Спецкурс 2020/2021: “Геометрические и комбинаторные свойства матриц и аппроксимация”
Блок лекций “Сложность матриц и аппроксимация”
Лекция 0: “Вспомогательные факты из линейной алгебры”

24 ноября 2020 г.

Матрицы и операторы

Пусть $\mathbb{F}$ — поле (нам интересны $\mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_p$).
$\mathbb{F}^{m\times n}$ — пространство матриц размера $m \times n$ над $\mathbb{F}$.
Матрица $A \in \mathbb{F}^{m\times n}$ задаёт линейное отображение (линейный оператор) $\mathbb{F}^n \to \mathbb{F}^m$ по правилу $x \mapsto Ax$.
Линейное отображение $\mathcal{A}\colon V \to W$ после выбора базисов $V = \langle v_1, \dots, v_n\rangle$ и $W = \langle w_1, \dots, w_m\rangle$ можно отождествить с матрицей A: $\mathcal{A}v_j = \sum A_{i,j} w_i$.
Иногда *оператором* называют только линейное отображение пространства *в себя*: $\mathcal{A}\colon V \to V$. В этом случае задаётся один базис $V = \langle v_1, \dots, v_n\rangle$, матрица будет квадратной. При замене базиса матрица преобразуется по правилу $A' = C^{-1}AC$.
Композиция операторов $\leftrightarrow$ умножение матриц (в подходящих базисах):

$$A \in \mathbb{F}^{m\times n},\ B \in \mathbb{F}^{n\times k} \Rightarrow AB \in \mathbb{F}^{m\times k},$$

$$(AB)_{i,j} = \sum^{n} A_{i,s}B_{s,j} = \langle A_i, B^j\rangle.$$

Матрицы и операторы

Некоторые величины корректно определены для *операторов* $\mathcal{A}\colon V \to V$. Например, $\operatorname{tr}\mathcal{A} := \operatorname{tr} A = \sum A_{i,i}$ не зависит от выбора базиса, поскольку $\operatorname{tr}(C^{-1}AC) = \operatorname{tr}(CC^{-1}A) = \operatorname{tr} A$. Более общий инвариант — характеристический многочлен $\chi_{\mathcal{A}}(t) := \det(A - tE)$. Другой пример — *ранг*: $\operatorname{rank}\mathcal{A} = \operatorname{rank} A = \dim \mathrm{Im}(\mathcal{A})$.

Нас, однако же, будут интересовать в основном характеристики, не инвариантные относительно замены базиса и потому относящиеся к *матрицам*, а не операторам. Пример: $\operatorname{rank}_\varepsilon(A)$ — минимально возможный ранг, который можно получить, изменив каждый элемент матрицы A не более чем на ε.

Нормы в пространствах векторов и матриц

ℓ_p^n — пространство $\mathbb{R}^n$ с нормой

$$\|x\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, & 1 \leqslant p < \infty, \\ \max_{1 \leqslant i \leqslant n} |x_i|, & p = \infty. \end{cases}$$

В евклидовом случае пишем сокращённо $|x| := \|x\|_2$.
Пусть $A \in \mathbb{R}^{m \times n}$. Определим нормы:

- норма Фробениуса $\|A\|_F := (\sum A_{i,j}^2)^{1/2}$;
- (p, q)-нормы $\|A\|_{p \to q} := \max\limits_{\|x\|_p \leqslant 1} \|Ax\|_q$;
- в частности, спектральная норма $\|A\|_{2 \to 2} := \max\limits_{|x| \leqslant 1} |Ax|$;
- в частности, максимум $\|A\|_\infty := \|A\|_{1 \to \infty} = \max_{i,j} |A_{i,j}|$;

Операторные нормы обладают свойством $\|AB\| \leqslant \|A\|\|B\|$; не все нормы таковы.

Ранг матрицы

Эквивалентные определения ранга матрицы $A \in \mathbb{F}^{m\times n}$:

- размерность образа $\dim \mathrm{Im}\mathcal{A}$ оператора с матрицей A (т.е. ранг это операторное понятие);
- размерность образа $\dim\{Ax \colon x \in \mathbb{F}^n\}$;
- размерность пространства столбцов $\dim\langle A^j\rangle$;
- размерность пространства строк $\dim\langle A_i\rangle$; отсюда $\operatorname{rank} A = \operatorname{rank} A^t$;
- максимальный размер невырожденного минора: $\max\{|I| = |J| \colon \det A[I, J] \neq 0\}$;
- минимальное число одноранговых матриц (т.е. вида $R_{i,j} = a_i b_j$) в представлении $A = R^{(1)} + R^{(2)} + \ldots + R^{(r)}$;
- минимальная размерность r, в которой найдутся вектора $x_i \in \mathbb{R}^r$, $y_j \in \mathbb{R}^r$, такие что $A_{i,j} = \langle x_i, y_j\rangle$ — **упражнение**.

