Теоретическое домашнее задание

Кузиной Екатерины

Во всех задачах основным полем считается $\mathbb{R}$, если не сказано обратного.

Задачи про матричные производные.

Задача 1 (1 балл)**.** Пусть f — функция на множестве квадратных матриц $n \times n$, а g — функция на множестве симметричных матриц $n \times n$, совпадающая с f на своей области определения. Докажите, что

$$\frac{\partial g}{\partial X} = \frac{\partial f}{\partial X} + \left(\frac{\partial f}{\partial X}\right)^T - \operatorname{diag}\left(\frac{\partial f}{\partial x_{11}}, \frac{\partial f}{\partial x_{22}}, \ldots, \frac{\partial f}{\partial x_{nn}}\right)$$

Зачем это надо? Представьте, что вам надо написать алгоритм градиентного спуска для функции, определённой только на подпространстве симметричных матриц — ясно, что направление наискорейшего спуска внутри этого подпространства может быть не таким, как на всём пространстве матриц.

Задача 2 (1 балл)**.** Найдите производную

$$\frac{\partial \operatorname{tr}\left(AX^2BX^{-T}\right)}{\partial X}$$

Решение:

$$d(trAX^2BX^{-T}) = tr(AXdXBX^{-T} + AdXXBX^{-T} + AX^2B(-X^{-1}dXX^{-1})^T) =$$

$$= tr(BX^{-T}AXdX + XBX^{-T}AdX - AX^2BX^{-T}dX^TX^{-T}) =$$

$$= tr(BX^{-T}AXdX + XBX^{-T}AdX - (X^{-T}AX^2BX^{-T})^TdX) =$$

$$= (X^TA^TX^{-1}B^T + A^TX^{-1}B^TX^T - X^{-T}AX^2BX^{-T}, dX) =$$

Ответ: $X^TA^TX^{-1}B^T + A^TX^{-1}B^TX^T - X^{-T}AX^2BX^{-T}$

Задача 3 (1 балл)**.** Найдите производную

$$\frac{\partial \ln\det\left(X^TAX\right)}{\partial X}$$

Решение:

$$d(ln|X^TAX|) = \frac{d|X^TAX|}{|X^TAX|} = \frac{(|X^TAX|(X^TAX)^{-T}, d(X^TAX))}{|X^TAX|} =$$

$$= ((X^TAX)^{-T}, d(X^TAX)) = tr((X^TAX)^{-1}d(X^TAX))) =$$

$$= tr((X^TAX)^{-1}(dX^TAX + X^TAdX)) = tr((AX(X^TAX)^{-1})^TdX + (X^TAX)^{-1}X^TAdX) =$$

$$= (AX(X^TAX)^{-1} + A^TX(X^TAX)^{-T}, dX)$$

Ответ: $AX(X^TAX)^{-1} + A^TX(X^TAX)^{-T}$

Задача 4 (2 балла)**.** Допустим, что векторы $y_1, \dots, y_m$ выбраны из многомерного нормального распределения с неизвестными вектором средних m и ковариационной матрицей Σ. В этом задании вам нужно будет найти оценки максимального правдоподобия $\hat{m}$ и $\hat{\Sigma}$.

Напомним вкратце, что такое оценка максимального правдоподобия в случае непрерывного распределения. Пусть $p(x|\theta_1, \dots, \theta_k)$ — функция плотности распределения с неизвестными нам параметрами $\theta_1, \dots, \theta_k$, а $y_1, \dots, y_m$ — выборка из этого распределения. *Функцией правдоподобия* назовём произведение $L(y_1 \dots, y_m|\theta_1, \dots, \theta_k) := \prod_{j=1}^{k} p(y_j|\theta_1, \dots, \theta_k)$; грубо говоря, это произведение показывает, насколько правдоподобно появление данной выборки $y_1, \dots, y_m$ при данных значениях параметров. В качестве оценки максимального правдоподобия выбирают те значения параметров, при которых функция правдоподобия достигает максимума. При этом как правило удобнее максимизировать не саму функцию правдоподобия, а *логарифмическую функцию правдоподобия* $l(y_1 \dots, y_m|\theta_1, \dots, \theta_k) = \ln L(y_1 \dots, y_m|\theta_1, \dots, \theta_k)$.

Подсказка. Постарайтесь превратить $\sum_i (x_i - m)^T \Sigma^{-1}(x_i - m)$ в функцию от матрицы X, столбцами которой являются векторы x_i.

