

Спецкурс 2020/2021: “Геометрические и комбинаторные свойства матриц и аппроксимация”
Блок лекций “Сложность матриц и аппроксимация”
Лекция 5: “Реализация матриц с большим отступом (margin)”

14 ноября 2020 г.

Определение

Пусть $S \in \{-1, 1\}^{m \times n}$ — сигнум-матрица.

Определение

Пусть $S \in \{-1, 1\}^{m \times n}$ — сигнум-матрица.

Определим “максимальный отступ” матрицы:

$$\text{margin}(S) := \max_{\{x_i\}, \{y_j\}} \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|},$$

где максимум берётся по всем *реализациям* матрицы, т.е. таким наборам векторов $x_1, \dots, x_m, y_1, \dots, y_n$, что $S_{i,j} = \text{sign} \langle x_i, y_j \rangle \quad \forall i, j$.

Определение

Пусть $S \in \{-1, 1\}^{m \times n}$ — сигнум-матрица.

Определим “максимальный отступ” матрицы:

$$\text{margin}(S) := \max_{\{x_i\}, \{y_j\}} \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|},$$

где максимум берётся по всем *реализациям* матрицы, т.е. таким наборам векторов $x_1, \dots, x_m, y_1, \dots, y_n$, что $S_{i,j} = \text{sign} \langle x_i, y_j \rangle \quad \forall i, j$.

Смысл в том, что равенство $S_{i,j} = \text{sign} \langle x_i, y_j \rangle$ должно выполняться с большим “запасом”.

Определение

Пусть $S \in \{-1, 1\}^{m \times n}$ — сигнум-матрица.

Определим “максимальный отступ” матрицы:

$$\text{margin}(S) := \max_{\{x_i\}, \{y_j\}} \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|},$$

где максимум берётся по всем *реализациям* матрицы, т.е. таким наборам векторов $x_1, \dots, x_m, y_1, \dots, y_n$, что $S_{i,j} = \text{sign} \langle x_i, y_j \rangle \quad \forall i, j$.

Смысл в том, что равенство $S_{i,j} = \text{sign} \langle x_i, y_j \rangle$ должно выполняться с большим “запасом”.

Margin — отступ, зазор.

Определение

Пусть $S \in \{-1, 1\}^{m \times n}$ — сигнум-матрица.

Определим “максимальный отступ” матрицы:

$$\text{margin}(S) := \max_{\{x_i\}, \{y_j\}} \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|},$$

где максимум берётся по всем *реализациям* матрицы, т.е. таким наборам векторов $x_1, \dots, x_m, y_1, \dots, y_n$, что $S_{i,j} = \text{sign} \langle x_i, y_j \rangle \quad \forall i, j$.

Смысл в том, что равенство $S_{i,j} = \text{sign} \langle x_i, y_j \rangle$ должно выполняться с большим “запасом”.

Margin — отступ, зазор.

Определим сложность реализации с отступом (*margin complexity*) сигнум матрицы S как

$$\text{mc}(S) := \text{margin}^{-1}(S).$$

Определение

Пусть $S \in \{-1, 1\}^{m \times n}$ — сигнум-матрица.

Определим “максимальный отступ” матрицы:

$$\text{margin}(S) := \max_{\{x_i\}, \{y_j\}} \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|},$$

где максимум берётся по всем *реализациям* матрицы, т.е. таким наборам векторов $x_1, \dots, x_m, y_1, \dots, y_n$, что $S_{i,j} = \text{sign} \langle x_i, y_j \rangle \quad \forall i, j$.

Смысл в том, что равенство $S_{i,j} = \text{sign} \langle x_i, y_j \rangle$ должно выполняться с большим “запасом”.

Margin — отступ, зазор.

Определим сложность реализации с отступом (*margin complexity*) сигнум матрицы S как

$$\text{mc}(S) := \text{margin}^{-1}(S).$$

(Чем сложнее реализовать матрицу, тем больше сложность.)

Базовые свойства

Оцените $mc(S)$:

Базовые свойства

Оцените $\text{tr}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.

Базовые свойства

Оцените $\text{tr}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.

Базовые свойства

Оцените $\text{tr}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.
- $S_{1,1} = 1$, $S_{i,j} = -1$ при остальных i, j .

Базовые свойства

Оцените $\text{mc}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.
- $S_{1,1} = 1$, $S_{i,j} = -1$ при остальных i, j .

Утверждение

Для любой матрицы $S \in \{-1, 1\}^{m \times n}$ имеем

$$1 \leq \text{mc}(S) \leq \min\{\sqrt{m}, \sqrt{n}\}.$$

Базовые свойства

Оцените $\text{mc}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.
- $S_{1,1} = 1$, $S_{i,j} = -1$ при остальных i, j .

Утверждение

Для любой матрицы $S \in \{-1, 1\}^{m \times n}$ имеем

$$1 \leq \text{mc}(S) \leq \min\{\sqrt{m}, \sqrt{n}\}.$$

Докажите оценку снизу.

Базовые свойства

Оцените $\text{mc}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.
- $S_{1,1} = 1$, $S_{i,j} = -1$ при остальных i, j .

Утверждение

Для любой матрицы $S \in \{-1, 1\}^{m \times n}$ имеем

$$1 \leq \text{mc}(S) \leq \min\{\sqrt{m}, \sqrt{n}\}.$$

Докажите оценку снизу. $\text{margin}(S) \leq 1$, т.к. $|\langle x_i, y_j \rangle| \leq |x_i| |y_j|$.

Базовые свойства

Оцените $\text{mc}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.
- $S_{1,1} = 1$, $S_{i,j} = -1$ при остальных i, j .

Утверждение

Для любой матрицы $S \in \{-1, 1\}^{m \times n}$ имеем

$$1 \leq \text{mc}(S) \leq \min\{\sqrt{m}, \sqrt{n}\}.$$

Докажите оценку снизу. $\text{margin}(S) \leq 1$, т.к. $|\langle x_i, y_j \rangle| \leq |x_i| |y_j|$.

Докажите оценку сверху.

Базовые свойства

Оцените $\text{mc}(S)$:

- “Диагональная” матрица $n \times n$: $S_{i,i} = 1$, $S_{i,j} = -1$ при $i \neq j$.
- $S_{i,i} = -1$, $S_{i,j} = 1$ при $i \neq j$.
- $S_{1,1} = 1$, $S_{i,j} = -1$ при остальных i, j .

Утверждение

Для любой матрицы $S \in \{-1, 1\}^{m \times n}$ имеем

$$1 \leq \text{mc}(S) \leq \min\{\sqrt{m}, \sqrt{n}\}.$$

Докажите оценку снизу. $\text{margin}(S) \leq 1$, т.к. $|\langle x_i, y_j \rangle| \leq |x_i| |y_j|$.

Докажите оценку сверху. Пусть $m \leq n$. Возьмём в качестве $\{y_j\}_{j=1}^n$ столбцы матрицы S , в качестве $\{x_i\}_{i=1}^m$ — базисные вектора. Тогда $|x_i| \cdot |y_j| \leq m^{1/2}$ и $\langle x_i, y_j \rangle = S_{i,j}$.

Классификация с максимальным отступом

Пусть $(x_1, t_1), \dots, (x_n, t_n)$ — выборка из некоторого множества объектов. Каждый объект принадлежит одному из двух классов: $t_i = 1$ или $t_i = -1$. Вектор $x_i \in \mathbb{R}^d$ состоит из d чисел (признаков), описывающих i -й объект.

