Спецкурс 2020/2021: “Геометрические и комбинаторные свойства матриц и аппроксимация” Блок лекций “Сложность матриц и аппроксимация” Лекция 2: “Коммуникационная сложность (продолжение)”

1 ноября 2020 г.

Меры сложности функции $f\colon \mathcal{X}\times\mathcal{Y}\to\{0,1\}$:

- $C(f)$ — коммуникационная сложность (детерминированная модель);
- $R(f)$ — сложность с "ограниченной ошибкой": $\mathsf{P}(Q(x,y)=f(x,y))\geqslant\frac{2}{3}, \quad \forall x,y.$
- $U(f)$ — сложность с "неограниченной ошибкой": $\mathsf{P}>\frac{1}{2}$. (Название не вполне удачное, смысл в том, что мы не ограничиваем ошибку и она может быть сколь угодно близка к $\frac{1}{2}$.)

Меры сложности функции $f\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$:

- $C(f)$ — коммуникационная сложность (детерминированная модель);
- $R(f)$ — сложность с "ограниченной ошибкой": $\mathsf{P}(Q(x,y) = f(x,y)) \geqslant \frac{2}{3}, \quad \forall x, y.$
- $U(f)$ — сложность с "неограниченной ошибкой": $\mathsf{P} > \frac{1}{2}$. (Название не вполне удачное, смысл в том, что мы не ограничиваем ошибку и она может быть сколь угодно близка к $\frac{1}{2}$.)

Примеры: $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$,

- EQ — равенство $x = y$;
- DISJ — дизъюнктность: отождествим x, y с подмножествами $\{1, \ldots, n\}$ и положим $\mathrm{DISJ}(x,y) = 0$ для $x \cap y = \varnothing$, и 1 иначе;
- $\mathrm{IP}(x,y) = \sum x_i y_i \bmod 2$.

Меры сложности функции $f\colon \mathcal{X}\times\mathcal{Y}\to\{0,1\}$:

- $C(f)$ — коммуникационная сложность (детерминированная модель);
- $R(f)$ — сложность с "ограниченной ошибкой": $\mathrm{P}(Q(x,y)=f(x,y))\geqslant\frac{2}{3},\quad \forall x,y.$
- $U(f)$ — сложность с "неограниченной ошибкой": $\mathrm{P}>\frac{1}{2}$. (Название не вполне удачное, смысл в том, что мы не ограничиваем ошибку и она может быть сколь угодно близка к $\frac{1}{2}$.)

Примеры: $\mathcal{X}=\mathcal{Y}=\{0,1\}^n$,

- EQ — равенство $x=y$;
- DISJ — дизъюнктность: отождествим x,y с подмножествами $\{1,\dots,n\}$ и положим $\mathrm{DISJ}(x,y)=0$ для $x\cap y=\varnothing$, и 1 иначе;
- $\mathrm{IP}(x,y)=\sum x_iy_i \bmod 2$.

f	$\mathrm{C}(f)$	$R(f)$	$U(f)$
EQ	$\asymp n$	$O(\log n)$	$O(1)$
DISJ	$\asymp n$	$\asymp n$	$O(\log n)$
IP	$\asymp n$	$\asymp n$	$\asymp n$

$C(\mathrm{DISJ}) \asymp n$

Матрица DISJ_n невырождена.

$C(\mathrm{DISJ}) \asymp n$

Матрица DISJ_n невырождена.

	$1 \in x$	$1 \notin x$
$1 \in x$	0	X
$1 \notin x$	X	$*$

При этом $X \sim \mathrm{DISJ}_{n-1}$.

Индукция по n.

$C(\mathrm{DISJ}) \asymp n$

Матрица DISJ_n невырождена.

	$1 \in x$	$1 \notin x$
$1 \in x$	0	X
$1 \notin x$	X	*

При этом $X \sim \mathrm{DISJ}_{n-1}$.
Индукция по n.

Таким образом, $\operatorname{rank} \mathrm{DISJ}_n = 2^n$,

$$C(\mathrm{DISJ}_n) \geqslant \log_2 \operatorname{rank}(\mathrm{DISJ}_n) = n.$$

$U(\mathrm{DISJ}) \ll \log n$

Анне дано множество $x \subset \{1, \ldots, n\}$, Борису — $y \in \{1, 2, \ldots, n\}$. Они должны определить, пересекаются ли $x \cap y$ (“appointment scheduling problem”).

$U(\mathrm{DISJ}) \ll \log n$

Анне дано множество $x \subset \{1, \dots, n\}$, Борису — $y \in \{1, 2, \dots, n\}$. Они должны определить, пересекаются ли $x \cap y$ (“appointment scheduling problem”).

Протокол: Анна выбирает случайный элемент $a \in x$ и отправляет его Борису.

$U(\mathrm{DISJ}) \ll \log n$

Анне дано множество $x \subset \{1, \dots, n\}$, Борису — $y \in \{1, 2, \dots, n\}$. Они должны определить, пересекаются ли $x \cap y$ (“appointment scheduling problem”).

Протокол: Анна выбирает случайный элемент $a \in x$ и отправляет его Борису.

Если Борис обнаруживает, что $a \in y$, то выдаёт ответ “пересечение непусто”. Иначе с вероятностью $\frac{1}{2} + \varepsilon$ выдаёт ответ “пусто” (соотв., с вероятностью $\frac{1}{2} - \varepsilon$ “непусто”).

$U(\mathrm{DISJ}) \ll \log n$

Анне дано множество $x \subset \{1, \ldots, n\}$, Борису — $y \in \{1, 2, \ldots, n\}$. Они должны определить, пересекаются ли $x \cap y$ (“appointment scheduling problem”).

Протокол: Анна выбирает случайный элемент $a \in x$ и отправляет его Борису.

