

Теоретическая работа №2

Котов Артем, МОиАД ВШЭ СПб

11 ноября 2021 г.

Содержание

Task 1	2
Task 2	2
Task 3	3
Task 4	3
Task 4	4

Task 1

Условие: Доказать тождество Вудбери: $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, где $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times m}$, $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{m \times n}$

Решение.

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

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

■

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:

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

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

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

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:

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

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:

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

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 некоего 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):

$$\begin{aligned} \mathbb{E}y|x &= \underbrace{\mathbb{E}Ax}_{Ax - \text{константа}} + \mathbb{E}z = Ax + 0 = Ax \\ \mathbb{V}y|x &= \underbrace{\mathbb{V}Ax}_{Ax - \text{константа}} + \mathbb{V}z = 0 + \Gamma \\ &\Downarrow \\ p(y|x) &= \mathcal{N}(y|Ax, \Gamma) \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}y &= \mathbb{E}Ax + \mathbb{E}z = A\mu + 0 = A\mu \\ \mathbb{V}y &= \mathbb{V}Ax + \mathbb{V}z = A\mathbb{V}xA^T + \Gamma = A\Sigma A^T + \Gamma \end{aligned}$$

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:

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

■

Task 4

Условие:

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

Решение.

Сначала определимся с размерностями, чтобы ничего не сломать: пусть $X \in \mathbb{R}^{n \times n}$, а $C \in \mathbb{R}^{n \times m}$, тогда $B \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{m \times n}$. Опять будем делать через дифференциал:

$$\begin{aligned}
 \text{Tr}(AX^{-T}BXC) &= \text{Tr}(CAX^{-T}BX) = \langle A^T C^T | X^{-T}BX \rangle \\
 d(\langle A^T C^T | X^{-T}BX \rangle) &= \langle A^T C^T | d(X^{-T}BX) \rangle = (*)
 \end{aligned}$$

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

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

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

Т.о. получаем:

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

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

$$\begin{aligned}
 &= \langle B^T X^{-1} A^T C^T - X^{-T} B X C A X^{-T} | dX \rangle \\
 &\quad \Downarrow \\
 &\frac{\partial \text{Tr}(A X^{-T} B X C)}{\partial X} = \underbrace{B^T X^{-1} A^T C^T}_{\textcircled{1}} - \underbrace{X^{-T} B X C A X^{-T}}_{\textcircled{2}}
 \end{aligned}$$

Проверим, что размерности сошлись:

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

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

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