Классификация с максимальным отступом

Пусть $(x_1, t_1), \dots, (x_n, t_n)$ — выборка из некоторого множества объектов. Каждый объект принадлежит одному из двух классов: $t_i = 1$ или $t_i = -1$. Вектор $x_i \in \mathbb{R}^d$ состоит из d чисел (признаков), описывающих i -й объект.

Задача классификации состоит в построении функции $\hat{f}: \mathbb{R}^d \rightarrow \{-1, 1\}$, позволяющей отличать объекты разных классов. То есть, ошибка $\text{Err} = P(t \neq \hat{f}(x))$ должна быть мала.

Классификация с максимальным отступом

Пусть $(x_1, t_1), \dots, (x_n, t_n)$ — выборка из некоторого множества объектов. Каждый объект принадлежит одному из двух классов: $t_i = 1$ или $t_i = -1$. Вектор $x_i \in \mathbb{R}^d$ состоит из d чисел (признаков), описывающих i -й объект.

Задача классификации состоит в построении функции $\hat{f}: \mathbb{R}^d \rightarrow \{-1, 1\}$, позволяющей отличать объекты разных классов. То есть, ошибка $\text{Err} = P(t \neq \hat{f}(x))$ должна быть мала.

В линейной классификации функция \hat{f} строится с помощью линейной: $\hat{f}(x) = \text{sign}\langle b, x \rangle$, $b \in \mathbb{R}^d$. (Случай аффинной функции $b_0 + \langle b, x \rangle$ можно свести к линейному, добавим признак, тождественно равный 1.)

Классификация с максимальным отступом

Пусть $(x_1, t_1), \dots, (x_n, t_n)$ — выборка из некоторого множества объектов. Каждый объект принадлежит одному из двух классов: $t_i = 1$ или $t_i = -1$. Вектор $x_i \in \mathbb{R}^d$ состоит из d чисел (признаков), описывающих i -й объект.

Задача классификации состоит в построении функции $\hat{f}: \mathbb{R}^d \rightarrow \{-1, 1\}$, позволяющей отличать объекты разных классов. То есть, ошибка $\text{Err} = P(t \neq \hat{f}(x))$ должна быть мала.

В линейной классификации функция \hat{f} строится с помощью линейной: $\hat{f}(x) = \text{sign}\langle b, x \rangle$, $b \in \mathbb{R}^d$. (Случай аффинной функции $b_0 + \langle b, x \rangle$ можно свести к линейному, добавим признак, тождественно равный 1.)

Пространство \mathbb{R}^d гиперплоскостью $\langle b, x \rangle = 0$ разделяется на два полупространства: $\langle b, x \rangle > 0$, классифицируется как $t = 1$, и $\langle b, x \rangle < 0$ (соответственно, $t = -1$).

Классификация с максимальным отступом

Разделяющая гиперплоскость $\langle b, x \rangle = 0$, строится по обучающей выборке $(x_1, t_1), \dots, (x_n, t_n)$.

Классификация с максимальным отступом

Разделяющая гиперплоскость $\langle b, x \rangle = 0$, строится по обучающей выборке $(x_1, t_1), \dots, (x_n, t_n)$.

Предположим, мы можем получить на обучении нулевую ошибку: существует ненулевой вектор b , такой что

$$t_i = \text{sign}\langle b, x_i \rangle, \quad i = 1, \dots, n.$$

Какой b взять, чтобы ошибка на тестовой выборке была поменьше?

Классификация с максимальным отступом

Разделяющая гиперплоскость $\langle b, x \rangle = 0$, строится по обучающей выборке $(x_1, t_1), \dots, (x_n, t_n)$.

Предположим, мы можем получить на обучении нулевую ошибку: существует ненулевой вектор b , такой что

$$t_i = \text{sign}\langle b, x_i \rangle, \quad i = 1, \dots, n.$$

Какой b взять, чтобы ошибка на тестовой выборке была поменьше?

Потребуем, чтобы b обеспечивал правильную классификацию с максимальным запасом, *отступом*:

$$\begin{cases} \min_i |\langle b, x_i \rangle| \rightarrow \max, \\ t_i = \text{sign}\langle b, x_i \rangle, \quad i = 1, \dots, n, \\ |b| = 1. \end{cases}$$

Классификация с максимальным отступом

Разделяющая гиперплоскость $\langle b, x \rangle = 0$, строится по обучающей выборке $(x_1, t_1), \dots, (x_n, t_n)$.

Предположим, мы можем получить на обучении нулевую ошибку: существует ненулевой вектор b , такой что

$$t_i = \text{sign}\langle b, x_i \rangle, \quad i = 1, \dots, n.$$

Какой b взять, чтобы ошибка на тестовой выборке была поменьше?

Потребуем, чтобы b обеспечивал правильную классификацию с максимальным запасом, *отступом*:

$$\begin{cases} \min_i |\langle b, x_i \rangle| \rightarrow \max, \\ t_i = \text{sign}\langle b, x_i \rangle, \quad i = 1, \dots, n, \\ |b| = 1. \end{cases}$$

Эквивалентная формулировка: мы проводим разделяющую гиперплоскость так, чтобы максимизировать расстояние от точек до края.

Опорные вектора

Можно показать, что b является линейной комбинацией *опорных* векторов x_i , для которых $|\langle b, x_i \rangle|$ минимально.

Опорные вектора

Можно показать, что b является линейной комбинацией *опорных векторов* x_i , для которых $|\langle b, x_i \rangle|$ минимально.

Отсюда название *Метод Опорных Векторов*, Support Vector Machine (SVM).

Опорные вектора

Можно показать, что b является линейной комбинацией *опорных векторов* x_i , для которых $|\langle b, x_i \rangle|$ минимально.

Отсюда название *Метод Опорных Векторов*, Support Vector Machine (SVM). На практике в SVM используется так называемый *soft margin*, когда мы не требуем, чтобы гиперплоскость правильно разделяла обучающую выборку; разрешаем “залезать за край”, но “штрафуем” за это:

Опорные вектора

Можно показать, что b является линейной комбинацией *опорных векторов* x_i , для которых $|\langle b, x_i \rangle|$ минимально.

Отсюда название *Метод Опорных Векторов*, Support Vector Machine (SVM). На практике в SVM используется так называемый *soft margin*, когда мы не требуем, чтобы гиперплоскость правильно разделяла обучающую выборку; разрешаем “залезать за край”, но “штрафуем” за это:

$$\begin{cases} |b|^2 + C \sum_{i=1}^n \xi_i \rightarrow \min, \\ t_i \langle b, x_i \rangle \geq 1 - \xi_i, & i = 1, \dots, n \\ \xi_i \geq 0, & i = 1, \dots, n. \end{cases}$$

Классификация с максимальным отступом

Вернёмся к классификации с максимальным отступом. Пусть

$\mathcal{O} = \{O_1, \dots, O_n\}$ — множество объектов

Классификация с максимальным отступом

Вернёмся к классификации с максимальным отступом. Пусть $\mathcal{O} = \{O_1, \dots, O_n\}$ — множество объектов, $\Phi: \mathcal{O} \rightarrow \mathbb{R}^d$ — отображение в пространство признаков (feature map), $x_i = \Phi(O_i)$.

Классификация с максимальным отступом

Вернёмся к классификации с максимальным отступом. Пусть $\mathcal{O} = \{O_1, \dots, O_n\}$ — множество объектов, $\Phi: \mathcal{O} \rightarrow \mathbb{R}^d$ — отображение в пространство признаков (feature map), $x_i = \Phi(O_i)$. Класс объекта задаётся (неизвестной нам) функцией $f: \mathcal{O} \rightarrow \{-1, 1\}$: $t_i = f(O_i)$.

