ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

Вспоминаем линейную алгебру. Скорости сходимости

МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

НЕДЕЛЯ 1

Даня Меркулов

Пётр Остроухов

Вспоминаем линейную алгебру. Скорости сходимости.

Семинар

Оптимизация для всех! ЦУ

Ключевые моменты лекции

Вспоминаем линейную алгебру

- Наивный matmul $\mathcal{O}(n^3)$, наивный matvec $\mathcal{O}(n^2)$

Вспоминаем линейную алгебру

- Наивный matmul $\mathcal{O}(n^3)$, наивный matvec $\mathcal{O}(n^2)$
- Все матрицы имеют SVD

$$A = U\Sigma V^T$$

Вспоминаем линейную алгебру

- Наивный matmul $\mathcal{O}(n^3)$, наивный matvec $\mathcal{O}(n^2)$
- Все матрицы имеют SVD

$$A = U\Sigma V^T$$

- $\text{tr}(ABCD) = \text{tr}(DABC) = \text{tr}(CDAB) = \text{tr}(BCDA)$ для любых матриц ABCD, если умножение определено.

Вспоминаем линейную алгебру

- Наивный matmul $\mathcal{O}(n^3)$, наивный matvec $\mathcal{O}(n^2)$
- Все матрицы имеют SVD

$$A = U\Sigma V^T$$

- $\text{tr}(ABCD) = \text{tr}(DABC) = \text{tr}(CDAB) = \text{tr}(BCDA)$ для любых матриц ABCD, если умножение определено.
- $\langle A, B \rangle = \text{tr}(A^T B)$

Скорости сходимости

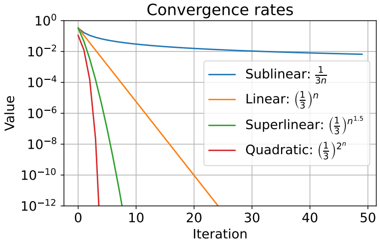

Рисунок 1. Иллюстрация различных скоростей сходимости

- Линейная (геометрическая, экспоненциальная) сходимость:

$$r_k \leq Cq^k, \quad 0 < q < 1, C > 0$$

Скорости сходимости

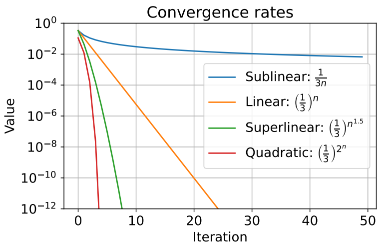

Рисунок 1. Иллюстрация различных скоростей сходимости

- Линейная (геометрическая, экспоненциальная) сходимость:

$$r_k \leq Cq^k, \quad 0 < q < 1, C > 0$$

- Любая сходящаяся последовательность, которая медленнее (быстрее) любой линейно сходящейся последовательности, имеет сублинейную (сверхлинейную) сходимость

Скорости сходимости

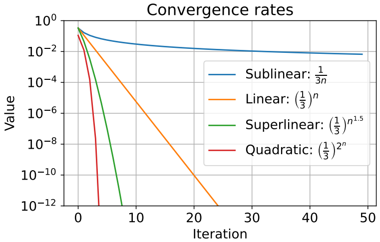

Рисунок 1. Иллюстрация различных скоростей сходимости

- Линейная (геометрическая, экспоненциальная) сходимость:

$$r_k \leq Cq^k, \quad 0 < q < 1, C > 0$$

- Любая сходящаяся последовательность, которая медленнее (быстрее) любой линейно сходящейся последовательности, имеет сублинейную (сверхлинейную) сходимость
- Инфимум всех $0 \leq q < 1$ таких, что $r_k \leq Cq^k$ называется **константой линейной сходимости**, и q^k называется **скоростью сходимости**.

Тест корней

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность неотрицательных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \sup_k \; r_k^{1/k}$$

- Если $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.

Тест корней

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность неотрицательных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k\to\infty} \sup_k \; r_k^{1/k}$$

- Если $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.

