Спецкурс “Многомерная аппроксимация и анализ булевых функций”. Лекции 4 и 5: Теоремы о влиянии переменных

Теорема FKN. На прошлой лекции мы доказали лемму Bonami. Из неё следует, что для функций ограниченной степени L_2 и L_4 нормы эквивалентны. В этом случае будут эквивалентны все L_p-нормы, $p \leqslant 4$ и, более того, такие функции большие на множестве большой меры. Сформулируем это в общем виде, для произвольных случайных величин.

Утверждение 1 (Неравенство Пэли–Зигмунда)**.** *Пусть случайная величина* $\xi \geqslant 0$, $t \in (0,1)$, *тогда*

$$\mathsf{P}(\xi > t\mathsf{E}\xi) \geqslant (1-t)^2 \frac{(\mathsf{E}\xi)^2}{\mathsf{E}\xi^2}.$$

Доказательство. Оценим мат. ожидание ξ, разбивая интеграл на два множества и используя неравенство Коши–Буняковского:

$$\mathsf{E}\xi = \mathsf{E}\xi\mathbf{1}\{\xi \leqslant t\mathsf{E}\xi\} + \mathsf{E}\xi\mathbf{1}\{\xi > t\mathsf{E}\xi\} \leqslant t\mathsf{E}\xi + (\mathsf{E}\xi^2)^{1/2}\mathsf{P}(\xi > t\mathsf{E}\xi)^{1/2}.$$

Перенося $t\mathsf{E}\xi$ в правую часть, получим требуемую оценку на вероятность. □

Пусть $f\colon \{-1,1\}^n \to \{-1,1\}$. По равенству Парсеваля, $\sum_S \hat{f}(S)^2 = 1$. Таким образом, числа $\hat{f}(S)^2$ задают распределение вероятностей. Иногда полезно сгруппировать их по количеству элементов:

$$W^k(f) := \sum_{|S|=k} \hat{f}(S)^2.$$

Теорема 1 (Friedgut, Kalai, Naor)**.** *Пусть* $f\colon \{-1,1\}^n \to \{-1,1\}$, *и* $W^1(f) = 1-\delta$, *где* $\delta \in (0,1)$. *Тогда найдутся индекс* $i \in [n]$ *и знак* $\sigma \in \{\pm 1\}$ *такие, что* $\operatorname{dist}(f, \sigma\chi_{\{i\}}) \leqslant C\delta$.

Напомним, что $\operatorname{dist}(f,g)$ это доля входов, на которых f и g отличаются.

Доказательство. Рассмотрим $g(x) := \sum_{i=1}^n \hat{f}(\{i\})\chi_{\{i\}}(x)$. Запишем это более кратко (допуская вольность в обозначениях): $g(x) = \sum_{i=1}^n \hat{f}(i)x_i$. Это — линейная часть функции f. Тогда

$$g^2(x) = \sum_{i=1}^n \hat{f}(i)^2 + \sum_{i\neq j} \hat{f}(i)\hat{f}(j)x_ix_j.$$

Из условия следует, что $\mathsf{E}g^2 = 1-\delta$. Докажем, что имеет место оценка

$$\operatorname{Var} g^2 \leqslant C\delta. \tag{1}$$

Используем неравенство Пэли–Зигмунда для $\xi = (g^2-(1-\delta))^2$ и лемму Бонами:

$$\mathsf{P}(|g^2-(1-\delta)| > \frac{1}{2}\sqrt{\operatorname{Var} g}) = \mathsf{P}((g^2-(1-\delta))^2 > \frac{1}{4}\operatorname{Var} g) \gtrsim \frac{[\mathsf{E}(g^2-(1-\delta))]^2}{\mathsf{E}(g^2-(1-\delta))^4} \gtrsim \text{const} > 0.$$

Если предположить, что $\operatorname{Var} g > C\delta$ с достаточно большим C, то $\mathsf{P}(|g^2-1| > C_1\sqrt{\delta}) > \text{const} > 0$. Но (проверьте!) если $|y_1^2-1| > \varepsilon$ для достаточно малого ε, и $|y_2| = 1$, то $|y_1 -$

$y_2|^2 > c\varepsilon^2$. Следовательно, если $|g^2 - 1| > C_1\sqrt{\delta}$, то $|f - g|^2 \geqslant C_2\delta$. Интегрируя, получаем $\mathsf{E}|f-g|^2 \geqslant \text{const}\cdot C_2\delta$, что больше δ, если C_2 достаточно велико. С другой стороны, $\mathsf{E}|f-g|^2 = 1 - W^1(f) = \delta$. Противоречие; значит, справедливо неравенство (1).

