



سؤال ۱. مقدار ویژه

- جمع مقادیر ویژه برابر با trace است.

اگر معادله مشخصه ماتریس را بنویسیم $(p(\lambda) = \det(A - \lambda I))$ ، داریم:

$$p(\lambda) = \det(A - \lambda I) = (-1)^n (\lambda^n - (\text{tr} A) \lambda^{n-1} + \dots + (-1)^n \det A).$$

از طرفی می‌دانیم:

$$p(\lambda) = (-1)^n (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$$

λ_j نشان‌دهنده‌ی مقدار ویژه‌ی ماتریس A است؛ بنابراین با مقایسه‌ی ضرایب، این رابطه اثبات می‌شود.

$$\text{tr}(A) = \lambda_1 + \dots + \lambda_n$$

- ضرب مقادیر ویژه برابر با دترمینان ماتریس است.

بار دیگر اگر معادله‌ی مشخصه را بنویسیم و فرض کنیم که λ ریشه‌های این معادله هستند، داریم:

$$\begin{aligned} \det(A - \lambda I) = p(\lambda) &= (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n) \\ &= (-1)(\lambda - \lambda_1)(-1)(\lambda - \lambda_2) \dots (-1)(\lambda - \lambda_n) \\ &= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda) \end{aligned}$$

در معادله‌ی اول، ضرب $(-1)^n$ از فاکتورگیری به دست آمده است. این ضرب را می‌توان به جای فاکتورگیری روی قطر گسترش داد و به معادله‌ی دوم رسید.

حال اگر λ را برابر با صفر قرار دهیم (چون متغیر است)، در سمت چپ معادله‌ی مشخصه $\det(A)$ و در سمت راست $\lambda_1 \dots \lambda_n$ به دست می‌آید، به بیان دیگر:

$$\det(A) = \lambda_1 \dots \lambda_n$$

- مقدار ویژه‌های AB و BA با یک‌دیگر برابرند.

اگر λ مقادیر ویژه‌ی AB باشند و آن را در بردار ناصفر x ضرب کنیم، داریم:

$$ABx = \lambda x$$

حال اگر $y = Bx$ باشد، آن‌گاه y هم ناصفر خواهد بود (در غیر این صورت در معادله‌ی بالا x و یا λ برابر با صفر می‌شوند که خلاف فرضیات ما است). با جای‌گذاری روابط داریم:

$$BAy = BABx = B(ABx) = B(\lambda x) = \lambda Bx = \lambda y$$

همان‌طور که مشاهده می‌شود، ثابت شد که λ مقدار ویژه‌ی BA نیز هست.

سؤال ۲. مشتق‌گیری

هر کدام از روابط زیر را ثابت کنید.

$$\frac{\partial a^t x}{\partial x} = a \quad \bullet$$

$$\frac{\partial a^t x}{\partial x} = \frac{\partial \left(a_1 \quad \dots \quad a_n \right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}{\partial x} = \frac{\partial (a_1 x_1 + \dots + a_n x_n)}{\partial x}$$

اگر $f = (a_1 x_1 + \dots + a_n x_n)$ داریم:

$$\frac{\partial a^t x}{\partial x} = \frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix} = a$$

$$\frac{\partial x^t A x}{\partial x} = 2x^T A \quad \bullet$$

$$y = f(x) = x^T A x = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$$

$$= \sum_{i=1}^n a_{i1} x_i x_1 + \sum_{j=1}^n a_{1j} x_1 x_j + \sum_{i=2}^n \sum_{j=2}^n a_{ij} x_i x_j$$

$$\frac{\partial f}{\partial x_1} = \sum_{i=1}^n a_{i1} x_i + \sum_{j=1}^n a_{1j} x_j$$