Если Борис обнаруживает, что $a \in y$, то выдаёт ответ “пересечение непусто”. Иначе с вероятностью $\frac{1}{2} + \varepsilon$ выдаёт ответ “пусто” (соотв., с вероятностью $\frac{1}{2} - \varepsilon$ “непусто”).

Если $x \cap y = \varnothing$, то вероятность успеха $\frac{1}{2} + \varepsilon$.

$U(\mathrm{DISJ}) \ll \log n$

Анне дано множество $x \subset \{1, \dots, n\}$, Борису — $y \in \{1, 2, \dots, n\}$. Они должны определить, пересекаются ли $x \cap y$ (“appointment scheduling problem”).

Протокол: Анна выбирает случайный элемент $a \in x$ и отправляет его Борису.

Если Борис обнаруживает, что $a \in y$, то выдаёт ответ “пересечение непусто”. Иначе с вероятностью $\frac{1}{2} + \varepsilon$ выдаёт ответ “пусто” (соотв., с вероятностью $\frac{1}{2} - \varepsilon$ “непусто”).

Если $x \cap y = \varnothing$, то вероятность успеха $\frac{1}{2} + \varepsilon$.

Если $|x \cap y| = k > 0$, то с вероятностью $q = k/|x| \geqslant 1/n$ Анна выберет элемент $a \in x \cap y$ и мы придём к успеху. Следовательно, вероятность успеха равна

$$q + (1-q)(\frac{1}{2} - \varepsilon) = \frac{1}{2} + q(\frac{1}{2} + \varepsilon) - \varepsilon \geqslant \frac{1}{2} + \frac{1}{2n} - \varepsilon,$$

что больше $\frac{1}{2}$ при достаточно малом ε.

Сигнум-ранг

Сигнум матрица = матрица с элементами ± 1.
Сигнум рангом (signum rank) сигнум-матрицы $S \in \{-1, 1\}^{m \times n}$ назовём минимальный ранг матриц A, таких что $\operatorname{sign} A = S$:

$$\operatorname{rank}_{\pm}(S) := \min\{\operatorname{rank} A \colon \operatorname{sign} A_{i,j} \equiv S_{i,j}\}.$$

Сигнум-ранг

Сигнум матрица = матрица с элементами ± 1.
Сигнум рангом (signum rank) сигнум-матрицы $S \in \{-1, 1\}^{m \times n}$ назовём минимальный ранг матриц A, таких что $\operatorname{sign} A = S$:

$$\operatorname{rank}_{\pm}(S) := \min\{\operatorname{rank} A \colon \operatorname{sign} A_{i,j} \equiv S_{i,j}\}.$$

Геометрическая интерпретация сигнум-ранга:

- Наборы векторов $X = \{x_1, \ldots, x_m\}$ и $Y = \{y_1, \ldots, y_n\}$ в $\mathbb{R}^d$ реализуют сигнум матрицу S, если

  $$\operatorname{sign}\langle x_i, y_j \rangle = S_{i,j}, \quad \forall i, j.$$

  $\operatorname{rank}_{\pm}(S)$ — минимальная размерность d, в которой существует реализация (X, Y) матрицы S.

Сигнум-ранг

Сигнум матрица = матрица с элементами ± 1.
Сигнум рангом (signum rank) сигнум-матрицы $S \in \{-1, 1\}^{m \times n}$ назовём минимальный ранг матриц A, таких что $\operatorname{sign} A = S$:

$$\operatorname{rank}_{\pm}(S) := \min\{\operatorname{rank} A \colon \operatorname{sign} A_{i,j} \equiv S_{i,j}\}.$$

Геометрическая интерпретация сигнум-ранга:

- Наборы векторов $X = \{x_1, \dots, x_m\}$ и $Y = \{y_1, \dots, y_n\}$ в $\mathbb{R}^d$ реализуют сигнум матрицу S, если

$$\operatorname{sign}\langle x_i, y_j \rangle = S_{i,j}, \quad \forall i, j.$$

  $\operatorname{rank}_{\pm}(S)$ — минимальная размерность d, в которой существует реализация (X, Y) матрицы S. Замечание: можно считать x_i и y_i *единичными* векторами и что они находятся в *общем положении* (т.е. любые d линейно независимы). Почему?

Сигнум-ранг

Сигнум матрица = матрица с элементами ± 1.
Сигнум рангом (signum rank) сигнум-матрицы $S \in \{-1, 1\}^{m \times n}$ назовём минимальный ранг матриц A, таких что $\operatorname{sign} A = S$:

$$\operatorname{rank}_{\pm}(S) := \min\{\operatorname{rank} A \colon \operatorname{sign} A_{i,j} \equiv S_{i,j}\}.$$

Геометрическая интерпретация сигнум-ранга:

- Наборы векторов $X = \{x_1, \ldots, x_m\}$ и $Y = \{y_1, \ldots, y_n\}$ в $\mathbb{R}^d$ реализуют сигнум матрицу S, если

  $$\operatorname{sign}\langle x_i, y_j \rangle = S_{i,j}, \quad \forall i, j.$$

  $\operatorname{rank}_{\pm}(S)$ — минимальная размерность d, в которой существует реализация (X, Y) матрицы S. Замечание: можно считать x_i и y_i *единичными* векторами и что они находятся в *общем положении* (т.е. любые d линейно независимы). Почему?
- Вместо векторов x_i и y_i можно говорить о точках P_i и гиперплоскостях H_i, разделяющих пространство на положительное полупространство H_i^+ и отрицательное H_i^-. Тогда $P_i \in H_i^+$, если $S_{i,j} = 1$ и $P_i \in H_i^-$, если $S_{i,j} = -1$.