Можно вычислить $\operatorname{Var} g^2$, используя явный вид g^2:

$$\operatorname{Var} g^2 = \sum_{i\neq j} \hat{f}(i)^2\hat{f}(j)^2 = (\sum \hat{f}(i)^2)^2 - \sum \hat{f}(i)^4 = (1-\delta)^2 - \sum \hat{f}(i)^4 \leqslant C\delta.$$

Отсюда $\sum \hat{f}(i)^4 \geqslant 1 - C'\delta$. С другой стороны, $\sum \hat{f}(i)^4 \leqslant \|\hat{f}\|_{\ell_2}^2\|\hat{f}\|_{\ell_\infty}^2 = \|\hat{f}\|_{\ell_\infty}^2$. Значит, найдётся большой коэффициент Фурье: $|\hat{f}(i^*)| > 1 - C''\delta$. Нетрудно видеть, что тогда $\operatorname{dist}(f, \operatorname{sign}\hat{f}(i^*)\cdot\chi_{\{i^*\}})| \leqslant C''\delta$. □

Функции "голосования". Функцию $f\colon \{-1,1\}^n \to \{-1,1\}$ можно воспринимать как способ скомбинировать голоса $x_i \in \{-1,1\}$, $i = 1,\ldots,n$, в единый результат.

Простейший пример функции голосования — *простым большинством* (majority vote). Пусть n нечётно. Положим $\mathrm{MAJ}(x_1,\ldots,x_n) = \operatorname{sign}(x_1+\ldots+x_n)$. Вычислим влияние каждого из x_i. В силу симметрии, достаточно рассмотреть $i = 1$. Голос x_1 является решающим (при фиксированных остальных), если $x_2 + \ldots + x_n = 0$. Следовательно,

$$\operatorname{Inf}_1 \mathrm{MAJ} = \mathsf{P}(x_2 + \ldots + x_n = 0) = \frac{\binom{n-1}{n/2}}{2^{n-1}} \asymp n^{-1/2}.$$

В следующей функции влияние каждого голоса меньше. *Голосование племён* (Tribes function). Пусть есть b племён, каждое размера s (size). Всего $n = bs$ переменных $x_1,\ldots,x_n$, разбитых на множества $I_1,\ldots,I_b$, $|I_k| = s$. Голосование устроено следующим образом: решение отвергается ($f(x) = -1$), если и только если какое-то племя его единогласно отвергает:

$$f(x) = -1 \quad \Leftrightarrow \quad \exists k\colon x_i = -1 \text{ для всех } i \in I_k.$$

Найдём $\mathsf{P}(f(x) = 1)$. В каждом племени k допускается любой набор голосов, кроме единогласного -1, вероятность равна $1 - 2^{-s}$. Поскольку все голоса независимы, получаем

$$\mathsf{P}(f(x) = 1) = (1 - 2^{-s})^b.$$

Возьмём $b = 2^s$, тогда $\mathsf{P}(f(x) = 1) = (1 - 1/b)^b \approx 1/e$. Это значит, что функция сбалансирована в том смысле, что $\mathsf{E}f$ отделено от -1 и 1 (эквивалентно, дисперсия $\operatorname{Var} f = 1 - (\mathsf{E}f(x))^2$ отделена от нуля). Вычислим влияние переменных.

Утверждение 2. *Пусть $b = 2^s$. Для любого i имеем $\operatorname{Inf}_i f \asymp \log(n)/n$.*

Доказательство. Достаточно рассмотреть Inf_1. Когда результат зависит от x_1:

- в племени I_1 все остальные голосуют "против": $x_i = -1$;
- в остальных племенах нет консенсуса "против".

