# 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 $A \in \mathbb{R}^{n \times n}$

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

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

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

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 $\alpha$, при которой уменьшение значения функционала будет наибольшим:

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

**Задача 6.** Найдите константу C, решающую следующую задачу ($0 < \tau < 1$ фиксировано):

$$\sum_{i=1}^{\ell} \rho_\tau(y_i - C) \to \min_C,$$

$$\rho_\tau(x) = \begin{cases} \tau x, & x > 0, \\ (\tau - 1)x, & x \leqslant 0. \end{cases}$$

**Задача 7.** Покажите, что если в задаче регрессии $p(y_i|x_i, w) = \frac{\alpha}{2} \exp\bigl(-\alpha \bigl|y_i - w^T x_i\bigr|\bigr)$ (распределение Лапласа, $\alpha$ фиксировано), то метод максимального правдоподобия эквивалентен оптимизации MAE для модели линейной регрессии.

**Задача 8.** Представим, что в некоторой задаче мы можем разбить признаки на $k$ непересекающихся групп (например, такие группы возникают при использовании one-hot кодирования — по одной группе бинарных признаков на каждый категориальный признак). Кроме того, мы хотим в модели линейной регрессии задать свой ненулевой коэффициент $L_2$—регуляризации для каждой группы. Какому априорному распределению на веса это будет соответствовать?

**Задача 9.** Убедитесь, что вы знаете ответы на следующие вопросы:

- Почему $L_1$-регуляризация производит отбор признаков?
- Почему коэффициент регуляризации нельзя подбирать по обучающей выборке?
- Что такое кросс-валидация, чем она лучше использования отложенной выборки?
- Почему категориальные признаки нельзя закодировать натуральными числами? Что такое one-hot encoding?

- Для чего нужно масштабировать матрицу объекты-признаки перед обучением моделей машинного обучения?
- Почему MSE чувствительно к выбросам?
- Что такое Huber Loss? В чем его преимущества по сравнению с MAE и MSE?
- Почему квантильная регрессия так называется?