U и Сигнум-ранг

Theorem (Paturi, Simon, 1986)

$$\lceil \log_2 \operatorname{rank}_{\pm}(f) \rceil \leqslant U(f) \leqslant \lceil \log_2 \operatorname{rank}_{\pm}(f) \rceil + 1.$$

U и Сигнум-ранг

Theorem (Paturi, Simon, 1986)

$$\lceil \log_2 \operatorname{rank}_\pm(f) \rceil \leqslant U(f) \leqslant \lceil \log_2 \operatorname{rank}_\pm(f) \rceil + 1.$$

Упрощённый протокол. Анна посылает Борису сообщение α_i с вероятностью $p_i(x)$, $1 \leqslant i \leqslant m$. Борис, получив α_i, выдаёт 1 с вероятностью $q_i(y)$ и 0 с вероятностью $1 - q_i(y)$. Таким образом, упрощённый протокол задаётся функциями

$$p_1, \dots, p_m \colon \mathcal{X} \to [0,1], \ \sum_{i=1}^{m} p_i(x) \equiv 1, \quad q_1, \dots, q_m \colon \mathcal{Y} \to [0,1].$$

Чему равна сложность протокола?

U и Сигнум-ранг

Theorem (Paturi, Simon, 1986)

$$\lceil \log_2 \operatorname{rank}_\pm(f) \rceil \leqslant U(f) \leqslant \lceil \log_2 \operatorname{rank}_\pm(f) \rceil + 1.$$

Упрощённый протокол. Анна посылает Борису сообщение α_i с вероятностью $p_i(x)$, $1 \leqslant i \leqslant m$. Борис, получив α_i, выдаёт 1 с вероятностью $q_i(y)$ и 0 с вероятностью $1 - q_i(y)$. Таким образом, упрощённый протокол задаётся функциями

$$p_1, \ldots, p_m \colon \mathcal{X} \to [0,1], \ \sum_{i=1}^{m} p_i(x) \equiv 1, \quad q_1, \ldots, q_m \colon \mathcal{Y} \to [0,1].$$

Чему равна сложность протокола?

Сложность равна $\lceil \log_2 m \rceil$. Сообщения α_i = бинарные строки, кодирующие m вариантов.

Теорема Paturi–Simon

Пусть f принимает (для удобства) значения ± 1. Отождествим f с сигнум матрицей $S = S_f$. Анна знает i, Борис знает j, нужно вычислить $S_{i,j}$.

Теорема Paturi–Simon

Пусть f принимает (для удобства) значения ± 1. Отождествим f с сигнум матрицей $S = S_f$. Анна знает i, Борис знает j, нужно вычислить $S_{i,j}$.

Пусть есть реализация сигнум-матрицы S в виде системы векторов x_i, y_j. Построим упрощённый протокол.

Теорема Paturi–Simon

Пусть f принимает (для удобства) значения ± 1. Отождествим f с сигнум матрицей $S = S_f$. Анна знает i, Борис знает j, нужно вычислить $S_{i,j}$.

Пусть есть реализация сигнум-матрицы S в виде системы векторов x_i, y_j. Построим упрощённый протокол.

"Причешем" реализацию: повышением размерности на 1 можно добиться выполнения неравенств:

$$x_{i,k} \geqslant 0, \quad \sum_{k=1}^{d} x_{i,k} = 1, \quad y_{j,k} \in [-\frac{1}{2}, \frac{1}{2}].$$

Итак, Анне выдали точку x_i, Борису — y_j. Анна отравляет сообщение α_k с вероятностью $x_{i,j}$. Борис отвечает, получив α_k:

$$\begin{cases} 1, & \text{с вероятностью } \frac{1}{2} + y_{j,k}, \\ -1, & \text{с вероятностью } \frac{1}{2} - y_{j,k}. \end{cases}$$

Теорема Paturi–Simon (продолжение)

Тогда разность вероятностей P(ответ=1) – P(ответ=0) равна

$$\sum_{k=1}^{d} x_{k,j}\left(\frac{1}{2}+y_{k,j}\right)-\sum_{k=1}^{d} x_{k,j}\left(\frac{1}{2}-y_{k,j}\right)=\langle x_i, y_j\rangle.$$

Теорема Paturi–Simon (продолжение)

Тогда разность вероятностей P(ответ=1) – P(ответ=0) равна

$$\sum_{k=1}^{d} x_{k,j}(\frac{1}{2} + y_{k,j}) - \sum_{k=1}^{d} x_{k,j}(\frac{1}{2} - y_{k,j}) = \langle x_i, y_j \rangle.$$

Значит, если $S_{i,j} = 1$, то $\langle x_i, y_j \rangle > 0$ и вероятность ответа "1" (правильного) выше, чем вероятность неправильного ответа.

Теорема Paturi–Simon (продолжение)

Тогда разность вероятностей P(ответ=1) – P(ответ=0) равна

$$\sum_{k=1}^{d} x_{k,j}(\frac{1}{2} + y_{k,j}) - \sum_{k=1}^{d} x_{k,j}(\frac{1}{2} - y_{k,j}) = \langle x_i, y_j \rangle.$$

Значит, если $S_{i,j} = 1$, то $\langle x_i, y_j \rangle > 0$ и вероятность ответа “1” (правильного) выше, чем вероятность неправильного ответа.

Обратно, вектора вероятностей упрощённого протокола задают реализацию сигнум-матрицы подходящей размерности (проверьте!). Теорема доказана.

Теорема Paturi–Simon (продолжение)

Тогда разность вероятностей P(ответ=1) – P(ответ=0) равна

$$\sum_{k=1}^{d} x_{k,j}\left(\frac{1}{2}+y_{k,j}\right)-\sum_{k=1}^{d} x_{k,j}\left(\frac{1}{2}-y_{k,j}\right)=\langle x_i, y_j\rangle.$$

Значит, если $S_{i,j}=1$, то $\langle x_i, y_j\rangle > 0$ и вероятность ответа “1” (правильного) выше, чем вероятность неправильного ответа.

Обратно, вектора вероятностей упрощённого протокола задают реализацию сигнум-матрицы подходящей размерности (проверьте!). Теорема доказана. Или нет?

