Московский государственный университет имени М. В. Ломоносова.

Теория 2. Матричные вычисления

Чванкина Дарья
Группа 417
Октябрь 2022

1.

$$\begin{gathered}(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1},\\ A\in\mathbb{R}^{n\times n}, C\in\mathbb{R}^{m\times m}, U\in\mathbb{R}^{n\times m}, V\in\mathbb{R}^{m\times n}, \det(A)\neq 0, \det(C)\neq 0\end{gathered} \tag{1}$$

Доказательство:

$$\begin{aligned}(A+UCV)(A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}) =\\ = I + UCVA^{-1} - (A+UCV)A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1} =\\ = I + UCVA^{-1} - (U+UCVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} =\\ = I + UCVA^{-1} - U(I+CVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} =\\ = I + UCVA^{-1} - UC(C^{-1}+VA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} =\\ = I + UCVA^{-1} - UCVA^{-1} = I\end{aligned} \tag{2}$$

Так как обратная матрица единственна, то выражение $A^{-1}-A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$ задает обратную матрицу к матрице $A+UCV$.

2. Пусть $p(x)=\mathcal{N}(x|\mu,\Sigma)$, $p(y|x)=\mathcal{N}(y|Ax,\Gamma), A\in\mathbb{R}^{m\times n}$. Найти $p(x|y)-?$

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx} = \text{Const } \mathcal{N}(y|Ax,\Gamma)\mathcal{N}(x|\mu,\Sigma) = \mathcal{N}(p,Q) \tag{3}$$

Нормальные распределения - сопряженные. Откуда $p(x|y)=\mathcal{N}(x|p,Q)$. Запишем распределение в общем виде и раскроем скобки в экспоненте.

$$\begin{gathered}p(x|y) = \mathcal{N}(p,Q) = \text{Const } exp[-\frac{1}{2}(x-p)^TQ^{-1}(x-p)] =\\ = \text{Const } exp[-\frac{1}{2}(x^TQ^{-1}x - p^TQ^{-1}x - x^TQ^{-1}p + p^TQ^{-1}p)]\end{gathered} \tag{4}$$

Из формулы (3) распишем произведение нормальных распределений:

$$\begin{aligned}p(x|y) = \text{Const } \mathcal{N}(y|Ax,\Gamma)\mathcal{N}(x|\mu,\Sigma) =\\ = \text{Const } exp[-\frac{1}{2}(y-Ax)^T\Gamma^{-1}(y-Ax)]exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)] =\\ = \text{Const } exp[-\frac{1}{2}(y^T\Gamma^{-1}y - y^T\Gamma^{-1}Ax - x^TA^T\Gamma^{-1}y + x^TA^T\Gamma^{-1}Ax+\\ +x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu)] = \text{Const } exp[\frac{1}{2}(x^T(A^T\Gamma^{-1}A+\Sigma^{-1})x-\\ -x^T(A^T\Gamma^{-1}y+\Sigma^{-1}\mu) + y^T\Gamma^{-1}y - y^T\Gamma^{-1}Ax - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu)]\end{aligned} \tag{5}$$

Из(4) и (5) получаем уравнения:

$$\begin{gathered}x^TQ^{-1}x = x^T(A^T\Gamma^{-1}A+\Sigma^{-1})x \Rightarrow Q = (A^T\Gamma^{-1}A+\Sigma^{-1})^{-1}\\ x^T(A^T\Gamma^{-1}y+\Sigma^{-1}\mu) = x^TQ^{-1}p \Rightarrow p = Q(A^T\Gamma^{-1}y+\Sigma^{-1}\mu)\end{gathered} \tag{6}$$

Итого: $p(x|y)=\mathcal{N}(x|Q(A^T\Gamma^{-1}y+\Sigma^{-1}\mu),Q)$, где $Q=(A^T\Gamma^{-1}A+\Sigma^{-1})^{-1}$

