Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Второе теоретическое задание: Матричные вычисления

Каратыщев Дмитрий Иванович, 417 группа

Дата выполнения работы:
6 ноября 2024 г.

1. Доказать тождество Вудбери:

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}.$$

Здесь $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times m}$, $U \in \mathbb{R}^{n\times m}$, $V \in \mathbb{R}^{m\times n}$.

Доказательство:

Преобразуем правую часть равенства. Пользуемся тем, что $(AB)^{-1} = B^{-1}A^{-1}$

$A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1} = A^{-1} - A^{-1}((C^{-1}+VA^{-1}U)U^{-1})^{-1}VA^{-1} =$
$= A^{-1} - A^{-1}(C^{-1}U^{-1}+VA^{-1})^{-1}VA^{-1} = A^{-1} - A^{-1}(V^{-1}(C^{-1}U^{-1}+VA^{-1}))^{-1}A^{-1} =$
$= A^{-1} - A^{-1}(V^{-1}C^{-1}U^{-1}+A^{-1})^{-1}A^{-1} = A^{-1} - ((UCV)^{-1}A+I)^{-1}A^{-1} =$
$= A^{-1} - [(UCV)^{-1}(A+UCV)]^{-1}A^{-1} = A^{-1} - (A+UCV)^{-1}UCVA^{-1} =$
$= (A+UCV)^{-1}[(A+UCV)A^{-1} - UCVA^{-1}] = (A+UCV)^{-1}(I+UCVA^{-1}-UCVA^{-1}) =$
$= (A+UCV)^{-1}I = (A+UCV)^{-1}$, что и требовалось доказать

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

Решение:

По теореме Байеса, получим, что $p(\boldsymbol{x} \mid \boldsymbol{y}) = \dfrac{p(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x})dx} =$
$= \dfrac{1}{\int p(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x})dx} \dfrac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{(\sqrt{2\pi})^n\sqrt{\det\Sigma}} \dfrac{\exp\left(-\frac{1}{2}(y-Ax)^T\Gamma^{-1}(y-Ax)\right)}{(\sqrt{2\pi})^m\sqrt{\det\Gamma}} \propto$
$\propto \exp\left(-\frac{1}{2}\left[(x-\mu)^T\Sigma^{-1}(x-\mu) - (y-Ax)^T\Gamma^{-1}(y-Ax)\right]\right) \propto$
$\propto \exp\left(-\frac{1}{2}\left[x^T(\Sigma^{-1}+A^T\Gamma^{-1}A)x - 2(\Sigma^{-1}\mu + A^T\Gamma^{-1}y)^Tx\right]\right) + C$

Отсюда делаем вывод, что в результате у нас тоже нормальное распределение с ковариационной матрицей $\hat{\Sigma} = (\Sigma^{-1}+A^T\Gamma^{-1}A)^{-1}$ и средним значением $\hat{\mu} = \hat{\Sigma}(\Sigma^{-1}\mu + A^T\Gamma^{-1}y)$

Ответ: $p(\boldsymbol{x} \mid \boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{x} \mid (\Sigma^{-1}+A^T\Gamma^{-1}A)^{-1}(\Sigma^{-1}\mu + A^T\Gamma^{-1}y),\ (\Sigma^{-1}+A^T\Gamma^{-1}A)^{-1}\right)$

3. Пусть $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma)$, $p(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{y} \mid A\boldsymbol{x}, \Gamma)$. Доказать, что $p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y} \mid A\boldsymbol{\mu}, \Gamma + A\Sigma A^T)$.

Доказательство: Нормальность распределения на $\boldsymbol{y}$ следует из того, что $p(\boldsymbol{x}, \boldsymbol{y})$ - это нормальное распределение, а маргинальные распределения нормального тоже являются нормальными. Теперь найдём среднее значение и ковариационную матрицу рассматриваемого распределения. Представим $\boldsymbol{y}$ как $A\boldsymbol{x}+\boldsymbol{\varepsilon}$, где $p(\boldsymbol{\varepsilon}) = \mathcal{N}(\boldsymbol{\varepsilon} \mid 0, \Gamma)$

$\mathbb{E}\boldsymbol{y} = \mathbb{E}(\mathbb{E}[\boldsymbol{y} \mid \boldsymbol{x}]) = \mathbb{E}[A\boldsymbol{x}] = A\mathbb{E}\boldsymbol{x} = A\boldsymbol{\mu}$

$Cov(\boldsymbol{y}) = \mathbb{E}\left[(\boldsymbol{y}-\mathbb{E}\boldsymbol{y})(\boldsymbol{y}-\mathbb{E}\boldsymbol{y})^T\right] = \mathbb{E}\left[(A\boldsymbol{x}+\boldsymbol{\varepsilon}-\mathbb{E}(A\boldsymbol{x}+\boldsymbol{\varepsilon}))(A\boldsymbol{x}+\boldsymbol{\varepsilon}-\mathbb{E}(A\boldsymbol{x}+\boldsymbol{\varepsilon}))^T\right] =$
$= \mathbb{E}\left[(A(\boldsymbol{x}-\boldsymbol{\mu})+\boldsymbol{\varepsilon})(A(\boldsymbol{x}-\boldsymbol{\mu})+\boldsymbol{\varepsilon})^T\right] = A\mathbb{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\right]A^T + \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right] + 2A\mathbb{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\varepsilon}^T\right] = A\Sigma A^T + \Gamma$