Теорема Paturi–Simon (продолжение)

Тогда разность вероятностей P(ответ=1) – P(ответ=0) равна

$$\sum_{k=1}^{d} x_{k,j}(\frac{1}{2}+y_{k,j})-\sum_{k=1}^{d} x_{k,j}(\frac{1}{2}-y_{k,j})=\langle x_i, y_j\rangle.$$

Значит, если $S_{i,j}=1$, то $\langle x_i, y_j\rangle > 0$ и вероятность ответа "1" (правильного) выше, чем вероятность неправильного ответа.

Обратно, вектора вероятностей упрощённого протокола задают реализацию сигнум-матрицы подходящей размерности (проверьте!).
Теорема доказана. Или нет?
Нужно доказать, что **задача коммуникации с неограниченной ошибкой сводится к упрощённым протоколам!**
Это сделано в той же работе Paturi и Simon.

Теорема Paturi–Simon (продолжение)

Как по обычному протоколу Q построить упрощённый протокол?

Теорема Paturi–Simon (продолжение)

Как по обычному протоколу Q построить упрощённый протокол?
Пусть H — множество всех возможных *историй*, т.е. последовательностей передаваемых Анной и Борисом сообщений:

$$h = (\alpha_1, \beta_1, \ldots, \beta_n),$$

где α_1 — сообщение Анны, β_1 — ответ Бориса и т.д. Последнее сообщение β_n состоит из одного бита и содержит ответ (для определенности считаем, что его посылает Борис). Обозначим это последнее сообщение через h_{last}.
В упрощённом протоколе Анна будет передавать $h \in H$ или специальное сообщение γ.
Грубо говоря, Анна предполагает, что Борис отправляет свои сообщения с равными вероятностями, рассчитывает историю коммуникации и отправляет её в соответствии со своими вероятностями. Борис "корректирует" вероятность и выдаёт ответ.

Теорема Paturi–Simon (продолжение)

Анна и Борис действуют в соответствии с вероятностным протоколом. Для Анны есть распределения: $p(\alpha_1|x)$ — вероятности отправить α_1 для заданного x, распределение $p(\alpha_2|\alpha_1, \alpha_1, x)$ и т.д. У Бориса это $q(\beta_1|\alpha_1, y)$ и т.д.

При заданных x, y вероятность истории h получается произведением вероятностей для Анны и Бориса:

$$P_A(h, x) = p(\alpha_1|x)p(\alpha_2|\beta_1, x) \cdots p(\alpha_n|...),$$

$$P_B(h, y) = q(\beta_1|y)q(\alpha_2|\beta_1, y) \cdots q(\beta_n|...).$$

Ключевой момент в том, что P_A не зависит от y, а P_B не зависит от x. Вероятность, того, что получим ответ b (при фиксированных x, y) равна

$$P(b|x, y) = \sum_{h\colon\, h_{\mathrm{last}}=b} P_A(h, x)P_B(h, y). \qquad (*)$$

Положим

$$d_x^b := \sum_{h\colon\, h_{\mathrm{last}}=b} P_A(h, x), \quad b \in \{0, 1\}, \quad d := \max_x d_x^1.$$

Теорема Paturi–Simon (продолжение)

Зададим вероятности для упрощённого протокола. Для Анны нужно задать вероятности отправить историю h или специальный символ γ:

$$p'(h|x) = \begin{cases} \frac{1}{2d} P_A(h,x), & h_{\text{last}} = 1, \\ \frac{1}{2d_x^0} P_A(h,x), & h_{\text{last}} = 0, \\ \frac{1}{2} - \frac{d_x^1}{2d}, & h = \gamma. \end{cases}$$

Нетрудно убедиться, что $\forall x \ \sum_h p'(h|x) = 1$. Для Бориса задаём вероятность ответа в зависимости от y и полученного h:

$$q'(1|y,h) = \begin{cases} P_B(h,y), & h_{\text{last}} = 1, \\ 1 - \frac{1}{2d}, & h_{\text{last}} = 0, \\ 0, & h = \gamma. \end{cases}$$

Здесь нужно убедиться, что $q' \in [0,1]$, для этого нужно, чтобы $d \geqslant 1/2$. Почему это так: при некотором x вероятность ответа 1 больше $1/2$ (иначе функция тождественно нулевая, не о чем говорить). Но эта вероятность не больше d_x^1, что видно из формулы (*).

Теорема Paturi–Simon (окончание)

Проверим, что упрощённый протокол даёт тот же результат. Фиксируем (x, y). Предположим, по обычному протоколу был ответ 1 (с вероятностью $P(1|x, y) > 1/2$). Докажем, что вероятность по упрощённому тоже $> 1/2$. Она равна:

$$\sum_{h_{\text{last}}=1} \frac{1}{2d} P_A(h, x) P_B(h, y) + \sum_{h_{\text{last}}=0} \frac{1}{2d_x^0} P_A(h, x)(1 - \frac{1}{2d}) =$$

$$= \frac{1}{2d} P(1|x, y) + \frac{1}{2} \cdot (1 - \frac{1}{2d}) = \frac{1}{2} + \frac{1}{2d}(P(1|x, y) - \frac{1}{2}) > \frac{1}{2}.$$

Аналогично разбирается случай $P(0|x, y) > 1/2$.

Отметим особо, что вероятности изменились! Остался неизменным лишь знак $\mathrm{sign}(P - 1/2)$. Следовательно, это рассуждение позволяет свести к упрощённому протоколу только коммуникацию с неограниченной ошибкой. В случае ограниченной ошибки доказано, что упрощённые протоколы слабее общих.

$U(\mathrm{EQ}) \ll \mathrm{const}$

Рассмотрим $2^n \times 2^n$ матрицу E, соответствующую EQ: на диагонали 1, вне диагонали -1. Оцените её сигнум-ранг.