از آنجایی که $a_{1i} = a_{i1}$ داریم:

$$= \sum_{i=1}^n a_{1i} x_i + \sum_{i=1}^n a_{1i} x_i$$

$$= 2 \sum_{i=1}^n a_{1i} x_i$$

$$\rightarrow \frac{\partial f}{\partial x} = \begin{pmatrix} 2 \sum_{i=1}^n a_{1i} x_i \\ \vdots \\ 2 \sum_{i=1}^n a_{ni} x_i \end{pmatrix} = 2 \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 2Ax = 2x^T A$$

$$\frac{\partial^2 x^T A x}{\partial x \partial x^T} = A + A^T \quad \bullet$$

با توجه به قاعده‌ی ضرب در بردارها داریم:

$$\frac{\partial u^T v}{\partial x} = u^T \frac{\partial v}{\partial x} + v^T \frac{\partial u}{\partial x}$$

برای راحتی کار در قسمت بعد رابطه‌ی $\frac{\partial Ax}{\partial x} = A$ را در این قسمت ثابت می‌کنیم:

$$s = Ax \rightarrow s_i = \sum_j a_{ij} x_j \rightarrow \frac{\partial s_i}{\partial x_j} = a_{ij} \rightarrow \frac{\partial s}{\partial x} = A$$

حال اگر قاعده‌ی بالا را استفاده کنیم، داریم:

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

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

$$\frac{\partial \text{tr}(X^T A X)}{\partial X} = X^T (A + A^T) \bullet$$

$$(X^T A X)_{ij} = \sum_k \sum_l x_{ki} a_{kl} x_{lj}$$

$$\text{tr}(X^T A X) = \sum_r \sum_k \sum_l x_{kr} a_{kl} x_{lr}$$

$$\frac{\partial \text{tr}(X^T A X)}{\partial x_{ij}} = \sum_l a_{il} x_{lj} + \sum_k x_{kj} a_{ki} = \sum_k x_{kj} a_{ki} + \sum_l x_{lj} a_{il}$$

$$\rightarrow \frac{\partial \text{tr}(X^T A X)}{\partial X} = X^T A + X^T A^T = X^T (A + A^T)$$

سؤال ۳. مقدار ویژه و رنک

• ثابت کنید اگر P یک ماتریس با رنک کامل باشد، ماتریس‌های M و $P^{-1}MP$ مجموعه مقدار ویژه‌ی یکسان دارند.

فرض کنید λ مجموعه مقدار ویژه‌ی $P^{-1}MP$ است. باید نشان دهیم که λ مقدار ویژه‌ی M نیز می‌باشد. حال اگر x بردار ویژه باشد داریم:

$$P^{-1}MPx = \lambda x \rightarrow M(Px) = \lambda Px$$

بنابراین Px یک بردار ویژه با مقدار ویژه‌ی λ است. به طور مشابه طرف دیگر آن اثبات می‌شود. اگر بردار ویژه‌ی x متعلق به M باشد، می‌توان نشان داد که $P^{-1}x$ یک بردار ویژه از $P^{-1}MP$ است:

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

ثابت شد که مقدار ویژه‌ی یکسان دارند.

نکته: از خاصیت رنک کامل بودن برای معکوس‌پذیر بودن P استفاده شده است.

• ثابت کنید که مجموع ابعاد فضای ویژه ماتریس $M_{n \times n}$ نمی‌تواند بیش‌تر از n شود. (M در بیش‌ترین حالت n تا مقدار ویژه‌ی متفاوت دارد.)

برای به دست آوردن مقادیر ویژه‌ی M باید معادله‌ی $\det(M - \lambda I) = 0$ را حل کنیم. چون این معادله، درجه‌ی n دارد، بنابراین حداکثر n مقدار مختلف دارد.

