

Quiz 2: матричные производные и градиентный спуск

Я понял, в чем ваша беда. Вы слишком серьезны. Умное лицо — это еще не признак ума, господа. Все глупости на Земле делаются именно с этим выражением лица. Вы улыбайтесь, господа. Улыбайтесь!

Барон Мюнхгаузен

Решите все задания. Все ответы должны быть обоснованы. Решения должны быть прописаны для каждого пункта. Рисунки должны быть чёткими и понятными. Все линии должны быть подписаны. При решении работы можно пользоваться чем угодно. Списывание карается обнулением работы. Удачи!

[3] Задание 1

У Аркадия есть матрицы A, B, X размера $n \times n$, $\det(X) \neq 0$. Помогите ей найти производную по матрице X функции

$$f(X) = \text{tr}(AXBX^{-1}),$$

[3] Задание 2

У Илона Маска есть одно наблюдение $x_1 = 1$, $y_1 = 2$. Он хочет обучить полносвязную нейронную сеть из одного нейрона с линейной функцией активации и l_2 регуляризацией на веса нейрона (Ridge-регрессию): $y = wx$. Для этого он использует функцию потерь

$$L(w) = \frac{1}{2n} \sum_{i=1}^n (y_i - wx_i)^2 + \frac{1}{2} \cdot w^2 \rightarrow \min_w.$$

Сделайте один шаг градиентного спуска. Используйте скорость обучения $\gamma = 0.1$. Вес w инициализирован нулём.

[2] Задание 3

- Как вы думаете, почему считается, что SGD лучше работает для оптимизации функций, имеющих больше одного экстремума?
- Предположим, что у функции потерь есть несколько локальных минимумов. Как можно адаптировать градиентный спуск так, чтобы он находил глобальный минимум чаще?

[2] Задание 4

Директор отдела по искусственному интеллекту Теслы Андрей Карпатый ищет минимум функции $f(x) = ax^2 + bx + c$, где $a > 0$, методом градиентного спуска. Андрей стартует из точки x_0 и настолько ленив, что не хочет делать больше одного шага. При каком значении длины шага γ Андрей за один шаг окажется точно в точке минимума?

[0.1] Задание 5

Перечислите три своих любимых сериала.