یادگیری ماشین

نیم‌سال اول ۰۱-۰۰

مدرس: سید عباس حسینی

دانشکده مهندسی کامپیوتر

تمرین اول

محمد جواد هزاره — شماره دانشجویی: ۹۸۱۰۱۰۷۴

۱ مقدار ویژه

۱. می‌دانیم که ماتریس A را می‌توان به صورت $S\Lambda S^{-1}$ نوشت که ماتریس Λ ماتریسی‌ست که مقدار ویژه‌های A روی قطر آن قرار گرفته‌اند و مابقی درایه‌های آن صفر است. S نیز شامل بردار ویژه‌های متناظر با مقدار ویژه‌های روی قطر Λ است که این بردارها به ترتیب قرار گرفتن مقدار ویژه‌ها، در ستون‌های S قرار گرفته‌اند. بنابراین داریم:

$$\begin{aligned}
trace(A) &= trace(S\Lambda S^{-1}) \\
&= trace(\Lambda S S^{-1}) \qquad (trace(AB) = trace(BA)) \\
&= trace(\Lambda) \\
&= \lambda_1 + \lambda_2 + \cdots + \lambda_n.
\end{aligned}$$

۲. به طور مشابه قسمت اول، ماتریس A را به صورت $S\Lambda S^{-1}$ می‌نویسیم، بنابراین:

$$\begin{aligned}
\det(A) &= \det(S\Lambda S^{-1}) \\
&= \det(S).\det(\Lambda).\det(S^{-1}) \qquad (\det(AB) = \det(A)\det(B)) \\
&= \det(\Lambda) \qquad \left(\det(A) = \frac{1}{\det(A^{-1})}\right) \\
&= \lambda_1\lambda_2\cdots\lambda_n.
\end{aligned}$$

۳. اگر $\lambda = 0$ یکی از مقدار ویژه‌های AB یا BA باشد، مشخصا مقدار ویژه دیگری نیز خواهد بود. بنابراین فرض کنیم $\lambda \neq 0$. اگر λ مقدار ویژه AB باشد، داریم:

$$ABx = \lambda x$$

$$B \times ABx = B \times \lambda x$$

$$BA(Bx) = \lambda(Bx)$$

بنابراین λ مقدار ویژه BA نیز خواهد بود. اگر λ مقدار ویژه BA باشد، به طور مشابه داریم:

$$
\begin{aligned}
BAx &= \lambda x \\
A \times BAx &= A \times \lambda x \\
AB(Ax) &= \lambda(Ax)
\end{aligned}
$$

بنابراین λ مقدار ویژه AB نیز است. پس از آن‌جایی که هر مقدار ویژه AB مقدار ویژه BA هم هست و بلعکس، این دو ماتریس مجموعه مقدار ویژه‌های یکسانی دارند.

۴. اگر λ مقدار ویژه ماتریس A باشد، داریم $\det(A - \lambda I) = 0$. نشان می‌دهیم این شرط برای ماتریس A^T نیز برقرار است. داریم:

$$
\begin{aligned}
\det(A^T - \lambda I) &= \det\left((A - \lambda I)^T\right) && ((A+B)^T = A^T + B^T, I^T = I) \\
&= \det(A - \lambda I) && (\det(A^T) = \det(A)) \\
&= 0.
\end{aligned}
$$

در نتیجه λ مقدار ویژه A^T نیز هست. به طور مشابه می‌توان نشان داد که هر مقدار ویژه A^T نیز مقدار ویژه A است و درنتیجه این دو ماتریس مجموعه مقدار ویژه‌های یکسانی دارند.