Классификация с максимальным отступом

Вернёмся к классификации с максимальным отступом. Пусть $\mathcal{O} = \{O_1, \dots, O_n\}$ — множество объектов, $\Phi: \mathcal{O} \rightarrow \mathbb{R}^d$ — отображение в пространство признаков (feature map), $x_i = \Phi(O_i)$. Класс объекта задаётся (неизвестной нам) функцией $f: \mathcal{O} \rightarrow \{-1, 1\}$: $t_i = f(O_i)$. Обозначим через $\text{margin}_\Phi(f)$ величину максимального отступа:

$$\begin{cases} \min_i |\langle b, x_i \rangle| \rightarrow \max, \\ t_i = \text{sign} \langle b, x_i \rangle, \quad i = 1, \dots, n, \quad |b| = 1. \end{cases}$$

Классификация с максимальным отступом

Вернёмся к классификации с максимальным отступом. Пусть $\mathcal{O} = \{O_1, \dots, O_n\}$ — множество объектов, $\Phi: \mathcal{O} \rightarrow \mathbb{R}^d$ — отображение в пространство признаков (feature map), $x_i = \Phi(O_i)$. Класс объекта задаётся (неизвестной нам) функцией $f: \mathcal{O} \rightarrow \{-1, 1\}$: $t_i = f(O_i)$. Обозначим через $\text{margin}_\Phi(f)$ величину максимального отступа:

$$\begin{cases} \min_i |\langle b, x_i \rangle| \rightarrow \max, \\ t_i = \text{sign} \langle b, x_i \rangle, \quad i = 1, \dots, n, \quad |b| = 1. \end{cases}$$

Теперь предположим, есть целый класс возможных функций:
 $\mathcal{F} = \{f_1, \dots, f_m\}$.

Классификация с максимальным отступом

Вернёмся к классификации с максимальным отступом. Пусть $\mathcal{O} = \{O_1, \dots, O_n\}$ — множество объектов, $\Phi: \mathcal{O} \rightarrow \mathbb{R}^d$ — отображение в пространство признаков (feature map), $x_i = \Phi(O_i)$. Класс объекта задаётся (неизвестной нам) функцией $f: \mathcal{O} \rightarrow \{-1, 1\}$: $t_i = f(O_i)$. Обозначим через $\text{margin}_\Phi(f)$ величину максимального отступа:

$$\begin{cases} \min_i |\langle b, x_i \rangle| \rightarrow \max, \\ t_i = \text{sign} \langle b, x_i \rangle, \quad i = 1, \dots, n, \quad |b| = 1. \end{cases}$$

Теперь предположим, есть целый класс возможных функций: $\mathcal{F} = \{f_1, \dots, f_m\}$. Минимальный отступ

$$\text{margin}_\Phi(\mathcal{F}) = \min(\text{margin}_\Phi(f_1), \dots, \text{margin}_\Phi(f_m))$$

характеризует сложность задачи классификации произвольной функции из \mathcal{F} .

Классификация с максимальным отступом

Если теперь минимизировать $\text{margin}_\Phi(\mathcal{F})$ по всевозможным отображениям $\Phi: \mathcal{O} \rightarrow S^{d-1}$ (нормируем признаки):

Классификация с максимальным отступом

Если теперь минимизировать $\text{margin}_\Phi(\mathcal{F})$ по всевозможным отображениям $\Phi: \mathcal{O} \rightarrow S^{d-1}$ (нормируем признаки):

$$\begin{cases} \min_{i,j} |\langle b^j, x_i \rangle| \rightarrow \max, \\ t_i^j = \text{sign} \langle b^j, x_i \rangle, \quad i = 1, \dots, n, j = 1, \dots, m, \\ |b^j| = 1, |x_i| = 1, \end{cases}$$

Классификация с максимальным отступом

Если теперь минимизировать $\text{margin}_\Phi(\mathcal{F})$ по всевозможным отображениям $\Phi: \mathcal{O} \rightarrow S^{d-1}$ (нормируем признаки):

$$\begin{cases} \min_{i,j} |\langle b^j, x_i \rangle| \rightarrow \max, \\ t_i^j = \text{sign} \langle b^j, x_i \rangle, \quad i = 1, \dots, n, j = 1, \dots, m, \\ |b^j| = 1, |x_i| = 1, \end{cases}$$

мы, получим в точности величину $\text{margin}(S)$ сигнум-матрицы

$$S = \{f_j(O_i)\}_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$$

Таблица различных мер сложности

Напомним определения:

$$\text{margin}(S) := \max\left\{\min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} : \text{sign}\langle x_i, y_j \rangle = S_{i,j}\right\},$$

Таблица различных мер сложности

Напомним определения:

$$\text{margin}(S) := \max\left\{\min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} : \text{sign}\langle x_i, y_j \rangle = S_{i,j}\right\},$$

$$\text{mc}(S) := \text{margin}^{-1}(S).$$

Таблица различных мер сложности

Напомним определения:

$$\text{margin}(S) := \max\left\{\min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} : \text{sign}\langle x_i, y_j \rangle = S_{i,j}\right\},$$

$$\text{mc}(S) := \text{margin}^{-1}(S).$$

	равенство	знак
размерность	rank	rank _±
длина	γ_2	mc

Таблица различных мер сложности

Напомним определения:

$$\text{margin}(S) := \max \left\{ \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} : \text{sign} \langle x_i, y_j \rangle = S_{i,j} \right\},$$

$$\text{mc}(S) := \text{margin}^{-1}(S).$$

	равенство	знак
размерность	rank	rank _±
длина	γ_2	mc

Величина $\log \text{rank}_{\pm}$ эквивалентна коммуникационной сложности в вероятностной модели с неограниченной ошибкой.

Таблица различных мер сложности

Напомним определения:

$$\text{margin}(S) := \max \left\{ \min_{i,j} \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} : \text{sign} \langle x_i, y_j \rangle = S_{i,j} \right\},$$

$$\text{mc}(S) := \text{margin}^{-1}(S).$$

	равенство	знак
размерность	rank	rank _±
длина	γ_2	mc

Величина $\log \text{rank}_{\pm}$ эквивалентна коммуникационной сложности в вероятностной модели с неограниченной ошибкой.

Величина γ_2 возникла в функциональном анализе (факторизация операторов через гильбертовы пространства).

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$ имеет место неравенство

$$\text{mc}(S) \geq \frac{\sqrt{mn}}{\|S\|_{2 \rightarrow 2}}.$$

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$ имеет место неравенство

$$\text{mc}(S) \geq \frac{\sqrt{mn}}{\|S\|_{2 \rightarrow 2}}.$$

Пусть $\{x_i\}, \{y_j\}$ — реализация S с помощью единичных векторов.
Рассмотрим величину

$$D := \sum_{j=1}^n \left(\sum_{i=1}^m |\langle x_i, y_j \rangle| \right)^2.$$

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$ имеет место неравенство

$$\text{mc}(S) \geq \frac{\sqrt{mn}}{\|S\|_{2 \rightarrow 2}}.$$

Пусть $\{x_i\}, \{y_j\}$ — реализация S с помощью единичных векторов.
Рассмотрим величину

$$D := \sum_{j=1}^n \left(\sum_{i=1}^m |\langle x_i, y_j \rangle| \right)^2.$$

Ранее было доказано (лекция №2), что $D \leq m \|S\|_{2 \rightarrow 2}^2$.