سؤال ۴. معین متقارن

ثابت کنید هر ماتریس مثبت معین متقارن A یک شکل یکتا به فرم $A = LL^T$ دارد که L پایین مثلثی با درایه‌های قطری مثبت است. (چولسکی)
برای اثبات آن از استقرا روی ابعاد آن استفاده می‌کنیم:

- اگر $n = 1$: ماتریس $A_{1 \times 1}$ مثبت معین است؛ اگر و تنها اگر به فرم $A = (a)$ با $a > 0$ باشد. اگر $l = \sqrt{a}$ و $L = (l)$ باشد، جلوتر ثابت می‌کنیم که L فاکتور چولسکی A می‌باشد.
- اگر $n \rightarrow n + 1$: فرض کنید برای همه‌ی ماتریس‌های متقارن و مثبت معین نشان داده‌ایم که فاکتور چولسکی دارند؛ بنابراین ماتریس A را می‌توان به فرم زیر نوشت:

$$A = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

که $a_{11} \in \mathbb{R}$ ، $A_{12} \in \mathbb{R}^{1 \times n}$ ، $A_{21} = A_{12}^T$ و $A_{22} \in \mathbb{R}^{n \times n}$ است. حال تعریف S را به صورت زیر تعریف می‌کنیم:

$$S = A_{22} - \frac{1}{a_{11}} A_{21} A_{12}$$

S ، مکمل Schur ماتریس A با توجه به a_{11} است. بدیهی است که ماتریس S متقارن است. بنابراین تنها کافی است نشان دهیم که مثبت معین است. فرض کنید $x \neq 0$ و $y \in \mathbb{R}^{n \times n}$ بوده و به صورت زیر باشد:

$$y = \begin{pmatrix} -\frac{1}{a_{11}} A_{12} x \\ x \end{pmatrix}$$

آنگاه $y \neq 0$ است و بنابراین با مثبت معین بودن A ثابت می‌کند که $y^T A y > 0$. علاوه بر آن داریم:

$$\begin{aligned} y^T A y &= \begin{pmatrix} -\frac{1}{a_{11}} A_{12} x & x^T \end{pmatrix} \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} -\frac{1}{a_{11}} A_{12} x \\ x \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{a_{11}} A_{12} x & x^T \end{pmatrix} \begin{pmatrix} -A_{12} x + A_{12} x \\ -\frac{1}{a_{11}} A_{21} A_{12} x + A_{22} x \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{a_{11}} A_{12} x & x^T \end{pmatrix} \begin{pmatrix} 0 \\ -\frac{1}{a_{11}} A_{21} A_{12} x + A_{22} x \end{pmatrix} \\ &= -\frac{1}{a_{11}} x^T A_{21} A_{12} x + x^T A_{22} x \\ &= x^T S x \end{aligned}$$

بنابراین $x^T S x > 0$ هرگاه $x \neq 0$ ثابت می‌شود که S مثبت معین است. با توجه به متقارن و مثبت معین بودن ماتریس S و با استفاده از فرض استقرا می‌توان نتیجه گرفت که تجزیه چولسکی دارد؛ یعنی می‌توان آن را به صورت $S = L_S L_S^T$ نوشت و L ماتریس پایین مثلثی است. L را به شکل زیر تعریف می‌کنیم:

$$L = \begin{pmatrix} \sqrt{a_{11}} & 0 \\ \frac{1}{\sqrt{a_{11}}} A_{21} & L_S \end{pmatrix}$$

مثبت معین بودن A به ما می فهماند که $a_{11} > 0$ ؛ بنابراین با استفاده از $A_{21}^T = A_{12}$ به راحتی می توان دید که L پایین مثلثی است.

$$\begin{aligned} LL^T &= \begin{pmatrix} \sqrt{a_{11}} & 0 \\ \frac{1}{\sqrt{a_{11}}}A_{21} & L_S \end{pmatrix} \begin{pmatrix} \sqrt{a_{11}} & \frac{1}{\sqrt{a_{11}}}A_{12} \\ 0 & L_S^T \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & \frac{1}{a_{11}}A_{21}A_{12} + L_S L_S^T \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & \frac{1}{a_{11}}A_{21}A_{12} + S \end{pmatrix} \\ &= A \end{aligned}$$

