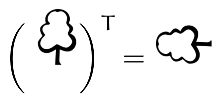
Yet Another Math for DS Course
Домашка №1

Матричное дифФфФфФференцирование
(продвинутая группа)



«Джек и бобовый стебель» (1890)

Добро пожаловать в первую домашку. Для вашего удобства она разбита на три части. В первой находятся совсем простые задачи. Если вы решите их идеально, вы наберёте 50 баллов из 100. Решая только первый раздел из каждой домашки, вы будете уверенно двигаться к троечке. Если вы претендуете на большее, для вас есть разделы с более сложными задачами.

Решение работы нужно сдать в виде pdf-файла. Решения должны быть оформлены на листочке аккуратным почерком либо затеханы на компьютере. Если у вас плохой почерк, домашка должна быть затехана. Затехать домашку можно в overleaf, typora, colab или другом любом удобном для вас сервисе.

Задачи на троечку

Если вы идеально решаете все задачи из этого раздела, вы получаете 50 баллов из 100.

Задача 1 (30 баллов). Найдите производные $\nabla_X f(X)$ следующих функций

1. $f(X) = \operatorname{tr}(AX^T XBX^{-T})$, где $A, X, B \in \mathbb{R}^{n\times n}$
2. $f(X) = \operatorname{tr}(AXB)$, где $A \in \mathbb{R}^{p\times m}, B \in \mathbb{R}^{n\times p}, X \in \mathbb{R}^{m\times n}$.
3. $f(X) = \det(X^T AX)$, где $A, X \in \mathbb{R}^{n\times n}$.

Задача 2 (10 баллов). Пусть $f(X) = \ln\det X$, где $X \in \mathbb{R}^{n\times n}$. Найдите производную $\nabla_X f(X)$.

Задача 3 (10 баллов). Решите следующую задачу матричной оптимизации. При решении считайте, что матрица A положительно определена.

$$f(x) = x^T A x - x^T b + c \to \min_x$$

Задачи на хор

Если вы идеально решаете ещё и этот раздел, вы получаете 70 баллов из 100.

Задача 4 (20 баллов). Предположим, что мы хотим обучить ridge-регрессию, но в качестве регуляризатора решили взять видоизменённую норму (S – некоторая фиксированная симметричная положительно-определённая матрица):

$$L(w) = (y - Xw)^T \cdot (y - Xw) + \|Sw\|_2^2.$$

Найдите градиент $\nabla_w L(w)$ получившейся функции потерь. Чему будет равно оптимальное значение весов в этом случае? Проверьте, что найденное значение действительно является точкой минимума.

Задачи на отл

Если вы идеально решаете ещё и этот раздел, вы получаете 90 баллов из 100.

Задача 5 (10 баллов). Рассмотрим симметричную матрицу $A \in \mathbb{R}^{n\times n}$ и ее спектральное разложение $A = Q\Lambda Q^T$. Пусть $\lambda \in \mathbb{R}^n$ - это диагональ матрицы Λ (то есть вектор, составленный из собственных значений A). Найдите производные:

1. $\nabla_\lambda \operatorname{tr}(A)$
2. $\nabla_Q \operatorname{tr}(A)$

Задача 6 (10 баллов). Пусть $x \in \mathbb{R}^n \setminus \{0\}$. Найдите $\nabla_x f(x)$ для

$$f(x) = \langle x, x\rangle^{\langle x, x\rangle}.$$

Задачи на десяточку

Если вы идеально решаете ещё и этот раздел, вы выбиваете 100 из 100. Вы большой молодец.

Задача 7 (10 баллов). Рассмотрим целевую функцию логистической регрессии

$$Q(w) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i \langle w, x_i \rangle)),$$

1. Найдите градиент ∇Q_w и упростите итоговое выражение таким образом, чтобы в нём участвовала сигмоидная функция

   $$\sigma(z) = \frac{1}{1 + \exp(-z)}.$$

   При решении данной задачи вам может понадобиться следующий факт (убедитесь, что он действительно выполняется):

   $$\sigma'(z) = \sigma(z)(1 - \sigma(z)).$$

2. Выпишите, как будет выглядеть шаг градиентного спуска.
3. Найдите вторую производную целевой функции по w.
4. Выпишите квадратичную аппроксимацию для $Q(w)$ в окрестности $w = 0$. Для этого разложите функцию потерь в ряд Тейлора до второго члена в окрестности точки $w = 0$. С какой задачей совпадает задача минимизации квадратичной аппроксимации?