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$ имеет место неравенство

$$\text{mc}(S) \geq \frac{\sqrt{mn}}{\|S\|_{2 \rightarrow 2}}.$$

Пусть $\{x_i\}, \{y_j\}$ — реализация S с помощью единичных векторов.
Рассмотрим величину

$$D := \sum_{j=1}^n \left(\sum_{i=1}^m |\langle x_i, y_j \rangle| \right)^2.$$

Ранее было доказано (лекция №2), что $D \leq m \|S\|_{2 \rightarrow 2}^2$.

С другой стороны, если это реализация с максимальным отступом, то $|\langle x_i, y_j \rangle| \geq \text{margin}(S)$, поэтому $D \geq nm^2 \text{margin}^2(S)$.

Theorem (Forster, 2002)

Для $S \in \{-1, 1\}^{m \times n}$ имеет место неравенство

$$\text{mc}(S) \geq \frac{\sqrt{mn}}{\|S\|_{2 \rightarrow 2}}.$$

Пусть $\{x_i\}, \{y_j\}$ — реализация S с помощью единичных векторов.
Рассмотрим величину

$$D := \sum_{j=1}^n \left(\sum_{i=1}^m |\langle x_i, y_j \rangle| \right)^2.$$

Ранее было доказано (лекция №2), что $D \leq m \|S\|_{2 \rightarrow 2}^2$.

С другой стороны, если это реализация с максимальным отступом, то $|\langle x_i, y_j \rangle| \geq \text{margin}(S)$, поэтому $D \geq nm^2 \text{margin}^2(S)$. Отсюда

$$nm^2 \text{margin}^2(S) \leq m \|S\|_{2 \rightarrow 2}^2, \quad \text{ч.т.д.}$$

Пусть H — матрица Адамара, т.е. $n \times n$ сигнум-матрица с ортогональными строками.

Пусть H — матрица Адамара, т.е. $n \times n$ сигнум-матрица с ортогональными строками. Чему равно $\text{tr}(H)$?

Пусть H — матрица Адамара, т.е. $n \times n$ сигнум-матрица с ортогональными строками. Чему равно $\text{tr}(H)$? Воспользуемся теоремой Форстера.

Пусть H — матрица Адамара, т.е. $n \times n$ сигнум-матрица с ортогональными строками. Чему равно $\text{тс}(H)$? Воспользуемся теоремой Форстера. Оценка снизу: $n^{1/2}$, сверху тоже $n^{1/2}$, следовательно, $\text{тс}(H) = n^{1/2}$.

Связь γ_2 и mc

Утверждение

$$mc(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i, j\}.$$

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i,j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$.

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i,j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$.

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i,j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$. Матрица $A = A_1 / \text{margin}(S)$ имеет $\gamma_2(A) \leq \text{mc}(S)$,

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i,j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$. Матрица $A = A_1 / \text{margin}(S)$ имеет $\gamma_2(A) \leq \text{mc}(S)$, при этом

$$A_{i,j}S_{i,j} = \frac{1}{\text{margin}(S)} \langle x_i, y_j \rangle \text{sign}\langle x_i, y_j \rangle \geq 1.$$

Мы доказали оценку $\min \gamma_2(A) \leq \text{mc}(S)$.

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i,j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$. Матрица $A = A_1 / \text{margin}(S)$ имеет $\gamma_2(A) \leq \text{mc}(S)$, при этом

$$A_{i,j}S_{i,j} = \frac{1}{\text{margin}(S)} \langle x_i, y_j \rangle \text{sign}\langle x_i, y_j \rangle \geq 1.$$

Мы доказали оценку $\min \gamma_2(A) \leq \text{mc}(S)$.

Обратно, пусть $A_{i,j}S_{i,j} \geq 1$.

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i, j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$. Матрица $A = A_1 / \text{margin}(S)$ имеет $\gamma_2(A) \leq \text{mc}(S)$, при этом

$$A_{i,j}S_{i,j} = \frac{1}{\text{margin}(S)} \langle x_i, y_j \rangle \text{sign}\langle x_i, y_j \rangle \geq 1.$$

Мы доказали оценку $\min \gamma_2(A) \leq \text{mc}(S)$.

Обратно, пусть $A_{i,j}S_{i,j} \geq 1$. Запишем $A_{i,j} = \langle x_i, y_j \rangle$ и $|x_i| \cdot |y_j| \leq \gamma_2(A)$.

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i, j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$. Матрица $A = A_1 / \text{margin}(S)$ имеет $\gamma_2(A) \leq \text{mc}(S)$, при этом

$$A_{i,j}S_{i,j} = \frac{1}{\text{margin}(S)} \langle x_i, y_j \rangle \text{sign}\langle x_i, y_j \rangle \geq 1.$$

Мы доказали оценку $\min \gamma_2(A) \leq \text{mc}(S)$.

Обратно, пусть $A_{i,j}S_{i,j} \geq 1$. Запишем $A_{i,j} = \langle x_i, y_j \rangle$ и $|x_i| \cdot |y_j| \leq \gamma_2(A)$. Тогда $\text{sign}\langle x_i, y_j \rangle = S_{i,j}$ и

$$\min \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} \geq \frac{|A_{i,j}|}{\gamma_2(A)} \geq \gamma_2^{-1}(A).$$

Связь γ_2 и mc

Утверждение

$$\text{mc}(S) = \min\{\gamma_2(A) : A_{i,j}S_{i,j} \geq 1 \forall i, j\}.$$

Доказательство. Пусть есть реализация $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с векторами $|x_i| = 1$, $|y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Матрица $A_1 = (\langle x_i, y_j \rangle)$ имеет $\gamma_2(A_1) \leq 1$. Матрица $A = A_1 / \text{margin}(S)$ имеет $\gamma_2(A) \leq \text{mc}(S)$, при этом

$$A_{i,j}S_{i,j} = \frac{1}{\text{margin}(S)} \langle x_i, y_j \rangle \text{sign}\langle x_i, y_j \rangle \geq 1.$$

Мы доказали оценку $\min \gamma_2(A) \leq \text{mc}(S)$.

Обратно, пусть $A_{i,j}S_{i,j} \geq 1$. Запишем $A_{i,j} = \langle x_i, y_j \rangle$ и $|x_i| \cdot |y_j| \leq \gamma_2(A)$. Тогда $\text{sign}\langle x_i, y_j \rangle = S_{i,j}$ и

$$\min \frac{|\langle x_i, y_j \rangle|}{|x_i| \cdot |y_j|} \geq \frac{|A_{i,j}|}{\gamma_2(A)} \geq \gamma_2^{-1}(A).$$

Следовательно, $\text{margin}(S) \geq \gamma_2^{-1}(A)$, т.е. $\text{mc}(S) \leq \gamma_2(A)$.

Утверждение

$$m\kappa(S) \geq mn/\gamma_2^*(S).$$

Утверждение

$$m\kappa(S) \geq mn/\gamma_2^*(S).$$

Пусть $m\kappa(S) = \gamma_2(A)$, $A_{i,j}S_{i,j} \geq 1$.

Утверждение

$$\text{mc}(S) \geq mn/\gamma_2^*(S).$$

Пусть $\text{mc}(S) = \gamma_2(A)$, $A_{i,j}S_{i,j} \geq 1$.

Нормируем $A' = A/\gamma_2(A)$, тогда

$$\gamma_2^*(S) \geq \langle S, A' \rangle = \sum S_{i,j}A'_{i,j} \geq \gamma_2(A)^{-1} = \text{mc}(S)^{-1}.$$

Ч.т.д.