Решение:

Пусть l - длина выборки (т.к. m является вектором средних).

$$\prod_{i=1}^{l} p(y_i; m, \Sigma) = \prod_{i=1}^{l} \frac{1}{\sqrt{(2\pi)^m det\Sigma}} exp\{-\frac{1}{2}(y_i - m)^T \Sigma^{-1}(y_i - m)\} =$$

$$= \frac{1}{((2\pi)^m det\Sigma)^{\frac{l}{2}}} exp\{-\frac{1}{2}\sum_{i=1}^{l}(y_i - m)^T \Sigma^{-1}(y_i - m)\} =$$

$$= \frac{1}{((2\pi)^m det\Sigma)^{\frac{l}{2}}} exp\{-\frac{1}{2} tr(\Sigma^{-1}\sum_{i=1}^{l}(y_i - m)(y_i - m)^T)\}$$

$$ln \prod_{i=1}^{l} p(y_i; m, \Sigma) = -\frac{lm}{2} ln(2\pi) - \frac{l}{2} ln|\Sigma| - \frac{1}{2} tr(\Sigma^{-1}\sum_{i=1}^{l}(y_i - m)(y_i - m)^T)$$

Продифференцировав логарифмическую функцию правдоподобия по m и Σ, получаем систему:

$$\begin{cases} -\frac{l}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\sum\limits_{i=1}^{l}(y_i - m)(y_i - m)^T \Sigma^{-1} = 0 \\ \sum\limits_{i=0}^{l}(y_i - m) = 0 \end{cases}$$

$$\begin{cases} \Sigma^{-1}(\sum\limits_{i=1}^{l}(y_i - m)(y_i - m)^T - l\Sigma)\Sigma^{-1} = 0 \\ \sum\limits_{i=0}^{l}(y_i - m) = 0 \end{cases}$$

Тогда решение данной системы:

$$\hat{m} = \frac{1}{l}\sum_{i=1}^{l} y_i \qquad \hat{\Sigma} = \sum_{i=1}^{l}(y_i - \hat{m})(y_i - \hat{m})^T$$

Ответ: $\hat{m} = \frac{1}{l}\sum\limits_{i=1}^{l} y_i \qquad \hat{\Sigma} = \sum\limits_{i=1}^{l}(y_i - \hat{m})(y_i - \hat{m})^T$

Задачи про всякие разные штуки.

Задача 5 (2 балла)**.** Для двух заданных матриц A и B одного размера найдите ортогональную матрицу Q, для которой норма Фробениуса разности $||QA - B||_F$ минимальна.

Решение:

$$\begin{cases} Q^TQ = E \\ ||QA - B||_F^2 \to min \end{cases}$$

Решаем данную систему с помощью метода множителей Лагранжа:
$L(Q, \lambda) = ||QA - B||_F^2 + tr(L^T(Q^TQ - E))$
$dL_Q = tr((2A(QA - B)^T + (L + L^T)Q^T)dQ)$
$\nabla f = 2(QA - B)A^T + Q(L + L^T) = 0$ $(*)$
1) $2Q^T(QA - B)A^T + (L + L^T) = 0$ (домножили $(*)$ на Q^T слева)
2) $2A(QA - B)^TQ + (L + L^T) = 0$ (транспонировали 1))
Приравняв 1) и 2) и раскрыв скобки, получаем: $Q^TBA^T = AB^TQ$.

Но как найти Q? Заметим, что AB^T является квадратной матрицей, и представим её через сингулярное разложение, т.е. $AB^T = U\Sigma V^T$. Т.к. эта матрица квадратная, то Σ - диагональная матрица такого же размера, поэтому $\Sigma = \Sigma^T$. Подставив это в уравнение, которое получили выше, получим, что $Q = VU^T$. Ясно, что $QQ^T = E$, т.к. U и V - ортогональные матрицы.

Ответ: $Q = VU^T$, где $AB^T = U\Sigma V^T$ - сингулярное разложение.

Задача 6 (1 балл)**.** Пусть A — матрица $m \times n$. Докажите, что каждая матрица X размера $n \times m$, удовлетворяющая условию $A^TAX = A^T$ является псевдообратной к A.

Доказательство:

Разложим матрицу A с помощью сингулярного разложения: $A = U\Sigma V^T$, где U и V - квадратные ортогональные матрицы размеров m*m и n*n соответственно, а Σ - некоторая "диагональная"прямоугольная матрица. Тогда $A^TAX = (U\Sigma V^T)^TU\Sigma V^TX = V\Sigma^T(U^TU)\Sigma V^TX = V\Sigma^T\Sigma V^TX$.

$A^TAX - A^T = V\Sigma^T\Sigma V^TX - V\Sigma^TU^T = V\Sigma^T(\Sigma V^TX - U^T) = 0 \Rightarrow \Sigma V^TX = U^T$

Рассмотрим $AXA = U\Sigma V^TXU\Sigma V^T = U(\Sigma V^TX)U\Sigma V^T = UU^TU\Sigma V^T = U\Sigma V^T = A$. Получаем, что матрица X, удовлетворяющая условию $A^TAX = A^T$ является псевдообратной к A.

Задача 7 (1,5 балла)**.** Пусть G — некоторая псевдообратная матрица для матрицы A. Докажите, что AG является проекцией на образ A. Для какой (или для каких) из псевдообратных матриц эта проекция будет ортогональной?