Собственные числа

Число λ называется собственным числом оператора $\mathcal{A}$, если найдётся ненулевой вектор v, для которого $\mathcal{A}v = v$. Собственные числа — корни многочлена $\chi_{\mathcal{A}}(t)$.

Пусть $\mathbb{F} = \mathbb{R}$, в пространстве V введено скалярное произведение $\langle \cdot, \cdot \rangle$. Оператор $\mathcal{A}\colon V \to V$ называется самосопряжённым, если $\langle Ax, y \rangle = \langle x, Ay \rangle$. Это равносильно тому, что его матрица симметрична: $A = A^t$ (не важно в каком базисе).

Самосопряжённый оператор диагонализуем в ортонормированном базисе. Следовательно, симметричная матрица представляется в виде $A = U^t \Lambda U$ с ортогональной U и диагональной Λ.

Сингулярное разложение

Сингулярное разложение, Singular Value Decomposition (SVD), матрицы $A \in \mathbb{R}^{m \times n}$:

$$A = U \Sigma V^t,$$

где U и V — ортогональные матрицы, Σ — матрица размера $m \times n$, на диагонали которой стоят неотрицательные числа $\Sigma_{i,i} = \sigma_i$, причём

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_{\min(m,n)} \geqslant 0,$$

а вне диагонали — нули. SVD всегда существует. То есть, любой оператор работает так: "поворачивает" вектор $(x \mapsto Vx)$, растягивает по осям, потом опять "поворачивает".
Числа (σ_j) определены однозначно; они называются *сингулярными числами* матрицы A. В отличие от собственных чисел, они определены (и вещественны!) для любой матрицы.

Сингулярное разложение

Определим нормы Шаттена (Schatten):

$$\|A\|_{S_p} := \|(\sigma_j(A))_{j=1}^{\min(m,n)}\|_p, \quad 1 \leqslant p \leqslant \infty.$$

Упражнения.

- Докажите существование SVD и однозначность Σ, рассмотрев матрицу AA^t. Выведите, что $\sigma_j^2(A) = \lambda_j(AA^t)$.
- Покажите, что сингулярные числа унитарно-инвариантны, т.е. $\sigma_j(U'A) = \sigma_j(AV')$ для любых ортогональных U', V'; сформулируйте следствие для S_p-норм.
- Докажите, что S_2-норма равна норме Фробениуса: $\sum \sigma_j^2(A) = \sum A_{i,j}^2$, а S_∞-норма равна $2 \to 2$ норме.
- Покажите, что $\operatorname{tr} A \leqslant \sum |A_{i,i}| \leqslant \|A\|_{S_1}$?
- Проверьте, что $\|A\|_{S_1} := \sum \sigma_j(A)$ является нормой (т.е. $\|A+B\|_{S_1} \leqslant \|A\|_{S_1} + \|B\|_{S_1}$);
- То же для S_p-нормы.

Сингулярное разложение и ранг

Пусть $\sigma_1(A) \geqslant \ldots \geqslant \sigma_r(A) > \sigma_{r+1}(A) = \ldots = 0$. Тогда $\operatorname{rank} A = r$. То есть, ранг = кол-во ненулевых сингулярных чисел.