Последнее слагаемое $2A\mathbb{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{\varepsilon}^T\right]$ равно нулю в силу того, что $\boldsymbol{x}$ и $\boldsymbol{\varepsilon}$ независимы, а также $\mathbb{E}\boldsymbol{\varepsilon} = 0$. Получили, что $\mathbb{E}\boldsymbol{y} = A\boldsymbol{\mu}$, $Cov(\boldsymbol{y}) = A\Sigma A^T + \Gamma$, что и требовалось доказать.

4. Вычислить $\frac{\partial}{\partial X}\det(X^{-1}+A)$ (все матрицы не являются симметричными);

Решение: Сначала выведем $\frac{\partial}{\partial A}\det(A)$. Разложим по теореме Лапласа $\det(A)$ по i строке: $\det(A) = \sum_j a_{ij}\left[adj(A)\right]_{ji}$. Здесь $adj(A)$ - это присоединённая матрица к A, которая является транспонированной матрицей алгебраических дополнений. Тогда $\frac{\partial}{\partial a_{ij}}\det(A) = \left[adj(A)\right]_{ji}$. Известно, что $adj(A) = \det(A)A^{-1}$. Отсюда получаем, что $\frac{\partial}{\partial A}\det(A) = \det(A)A^{-T}$. Из этого также следует, что $d\det(A)[H] = \left\langle\det(A)A^{-T}, H\right\rangle$.

Теперь рассмотрим $\frac{\partial}{\partial A}A^{-1}$. В явном виде данная производная является тензором 4-го порядка, поэтому проще записать её дифференциал. $AA^{-1} = I \Rightarrow d(AA) = dI \Rightarrow dAA^{-1} + AdA^{-1} = 0 \Rightarrow dA^{-1} = -A^{-1}dAA^{-1}$

Теперь приступим к основному заданию. Запишем дифференциал: $d\det(X^{-1}+A) = \det(X^{-1}+A)\operatorname{tr}\left((X^{-1}+A)^{-1}d(X^{-1}+A)\right) = \det(X^{-1}+A)\operatorname{tr}\left((X^{-1}+A)^{-1}(-X^{-1}dXX^{-1})\right) =$
$= -\det(X^{-1}+A)\operatorname{tr}\left((X^{-1}+A)^{-1}X^{-1}dXX^{-1}\right) = -\det(X^{-1}+A)\operatorname{tr}\left(X^{-1}(X^{-1}+A)^{-1}X^{-1}dX\right) =$
$= -\det(X^{-1}+A)\operatorname{tr}\left((I+AX)^{-1}X^{-1}dX\right) = -\det(X^{-1}+A)\operatorname{tr}\left((X+XAX)^{-1}dX\right) =$
$= \left\langle-\det(X^{-1}+A)(X+XAX)^{-T}, dX\right\rangle \Rightarrow \frac{\partial}{\partial X}\det(X^{-1}+A) = -\det(X^{-1}+A)(X+XAX)^{-T}$

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

5. Вычислить $\frac{\partial}{\partial X}\operatorname{tr}(AX^{-T}BXC)$ (все матрицы не являются симметричными, матрицы A, C не являются квадратными).

Решение: $d\operatorname{tr}(AX^{-T}BXC) = \operatorname{tr}\left(d(AX^{-T}BXC)\right) = \operatorname{tr}\left(d(AX^{-T})BXC + AX^{-T}d(BXC)\right) =$
$= -\operatorname{tr}\left(AX^{-T}(dX)^T X^{-T}BXC\right) + \operatorname{tr}\left(AX^{-T}BdXC\right) = -\operatorname{tr}\left((dX)^T X^{-T}BXCAX^{-T}\right) + \operatorname{tr}\left(CAX^{-T}BdX\right) =$
$= -\operatorname{tr}\left((X^{-T}BXCAX^{-T})^T dX\right) + \operatorname{tr}\left((B^T X^{-1}A^T C^T)^T dX\right) = \left\langle -X^{-T}BXCAX^{-T} + B^T X^{-1}A^T C^T, dX\right\rangle$. Отсюда получаем, что $\frac{\partial}{\partial X}\operatorname{tr}(AX^{-T}BXC) = B^T X^{-1}A^T C^T - X^{-T}BXCAX^{-T}$

Ответ: $\frac{\partial}{\partial X}\operatorname{tr}(AX^{-T}BXC) = B^T X^{-1}A^T C^T - X^{-T}BXCAX^{-T}$