

Quiz 2: матричные производные и градиентный спуск

Во многой мудрости, много печали, и кто
умножает познания, умножает скорбь.

Царь Соломон (Екклесиаст 1:17,18)

[4] Задание 1

Пусть $y_i \geq 0$. Илон Маск обучает линейную регрессию $y = wx$ с квадратической логарифмической функцией ошибки (mean squared logarithmic error) и l_2 регуляризацией. Для этого он использует функцию потерь

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

Найдите значение градиента $\nabla_w L(w)$, выпишите формулу для шага градиентного спуска. Как думаете, в каких ситуациях на практике используют эту функцию ошибки? Зачем к y_i прибавляют единицу?

[3] Задание 2

Выберите все верные утверждения про градиентный спуск и предложите, как исправить остальные пункты, чтобы они стали правильными:

- а. Градиентный спуск используется для оптимизации весов в линейной регрессии с MAE, потому что аналитический минимум посчитать нельзя.
- б. Если взять длину шага $\eta_k = e^{-\lambda k}$, $\lambda > 0$, то НЕ гарантируется, что градиентный спуск сойдется.
- в. Для выпуклой функции нет смысла делать мульти-старт градиентного спуска.
- г. В SGD, скорее всего, потребуется больше итераций для сходимости, чем в обычном градиентном спуске.
- д. AdaGrad — это пример метода, который реализует инерцию (momentum).
- е. Чем больше размер mini-batch в SGD, тем больше дисперсия оценки градиента.

[3] Задание 3

Найдите $\frac{\partial y}{\partial x}$ и $\frac{\partial y}{\partial A}$ для функции

$$y = x^T A c, \quad A \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n, c \in \mathbb{R}^n$$

[0.1] Задание 4

Расскажи о своей любимой еде, почему она тебе нравится?