Утверждение

$$\text{mc}(S) \geq mn/\gamma_2^*(S).$$

Пусть $\text{mc}(S) = \gamma_2(A)$, $A_{i,j}S_{i,j} \geq 1$.

Нормируем $A' = A/\gamma_2(A)$, тогда

$$\gamma_2^*(S) \geq \langle S, A' \rangle = \sum S_{i,j}A'_{i,j} \geq \gamma_2(A)^{-1} = \text{mc}(S)^{-1}.$$

Ч.т.д. Оказывается, эта оценка усиливает теорему Форстера

($\text{mc}(S) \geq \sqrt{mn}/\|S\|_{2 \rightarrow 2}$), поскольку $\gamma_2^*(S) \leq \sqrt{mn}\|S\|_{2 \rightarrow 2}$

(Упражнение).

Мы доказали, что величины mc , γ_2 и rank связаны следующим образом:

$$mc(S) \leq \gamma_2(S) \leq \sqrt{\text{rank}(S)}.$$

Для матриц Адамара достигается равенство.

Мы доказали, что величины mc , γ_2 и rank связаны следующим образом:

$$mc(S) \leq \gamma_2(S) \leq \sqrt{\text{rank}(S)}.$$

Для матриц Адамара достигается равенство.

Установим связь между mc и rank_{\pm} .

Мы доказали, что величины mc , γ_2 и rank связаны следующим образом:

$$mc(S) \leq \gamma_2(S) \leq \sqrt{\text{rank}(S)}.$$

Для матриц Адамара достигается равенство.

Установим связь между mc и rank_\pm .

Утверждение

$$\text{rank}_\pm(S) \leq C mc(S)^2 \log(m+n),$$

где C — абсолютная постоянная.

Johnson–Lindenstrauss

Нам потребуется лемма Johnson–Lindenstrauss.

Johnson–Lindenstrauss

Нам потребуется лемма Johnson–Lindenstrauss.

Утверждение

Пусть R — матрица $d \times N$ со стандартными гауссовыми элементами, т.е. $R_{i,j} \sim \mathcal{N}(0, 1)$. Тогда для любых векторов $x, y \in \mathbb{R}^N$, $|x|, |y| \leq 1$ и $\varepsilon \in (0, 1/2)$ имеем

$$P\left(\left|\left\langle \frac{1}{\sqrt{d}}Rx, \frac{1}{\sqrt{d}}Ry \right\rangle - \langle x, y \rangle\right| \geq \varepsilon\right) \leq 2 \exp(-d\varepsilon^2/8).$$

Johnson–Lindenstrauss

Нам потребуется лемма Johnson–Lindenstrauss.

Утверждение

Пусть R — матрица $d \times N$ со стандартными гауссовыми элементами, т.е. $R_{i,j} \sim \mathcal{N}(0, 1)$. Тогда для любых векторов $x, y \in \mathbb{R}^N$, $|x|, |y| \leq 1$ и $\varepsilon \in (0, 1/2)$ имеем

$$P(|\langle \frac{1}{\sqrt{d}}Rx, \frac{1}{\sqrt{d}}Ry \rangle - \langle x, y \rangle| \geq \varepsilon) \leq 2 \exp(-d\varepsilon^2/8).$$

Отметим, что от размерности N ничего не зависит.

Johnson–Lindenstrauss

Нам потребуется лемма Johnson–Lindenstrauss.

Утверждение

Пусть R — матрица $d \times N$ со стандартными гауссовыми элементами, т.е. $R_{i,j} \sim \mathcal{N}(0, 1)$. Тогда для любых векторов $x, y \in \mathbb{R}^N$, $|x|, |y| \leq 1$ и $\varepsilon \in (0, 1/2)$ имеем

$$P\left(\left|\left\langle \frac{1}{\sqrt{d}}Rx, \frac{1}{\sqrt{d}}Ry \right\rangle - \langle x, y \rangle\right| \geq \varepsilon\right) \leq 2 \exp(-d\varepsilon^2/8).$$

Отметим, что от размерности N ничего не зависит.

Следствие (Johnson–Lindenstrauss): M точек в единичном шаре можно (линейно) отобразить в пространство размерности $d \asymp \log(M)\varepsilon^{-2}$, изменив попарные расстояния не более чем на ε .

Johnson–Lindenstrauss

Нам потребуется лемма Johnson–Lindenstrauss.

Утверждение

Пусть R — матрица $d \times N$ со стандартными гауссовыми элементами, т.е. $R_{ij} \sim \mathcal{N}(0, 1)$. Тогда для любых векторов $x, y \in \mathbb{R}^N$, $|x|, |y| \leq 1$ и $\varepsilon \in (0, 1/2)$ имеем

$$P(|\langle \frac{1}{\sqrt{d}}Rx, \frac{1}{\sqrt{d}}Ry \rangle - \langle x, y \rangle| \geq \varepsilon) \leq 2 \exp(-d\varepsilon^2/8).$$

Отметим, что от размерности N ничего не зависит.

Следствие (Johnson–Lindenstrauss): M точек в единичном шаре можно (линейно) отобразить в пространство размерности $d \asymp \log(M)\varepsilon^{-2}$, изменив попарные расстояния не более чем на ε . Действительно, нам требуется сохранить M^2 скалярных произведений; если $2M^2 \exp(-d\varepsilon^2/8) < 1$, то случайная матрица делает это с положительной вероятностью.

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll mc(S)^2 \log(m+n)$.

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll mc(S)^2 \log(m+n)$.

Возьмём реализацию $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с $|x_i| = |y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$.

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll mc(S)^2 \log(m+n)$.

Возьмём реализацию $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с $|x_i| = |y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Применим лемму Johnson–Lindenstrauss:

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll mc(S)^2 \log(m+n)$.

Возьмём реализацию $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с $|x_i| = |y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Применим лемму Johnson–Lindenstrauss:

$$P(|\langle x'_i, y'_j \rangle - \langle x_i, y_j \rangle| \geq \varepsilon) \leq 2 \exp(-d\varepsilon^2/8). \quad (*)$$

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll \text{mc}(S)^2 \log(m+n)$.

Возьмём реализацию $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с $|x_i| = |y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Применим лемму Johnson–Lindenstrauss:

$$P(|\langle x'_i, y'_j \rangle - \langle x_i, y_j \rangle| \geq \varepsilon) \leq 2 \exp(-d\varepsilon^2/8). \quad (*)$$

Если $2mn \exp(-d\varepsilon^2/8) < 1$, то найдутся вектора $\{x'_i\}, \{y'_j\}$ в d -мерном пространстве, такие что (*) выполнено для всех $i = 1, \dots, m$, $j = 1, \dots, n$.

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll \text{mc}(S)^2 \log(m+n)$.

Возьмём реализацию $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с $|x_i| = |y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Применим лемму Johnson–Lindenstrauss:

$$P(|\langle x'_i, y'_j \rangle - \langle x_i, y_j \rangle| \geq \varepsilon) \leq 2 \exp(-d\varepsilon^2/8). \quad (*)$$

Если $2mn \exp(-d\varepsilon^2/8) < 1$, то найдутся вектора $\{x'_i\}, \{y'_j\}$ в d -мерном пространстве, такие что (*) выполнено для всех $i = 1, \dots, m$, $j = 1, \dots, n$.

При этом, если $\varepsilon < \text{margin}(S)$, например, $\varepsilon = \frac{1}{2} \text{margin}(S)$, то знак $\langle x_i, y_j \rangle$ не поменяется: $\text{sign}\langle x'_i, y'_j \rangle = S_{i,j}$. Тогда $\text{rank}_{\pm}(S) \leq d$.