$U(\mathrm{EQ}) \ll \mathrm{const}$

Рассмотрим $2^n \times 2^n$ матрицу E, соответствующую EQ: на диагонали 1, вне диагонали -1. Оцените её сигнум-ранг.

Легко видеть, что $\mathrm{rank}_{\pm}(E) \leqslant 3$: рассмотрим функции $a + bt + ct^2$, ясно, что на точках $\{1, \ldots, N\}$ можно получить ими любую последовательность знаков вида $(-1, -1, \ldots, -1, 1, -1, \ldots, -1)$.

$U(\mathrm{EQ}) \ll \mathrm{const}$

Рассмотрим $2^n \times 2^n$ матрицу E, соответствующую EQ: на диагонали 1, вне диагонали -1. Оцените её сигнум-ранг.

Легко видеть, что $\mathrm{rank}_{\pm}(E) \leqslant 3$: рассмотрим функции $a + bt + ct^2$, ясно, что на точках $\{1, \ldots, N\}$ можно получить ими любую последовательность знаков вида $(-1, -1, \ldots, -1, 1, -1, \ldots, -1)$.

Упражнение. Найти $\mathrm{rank}_{\pm}(E) \in \{1, 2, 3\}$.

Оценка Forster-а

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$,

$$\operatorname{rank}_{\pm}(S) \geqslant \frac{\sqrt{mn}}{\|S\|_{2\to 2}}.$$

Оценка Forster-а

Theorem (Forster, 2002)

Для $S \in \{-1,1\}^{m\times n}$,

$$\operatorname{rank}_{\pm}(S) \geqslant \frac{\sqrt{mn}}{\|S\|_{2\to 2}}.$$

Вспомним, что в матрице Уолша–Адамара $H^n(x,y) = (-1)^{\mathrm{IP}(x,y)}$ строки ортогональны и их длина равна $2^{n/2}$, поэтому $\|H\| = 2^{n/2}$.
Следствие: $U(\mathrm{IP}) \asymp \log \operatorname{rank}_{\pm} H \geqslant n/2$.

Оценка Forster-a

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$,

$$\operatorname{rank}_{\pm}(S) \geqslant \frac{\sqrt{mn}}{\|S\|_{2\to 2}}.$$

Вспомним, что в матрице Уолша–Адамара $H^n(x, y) = (-1)^{\mathrm{IP}(x,y)}$ строки ортогональны и их длина равна $2^{n/2}$, поэтому $\|H\| = 2^{n/2}$.
Следствие: $U(\mathrm{IP}) \asymp \log \operatorname{rank}_{\pm} H \geqslant n/2$.
Доказательство теоремы. Пусть $x_i, y_j \in \mathbb{R}^d$ — вектора, реализующие S, при этом $d = \operatorname{rank}_{\pm}(S)$.

Оценка Forster-a

Theorem (Forster, 2002)

Для $S \in \{-1,1\}^{m\times n}$,

$$\mathrm{rank}_{\pm}(S) \geqslant \frac{\sqrt{mn}}{\|S\|_{2\to 2}}.$$

Вспомним, что в матрице Уолша–Адамара $H^n(x,y) = (-1)^{\mathrm{IP}(x,y)}$ строки ортогональны и их длина равна $2^{n/2}$, поэтому $\|H\| = 2^{n/2}$.
Следствие: $U(\mathrm{IP}) \asymp \log \mathrm{rank}_{\pm} H \geqslant n/2$.
Доказательство теоремы. Пусть $x_i, y_j \in \mathbb{R}^d$ — вектора, реализующие S, при этом $d = \mathrm{rank}_{\pm}(S)$.
"Причешем" их, применив подходящий невырожденный линейный оператор $B\colon \mathbb{R}^d \to \mathbb{R}^d$ и положив $\widetilde{x}_i := Bx_i/|Bx_i|$. При этом, если $\widetilde{y}_j := B^{-t}y_j/|B^{-t}y_j|$, то

$$\mathrm{sign}\langle \widetilde{x}_i, \widetilde{y}_j\rangle = \mathrm{sign}\langle Bx_i, (B^{-t}y_j)\rangle = \mathrm{sign}\langle B^{-1}Bx_i, y_j\rangle = S_{i,j}.$$

Доказательство теоремы Forster-а

Lemma (без доказательства)

Пусть $x_1, \ldots, x_m \in \mathbb{R}^d$ *— вектора в общем положении. Существует невырожденный линейный оператор* $B\colon \mathbb{R}^d \to \mathbb{R}^d$, *такой что для* $\widetilde{x}_i := Bx_i/|Bx_i|$ *выполнено*

$$\sum_{i=1}^{m} \widetilde{x}_i \widetilde{x}_i^t = \frac{m}{d} I_d.$$

Доказательство теоремы Forster-a

Lemma (без доказательства)

Пусть $x_1, \ldots, x_m \in \mathbb{R}^d$ *— вектора в общем положении. Существует невырожденный линейный оператор* $B\colon \mathbb{R}^d \to \mathbb{R}^d$*, такой что для* $\widetilde{x}_i := Bx_i/|Bx_i|$ *выполнено*

$$\sum_{i=1}^{m} \widetilde{x}_i \widetilde{x}_i^t = \frac{m}{d} I_d.$$

Применим B для теоремы Форстера. Итак, мы можем считать, что $x_i, y_j \in \mathbb{R}^d$, реалзующие S, являются единичными векторами, причём $\sum x_i x_i^t = (m/d) I_d$.