$\square$

۲ کوواریانس و امید ریاضی

۱. با استفاده از رابطه $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ برای جمله اول سمت راست تساوی داریم:

$$
\begin{aligned}
\mathrm{Var}(\mathbb{E}[X|Y]) &= \mathbb{E}\left[\mathbb{E}[X|Y]^2\right] - \mathbb{E}\left[\mathbb{E}[X|Y]\right]^2 \\
&= \mathbb{E}\left[\mathbb{E}[X|Y]^2\right] - \mathbb{E}[X]^2
\end{aligned}
$$

با استفاده از رابطه $\mathrm{Var}(X|Y) = \mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2$ برای جمله دوم سمت راست تساوی داریم:

$$
\begin{aligned}
\mathbb{E}[\mathrm{Var}(X|Y)] &= \mathbb{E}\left[\mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2\right] \\
&= \mathbb{E}\left[\mathbb{E}[X^2|Y]\right] - \mathbb{E}\left[\mathbb{E}[X|Y]^2\right] \\
&= \mathbb{E}[X^2] - \mathbb{E}\left[\mathbb{E}[X|Y]^2\right]
\end{aligned}
$$

بنابراین برای جمع این عبارت‌ها داریم:

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[X|Y]) + \mathbb{E}[\mathrm{Var}(X|Y)] &= \left(\mathbb{E}\left[\mathbb{E}[X|Y]^2\right] - \mathbb{E}[X]^2\right) + \left(\mathbb{E}[X^2] - \mathbb{E}\left[\mathbb{E}[X|Y]^2\right]\right) \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \boxed{\mathrm{Var}(X)}\end{aligned}$$

۲. با استفاده از رابطه داده شده، سمت چپ تساوی را باز می‌کنیم:

$$\begin{aligned}\mathrm{Cov}(X,Y|Z) &= \mathbb{E}\left[(X-\mathbb{E}[X|Z])(Y-\mathbb{E}[Y|Z])|Z\right] \\ &= \mathbb{E}\left[(XY - X\mathbb{E}[Y|Z] - Y\mathbb{E}[X|Z] + \mathbb{E}[X|Z]\mathbb{E}[Y|Z])\,|Z\right] \\ &= \mathbb{E}[XY|Z] - \mathbb{E}\left[(X\mathbb{E}[Y|Z])|Z\right] - \mathbb{E}\left[(Y\mathbb{E}[X|Z])|Z\right] + \mathbb{E}\left[(\mathbb{E}[X|Z]\mathbb{E}[Y|Z])|Z\right] \\ &= \mathbb{E}[XY|Z] - \mathbb{E}[X|Z]\mathbb{E}[Y|Z] - \mathbb{E}[Y|Z]\mathbb{E}[X|Z] + \mathbb{E}[X|Z]\mathbb{E}[Y|Z] \\ &= \boxed{\mathbb{E}[XY|Z] - \mathbb{E}[X|Z]\mathbb{E}[Y|Z]}\end{aligned}$$

۳. با استفاده از قسمت دوم، برای جمله اول سمت راست تساوی داریم:

$$\begin{aligned}\mathbb{E}\left[\mathrm{Cov}(X,Y|Z)\right] &= \mathbb{E}\left[\mathbb{E}[XY|Z] - \mathbb{E}[X|Z]\mathbb{E}[Y|Z]\right] \\ &= \mathbb{E}\left[\mathbb{E}[XY|Z]\right] - \mathbb{E}\left[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]\right] \\ &= \mathbb{E}[XY] - \mathbb{E}\left[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]\right] \qquad (*)\end{aligned}$$

برای جمله دوم سمت راست تساوی از رابطه مشابه برای کوواریانس غیر شرطی استفاده می‌کنیم:

$$\begin{aligned}\mathrm{Cov}\left(\mathbb{E}[X|Z], \mathbb{E}[Y|Z]\right) &= \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] - \mathbb{E}[\mathbb{E}[X|Z]]\mathbb{E}[\mathbb{E}[Y|Z]] \\ &= \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] - \mathbb{E}[X]\mathbb{E}[Y] \qquad (**)\end{aligned}$$

بنابراین با توجه به $(*)$ و $(**)$ داریم:

$$\begin{aligned}\mathbb{E}\left[\mathrm{Cov}(X,Y|Z)\right] + \mathrm{Cov}\left(\mathbb{E}[X|Z], \mathbb{E}[Y|Z]\right) &= \mathbb{E}[XY] - \mathbb{E}\left[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]\right] + \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \boxed{\mathrm{Cov}(X,Y)}\end{aligned}$$