Тест корней

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность неотрицательных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \sup_k \; r_k^{1/k}$$

- Если $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.
- Если $q = 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет сублинейную сходимость.

Тест корней

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность неотрицательных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \sup_k \; r_k^{1/k}$$

- Если $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.
- Если $q = 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет сублинейную сходимость.
- Случай $q > 1$ невозможен.

Тест отношений

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность строго положительных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- Если существует q и $0 \le q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.

Тест отношений

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность строго положительных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- Если существует q и $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.

Тест отношений

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность строго положительных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- Если существует q и $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.
- Если q не существует, но $q = \lim_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой, не превышающей q.

Тест отношений

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность строго положительных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- Если существует q и $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.
- Если q не существует, но $q = \lim_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой, не превышающей q.
- Если $\lim_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет сублинейную сходимость.

Тест отношений

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность строго положительных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- Если существует q и $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.
- Если q не существует, но $q = \lim_{k \to \infty} \sup_k \frac{r_{k+1}}{r_k} < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой, не превышающей q.
- Если $\lim_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} = 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет сублинейную сходимость.
- Случай $\lim_{k \to \infty} \inf_k \frac{r_{k+1}}{r_k} > 1$ невозможен.

Тест отношений

Пусть $\{r_k\}_{k=m}^{\infty}$ - последовательность строго положительных чисел, сходящаяся к нулю, и пусть

$$q = \lim_{k \to \infty} \frac{r_{k+1}}{r_k}$$

- Если существует q и $0 \leq q < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой q.
- В частности, если $q = 0$, то $\{r_k\}_{k=m}^{\infty}$ имеет сверхлинейную сходимость.
- Если q не существует, но $q = \lim\limits_{k \to \infty} \sup_k \dfrac{r_{k+1}}{r_k} < 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет линейную сходимость с константой, не превышающей q.
- Если $\lim\limits_{k \to \infty} \inf_k \dfrac{r_{k+1}}{r_k} = 1$, то $\{r_k\}_{k=m}^{\infty}$ имеет сублинейную сходимость.
- Случай $\lim\limits_{k \to \infty} \inf_k \dfrac{r_{k+1}}{r_k} > 1$ невозможен.
- В остальных случаях (т.е., когда $\lim\limits_{k \to \infty} \inf_k \dfrac{r_{k+1}}{r_k} < 1 \leq \lim\limits_{k \to \infty} \sup_k \dfrac{r_{k+1}}{r_k}$) мы не можем сделать никаких конкретных утверждений о скорости сходимости $\{r_k\}_{k=m}^{\infty}$.

Задачи

Задача 1. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3\times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b.

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)

Проверьте простой 🐍 код после вашего интуитивного ответа.

Задача 1. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3\times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b.

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)
2. $\left(A_1 \left(A_2 \left(A_3 x\right)\right)\right)$ (справа налево)

Проверьте простой код после вашего интуитивного ответа.

Задача 1. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3 \times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b.

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)
2. $\left(A_1 \left(A_2 \left(A_3 x\right)\right)\right)$ (справа налево)
3. Не имеет значения

Проверьте простой 🐍 код после вашего интуитивного ответа.

Задача 1. Простая, но важная идея о матричных вычислениях.

Предположим, у вас есть следующее выражение

$$b = A_1 A_2 A_3 x,$$

где $A_1, A_2, A_3 \in \mathbb{R}^{3\times 3}$ - случайные квадратные плотные матрицы, и $x \in \mathbb{R}^n$ - вектор. Вам нужно вычислить b.

Какой способ лучше всего использовать?

1. $A_1 A_2 A_3 x$ (слева направо)
2. $(A_1 (A_2 (A_3 x)))$ (справа налево)
3. Не имеет значения
4. Результаты первых двух вариантов не будут одинаковыми.

Проверьте простой код после вашего интуитивного ответа.

Задача 2. Связь между Фробениусовой нормой и сингулярными значениями.