Докажем теперь оценку $\text{rank}_{\pm}(S) \ll \text{mc}(S)^2 \log(m+n)$.

Возьмём реализацию $S_{i,j} = \text{sign}\langle x_i, y_j \rangle$ с $|x_i| = |y_j| = 1$ и максимальным отступом $\text{margin}(S) = \min |\langle x_i, y_j \rangle|$. Применим лемму Johnson–Lindenstrauss:

$$P(|\langle x'_i, y'_j \rangle - \langle x_i, y_j \rangle| \geq \varepsilon) \leq 2 \exp(-d\varepsilon^2/8). \quad (*)$$

Если $2mn \exp(-d\varepsilon^2/8) < 1$, то найдутся вектора $\{x'_i\}, \{y'_j\}$ в d -мерном пространстве, такие что (*) выполнено для всех $i = 1, \dots, m$, $j = 1, \dots, n$.

При этом, если $\varepsilon < \text{margin}(S)$, например, $\varepsilon = \frac{1}{2} \text{margin}(S)$, то знак $\langle x_i, y_j \rangle$ не поменяется: $\text{sign}\langle x'_i, y'_j \rangle = S_{i,j}$. Тогда $\text{rank}_{\pm}(S) \leq d$. Условие на d :

$$d > 8\varepsilon^{-2} \ln(2mn) \asymp \text{mc}(S)^2 \log(2mn).$$

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Пусть μ — мера на множестве индексов $[m] \times [n]$.

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Пусть μ — мера на множестве индексов $[m] \times [n]$. Для $R \subset [m] \times [n]$ обозначим через R_+ множество индексов $(i, j) \in R$ в которых $S_{i,j} = 1$, и R_- соответственно.

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Пусть μ — мера на множестве индексов $[m] \times [n]$. Для $R \subset [m] \times [n]$ обозначим через R_+ множество индексов $(i, j) \in R$ в которых $S_{i,j} = 1$, и R_- соответственно. Тогда

$$\text{disc}_\mu(S) := \max_R |\mu(R_+) - \mu(R_-)|,$$

где максимум берётся по всевозможным комбинаторным прямоугольникам $R = R' \times R'' \subset [m] \times [n]$.

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Пусть μ — мера на множестве индексов $[m] \times [n]$. Для $R \subset [m] \times [n]$ обозначим через R_+ множество индексов $(i, j) \in R$ в которых $S_{i,j} = 1$, и R_- соответственно. Тогда

$$\text{disc}_\mu(S) := \max_R |\mu(R_+) - \mu(R_-)|,$$

где максимум берётся по всевозможным комбинаторным прямоугольникам $R = R' \times R'' \subset [m] \times [n]$. Положим также $\text{disc}(S) := \min_\mu \text{disc}_\mu(S)$, минимум берётся по всем *вероятностным* распределениям μ .

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Пусть μ — мера на множестве индексов $[m] \times [n]$. Для $R \subset [m] \times [n]$ обозначим через R_+ множество индексов $(i, j) \in R$ в которых $S_{i,j} = 1$, и R_- соответственно. Тогда

$$\text{disc}_\mu(S) := \max_R |\mu(R_+) - \mu(R_-)|,$$

где максимум берётся по всевозможным комбинаторным прямоугольникам $R = R' \times R'' \subset [m] \times [n]$. Положим также $\text{disc}(S) := \min_\mu \text{disc}_\mu(S)$, минимум берётся по всем *вероятностным* распределениям μ .

В лекции №1 мы установили неравенство $C(f) \geq \log_2(1/\text{disc}(f))$ для коммуникационной сложности в детерминированной модели

Дискрепанс

Напомним понятие дискрепанса матрицы $S \in \{-1, 1\}^{m \times n}$.

Пусть μ — мера на множестве индексов $[m] \times [n]$. Для $R \subset [m] \times [n]$ обозначим через R_+ множество индексов $(i, j) \in R$ в которых $S_{i,j} = 1$, и R_- соответственно. Тогда

$$\text{disc}_\mu(S) := \max_R |\mu(R_+) - \mu(R_-)|,$$

где максимум берётся по всевозможным комбинаторным прямоугольникам $R = R' \times R'' \subset [m] \times [n]$. Положим также $\text{disc}(S) := \min_\mu \text{disc}_\mu(S)$, минимум берётся по всем *вероятностным* распределениям μ .

В лекции №1 мы установили неравенство $C(f) \geq \log_2(1/\text{disc}(f))$ для коммуникационной сложности в детерминированной модели ($\chi \leq 2^C$, $\text{disc} \geq 1/\chi$).

Дискрепанс и CUT-норма

В случае равномерной меры μ величина $\text{disc}_\mu(S)$ выражается следующим образом:

$$mn \text{disc}_\mu(S) = \max_{R', R''} \left| \sum_{i \in R', j \in R''} S_{i,j} \right|.$$

Дискрепанс и CUT-норма

В случае равномерной меры μ величина $\text{disc}_\mu(S)$ выражается следующим образом:

$$mn \text{disc}_\mu(S) = \max_{R', R''} \left| \sum_{i \in R', j \in R''} S_{i,j} \right|.$$

Величина в правой части равенства называется также CUT-нормой: $\|S\|_{\text{CUT}}$. CUT — т.к. мы “вырезаем” из матрицы прямоугольник и суммируем элементы. Можно считать её дискрепансом по отношению к считающей мере $\mu_{i,j} \equiv 1$. Кроме того, CUT-норма определена для всех вещественных матриц (и является нормой).

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1,1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1,1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_u(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1,1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_u(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Пусть Ω — множество и \mathcal{A} — некоторое семейство его подмножеств.

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1,1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_u(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Пусть Ω — множество и \mathcal{A} — некоторое семейство его подмножеств. Задача: раскрасить Ω в два цвета так, чтобы в каждом $A \in \mathcal{A}$ было примерно поровну точек обоих цветов.

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1, 1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_u(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Пусть Ω — множество и \mathcal{A} — некоторое семейство его подмножеств. Задача: раскрасить Ω в два цвета так, чтобы в каждом $A \in \mathcal{A}$ было примерно поровну точек обоих цветов. Раскраска $\chi: \Omega \rightarrow \{-1, 1\}$ имеет дискрепанс

$$\text{disc}(\chi, \mathcal{A}) = \max_{A \in \mathcal{A}} \left| \sum_{x \in A} \chi(x) \right|.$$

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1, 1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_U(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Пусть Ω — множество и \mathcal{A} — некоторое семейство его подмножеств. Задача: раскрасить Ω в два цвета так, чтобы в каждом $A \in \mathcal{A}$ было примерно поровну точек обоих цветов. Раскраска $\chi: \Omega \rightarrow \{-1, 1\}$ имеет дискрепанс

$$\text{disc}(\chi, \mathcal{A}) = \max_{A \in \mathcal{A}} \left| \sum_{x \in A} \chi(x) \right|.$$

Дискрепанс семейства \mathcal{A} определяется как $\min_{\chi} \text{disc}(\chi, \mathcal{A})$.

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1, 1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_u(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Пусть Ω — множество и \mathcal{A} — некоторое семейство его подмножеств. Задача: раскрасить Ω в два цвета так, чтобы в каждом $A \in \mathcal{A}$ было примерно поровну точек обоих цветов. Раскраска $\chi: \Omega \rightarrow \{-1, 1\}$ имеет дискрепанс

$$\text{disc}(\chi, \mathcal{A}) = \max_{A \in \mathcal{A}} \left| \sum_{x \in A} \chi(x) \right|.$$

Дискрепанс семейства \mathcal{A} определяется как $\min_{\chi} \text{disc}(\chi, \mathcal{A})$.