$\square$

۳ مشتق ماتریس

۱. برای عبارت داده شده داریم:

$$x^T Ax = \sum_{i,j} A_{ij} x_i x_j$$

بنابراین برای درایه‌های مشتق داریم:

$$\begin{aligned}
\frac{\partial}{\partial x_m}(x^T Ax) &= \frac{\partial}{\partial x_m} \sum_{i,j} A_{ij} x_i x_j \\
&= \frac{\partial}{\partial x_m} \left(\sum_{j \neq m} A_{mj} x_m x_j + \sum_{i \neq m} A_{im} x_i x_m + A_{mm} x_m x_m + \sum_{o.w.} A_{ij} x_i x_j \right) \\
&= \sum_{j \neq m} A_{mj} x_j + \sum_{i \neq m} A_{im} x_m + 2A_{mm} x_m + 0 \\
&= \left(\sum_{j \neq m} A_{mj} x_j + A_{mm} x_m \right) + \left(\sum_{i \neq m} A_{im} x_i + A_{mm} x_m \right) \\
&= \sum_{j} A_{mj} x_j + \sum_{i} A_{im} x_i \\
&= (Ax)_m + (A^T x)_m
\end{aligned}$$

بنابراین اگر فرض کنیم ماتریس A ماتریسی متقارن است، آنگاه خواهیم داشت:

$$\boxed{\frac{\partial}{\partial x}(x^T Ax) = (A + A^T)x = 2Ax}$$

۲. ابتدا تریس $X^T AX$ را به صورت جمع عناصر می‌نویسیم: (هر جا اندیسی دوبار تکرار شده است، روی آن جمع انجام می‌شود.)

$$\begin{aligned}
(X^T AX)_{ij} &= X^T_{ik} (AX)_{kj} \\
&= X_{ki} (A_{kl} X_{lj}) \\
&= A_{kl} X_{ki} X_{lj}
\end{aligned}$$

بنابراین برای تریس ماتریس داریم:

$$trace(X^T AX) = A_{kl} X_{ki} X_{li}$$

حال برای درایه‌های ماتریس مشتق داریم:

$$\begin{aligned}
\frac{\partial}{\partial X_{mn}} trace(X^T AX) &= \frac{\partial}{\partial X_{mn}} (A_{kl}X_{ki}X_{li}) \\
&= \frac{\partial}{\partial X_{mn}} \left(\sum_{l \neq m} A_{ml}X_{mn}X_{ln} + \sum_{k \neq m} A_{km}X_{kn}X_{mn} + A_{mm}X_{mn}^2 + \sum_{o.w.} A_{kl}X_{ki}X_{li} \right) \\
&= \sum_{l \neq m} A_{ml}X_{ln} + \sum_{k \neq m} A_{km}X_{kn} + 2A_{mm}X_{mn} + 0 \\
&= \left(\sum_{l \neq m} A_{ml}X_{ln} + A_{mm}X_{mn} \right) + \left(\sum_{k \neq m} A_{km}X_{kn} + A_{mm}X_{mn} \right) \\
&= \sum_{l} A_{ml}X_{ln} + \sum_{k} A_{km}X_{kn} \\
&= (AX)_{mn} + \sum_{k} A^T_{mk}X_{kn} \\
&= (AX)_{mn} + (A^T X)_{mn}
\end{aligned}$$

بنابراین برای مشتق داریم:

$$\boxed{\frac{\partial}{\partial X} trace(X^T AX) = (A + A^T)X}$$

$\square$

۴ متغیر تصادفی

نخست CDF را برای $\sqrt{X}$ محاسبه می‌کنیم:

$$\begin{aligned}
\mathbb{P}(\sqrt{X} \leq x) &= \mathbb{P}(\sqrt{X} < 0) + \mathbb{P}(0 \geq \sqrt{X} \leq x) \\
&= 0 + \mathbb{P}(0 \leq X \leq x^2) \\
&= \int_0^{x^2} p_X(t)dt \\
&= x^2.
\end{aligned}$$