Пусть $A \in \mathbb{R}^{m \times n}$, и пусть $q := \min\{m, n\}$. Докажите, что

$$\|A\|_F^2 = \sum_{i=1}^{q} \sigma_i^2(A),$$

где $\sigma_1(A) \geq ... \geq \sigma_q(A) \geq 0$ - сингулярные значения матрицы A. Подсказка: используйте связь между Фробениусовой нормой и скалярным произведением и SVD.

Задача 3. Знайте свое скалярное произведение.

Упростите следующее выражение:

$$\sum_{i=1}^{n} \langle S^{-1} a_i, a_i \rangle,$$

где $S = \sum\limits_{i=1}^{n} a_i a_i^T, a_i \in \mathbb{R}^n,$ det$(S) \neq 0$

Задача 4. Простые скорости сходимости.

Определите сходимость (и её скорость) или расходимость следующих последовательностей:

- $r_k = \frac{1}{3^k}$

Задача 4. Простые скорости сходимости.

Определите сходимость (и её скорость) или расходимость следующих последовательностей:

- $r_k = \frac{1}{3^k}$
- $r_k = \frac{4}{3^k}$

Задача 4. Простые скорости сходимости.

Определите сходимость (и её скорость) или расходимость следующих последовательностей:

- $r_k = \frac{1}{3^k}$
- $r_k = \frac{4}{3^k}$
- $r_k = \frac{1}{k^{10}}$

Задача 4. Простые скорости сходимости.

Определите сходимость (и её скорость) или расходимость следующих последовательностей:

- $r_k = \frac{1}{3^k}$
- $r_k = \frac{4}{3^k}$
- $r_k = \frac{1}{k^{10}}$
- $r_k = 0.707^k$

Задача 4. Простые скорости сходимости.

Определите сходимость (и её скорость) или расходимость следующих последовательностей:

- $r_k = \frac{1}{3^k}$
- $r_k = \frac{4}{3^k}$
- $r_k = \frac{1}{k^{10}}$
- $r_k = 0.707^k$
- $r_k = 0.707^{2^k}$

Задача 5. Один тест проще, чем другой.

Определите сходимость (и её скорость) или расходимость следующей последовательности:

$$r_k = \frac{1}{k^k}$$

Задача 6. Сверхлинейно, но не квадратично.

Покажите, что следующая последовательность не имеет квадратичной сходимости.

$$r_k = \frac{1}{3^{k^2}}$$

А где это нужно в реальной жизни?

LoRA: Low-Rank Adaptation of Large Language Models (arXiv:2106.09685)

Поскольку современные LLM слишком большие, чтобы вместиться в память среднего пользователя, мы используем некоторые трюки, чтобы сделать их потребление памяти меньше. Одним из наиболее популярных трюков является LoRA (Low-Rank Adaptation of Large Language Models).

Предположим, у нас есть матрица $W \in \mathbb{R}^{d \times k}$ и мы хотим выполнить следующее обновление:

$$W = W_0 + \Delta W.$$

Основная идея LoRA состоит в том, чтобы разложить обновление ΔW на две низкоранговые матрицы:

$$W = W_0 + \Delta W = W_0 + BA, \quad B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times k},$$
$$rank(A) = rank(B) = r \ll \min\{d, k\}.$$

Проверьте ноутбук для примера реализации LoRA.

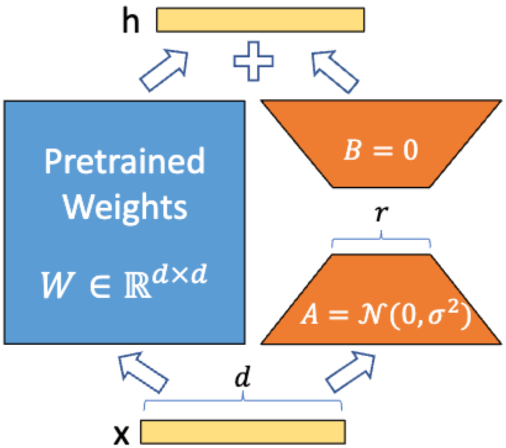

Рисунок 2. Иллюстрация LoRA