Доказательство теоремы Forster-a

Lemma (без доказательства)

Пусть $x_1, \ldots, x_m \in \mathbb{R}^d$ *— вектора в общем положении. Существует невырожденный линейный оператор* $B\colon \mathbb{R}^d \to \mathbb{R}^d$*, такой что для* $\widetilde{x}_i := Bx_i/|Bx_i|$ *выполнено*

$$\sum_{i=1}^{m} \widetilde{x}_i \widetilde{x}_i^t = \frac{m}{d} I_d.$$

Применим B для теоремы Форстера. Итак, мы можем считать, что $x_i, y_j \in \mathbb{R}^d$, реалзующие S, являются единичными векторами, причём $\sum x_i x_i^t = (m/d) I_d$.
В вопросах реализации важен margin (зазор), т.е. расстояние от P_i до гиперплоскости H_j. Чем он больше, т.е. чем дальше точки от края и реализация "лучше". Это применяется в оценках ML алгоритмов типа SVM (будет разобрано в следующих лекциях). Имеем

$$\mathrm{dist}(P_i, H_j) = |\langle x_i, y_j \rangle|, \quad \text{т.к. } |y_j| = 1.$$

Доказательство теоремы Forster-a (продолжение)

Рассмотрим величину

$$D = \sum_{j=1}^{n} (\sum_{i=1}^{m} \mathrm{dist}(P_i, H_j))^2.$$

Доказательство теоремы Forster-a (продолжение)

Рассмотрим величину

$$D = \sum_{j=1}^{n} (\sum_{i=1}^{m} \mathrm{dist}(P_i, H_j))^2.$$

Мы покажем, что если размерность мала, то при условии

$$\sum_{i=1}^{m} x_i x_i^t = (m/d) I_d$$

величина D не может быть маленькой.

Доказательство теоремы Forster-а (продолжение)

Рассмотрим величину

$$D = \sum_{j=1}^{n} (\sum_{i=1}^{m} \mathrm{dist}(P_i, H_j))^2.$$

Мы покажем, что если размерность мала, то при условии

$$\sum_{i=1}^{m} x_i x_i^t = (m/d) I_d$$

величина D не может быть маленькой.

$$\sum_{i=1}^{m} |\langle x_i, y_j \rangle| \geqslant \sum_{i=1}^{m} \langle x_i, y_j \rangle^2 = \sum_{i=1}^{m} y_j^t x_i x_i^t y_j =$$

$$= y_j^t \sum_{i=1}^{m} x_i x_i^t y_j = y_j^t \frac{m}{d} I_d y_j = \frac{m}{d}.$$

Доказательство теоремы Forster-a (продолжение)

Рассмотрим величину

$$D = \sum_{j=1}^{n} \Big(\sum_{i=1}^{m} \mathrm{dist}(P_i, H_j)\Big)^2.$$

Мы покажем, что если размерность мала, то при условии

$$\sum_{i=1}^{m} x_i x_i^t = (m/d) I_d$$

величина D не может быть маленькой.

$$\sum_{i=1}^{m} |\langle x_i, y_j\rangle| \geqslant \sum_{i=1}^{m} \langle x_i, y_j\rangle^2 = \sum_{i=1}^{m} y_j^t x_i x_i^t y_j =$$

$$= y_j^t \sum_{i=1}^{m} x_i x_i^t y_j = y_j^t \frac{m}{d} I_d y_j = \frac{m}{d}.$$

Отсюда $D \geqslant nm^2/d^2$.

Доказательство теоремы Forster-a (продолжение)

Рассмотрим величину

$$D = \sum_{j=1}^{n} (\sum_{i=1}^{m} \mathrm{dist}(P_i, H_j))^2.$$

Мы покажем, что если размерность мала, то при условии

$$\sum_{i=1}^{m} x_i x_i^t = (m/d) I_d$$

величина D не может быть маленькой.

$$\sum_{i=1}^{m} |\langle x_i, y_j \rangle| \geqslant \sum_{i=1}^{m} \langle x_i, y_j \rangle^2 = \sum_{i=1}^{m} y_j^t x_i x_i^t y_j =$$

$$= y_j^t \sum_{i=1}^{m} x_i x_i^t y_j = y_j^t \frac{m}{d} I_d y_j = \frac{m}{d}.$$

Отсюда $D \geqslant nm^2/d^2$. Где использовалось, что это реализация S?

Доказательство теоремы Forster-a (продолжение)

Теперь оценим D сверху, используя $\|S\| := \|S\|_{2\to 2}$.

Доказательство теоремы Forster-а (продолжение)

Теперь оценим D сверху, используя $\|S\| := \|S\|_{2\to 2}$.

$$\sum_{i=1}^{m} |\langle x_i, y_j\rangle| = \sum_{i=1}^{m} S_{i,j}\langle x_i, y_j\rangle \leqslant \sum_{i=1}^{m}\langle S_{i,j}x_i, y_j\rangle \leqslant |\sum_{i=1}^{m} S_{i,j}x_i|.$$

Доказательство теоремы Forster-a (продолжение)

Теперь оценим D сверху, используя $\|S\| := \|S\|_{2\to 2}$.

$$\sum_{i=1}^{m} |\langle x_i, y_j\rangle| = \sum_{i=1}^{m} S_{i,j}\langle x_i, y_j\rangle \leqslant \sum_{i=1}^{m}\langle S_{i,j}x_i, y_j\rangle \leqslant |\sum_{i=1}^{m} S_{i,j}x_i|.$$

Суммируем по j:

$$D \leqslant \sum_{j=1}^{n} |\sum_{i=1}^{m} S_{i,j}x_j|^2 = \sum_{j=1}^{n}(\sum_{k=1}^{m} S_{k,j}x_k^t)(\sum_{l=1}^{m} S_{l,j}x_l) =$$

$$\sum_{1\leqslant k,l\leqslant m} x_k^t x_l \sum_{j=1}^{n} S_{k,j}S_{l,j} = \sum_{1\leqslant k,l\leqslant m} \langle x_k, x_l\rangle (SS^t)_{k,l}.$$

Доказательство теоремы Forster-a (продолжение)

У нас возникло скалярное произведение двух $m \times m$ матриц:

$$\langle G, H \rangle := \sum_{i,j} G_{i,j} H_{i,j}.$$

Первая матрица = матрица Грама системы $\{x_k\}$, вторая матрица = SS^t. Заметим, что обе матрицы являются неотрицательно определёнными (напомним, $H \geqslant 0$, если $H = H^t$ и $x^t H x \geqslant 0$ для всех x).