سؤال ۵. مفاهیم پایه‌ای

۱. متغیر تصادفی

(آ •

$$\sum_{x=2}^{\infty} f(x) = 1 \rightarrow a + \frac{a}{3} + \frac{a}{9} + \dots = 1 \rightarrow \frac{a}{a - \frac{1}{3}} = 1 \rightarrow \frac{3}{2}a = 1 \rightarrow a = \frac{2}{3}$$

(ب •

$$F_X(x) = \begin{cases} \frac{2}{3} + \frac{2}{3^2} + \dots + \frac{2}{3^{x-1}} & x = 2, 3, \dots \\ 0 & otherwise \end{cases}$$

$$\rightarrow F_X(x) = \begin{cases} 1 - \frac{1}{3^{x-1}} & x = 2, 3, \dots \\ 0 & otherwise \end{cases}$$

(پ •

$$P(X > 5) = 1 - P(X \leq 5) = 1 - F_X(5) = 1 - \frac{80}{81} = \frac{1}{81}$$

(ت •

$$P(X > 5 | X < 7) = \frac{P(5 < X < 7)}{P(X < 7)} = \frac{F_X(6) - (1 - F_X(5))}{F_X(6)} = \frac{\frac{242}{243} - \frac{80}{81}}{\frac{242}{243}} = \frac{2}{242}$$

۲. اثبات

(آ •

$$E[E[X|Y]] = \sum_y E[X|Y=y]P(Y=y)$$

با قانون partition می‌دانیم که عبارت بالا برابر با $E[X]$ می‌باشد. بنابراین:

$$E[E[X|Y]] = E[X]$$

(ب •

$$\begin{aligned} E[\text{var}(X|Y)] + \text{var}(E[X|Y]) &= E[E[X^2|Y] - E[X|Y]^2] + (E[E[X|Y]^2] - E[E[X|Y]]^2) \\ &= E[E[X^2|Y]] - E[E[X|Y]^2] + E[E[X|Y]^2] - E[X]^2 \\ &= E[X^2] - E[X]^2 \\ &= \text{var}(X) \end{aligned}$$

سؤال ۶. دیریکله

۱. توزیع Prior یک توزیع دیریکله با پارامترهای $(4, 2, 2, 3)$ است.

میانگین prior ها برابر با $\frac{a_{0,i}}{A_0}$ می باشد، بنابراین برای هر کدام داریم:

$$\theta_{11} = \frac{4}{11}, \theta_{10} = \frac{2}{11}, \theta_{01} = \frac{2}{11}, \theta_{00} = \frac{3}{11}$$

واریانس prior ها برابر با $\frac{a_{0,i}}{(A_0+1)A_0} - \frac{a_{0,i}^2}{A_0^2(A_0+1)}$ می باشد؛ بنابراین برای هر کدام داریم:

$$\theta_{11} = \frac{4}{12 \times 11} - \frac{4^2}{11^2 \times 12}$$

$$\theta_{10} = \frac{2}{12 \times 11} - \frac{2^2}{11^2 \times 12}$$

$$\theta_{01} = \frac{2}{12 \times 11} - \frac{2^2}{11^2 \times 12}$$

$$\theta_{00} = \frac{3}{12 \times 11} - \frac{3^2}{11^2 \times 12}$$

۲. توزیع Posterior یک توزیع دیریکله با پارامترهای $(29, 9, 8, 16) = (4 + 25, 2 + 7, 2 + 6, 3 + 13)$ است.

۳. میانگین posterior ها برابر با $\frac{a_{1,i}}{A_1}$ است. در نتیجه برای هر کدام داریم:

$$\theta_{11} = \frac{29}{62}, \theta_{10} = \frac{9}{62}, \theta_{01} = \frac{8}{62}, \theta_{00} = \frac{16}{62}$$

