Теоретическая работа №2

Котов Артем, МОиАД ВШЭ СПб

11 ноября 2021 г.

Содержание

Task 1

Условие: Доказать тождество Вудбери: $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$, где $A \in \mathbb{R}^{nxn}$, $C \in \mathbb{R}^{mxm}$, $U \in \mathbb{R}^{nxm}$, $V \in \mathbb{R}^{mxn}$

Решение.

Домножим правую часть равенства на обратное от левой части, должны получить единичную матрицу:

$$
\begin{gathered}
(A+UCV)(A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}) = \\
\text{[усердно раскрываем скобки]} \\
= I + UCVA^{-1} - (U + UCVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\
= I + UCVA^{-1} - UC\underbrace{(C^{-1}+VA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}}_{=I}VA^{-1} \\
= I + UCVA^{-1} - UCVA^{-1} = I
\end{gathered}
$$

■

Task 2

Условие: Given $p(x) = \mathcal{N}(x|\mu,\Sigma)$ and $p(y|x) = \mathcal{N}(y|Ax,\Gamma)$, $p(x|y) =?$

Решение.

$$p(x|y) = \frac{1}{Z}p(y|x)p(x)$$

Нормировочная константа Z не содержит x, поэтому не будет нас сейчас интересовать в смысле вида распределения на x. Разберемся с числителем:

$$p(y|x)p(x) = \text{Const}\exp[-\frac{1}{2}((x-\mu)^T\Sigma^{-1}(x-\mu) + (y-Ax)^T\Gamma^{-1}(y-Ax))]$$

Разберемся с показателем экспоненты (опускаю $-\frac{1}{2}$):

$$
\begin{gathered}
(x-\mu)^T\Sigma^{-1}(x-\mu) + (y-Ax)^T\Gamma^{-1}(y-Ax) \\
= \text{[очередное раскрытие скобок; } \Sigma^T = \Sigma; \Gamma^T = \Gamma] = \\
= x^T(\Sigma^{-1} + A^T\Gamma^{-1}A)x - 2(\mu^T\Sigma^{-1} + y^T\Gamma^{-1}A)x + C,
\end{gathered}
$$

т.е. с учетом $-\frac{1}{2}$ получим экспоненту, у которой в показателе стоит отрицательно направленная парабола, после выделения полного квадрата которой получим, что это гауссово распределение с параметрами:

$$
\begin{gathered}
\mu' = (\Sigma^{-1} + A^T\Gamma^{-1}A)^{-1}(\Sigma^{-1}\mu + A^T\Gamma^{-1}y) \\
\Sigma' = (\Sigma^{-1} + A^T\Gamma^{-1}A)^{-1},
\end{gathered}
$$

т.е.

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

■

Task 3

Условие: Доказать, что $p(y) = \mathcal{N}(y|A\mu, \Gamma + A\Sigma A^{-1})$

Решение.

Мы знаем, что $p(y)$ можно выразить, например, так:

$$p(y) = \int\limits_{\Omega_x} p(y|x)p(x)dx$$

Еще в предыдущем номере мы могли заметить, что зависимость от y в таком выражении – некая экспонента с отрицательной параболой в показателе степени, т.е. мы действительно можем искать решение в виде некого гауссового распределения.

Предположим, что $y = Ax + z$, где $x \sim \mathcal{N}(\mu, \Sigma)$, а $z \sim \mathcal{N}(0, \Gamma)$ и покажем, что такой выбор удовлетворяет тому, что дано в условии, а именно $p(y|x) = \mathcal{N}(y|Ax, \Gamma)$ (к тому же сумма двух нормальных независимых величины есть величина нормальная):

$$\mathbb{E}y|x = \underbrace{\mathbb{E}Ax}_{Ax - \text{константа}} + \mathbb{E}z = Ax + 0 = Ax$$

$$\mathbb{V}y|x = \underbrace{\mathbb{V}Ax}_{Ax - \text{константа}} + \mathbb{V}z = 0 + \Gamma$$

$$\Downarrow$$

$$p(y|x) = \mathcal{N}(y|Ax, \Gamma)$$

Чтобы задать гауссово распределение, нам достаточно занять матожидание и матрицу ковариации:

$$\mathbb{E}y = \mathbb{E}Ax + \mathbb{E}z = A\mu + 0 = A\mu$$

$$\mathbb{V}y = \mathbb{V}Ax + \mathbb{V}z = A\mathbb{V}xA^T + \Gamma = A\Sigma A^T + \Gamma$$