3. Пусть $p(x) = \mathcal{N}(x|\mu, \Sigma), p(y|x) = \mathcal{N}(y|Ax, \Gamma)$. Доказать: $p(y) = \mathcal{N}(y|A\mu, \Gamma + A\Sigma A^T)$.
   Представим $p(y|x)$ в виде $p(y|x) = Ax + \epsilon$, где $\epsilon = \mathcal{N}(0, \Gamma)$.

$$\begin{aligned} p(y) = \int p(x,y)dx = \int p(y|x)p(x)dx = \int \mathcal{N}(y|Ax,\Gamma)\mathcal{N}(x|\mu,\Sigma)dx = \mathcal{N}(\mathbb{E}y, \mathbb{D}y) \\ \mathbb{E}y = \mathbb{E}(Ax+\epsilon) = A\mathbb{E}x + 0 = A\mu \\ \mathbb{D}y = \mathbb{D}(Ax+\epsilon) = \mathbb{D}(Ax) + \mathbb{D}\epsilon + 2Cov(Ax,\epsilon) = A\Sigma A^T + \Gamma + 0 = \Gamma + A\Sigma A^T \end{aligned} \tag{7}$$

   Откуда следует, что $p(y) = \mathcal{N}(y|A\mu, \Gamma + A\Sigma A^T)$.

4. $\frac{\partial}{\partial X} det(X^{-1} + A)$ - ?
   Воспользуемся формулами:

$$d(\text{Det } X) = \text{Det } X\langle X^{-T}, dX\rangle \tag{8}$$

$$d(X^{-1}) = -X^{-1}(dX)X^{-1} \tag{9}$$

$$df(x) = \langle \nabla f(x), dX\rangle \tag{10}$$

$$\begin{aligned} d(\text{Det}(X^{-1}+A)) = \{(8)\} = \text{Det}(X^{-1}+A)\langle (X^{-1}+A)^{-T}, d(X^{-1}+A)\rangle = \\ = \{(9)\} = -\text{Det}(X^{-1}+A)\langle (X^{-1}+A)^{-T}, X^{-1}(dX)X^{-1}\rangle = \\ = -\text{Det}(X^{-1}+A)\langle X^{-T}(X^{-1}+A)^{-T}X^{-T}, dX\rangle \end{aligned} \tag{11}$$

   Из (10) следует, что

$$\frac{\partial}{\partial X}\text{Det } (X^{-1}+A) = -\text{Det } (X^{-1}+A)X^{-T}(X^{-1}+A)^{-T}X^{-T} \tag{12}$$

5. $\frac{\partial}{\partial X} tr(AX^{-T}BXC)$ - ?
   Дополнительные формулы:

$$d(XY) = (dX)Y + X(dY) \tag{13}$$

$$d\langle X, Y\rangle = \langle dX, Y\rangle + \langle X, dY\rangle \tag{14}$$

$$d(X^T) = (dX)^T \tag{15}$$

Перепишем след через скалярное произведение:

$$tr(AX^{-T}BXC) = \langle I_n, AX^{-T}BXC\rangle \tag{16}$$

$$\begin{aligned} d\langle I_n, AX^{-T}BXC\rangle = \{(13),(14)\} = \langle I_n, d(AX^{-T}B)XC + AX^{-T}Bd(XC)\rangle = \\ = \{(9),(15)\} = \langle I_n, -AX^{-T}(dX)^T X^{-T}BXC\rangle + \langle I_n, AX^{-T}B(dX)C\rangle = \\ = \langle -X^{-1}A^T(X^{-T}BXC)^T, (dX)^T\rangle + \langle B^T X^{-1}A^T C^T, dX\rangle = \\ = \langle -X^{-T}BXCAX^{-T}, dX\rangle + \langle B^T X^{-1}A^T C^T, dX\rangle = \\ = \langle -X^{-T}BXCAX^{-T} + B^T X^{-1}A^T C^T, dX\rangle \end{aligned} \tag{17}$$

Из (10) следует, что

$$\frac{\partial}{\partial X} tr(AX^{-T}BXC) = -X^{-T}BXCAX^{-T} + B^T X^{-1}A^T C^T \tag{18}$$