که در بازه $[0,1]$ تعریف شده است. بنابراین برای PDF داریم:

$$p_{X^2}(x) = \begin{cases} 2x & x \in [0,1] \\ 0 & o.w. \end{cases}$$

برای X^2 نیز به طور مشابه عمل می‌کنیم:

$$\begin{aligned} \mathbb{P}(X^2 \leq x) &= \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{X}) \\ &= \mathbb{P}(-\sqrt{x} \leq X < 0) + \mathbb{P}(0 \geq X \leq \sqrt{x}) \\ &= 0 + \int_0^{\sqrt{x}} p_X(t)dt \\ &= \sqrt{x}. \end{aligned}$$

که در بازه $[0,1]$ تعریف شده است. بنابراین برای PDF داریم:

$$p_{\sqrt{x}}(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & x \in [0,1] \\ 0 & o.w. \end{cases}$$

□

۵ رنک

اگر λ مقدار ویژه $P^{-1}MP$ باشد، نشان می‌دهیم مقدار ویژه M نیز است. برای این منظور داریم:

$$P^{-1}MPx = \lambda x$$

$$P \times P^{-1}MPx = P \times \lambda x$$

$$M(Px) = \lambda(Px).$$

و چون Px صفر نیست چرا که می‌دانیم $x \neq 0$ و P فول رنک است، λ مقدار ویژه M نیز است.
حال اگر λ را مقدار ویژه M در نظر بگیریم، داریم:

$$Mx = \lambda x$$

$$M(PP^{-1})x = \lambda x$$

$$P^{-1} \times M(PP^{-1})x = P^{-1} \times \lambda x$$

$$P^{-1}MP(P^{-1}x) = \lambda(P^{-1}x).$$

و با استدلالی مشابه قسمت قبل داریم $P^{-1}x \neq 0$ ، پس λ مقدار ویژه $P^{-1}MP$ نیز است.
در نتیجه دو ماتریس M و $P^{-1}MP$ مجموعه مقدار ویژه‌های یکسانی دارند. $\square$

۶ فاکتورگیری ماتریس

۱. از آن‌جایی که A مثبت معین است، تمام مقدار ویژه‌های آن مثبت بوده و در نتیجه حاصل‌ضرب آن‌ها که همان مقدار دترمینان A را مشخص می کند مقداری ناصفر است. بنابراین A ماتریسی وارون‌پذیر خواهد بود. از طرفی می‌دانیم ماتریس‌های وارون‌پذیر تجزیه LU یکتایی دارند به شرطی که همه‌ی عناصر روی قطر ماتریس L را عدد یک نگه داریم. حال فرض کنیم ماتریس A را به صورت LU تجزیه کرده و داریم $A = LU$. همینطور از آنجایی که A ماتریسی متقارن است، برای ترانهاده‌ی آن نیز داریم $A^T = A = LU$. از طرفی با توجه به ترانهاده‌ی حاصل‌ضرب ماتریس‌ها داریم $A = LU \implies A^T = U^T L^T$. بنابراین از آنجایی که تجزیه LU ماتریس A^T نیز یکتاست، داریم:

$$\left.\begin{aligned} A^T &= LU \\ A^T &= U^T L^T \end{aligned}\right\} \implies U = L^T$$

بنابراین ماتریس A را می‌توان به صورت $A = LL^T$ نوشت که با توجه به یکتا بودن L و U این تجزیه نیز یکتاست و عناصر روی قطر L همگی مثبت و برابر یک هستند.