Таким образом, $p(y) = \mathcal{N}(y|A\mu, A\Sigma A^T + \Gamma)$ ■

Task 4

Условие:

$$\frac{\partial \det(X^{-1} + A)}{\partial X} = ?$$

Решение.

Будем работать в терминах дифференциала, сведем к каноничной форме, из которой и получим необходимое:

$$\begin{aligned} d(\det(X^{-1}+A)) &= \det(X^{-1}+A)\,\langle (X^{-1}+A)^{-T} | d(X^{-1}+A)\rangle \\ &= \det(X^{-1}+A)\,\langle (X^{-1}+A)^{-T} | -X^{-1}(dX)X^{-1}\rangle \\ &= \det(X^{-1}+A)\,\langle -X^{-T}(X^{-1}+A)^{-T}X^{-T} | dX\rangle \end{aligned}$$

$$\Downarrow$$

$$\frac{\partial \det(X^{-1}+A)}{\partial X} = \det(X^{-1}+A)(-X^{-T}(X^{-1}+A)^{-T}X^{-T})$$

■

Task 4

Условие:

$$\frac{\partial Tr(AX^{-T}BXC)}{\partial X} = ?$$

Решение.

Сначала определимся с размерностями, чтобы ничего не сломать: пусть $X \in \mathbb{R}^{nxn}$, а $C \in \mathbb{R}^{nxm}$, тогда $B \in \mathbb{R}^{nxn}$, $A \in \mathbb{R}^{mxn}$.

Опять будем делать через дифференциал:

$$Tr(AX^{-T}BXC) = Tr(CAX^{-T}BX) = \langle A^TC^T | X^{-T}BX\rangle$$

$$d(\langle A^TC^T | X^{-T}BX\rangle) = \langle A^TC^T | d(X^{-T}BX)\rangle = (*)$$

1) $d(X^{-T}BX) = d(X^{-T}B)X + X^{-T}BdX$

2) $d(X^{-T}B) = (-X^{-1}dXX^{-1})^TB$

$$\begin{aligned} (*) &= (-X^{-1}dXX^{-1})^TBX + X^{-T}BdX \\ &= -X^{-T}dX^TX^{-T}BX + X^{-T}BdX \\ &= X^{-T}(-dX^TX^{-T}BX + (dX^TB^T)^T) \\ &= -X^{-T}((X^TB^TX^{-1}dX)^T - BdX) \end{aligned}$$

Т.о. получаем:

$$\begin{aligned} &\langle A^TC^T | -X^{-T}((X^TB^TX^{-1}dX)^T - BdX)\rangle \\ &= \langle A^TC^T | -X^{-T}(X^TB^TX^{-1}dX)^T\rangle + \langle A^TC^T | X^{-T}BdX\rangle = \end{aligned}$$

перекидываем от dX все, собираем скалярное произведение обратно в кучу (честно, не очень хочется это набирать в техе):

$$= \langle B^T X^{-1} A^T C^T - X^{-T} BXCAX^{-T} | dX \rangle$$

$$\Downarrow$$

$$\frac{\partial Tr(AX^{-T}BXC)}{\partial X} = \underbrace{B^T X^{-1} A^T C^T}_{①} - \underbrace{X^{-T} BXCAX^{-T}}_{②}$$

Проверим, что размерности сошлись:

1) $(n, n)(n, n)(n, m)(m, n) \Longrightarrow (n, n)$

2) $(n, n)(n, n)(n, n)(n, m)(m, n)(n, n) \Longrightarrow (n, n)$

В итоге, получим подходящую размерность, т.к. дифференцировали скаляр по матрице и ожидали получить матрицу размера (n, n). ■