Ранг — разрывная функция на $\mathbb{R}^{m\times n}$, с ней сложно работать. Сингулярные числа позволяют связать ранг с непрерывными характеристиками матрицы.
Пример. Мы знаем, что $\|\sigma(A)\|_\infty = \|A\|_{2\to 2}$ и $\|\sigma(A)\|_2 = \|A\|_F$. Поскольку $\|x\|_2 \leqslant r^{1/2}\|x\|_\infty$ для $x \in \mathbb{R}^r$, получаем следствие:

$$\operatorname{rank} A \geqslant \left(\frac{\|A\|_F}{\|A\|_{2\to 2}}\right)^2.$$

Отображение $A \mapsto \sigma(A)$ липшицево в евклидовой норме (Wielandt-Hoffman):

$$|\sigma(A) - \sigma(B)| = \Big(\sum_k (\sigma_k(A) - \sigma_k(B))^2\Big)^{1/2} \leqslant \|A - B\|_F.$$

Сингулярное разложение и ранг

Theorem (Ecckart-Young, 1936)

$$\min_{\operatorname{rank} B\leqslant r} \|A-B\|_F = \Big(\sum_{k>r} \sigma_k^2(A)\Big)^{1/2}.$$

Напомним доказательство теоремы. Общий случай сводится к квадратным матрицам (всегда можно дополнить нулями). В силу унитарной инвариантности можно считать матрицу A диагональной: $A=\Sigma=\mathrm{diag}(\sigma_1,\ldots,\sigma_n)$. Ясно, что для оценки сверху можно взять $B=\mathrm{diag}(\sigma_1,\ldots,\sigma_r,0,\ldots)$.
Оценка снизу. Величина $\|A-B\|_F^2$ равна $\sum_j |A^j-B^j|^2$ (по столбцам). Рассмотрим пространство L_r, натянутое на столбцы B^j, тогда $|A^j-B^j|$ не меньше расстояния $\rho(A^j,L_r)$, и

$$\min_{\operatorname{rank} B\leqslant r} \|A-B\|_F^2 = \min_{\dim L_r\leqslant r} \sum_{j=1}^n \rho^2(A^j,L_r) = \min_{\dim L_r\leqslant r} \sum_{j=1}^n \sigma_j^2\rho^2(e_j,L_r).$$

Eckart-Young, продолжение

Выберем в L_r ортонормированный базис $\{v_1, \dots, v_r\}$. Тогда проекция на L_r имеет вид $P_{L_r}x = \sum_{k=1}^r v_k \langle v_k, x\rangle$, откуда

$$w_j := \rho(e_j, L_r)^2 = |e_j|^2 - |P_{L_r} e_j|^2 = 1 - \sum_{k=1}^{r} v_{k,j}^2,$$

$$\sum_{j=1}^{n} \sigma_j^2 \rho^2(e_j, L_r) = \sum_{j=1}^{n} w_j \sigma_j^2.$$

Проанализируем коэффициенты w_j, $j = 1, \dots, n$. Имеем:

(i) $w_j \in [0, 1]$;

(ii) $\sum_{j=1}^n w_j = n - r$, так как $\sum_{j=1}^n \sum_{k=1}^r v_{k,j}^2 = r$.

В силу монотонности σ_j, при ограничениях (i), (ii) сумма $\sum w_j \sigma_j^2$ будет минимальна, если $w_1 = \dots = w_r = 0$ и $w_{r+1} = \dots = w_n = 1$. При таких w_j получаем как раз оценку $\sum_{j>r} \sigma_j^2$.

Положительная определённость

Пусть $A \in \mathbb{R}^{n \times n}$ — симметричная матрица.
Назовём A неотрицательно определённой (positive semi-definite, PSD), если $x^t A x \geqslant 0$ для любого вектора x. Положительно определённая, если $x^t A x > 0$ для $x \neq 0$. Обозначение: $A \geqslant 0$ (соотв., $A > 0$).
Можно ввести частичный порядок на матрицах:
$A \geqslant B \Leftrightarrow A - B \geqslant 0$.

Упражнения.

- Если $A \geqslant 0$, то $A = BB^t$ для некоторой B (т.е. A это *матрица Грама*).
- Если $A \geqslant 0$, то $A = R^1 + \ldots + R^r$, где R^j имеют вид aa^t.
- Если $A \geqslant 0$ и $B \geqslant 0$, то $A \circ B \geqslant 0$ и $A \otimes B$. Где $A \circ B$ — произведение Шура-Адамара, поэлементное, $(A \circ B)_{i,j} = A_{i,j} B_{i,j}$. Произведение Кронекера $A \otimes B$ состоит из блоков вида $A_{i,j} B$.

Разное

Матрица = функция двух аргументов (дискретных).
Если аргументы $x, y \in \{0, 1\}^n$, эта структура может быть удобна.
Например, для $M(x, y) = p(x, y)$, где p — полином из m мономов, имеем $\operatorname{rank} M \leqslant m$ (**упражнение**!).