Доказательство теоремы Forster-а (продолжение)

У нас возникло скалярное произведение двух $m \times m$ матриц:

$$\langle G, H \rangle := \sum_{i,j} G_{i,j} H_{i,j}.$$

Первая матрица = матрица Грама системы $\{x_k\}$, вторая матрица = SS^t. Заметим, что обе матрицы являются неотрицательно определёнными (напомним, $H \geqslant 0$, если $H = H^t$ и $x^t H x \geqslant 0$ для всех x).

Хочется воспользоваться неравенством

$$\langle H, G \rangle \geqslant 0, \quad \text{если } H \geqslant 0 \text{ и } G \geqslant 0.$$

Почему это так? См., например, Лекцию №0.

Доказательство теоремы Forster-а (продолжение)

У нас возникло скалярное произведение двух $m \times m$ матриц:

$$\langle G, H\rangle := \sum_{i,j} G_{i,j} H_{i,j}.$$

Первая матрица = матрица Грама системы $\{x_k\}$, вторая матрица = SS^t. Заметим, что обе матрицы являются неотрицательно определёнными (напомним, $H \geqslant 0$, если $H = H^t$ и $x^t H x \geqslant 0$ для всех x).

Хочется воспользоваться неравенством

$$\langle H, G\rangle \geqslant 0, \quad \text{если } H \geqslant 0 \text{ и } G \geqslant 0.$$

Почему это так? См., например, Лекцию №0.

Например, представим H и G в виде сумм одноранговых матриц вида aa^t. Для такой пары матриц утверждение очевидно: $\sum_{k,l} a_k a_l b_k b_l = (\sum a_k b_k)^2$.

Доказательство теоремы Forster-а (продолжение)

У нас возникло скалярное произведение двух $m \times m$ матриц:

$$\langle G, H\rangle := \sum_{i,j} G_{i,j} H_{i,j}.$$

Первая матрица = матрица Грама системы $\{x_k\}$, вторая матрица = SS^t. Заметим, что обе матрицы являются неотрицательно определёнными (напомним, $H \geqslant 0$, если $H = H^t$ и $x^t Hx \geqslant 0$ для всех x).

Хочется воспользоваться неравенством

$$\langle H, G\rangle \geqslant 0, \quad \text{если } H \geqslant 0 \text{ и } G \geqslant 0.$$

Почему это так? См., например, Лекцию №0.

Например, представим H и G в виде сумм одноранговых матриц вида aa^t. Для такой пары матриц утверждение очевидно: $\sum_{k,l} a_k a_l b_k b_l = (\sum a_k b_k)^2$.

Верно ли неравенство:

$$\langle H, G_1\rangle \leqslant \langle H, G_2\rangle, \quad \text{если } H \geqslant 0 \text{ и } G_1 \leqslant G_2.$$

Доказательство теоремы Forster-а (окончание)

Заметим, что $AA^t \leqslant \|A\|^2 I_m$ для любой $m \times n$ матрицы:

$$x^t AA^t x = |A^t x|^2 \leqslant \|A\|^2 |x|^2, \quad x^t(\|A\|^2 I_m - AA^t)x \geqslant 0.$$

Следовательно,

$$\sum_{k,l} \langle x_k, x_l \rangle (SS^t)_{k,l} \leqslant \sum_{k,l} \langle x_k, x_l \rangle \|S\|^2 (I_m)_{k,l} = \|S\|^2 \sum_k |x_k|^2 = \|S\|^2 m.$$

Доказательство теоремы Forster-a (окончание)

Заметим, что $AA^t \leqslant \|A\|^2 I_m$ для любой $m \times n$ матрицы:

$$x^t A A^t x = |A^t x|^2 \leqslant \|A\|^2 |x|^2, \quad x^t(\|A\|^2 I_m - AA^t)x \geqslant 0.$$

Следовательно,

$$\sum_{k,l} \langle x_k, x_l \rangle (SS^t)_{k,l} \leqslant \sum_{k,l} \langle x_k, x_l \rangle \|S\|^2 (I_m)_{k,l} = \|S\|^2 \sum_k |x_k|^2 = \|S\|^2 m.$$

Итого, $D \leqslant \|S\|^2 m$.

Доказательство теоремы Forster-a (окончание)

Заметим, что $AA^t \leqslant \|A\|^2 I_m$ для любой $m \times n$ матрицы:

$$x^t A A^t x = |A^t x|^2 \leqslant \|A\|^2 |x|^2, \quad x^t(\|A\|^2 I_m - AA^t)x \geqslant 0.$$

Следовательно,

$$\sum_{k,l} \langle x_k, x_l \rangle (SS^t)_{k,l} \leqslant \sum_{k,l} \langle x_k, x_l \rangle \|S\|^2 (I_m)_{k,l} = \|S\|^2 \sum_k |x_k|^2 = \|S\|^2 m.$$

Итого, $D \leqslant \|S\|^2 m$.
Сравнивая с неравенством $D \geqslant nm^2/d^2$, получим оценку на d.

Сигнум-ранг для случайных матриц

Для матрицы Уолша–Адамара $H^n_{x,y} = (-1)^{\langle x,y \rangle}$ имеем $\mathrm{rank}_\pm H^n \geqslant N^{1/2}$, где $N = 2^n$, что даёт оптимальную оценку U-сложности

$$U(\mathrm{IP}) \asymp \log \mathrm{rank}_\pm(H^n) \asymp \log N.$$

Однако, ранг сигнум-матрицы из $\{-1, 1\}^N$ теоретически, может быть порядка N. Существуют ли такие матрицы?

