Машинное обучение
ФКН ВШЭ
Теоретическое домашнее задание №1

Задача 1. Найдите производную по матрице $A \in \mathbb{R}^{n \times n}$

$$\frac{\partial}{\partial A} \sum_{i=1}^{n} \lambda_i,$$

где $\lambda_1, \dots, \lambda_n$ — набор собственных значений матрицы A.

Задача 2. Найдите производную по матрице $A \in \mathbb{R}^{n \times n}$, $x^T A x > 0 \; \forall x \in \mathbb{R}^n, x \neq 0$

$$\frac{\partial}{\partial A} \log \det A.$$

Задача 3. Найдите производную по вектору $a \in \mathbb{R}^n$

$$\frac{\partial}{\partial a} \left(a^T \exp(a a^T) a \right),$$

где $\exp(B)$ — матричная экспонента, $B \in \mathbb{R}^{n \times n}$. Матричной экспонентой обозначают ряд

$$1 + \frac{B}{1!} + \frac{B^2}{2!} + \frac{B^3}{3!} + \frac{B^4}{4!} + \ldots = \sum_{k=0}^{\infty} \frac{B^k}{k!}.$$

Задача 4. В методе t-SNE, который широко используется для визуализации данных, задача построения низкоразмерных представлений объектов сводится к минимизации функционала

$$C = \sum_{i=1}^{\ell} \sum_{j \neq i}^{\ell} p_{ij} \log \frac{p_{ij}}{q_{ij}} \to \min_{\{z_1, \dots, z_\ell\}},$$

$$q_{ij} = \frac{(1 + ||z_i - z_j||^2)^{-1}}{\sum_{k \neq m}^{\ell} (1 + ||z_k - z_m||^2)^{-1}},$$

$$p_{i|j} = \frac{\exp(-||x_i - x_j||^2 / 2\sigma_j^2)}{\sum_{k \neq j} \exp(-||x_k - x_j||^2 / 2\sigma_j^2)}, \quad p_{i|i} = 0,$$

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2\ell},$$

где $x_i \in \mathbb{R}^D$, $z_i \in \mathbb{R}^d$. По непрерывности будем считать, что $0 \log \frac{0}{0} = 0$.

Найдите производную $\frac{\partial C}{\partial z_i}$, которая необходима для решения задачи градиентным спуском.

Задача 5. Рассмотрим задачу обучения линейной регрессии

$$Q(w) = (y - Xw)^T (y - Xw) \to \min_w$$

Будем решать её с помощью градиентного спуска. Допустим, мы находимся на некоторой итерации k, и хотим выполнить очередной шаг

$$w^{(k)} = w^{(k-1)} - \alpha \nabla_w Q(w^{(k-1)}).$$

При известных y, X, $w^{(k-1)}$ найдите длину шага α, при которой уменьшение значения функционала будет наибольшим:

$$Q(w^{(k-1)} - \alpha \nabla_w Q(w^{(k-1)})) \to \min_\alpha .$$