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

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

Задача 4. 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:

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

Задача 5. 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 6. Рассмотрим задачу регрессии для выборки X^ℓ, каждый объект которой описывается d признаками. Пусть все признаки принимают значения из интервала $[0, 1]$. Разобьем гиперкуб $[0, 1]^d$ равномерно на $M = a^d$ областей, каждая из которых представляет собой подкуб со стороной $1/a$. Пронумеруем все области. Обозначим через B_m множество индексов объектов обучающей выборки, попавших в области с номером m, через n_m — число таких объектов, а через $r(x)$ — номер области, в которую попадает объект x. *Гистограммным* называется алгоритм $a(x)$, который возвращает для объекта x средний ответ по всем обучающим объектам из области $r(x)$:

$$a(x) = \begin{cases} \dfrac{1}{n_{r(x)}} \sum\limits_{i \in B_{r(x)}} y_i, & n_{r(x)} \neq 0; \\ 0, & n_{r(x)} = 0. \end{cases}$$

Покажите, что в случае с гистограммным алгоритмом leave-one-out-оценка (то есть оценка кросс-валидации с числом блоков $k = \ell$) для квадратичной функции потерь может быть выписана аналитически. Ее вычисление по найденной вами формуле должно требовать не более $O(\ell)$ операций.

Задача 7. Убедитесь, что вы знаете ответы на следующие вопросы:

- Почему L_1-регуляризация производит отбор признаков?
- Почему коэффициент регуляризации нельзя подбирать по обучающей выборке?
- Почему накладывать регуляризатор на свободный коэффициент w_0 может быть плохой идеей?
- 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?