Вероятность этого равна $2^{-(s-1)}\cdot(1-2^{-s})^{b-1}$. Второй множитель $\asymp$ const. Первый множитель $\asymp 2^{-s}$, выразим эту величину через n. Имеем $n = bs = s2^s$, откуда $s \asymp \log(n)$, $2^s = n/s \asymp n/\log n$. Что и требовалось. □

Приведём ещё два примера функций голосования. *Диктатор* это функция $f(x) = x_i$, т.е. всё определяет i-й голос. Функция называется *k-хунтой*, если она зависит от каких-то k голосов, т.е. $f(x_1,\ldots,x_n) = g(x_{i_1},\ldots,x_{i_k})$ для некоторой g и индексов $i_1,\ldots,i_k$.

Оценки влияния переменных. Приведём сначала элементарную оценку. Ранее мы уже видели, что $\operatorname{Inf}_i f = \sum_{i\in S} \hat{f}(S)^2$, и

$$I(f) = \sum_{i=1}^{n} \operatorname{Inf}_i f = \sum |S| \hat{f}(S)^2 \geqslant \sum_{S\neq\varnothing} \hat{f}(S)^2 = \operatorname{Var} f.$$

Следовательно, $\max_i \operatorname{Inf}_i(f) \geqslant I(f)/n \geqslant (1/n)\operatorname{Var} f$.

Теорема 2 (Kahn, Kalai, Linial)**.** *Для любой функции* $f\colon \{-1,1\}^n \to \{-1,1\}$ *найдётся индекс* $i \in [n]$ *такой, что*

$$\operatorname{Inf}_i f \geqslant c\frac{\log n}{n} \operatorname{Var} f.$$

KKL-теорема даёт оптимальную оценку, что видно из примера функции Tribes. Нам потребуется следующая

Лемма 1. *Пусть* $F\colon \{-1,1\}^n \to \{-1,0,1\}$, $\mathsf{P}(F \neq 0) = \alpha$, *тогда*

$$\sum_{|S|\leqslant d} \hat{F}(S)^2 \leqslant 3^{d/2}\alpha^{5/4}.$$

Доказательство. Рассмотрим $F^{\leqslant d} := \sum_{|S|\leqslant d} \hat{F}(S)\chi_S$. Воспользуемся неравенством Гёльдера $\langle f, g\rangle \leqslant \|f\|_{L_p}\|g\|_{L_q}$ и доказанной на прошлой лекции леммой Bonami: $1/p + 1/q = 1$:

$$\|F^{\leqslant d}\|_{L_2}^2 = \langle F^{\leqslant d}, F^{\leqslant d}\rangle \leqslant \langle F^{\leqslant d}, F\rangle \leqslant \|F^{\leqslant d}\|_{L_4}\|F\|_{L_{4/3}} \leqslant 3^{d/2}\|F^{\leqslant d}\|_{L_2}\|F\|_{L_{4/3}} \leqslant$$
$$\leqslant 3^{d/2}\|F\|_{L_2}\|F\|_{L_{4/3}} = 3^{d/2}\alpha^{1/2}\alpha^{3/4}.$$

□

Перейдём теперь к доказательству теоремы.

Доказательство. Если $I(f) \geqslant c\log n \cdot \operatorname{Var} f$, то всё доказано. Пусть

$$I(f) \leqslant K \operatorname{Var} f \tag{2}$$

(где $K = c\log n$), докажем, что

$$\max \operatorname{Inf}_i f \geqslant \exp(-c_0 K) \tag{3}$$

с некоторой абсолютной постоянной c_0. Это даст $\max \operatorname{Inf}_i f \geqslant \exp(-c_0 c\log n) \geqslant n^{-1/2} \geqslant \log(n)/n$, если исходное c достаточно мало.

Итак, нужно вывести из (2) неравенство (3). Предположим противное, $\operatorname{Inf}_i f < \alpha := \exp(-c_0 K)$.

