‌ی آن داریم

$$A^T A = \begin{bmatrix} 30 & -24 & 7 \\ -24 & 21 & -2 \\ 7 & -2 & 13 \end{bmatrix} \rightarrow l2(A) = \sqrt{\lambda_{max}} \approx 7.147$$

۲ احتمال

۱.۲ Expectation

با استفاده از قانون lotus داریم

$$\mathbb{E}(X^3 + 2X - 7) = \int_1^2 (x^3 + 2x - 7)(2x - 2)dx$$

$$= \int_1^2 (2x^4 + 4x^2 - 18x - 2x^3 + 14)dx = \frac{2}{5} \times 31 + \frac{4}{3} \times 7 - 27 + 14 - \frac{1}{2} \times 15 = 1.2\bar{3}$$

۲.۲ Function of random variables

max

ابتدا $F_{Y_1}(y)$ را بدست می‌آوریم و با مشتق گرفتن از آن به $f_{Y_1}(y)$ می‌رسیم

$$P(Y_1 \leq y) = P(X_1 \leq y, \ldots, X_n \leq y) = \prod_{i=1}^{n} P(X_i \leq y) = \prod_{i=1}^{n} F_X(y) = F_X(y)^n$$

$$f_{Y_1}(y) = \frac{d}{dy}P(Y_1 \leq y) = nf_X(y)F_X(y)^{n-1}$$

min

همانند بخش قبل عمل میکنیم

$$1 - F_{Y_2}(y) = P(Y_2 > y) = \prod_{i=1}^{n}(1 - F_X(y)) = (1 - F_X(y))^n$$
$$\rightarrow F_{Y_2}(y) = 1 - (1 - F_X(y))^n$$

$$f_{Y_2}(y) = \frac{d}{dy}P(Y_2 \leq y) = nf_X(y)(1 - F_X(y))^{n-1}$$

۳.۲ Estimation

الف

برای توزیع joint با توجه به رابطه‌ی قانون شرطی داریم.

$$f_{X,Y,\mu}(t,x,y) = f_{X,Y|\mu}(x,y|\mu = t) \times f_\mu(t)$$

دقت کنید متغیرهای تصادفی X و Y مستقل هستند و می‌توانیم احتمال شرطی بالا را جدا کنیم و بنویسیم:

$$f_{X,Y,\mu}(t,x,y) = f_{X|\mu}(x|\mu = t) \times f_{Y|\mu}(y|\mu = t) \times f_\mu(t)$$

با استفاده از روابط pdf داریم:

$$f_{X,Y,\mu}(t,x,y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t)^2} \times \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-t)^2} \times 1$$

در نهایت توزیع joint ما به صورت مقابل خواهد بود:

$$f_{X,Y,\mu}(t,x,y) = \begin{cases} \frac{1}{2\pi}e^{-\frac{(x-t)^2+(y-t)^2}{2}} & 0 \leq t \leq 1 \\ 0 & O.w \end{cases}$$

ب:

برای پیدا کردن تخمین MAP نیاز به تعدادی Sample از توزیع‌های X و Y داریم و در نهایت عبارت مقابل را ماکسیمم کنیم:

$$f(\mu|D) \propto f(D|\mu)f(\mu)$$

حال با استفاده از نتایج بخش الف می‌توانیم بنویسیم:

$$f(D|\mu)f(\mu) = \prod_{i=1}^{n} \frac{1}{2\pi} e^{-\frac{(x_i-\mu)^2+(y_i-\mu)^2}{2}} \qquad 0 \leq \mu \leq 1$$

برای ماکسیمم کردن عبارت بالا نیاز به ضریب $\frac{1}{2\pi}$ نداریم و می‌توانیم تابع لگاریتم آن را ماکسیمم کنیم.

$$log(f(D|\mu)f(\mu)) \propto \sum_{i=1}^{n} -((x_i-\mu)^2 + (y_i-\mu)^2) = g(\mu)$$

با مشتق گرفتن نسبت به μ خواهیم داشت:

$$\frac{dg}{d\mu} = \sum_{i=1}^{n} 2(x_i + y_i - 2\mu) = 0 \equiv \sum_{i=1}^{n} (x_i + y_i - 2\mu) = 0$$

$$\Rightarrow \mu_{MAP} = \frac{\sum_{i=1}^{n} x_i + y_i}{2n}$$

دقت کنید μ حتما باید بین 0 و 1 باشد و در صورتی که رابطه‌ی بالا از 0 کمتر و یا از 1 بیشتر شود خواهیم داشت:

$$\mu_{MAP} = \begin{cases} 1 & \frac{\sum_{i=1}^{n} x_i+y_i}{2n} > 1 \\ 0 & \frac{\sum_{i=1}^{n} x_i+y_i}{2n} < 0 \\ \frac{\sum_{i=1}^{n} x_i+y_i}{2n} & O.w \end{cases}$$