واریانس posterior ها برابر با $\frac{a_{1,i}}{(A_1+1)A_1} - \frac{a_{1,i}^2}{A_1^2(A_1+1)}$ می باشد؛ بنابراین برای هر کدام داریم:

$$\theta_{11} = \frac{29}{63 \times 62} - \frac{29^2}{62^2 \times 63}$$

$$\theta_{10} = \frac{9}{63 \times 62} - \frac{9^2}{62^2 \times 63}$$

$$\theta_{01} = \frac{8}{63 \times 62} - \frac{8^2}{62^2 \times 63}$$

$$\theta_{00} = \frac{16}{63 \times 62} - \frac{16^2}{62^2 \times 63}$$

سؤال ۷. گاوسی چندمتغیره

۱. اگر $a = \begin{pmatrix} 3 & -2 & 1 \end{pmatrix}^T$ آن‌گاه داریم:

$$a^T X = 3X_1 - 2X_2 + X_3 \rightarrow a^T X \sim N(a^T \mu, a^T \Sigma a)$$

$$a^T \mu = \begin{pmatrix} 3 & -2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} = 13$$

$$a^T \Sigma a = \begin{pmatrix} 3 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} = 9$$

در نتیجه توزیع آن برابر با $N_3(13, 9)$ است.

۲. اگر $a = \begin{pmatrix} a_1 & a_2 \end{pmatrix}^T$ آن‌گاه داریم:

$$Y = X_2 - a^T \begin{pmatrix} X_1 \\ X_3 \end{pmatrix} = -a_1 X_1 + X_2 - a_2 X_3$$

حال اگر $A = \begin{pmatrix} 0 & 1 & 0 \\ -a_1 & 1 & -a_2 \end{pmatrix}$ آن‌گاه داریم:

$$AX = \begin{pmatrix} X_2 \\ Y \end{pmatrix} \sim N(A\mu, A\Sigma A^T)$$

که

$$\begin{aligned} A\Sigma A^T &= \begin{pmatrix} 0 & 1 & 0 \\ -a_1 & 1 & -a_2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 0 & -a_1 \\ 1 & 1 \\ 0 & -a_2 \end{pmatrix} \\ &= \begin{pmatrix} 3 & -a_1 - 2a_2 + 3 \\ -a_1 - 2a_2 + 3 & a_1^2 - 2a_1 - 4a_2 + 2a_1 a_2 + 2a_2^2 + 3 \end{pmatrix} \end{aligned}$$

. چون می‌خواهیم X_2 و Y مستقل باشند، باید $-a_1 - 2a_2 + 3 = 0$. بنابراین، برای $v \in \mathbb{R}$ داریم:

$$a = \begin{pmatrix} 3 \\ 0 \end{pmatrix} + v \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

سؤال ۸. ML & MAP

۱. پیدا کردن joint

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

فرمول توزیع نرمال به صورت زیر است:

$$N(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

حال با توجه به این که واریانس برابر با ۱ است، داریم:

$$f_{\mu, X, Y}(t, x, y) = f(x, y|t)f(t) = \begin{cases} \frac{1}{2\pi} e^{-\frac{(x-t)^2 + (y-t)^2}{2}} & t \in [0, 1] \\ 0 & otherwise \end{cases}$$

۲. پیدا کردن posterior

$$MAP = \operatorname{argmax}_t f(t|x, y) = \frac{f(x, y|t)f(t)}{f(x, y)} \sim \operatorname{argmax}_t \log f(t|x, y) \times 1$$

با مشتق‌گیری از رابطه‌ی بالا و قرار دادن آن با صفر به رابطه‌ی $t = \frac{x+y}{2}$ می‌رسیم.

۳. ارزیابی:

• آ

$$x = \frac{3}{4}, y = 1 \rightarrow t = \frac{\frac{3}{4} + 1}{2} = \frac{7}{8}$$

• ب

$$x = \frac{1}{2}, y = 2 \rightarrow t = \frac{\frac{1}{2} + 2}{2} = \frac{5}{4} > 1$$

از آن جایی که تابع لگاریتم یک تابع یک‌به‌یک و صعودی است و مقدار احتمال بیشتر از یک نداریم، باید این احتمال را ۱ در نظر بگیریم.