Пусть d — параметр, который мы подберём позже. Фиксируем индекс i и рассмотрим $F_i := D_i f$. Заметим, что F_i принимает значения $0, -1, 1$, поэтому к этой функции применима доказанная лемма. Кроме того, $\mathsf{P}(F_i \neq 0) = \operatorname{Inf}_i f < \alpha$ в силу нашего предположения. Имеем

$$\sum_{\substack{|S|\leqslant d+1\\ i\in S}} \hat{f}(S)^2 = \sum_{\substack{|S|\leqslant d\\ i\notin S}} \hat{f}(S\cup\{i\})^2 = \sum_{|S|\leqslant d} \hat{F}_i(S)^2 \leqslant 3^{d/2}(\operatorname{Inf}_i f)^{5/4} \leqslant 3^{d/2}\alpha^{1/4}\operatorname{Inf}_i f.$$

Суммируя по всем $i = 1, \ldots, n$, получим:

$$\sum_{0<|S|\leqslant d+1} \hat{f}(S)^2 \leqslant \sum_{i=1}^{n} \sum_{\substack{|S|\leqslant d+1\\ i\in S}} \hat{f}(S)^2 \leqslant 3^{d/2}\alpha^{1/4}\sum_{i=1}^{n} \operatorname{Inf}_i f = 3^{d/2}\alpha^{1/4} I(f).$$

Теперь рассмотрим сумму

$$\sum_{|S|>d+1} \hat{f}(S)^2 \leqslant \frac{1}{d} \sum_{|S|>d+1} |S| \hat{f}(S)^2 \leqslant \frac{I(f)}{d}.$$

Итого, получаем

$$\operatorname{Var} f = \sum_{|S|>0} \hat{f}(S)^2 \leqslant 3^{d/2} \alpha^{1/4} I(f) + \frac{I(f)}{d}.$$

Воспользуемся предположением (2):

$$\operatorname{Var} f \leqslant \operatorname{Var} f \cdot K \cdot \left(3^{d/2} \alpha^{1/4} + 1/d\right).$$

Покажем, что множитель в скобках меньше $1/K$, что даст противоречие. Возьмём $d = 2K$, тогда $3^{d/2}\alpha^{1/4} = 3^K \exp(-(1/4)c_0 K) < 1/(2K)$, если число c_0 достаточно велико. □

В теории голосования естественно требовать *монотонность* функции $f\colon \{-1,1\}^n \to \{-1,1\}$. Это означает, что если $x_i \leqslant x'_i$ при всех i, то и $f(x) \leqslant f(x')$. (Хотя, в KKL теореме это не требуется.)

Следствие 1. *Пусть $f\colon \{-1,1\}^n \to \{-1,1\}$ монотонна и $\mathsf{E} f \geqslant -0.99$. Существует подмножество $J \subset [n]$, $|J| \leqslant Cn/\log n$, такое что если "подкупить голосующих из J", то результат будет близок к 1, а именно,*

$$\mathsf{E}(f(x) \mid x|_J \equiv 1) \geqslant 0.99.$$

Доказательство. Сначала подкупаем i_1 с максимальным влиянием на $f_0 := f$, т.е. $i_1 = \arg\max \operatorname{Inf}_i f$. Далее подкупаем i_2 с максимальным влиянием на $f_1 := f(x^{(i\mapsto 1)})$. И так далее, пока не выполнится требование $\mathsf{E} f_k \geqslant 0.99$. Заметим, что на каждом шаге

$$\mathsf{E} f_{t+1} \geqslant \mathsf{E} f_t + \max \operatorname{Inf} f_t,$$

поскольку замена переменную на 1 сдвигает среднее как раз на информативность этой переменной. По теореме KKL каждый шаг даёт $\gtrsim \log(n)/n$, откуда количество шагов оценивается как $O(n/\log n)$. □

Близость к хунте.

Утверждение 3. *Пусть $f\colon \{-1,1\}^n \to \{-1,1\}$ имеет степень $\deg f \leqslant k$. Тогда f является $k2^{k-1}$-хунтой.*

Другими словами, многочлен малой степени, принимающий значения ± 1, не может зависеть от большого количества переменных.