$\text{disc}_u(S)$ это дискрепанс для множества $\Omega = [m] \times [n]$, семейства $\mathcal{R} = \{R' \times R''\}$ комбинаторных прямоугольников и $\chi = S$.

Комбинаторный дискрепанс

Ранее нам уже встречался дискрепанс матрицы M вида $\text{disc}_{old}(M) = \min_{x \in \{-1, 1\}^n} \|Mx\|_\infty$. В чём связь между этими понятиями?

$\text{disc}_u(S)$ и $\text{disc}_{old}(M)$ происходят из общего понятия — *комбинаторного дискрепанса*.

Пусть Ω — множество и \mathcal{A} — некоторое семейство его подмножеств. Задача: раскрасить Ω в два цвета так, чтобы в каждом $A \in \mathcal{A}$ было примерно поровну точек обоих цветов. Раскраска $\chi: \Omega \rightarrow \{-1, 1\}$ имеет дискрепанс

$$\text{disc}(\chi, \mathcal{A}) = \max_{A \in \mathcal{A}} \left| \sum_{x \in A} \chi(x) \right|.$$

Дискрепанс семейства \mathcal{A} определяется как $\min_\chi \text{disc}(\chi, \mathcal{A})$.

$\text{disc}_u(S)$ это дискрепанс для множества $\Omega = [m] \times [n]$, семейства $\mathcal{R} = \{R' \times R''\}$ комбинаторных прямоугольников и $\chi = S$.

$\text{disc}_{old}(S)$ это дискрепанс семейства множеств $A_i \subset \Omega = [n]$, задаваемых строками S_i (т.е. $j \in A_i$, если $S_{i,j} = 1$).

Дискрепанс и $m\epsilon$

Theorem (Linial, Shraibman, 2008)

$$\frac{1}{8} \text{margin}(S) \leq \text{disc}(S) \leq 8 \text{margin}(S).$$

Дискрепанс и $m\epsilon$

Theorem (Linial, Shraibman, 2008)

$$\frac{1}{8} \text{margin}(S) \leq \text{disc}(S) \leq 8 \text{margin}(S).$$

Доказательство.

Дискрепанс и $m\epsilon$

Theorem (Linial, Shraibman, 2008)

$$\frac{1}{8} \text{margin}(S) \leq \text{disc}(S) \leq 8 \text{margin}(S).$$

Доказательство.

1 шаг. Заменяем в определении margin произвольные вектора $\{x_i\}, \{y_j\} \in \mathbb{R}^N$ на знаковые $\{x_i\}, \{y_j\} \in \{-1, 1\}^N$. Получится величина $\text{margin}_{\pm}(S) \leq \text{margin}(S)$.

Дискрепанс и m_S

Theorem (Linial, Shraibman, 2008)

$$\frac{1}{8} \text{margin}(S) \leq \text{disc}(S) \leq 8 \text{margin}(S).$$

Доказательство.

1 шаг. Заменяем в определении margin произвольные вектора $\{x_i\}, \{y_j\} \in \mathbb{R}^N$ на знаковые $\{x_i\}, \{y_j\} \in \{-1, 1\}^N$. Получится величина $\text{margin}_{\pm}(S) \leq \text{margin}(S)$.

Мы докажем (используя неравенство Гротендика), что эти величины эквивалентны:

$$\text{margin}_{\pm}(S) \leq \text{margin}(S) \leq K_G \text{margin}_{\pm}(S).$$

Дискрепанс и m_S

Theorem (Linial, Shraibman, 2008)

$$\frac{1}{8} \text{margin}(S) \leq \text{disc}(S) \leq 8 \text{margin}(S).$$

Доказательство.

1 шаг. Заменяем в определении margin произвольные вектора $\{x_i\}, \{y_j\} \in \mathbb{R}^N$ на знаковые $\{x_i\}, \{y_j\} \in \{-1, 1\}^N$. Получится величина $\text{margin}_{\pm}(S) \leq \text{margin}(S)$.

Мы докажем (используя неравенство Гротендика), что эти величины эквивалентны:

$$\text{margin}_{\pm}(S) \leq \text{margin}(S) \leq K_G \text{margin}_{\pm}(S).$$

Левое неравенство очевидно (строже требования к x_i, y_j , меньше отступ).

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Рассмотрим матрицу $B = \left(\frac{\langle x_i, y_j \rangle}{|x_i| |y_j|} \right)$, где $x_i, y_j \in \{-1, 1\}^N$. Имеем

$B_{i,j} = N^{-1} \sum_{p=1}^N x_{i,p} y_{j,p}$. Тем самым, B есть выпуклая комбинация одноранговых сигнум матриц $B_p = (x_{i,p} y_{j,p})_{i,j}$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Рассмотрим матрицу $B = \left(\frac{\langle x_i, y_j \rangle}{|x_i| |y_j|} \right)$, где $x_i, y_j \in \{-1, 1\}^N$. Имеем

$B_{i,j} = N^{-1} \sum_{p=1}^N x_{i,p} y_{j,p}$. Тем самым, B есть выпуклая комбинация одноранговых сигнум матриц $B_p = (x_{i,p} y_{j,p})_{i,j}$.

Обозначим через M_{\pm} выпуклую оболочку всех одноранговых $m \times n$ сигнум-матриц.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Рассмотрим матрицу $B = \left(\frac{\langle x_i, y_j \rangle}{|x_i| |y_j|} \right)$, где $x_i, y_j \in \{-1, 1\}^N$. Имеем

$B_{i,j} = N^{-1} \sum_{p=1}^N x_{i,p} y_{j,p}$. Тем самым, B есть выпуклая комбинация одноранговых сигнум матриц $B_p = (x_{i,p} y_{j,p})_{i,j}$.

Обозначим через M_{\pm} выпуклую оболочку всех одноранговых $m \times n$ сигнум-матриц. Тогда $B \in M_{\pm}$. Обратно, любую матрицу из выпуклой оболочки можно приблизить такой матрицей B (с коэффициентами $1/N$).

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Рассмотрим матрицу $B = \left(\frac{\langle x_i, y_j \rangle}{|x_i| |y_j|} \right)$, где $x_i, y_j \in \{-1, 1\}^N$. Имеем

$B_{i,j} = N^{-1} \sum_{p=1}^N x_{i,p} y_{j,p}$. Тем самым, B есть выпуклая комбинация одноранговых сигнум матриц $B_p = (x_{i,p} y_{j,p})_{i,j}$.

Обозначим через M_{\pm} выпуклую оболочку всех одноранговых $m \times n$ сигнум-матриц. Тогда $B \in M_{\pm}$. Обратно, любую матрицу из выпуклой оболочки можно приблизить такой матрицей B (с коэффициентами $1/N$).

Следовательно,

$$\text{margin}_{\pm}(S) = \max_{B \in M_{\pm}} \min_{i,j} S_{i,j} B_{i,j}.$$

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Рассмотрим матрицу $B = \left(\frac{\langle x_i, y_j \rangle}{|x_i| |y_j|} \right)$, где $x_i, y_j \in \{-1, 1\}^N$. Имеем

$B_{i,j} = N^{-1} \sum_{p=1}^N x_{i,p} y_{j,p}$. Тем самым, B есть выпуклая комбинация одноранговых сигнум матриц $B_p = (x_{i,p} y_{j,p})_{i,j}$.