Сигнум-ранг для случайных матриц

Для матрицы Уолша–Адамара $H^n_{x,y} = (-1)^{\langle x,y\rangle}$ имеем $\operatorname{rank}_{\pm} H^n \geqslant N^{1/2}$, где $N = 2^n$, что даёт оптимальную оценку U-сложности

$$U(\mathrm{IP}) \asymp \log \operatorname{rank}_{\pm}(H^n) \asymp \log N.$$

Однако, ранг сигнум-матрицы из $\{-1,1\}^N$ теоретически, может быть порядка N. Существуют ли такие матрицы?

Явные конструкции таких матриц науке неизвестны (лучший результат — $N^{1/2}$). Однако, можно доказать, что сигнум-матриц сигнум ранга $\leqslant \varepsilon N$ мало (при маленьком, но фиксированном $\varepsilon > 0$) и, следовательно, существуют матрицы с $\operatorname{rank}_{\pm}(S) \geqslant \varepsilon N$.

Algebraic method (Alon, Frankl, Rödl, 1985)

Denote by $S_r(n_1, n_2)$ the number of $n_1 \times n_2$ signum-matrices $(\operatorname{sign} B_{i,j})$, for all $\operatorname{rank} B \leqslant r$ with nonzero elements.

Algebraic method (Alon, Frankl, Rödl, 1985)

Denote by $S_r(n_1, n_2)$ the number of $n_1 \times n_2$ signum-matrices $(\operatorname{sign} B_{i,j})$, for all $\operatorname{rank} B \leqslant r$ with nonzero elements.

Let $p_1, \dots, p_M$ be polynomials of N variables. Each point $x \in \mathbb{R}^N$ with $p_i(x) \neq 0$, $i = 1, \dots, M$, yields the signum-vector $(\operatorname{sign} p_1(x), \dots, \operatorname{sign} p_M(x)) \in \{-1, 1\}^M$. Denote by $z(p_1, \dots, p_M)$ the number of such signum vectors.
Denote by $Z(N, M, D)$ the maximum possible value of $z(p_1, \dots, p_M)$ over polynomials $p_1, \dots, p_M$ of degree $\leqslant D$.

Algebraic method (Alon, Frankl, Rödl, 1985)

Denote by $S_r(n_1, n_2)$ the number of $n_1 \times n_2$ signum-matrices $(\operatorname{sign} B_{i,j})$, for all $\operatorname{rank} B \leqslant r$ with nonzero elements.

Let $p_1, \dots, p_M$ be polynomials of N variables. Each point $x \in \mathbb{R}^N$ with $p_i(x) \neq 0$, $i = 1, \dots, M$, yields the signum-vector $(\operatorname{sign} p_1(x), \dots, \operatorname{sign} p_M(x)) \in \{-1, 1\}^M$. Denote by $z(p_1, \dots, p_M)$ the number of such signum vectors.
Denote by $Z(N, M, D)$ the maximum possible value of $z(p_1, \dots, p_M)$ over polynomials $p_1, \dots, p_M$ of degree $\leqslant D$.

Statement (Alon, Frankl, Rödl)

$$S_r(n_1, n_2) \leqslant Z(r(n_1 + n_2), n_1 n_2, 2).$$

Доказательство.

If a matrix M has rank $\leqslant r$ and yields signum-matrix $\sigma = (\operatorname{sign} M_{i,j})$, then for some vectors $u^s \in \mathbb{R}^{n_1}$, $v^s \in \mathbb{R}^{n_2}$, $s = 1, \ldots, r$, we have

$$M = \sum_{s=1}^{r} u^s \otimes v^s, \quad \sigma_{i,j} = \operatorname{sign}\Big(\sum_{s=1}^{r} u_i^s v_j^s\Big).$$

Доказательство.

If a matrix M has rank $\leqslant r$ and yields signum-matrix $\sigma = (\operatorname{sign} M_{i,j})$, then for some vectors $u^s \in \mathbb{R}^{n_1}$, $v^s \in \mathbb{R}^{n_2}$, $s = 1, \ldots, r$, we have

$$M = \sum_{s=1}^{r} u^s \otimes v^s, \quad \sigma_{i,j} = \operatorname{sign}(\sum_{s=1}^{r} u_i^s v_j^s).$$

Let us look at this in the following way: we have variables x_i^s and y_j^s, $i \in [n_1]$, $j \in [n_2]$, $s \in [r]$, $r(n_1 + n_2)$ of them. There are fixed polynomials in these variables:

$$q_{i,j}(x, y) = \sum_{s=1}^{r} x_i^s y_j^s.$$

So, the existence of M that yields σ is equivalent to existence of x such that $(\operatorname{sign} q_{i,j}(x, y))$ equals σ. Hence $S_r(n_1, n_2) = z(\{q_{i,j}\})$. □

Warren's bound

It is convenient to use Warren's bound (1968) on the number of connected components of the set $\mathbb{R}^N \setminus \cup_{i=1}^{M}\{p_i(x) = 0\}$, which gives

$$Z(N, M, D) \leqslant (4eDM/N)^N, \quad \text{for } M \geqslant N.$$

Warren's bound

It is convenient to use Warren's bound (1968) on the number of connected components of the set $\mathbb{R}^N \setminus \cup_{i=1}^M \{p_i(x) = 0\}$, which gives

$$Z(N, M, D) \leqslant (4eDM/N)^N, \quad \text{for } M \geqslant N.$$

We use bound on S_r together with Warren's bound for $n \times n$ matrices and rank $r = \varepsilon n$:

$$\log S_r(n, n) \leqslant \log Z(2rn, n^2, 2) \leqslant 2rn \log(cn/r) \asymp n^2 \varepsilon \log(1/\varepsilon).$$

There totally 2^{n^2} signum matrices and only $2^{cn^2 \varepsilon \log(1/\varepsilon)}$ of low signum-rank.