۲. با توجه به متعامد بودن Q داریم $A^TA = R^TQ^TQR = R^TR$ و در نتیجه مقدار تکین‌های R همان مقدار تکین‌های A خواهند بود چرا که این مقادیر مقدار ویژه‌های ماتریس R^TR یا A^TA هستند. بنابراین Σ در تجزیه SVD هر دو ماتریس R و A یکسان خواهد بود. هم‌چنین ماتریس V نیز در هر دو تجزیه یکسان است، چرا که این ماتریس در تجزیه A شامل بردار ویژه‌های A^TA بوده که با R^TR یکسان است. تنها تفاوت تجزیه SVD ماتریس R و A در ماتریس U این تجزیه خواهد بود. این ماتریس برای تجزیه A شامل بردار ویژه‌های AA^T است و برای تجزیه R شامل بردار ویژه‌های RR^T. اگر فرض کنیم x یکی از بردار ویژه‌های AA^T باشد، آنگاه داریم:

$$\begin{gathered} AA^T x = \lambda x \\ QRR^TQ^Tx = \lambda x \\ Q^T \times \Big(QRR^TQ^T x\Big) = Q^T \times \Big(\lambda x\Big) \\ RR^T(Q^Tx) = \lambda(Q^Tx) \end{gathered}$$

بنابراین اگر x یک بردار ویژه AA^T باشد، آنگاه Q^Tx یک بردار ویژه RR^T خواهد بود. در نتیجه اگر ماتریس U در تجزیه A را با U_A نشان دهیم، آنگاه این ماتریس در تجزیه R برابر خواهد بود با Q^TU_A.

پس اگر تجزیه SVD ماتریس A به صورت $U\Sigma V^T$ باشد، تجزیه SVD ماتریس R به صورت $Q^T U\Sigma V^T$ خواهد بود.

$\square$

۷ ماتریس Nilpotent

برای نشان دادن وارون‌پذیر بودن ماتریس $I-A$ وارون آن را ارائه می‌کنیم. در این صورت هم نشان داده شده که ماتریس وارون‌پذیر است و هم وارون آن را پیدا کرده‌ایم. با توجه به این‌که می‌دانیم A^k ماتریس صفر است، سعی می‌کنیم ماتریسی در $I-A$ ضرب کنیم که به عبارت $I-A^k$ برسیم. برای این کار می‌توان مشابه اتحادهای جبری، از ماتریس زیر استفاده کرد:

$$A^{k-1}+A^{k-2}+\cdots+I$$

برای نشان دادن این‌که این ماتریس وارون $I-A$ است داریم:

$$\begin{aligned}(I-A)\times(A^{k-1}+A^{k-2}+\cdots+I) &= (A^{k-1}+A^{k-2}+\cdots+I)-(A^k+A^{k-1}+\cdots+A)\\ &= I-A^k\\ &= I \qquad\qquad (A^k=0)\end{aligned}$$

و اگر این ماتریس را از طرف چپ در $I-A$ ضرب کنیم به نتیجه مشابهی می‌رسیم. بنابراین ماتریس ارائه شده وارون $I-A$ خواهد بود. $\square$

۸ MAP

۱. برای محاسبه توزیع مشترک داریم:

$$\begin{aligned}f_{\mu,X,Y}(t,x,y) &= f_\mu(t).f_{X,Y|\mu}(x,y|t)\\ &= f_\mu(t).f_{X|\mu}(x|t).f_{Y|\mu}(y|t) \qquad\qquad (X \text{ و } Y \text{ مستقل هستند})\\ &= 1\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-t)^2}{2}}\times\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-t)^2}{2}}\\ &= \frac{1}{2\pi}e^{-\frac{(x-t)^2+(y-t)^2}{2}}\end{aligned}$$

۲. برای پیدا کردن تخمین‌گر MAP پارامتر μ با استفاده از داده‌های X و Y داریم:

$$\mu_{MAP} = \arg\max_{\mu} f_{\mu|X,Y} = \arg\max_{\mu} f_{\mu,X,Y}$$

که با استفاده از قسمت یک داریم:

$$\begin{aligned}\frac{d}{d\mu}\ln(f_{\mu,X,Y}) &= \frac{d}{d\mu}\left(-\ln(2\pi) - \frac{(x-\mu)^2+(y-\mu)^2}{2})\right)\\ &= 0 + (x-\mu) + (y-\mu)\end{aligned}$$

$$\Longrightarrow x - \mu_{MAP} + y - \mu_{MAP} = 0$$

$$\Longrightarrow \mu_{MAP} = \frac{x+y}{2}$$

$\square$