Доказательство. Докажем, что если $g\colon \{-1,1\}^k \to \mathbb{R}$ и $\deg g \leqslant k$, то

$$g \equiv 0 \quad \text{или} \quad \mathsf{P}(g(x) \neq 0) \geqslant 2^{-k}. \tag{4}$$

Рассмотрим для этого две функции: $g_1(x) = g(x^{(n\mapsto 1)})$ и $g_{-1}(x) = g(x^{(n\mapsto -1)})$. Если $g_1 \not\equiv 0$ и $g_{-1} \not\equiv 0$, пользуемся индукцией (по числу переменных). Если, скажем, $g_1 \equiv 0$, то покажем, что $\deg g_{-1} \leqslant k-1$. Запишем $g(x) = x_n h(\bar{x}) + v(\bar{x})$, $\deg h \leqslant k-1$. Если $h(\bar{x}) + v(\bar{x}) \equiv 0$, то тогда и $\deg v \leqslant k-1$, откуда $\deg g_{-1} = \deg(-h+v) \leqslant k-1$. Пользуясь индукцией для g_{-1}, получаем (4).

Теперь покажем, что для нашей функции f либо $\operatorname{Inf}_i f = 0$ (т.е. f не зависит вообще от этой переменной), либо $\operatorname{Inf}_i f \geqslant 2^{1-k}$. Действительно, $\operatorname{Inf}_i f = \mathsf{P}(D_i f \neq 0)$ и $\deg D_i \leqslant k-1$, так что остаётся воспользоваться (4).

С другой стороны, $I(f) = \sum |S| \hat{f}(S)^2 \leqslant \deg f \leqslant k$, поэтому количество существенных переменных ($\operatorname{Inf}_i f \neq 0$) не больше $k2^{k-1}$. □

Недостаток этого утверждения в том, что здесь требуется оценка на степень вместо спектральной концентрации, как в теореме FKN. В общем случае известна следующая

Теорема 3 (Friedgut junta theorem)**.** *Пусть* $f\colon \{-1,1\}^n \to \{-1,1\}$*,* $\varepsilon \in (0,1)$*. Тогда существует* k*-хунта* h *со значениями* ± 1*, такая что* $k \leqslant \exp(C\, I(f)/\varepsilon)$*, и* $\operatorname{dist}(f,h) \leqslant \varepsilon$*.*

Доказательство. Выберем влиятельные координаты: $J := \{i\colon \operatorname{Inf}_i f \geqslant \tau\}$, где параметр τ будет определён позже. Очевидно, что $|J| \leqslant I(f)/\tau$.

Построим промежуточное приближение

$$g(x) = \sum_{S \subset J} \hat{f}(S)\chi_S(x).$$

Тогда

$$\|f-g\|_{L_2}^2 = \sum_{S \not\subset J} \hat{f}(S)^2 \leqslant \sum_{|S|>d} \hat{f}(S)^2 + \sum_{\substack{|S|\leqslant d \\ S\not\subset J}} \hat{f}(S)^2,$$

где параметр d будет выбран позже. Первое слагаемое легко оценивается величиной $I(f)/d$, как и раньше. Для оценки второго слагаемого применим лемму (1) к функции $F_i := D_i f$, как в доказательстве теоремы KKL. Получим:

$$\sum_{\substack{|S|\leqslant d \\ i\in S}} \hat{f}(S)^2 \leqslant 3^{d/2}(\operatorname{Inf}_i f)^{5/4},$$

откуда

$$\sum_{\substack{|S|\leqslant d \\ S\not\subset J}} \hat{f}(S)^2 \leqslant \sum_{i\notin J}\sum_{\substack{|S|\leqslant d \\ i\in S}} \hat{f}(S)^2 \leqslant 3^{d/2}\sum_{i\notin J}(\operatorname{Inf}_i f)^{5/4} \leqslant 3^{d/2}\tau^{1/4} I(f).$$