Обозначим через M_{\pm} выпуклую оболочку всех одноранговых $m \times n$ сигнум-матриц. Тогда $B \in M_{\pm}$. Обратно, любую матрицу из выпуклой оболочки можно приблизить такой матрицей B (с коэффициентами $1/N$).

Следовательно,

$$\text{margin}_{\pm}(S) = \max_{B \in M_{\pm}} \min_{i,j} S_{i,j} B_{i,j}.$$

По определению,

$$\text{margin}(S) = \max_{\gamma_2(A) \leq 1} \min_{i,j} S_{i,j} A_{i,j}.$$

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Рассмотрим матрицу $B = \left(\frac{\langle x_i, y_j \rangle}{|x_i| |y_j|} \right)$, где $x_i, y_j \in \{-1, 1\}^N$. Имеем

$B_{i,j} = N^{-1} \sum_{p=1}^N x_{i,p} y_{j,p}$. Тем самым, B есть выпуклая комбинация одноранговых сигнум матриц $B_p = (x_{i,p} y_{j,p})_{i,j}$.

Обозначим через M_{\pm} выпуклую оболочку всех одноранговых $m \times n$ сигнум-матриц. Тогда $B \in M_{\pm}$. Обратно, любую матрицу из выпуклой оболочки можно приблизить такой матрицей B (с коэффициентами $1/N$).

Следовательно,

$$\text{margin}_{\pm}(S) = \max_{B \in M_{\pm}} \min_{i,j} S_{i,j} B_{i,j}.$$

По определению,

$$\text{margin}(S) = \max_{\gamma_2(A) \leq 1} \min_{i,j} S_{i,j} A_{i,j}.$$

Из неравенства Гротендика мы вывели, что $\gamma_2^*(A) \leq K_G \|A\|_{\infty \rightarrow 1}$. Следовательно, для сопряжённой нормы $\gamma_2(A) \geq K_G^{-1} \|A\|_{\nu}$. Нетрудно проверить, что M_{\pm} это шар в норме $\|A\|_{\nu}$, сопряженной к $\|A\|_{\infty \rightarrow 1}$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

2 шаг. Докажем, что $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

2 шаг. Докажем, что $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$.

Здесь $(S \circ \mu)_{i,j} = S_{i,j} \mu_{i,j}$. Ясно, что неравенство сводится к матрице $T = (S \circ \mu)$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

2 шаг. Докажем, что $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$.

Здесь $(S \circ \mu)_{i,j} = S_{i,j} \mu_{i,j}$. Ясно, что неравенство сводится к матрице $T = (S \circ \mu)$.

Имеем $\text{disc}_\mu(S) = \|T\|_{\text{CUT}}$, нужно доказать:

$$\|T\|_{\text{CUT}} \leq \|T\|_{\infty \rightarrow 1} \leq 4\|T\|_{\text{CUT}}.$$

Левое неравенство: $\sum_{i \in R', j \in R''} T_{i,j} = \langle T \mathbf{1}_{R''}, \mathbf{1}_{R'} \rangle \leq \|T\|_{\infty \rightarrow 1}$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

2 шаг. Докажем, что $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$.

Здесь $(S \circ \mu)_{i,j} = S_{i,j} \mu_{i,j}$. Ясно, что неравенство сводится к матрице $T = (S \circ \mu)$.

Имеем $\text{disc}_\mu(S) = \|T\|_{\text{CUT}}$, нужно доказать:

$$\|T\|_{\text{CUT}} \leq \|T\|_{\infty \rightarrow 1} \leq 4\|T\|_{\text{CUT}}.$$

Левое неравенство: $\sum_{i \in R', j \in R''} T_{i,j} = \langle T \mathbf{1}_{R''}, \mathbf{1}_{R'} \rangle \leq \|T\|_{\infty \rightarrow 1}$. Правое

неравенство: для любых $x_i, y_j \in \{-1, 1\}$ имеем

$$\sum t_{i,j} x_i y_j = \sum_{\substack{x_i=1 \\ y_j=1}} t_{i,j} - \sum_{\substack{x_i=1 \\ y_j=-1}} t_{i,j} - \sum_{\substack{x_i=-1 \\ y_j=1}} t_{i,j} + \sum_{\substack{x_i=-1 \\ y_j=-1}} t_{i,j}.$$

Ясно, что эта величина не превосходит $4\|T\|_{\text{CUT}}$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Итак, $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$, следовательно,

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Итак, $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$, следовательно,

$$\text{disc}(S) \leq \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}(S).$$

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Итак, $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$, следовательно,

$$\text{disc}(S) \leq \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}(S).$$

3 шаг. Остаётся доказать, что

$$\text{margin}_\pm(S) = \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1}.$$

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Итак, $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$, следовательно,

$$\text{disc}(S) \leq \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}(S).$$

3 шаг. Остаётся доказать, что

$$\text{margin}_{\pm}(S) = \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1}.$$

Мы выяснили, что $\text{margin}_{\pm}(S) = \max_{B \in M_{\pm}} \min_{i,j} S_{i,j} B_{i,j}$. Множество M_{\pm} это многогранник, вершины которого – одноранговые сингум-матрицы. Будем обозначать такие матрицы как X^q , $q \in Q$. Таким образом, $B = \sum_{q \in Q} \lambda_q X^q$, где $\sum \lambda_q = 1$, $\lambda_q \geq 0$.

Доказательство $\text{disc} \asymp \text{margin}$ (продолжение)

Итак, $\text{disc}_\mu(S) \leq \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}_\mu(S)$, следовательно,

$$\text{disc}(S) \leq \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1} \leq 4 \text{disc}(S).$$

3 шаг. Остаётся доказать, что

$$\text{margin}_{\pm}(S) = \inf_{\mu} \|S \circ \mu\|_{\infty \rightarrow 1}.$$

Мы выяснили, что $\text{margin}_{\pm}(S) = \max_{B \in M_{\pm}} \min_{i,j} S_{i,j} B_{i,j}$. Множество M_{\pm} это многогранник, вершины которого – одноранговые сингум-матрицы. Будем обозначать такие матрицы как X^q , $q \in Q$. Таким образом, $B = \sum_{q \in Q} \lambda_q X^q$, где $\sum \lambda_q = 1$, $\lambda_q \geq 0$.

Обозначим $\delta := \min_{i,j} S_{i,j} B_{i,j}$. Нам нужно максимизировать δ при условии $S_{i,j} \sum_q \lambda_q X_{i,j}^q \geq \delta$.

Доказательство $\text{disc} \asymp \text{margin}$ (окончание)

Мы приходим к задаче линейного программирования:

$$\begin{cases} \delta \rightarrow \max, \\ \sum_q \lambda_q (X^q \circ S)_{i,j} \geq \delta, \\ \sum_q \lambda_q = 1, \lambda_q \geq 0. \end{cases}$$

Доказательство $\text{disc} \asymp \text{margin}$ (окончание)

Мы приходим к задаче линейного программирования:

$$\begin{cases} \delta \rightarrow \max, \\ \sum_q \lambda_q (X^q \circ S)_{i,j} \geq \delta, \\ \sum_q \lambda_q = 1, \lambda_q \geq 0. \end{cases}$$

Упражнение: найдите двойственную задачу и докажите, что её значение равно $\|S\|_{\infty \rightarrow 1}$.



N. Linial, S. Mendelson, G. Schechtman, A. Shraibman,
“COMPLEXITY MEASURES OF SIGN MATRICES” (2007).



S.V. Lokam, “Complexity Lower Bounds using Linear Algebra” (2009).