Итого, $\|f-g\|_{L_2}^2 \leqslant I(f)\cdot(1/d + 3^{d/2}\tau^{1/4})$. Берём $d = 2I(f)/\varepsilon$, $\tau = (\varepsilon/(2I(f)3^{d/2}))^4$, это даёт оценку $\|f-g\|_{L_2}^2 \leqslant \varepsilon$. Отсюда для $h(x) := \operatorname{sign} g(x)$ имеем

$$\operatorname{dist}(f,h) \leqslant \mathsf{E}|f-g|^2 \leqslant \varepsilon.$$

При этом h является k-хунтой с $k \leqslant |J| \leqslant I(f)/\tau$, ясно, что эта величина экспоненциально зависит от $I(f)/\varepsilon$. □

Точные пороги (sharp thresholds). Одно из применений техники с влиянием переменных — точные пороги для случайных объектов. Рассмотрим случайный граф $G(m,p)$ с m вершинами, такой что каждое ребро проводится в нём с вероятностью p и независимо от других. Рассмотрим какое-либо свойство этого случайного графа, например, связность. Ясно, что при малых p оно выполнено с малой вероятностью, а при больших p — с большой. Оказывается, этот "фазовый переход" происходит всегда очень быстро!

Формально, закодируем граф $n = \binom{m}{2}$ булевыми переменными (0 если нет ребра и 1 если есть) и пусть $f\colon \{0,1\}^n \to \{0,1\}$ — монотонная симметричная (не меняется от перестановок координат) функция. Граф $G(m,p)$ соответствует распределению на $x \in \{0,1\}^n$, при котором $\mathsf{P}(x_i = 1) = p$. Вероятность с этим распределением обозначим P_p, а мат.ожидание E_p. Тогда $\phi(p) := \mathsf{E}_p f(x)$ и нас интересует величина $p''-p'$, где $\phi(p') = \varepsilon$, $\phi(p'') = 1-\varepsilon$, и ε мало. Обозначим также $\operatorname{Inf}_{i,p} f := \mathsf{P}_p(f(x^{\oplus i}) \neq f(x))$ — аналог влияния для p-распределения.

Докажем формулу Руссо:

$$\phi'(p) = \sum_{i=1}^n \operatorname{Inf}_{i,p}(f).$$

Введём ещё более общее распределение, при котором $\mathsf{P}(x_i = 1) = p_i$, $\bar{p} = (p_1, \ldots, p_n)$, и рассмотрим $\Phi(\bar{p}) = \mathsf{E}_{\bar{p}} f(x)$. Тогда $\Phi(\bar{p}) = p_i \mathsf{E} f(x^{(i \mapsto 1)}) + (1 - p_i)\mathsf{E} f(x^{(i \mapsto -1)})$, откуда

$$\frac{\partial \Phi}{\partial p_i} = \mathsf{E}(f(x^{i \mapsto 1}) - f(x^{i \mapsto 0})) = \mathrm{Inf}_{i,p}(f).$$

Дальше применяем правило дифференцирования функции многих переменных:

$$\frac{d}{dp}\phi(p) = \frac{d}{dp}\Phi(p, \ldots, p) = \sum_{i=1}^{n} \frac{\partial \Phi}{\partial p_i} = \sum_{i=1}^{n} \mathrm{Inf}_{i,p}(f) = I_p(f).$$

При доказательстве теоремы KKL мы доказали оценку $\max \mathrm{Inf}_i f \geqslant \exp(-cI(f)/\operatorname{Var} f)$. Её можно перенести на p-распределение (см. [FK]): $\max \mathrm{Inf}_{i,p} f \geqslant \exp(-cI_p(f)/\operatorname{Var}_p f)$. Прологарифмируем:

$$I_p(f) \geqslant c \operatorname{Var}_p f \cdot \log(1/\max \mathrm{Inf}_{i,p} f). \tag{5}$$

Если f симметрична, то $\mathrm{Inf}_{i,p} f = I_p(f)/n$. Тогда из (5) нетрудно вывести неравенство

$$I_p(f) \geqslant c \operatorname{Var}_p f \cdot \log n.$$

Вспоминая, что $\phi'(p) = I_p$ и $\operatorname{Var}_p f = \phi(p)(1 - \phi(p))$, получаем

$$\frac{\phi'(p)}{\phi(p)(1 - \phi(p))} \geqslant c \log n.$$

Интегрируя это неравенство по отрезку $[p', p'']$, где $\phi(p') = \varepsilon$, $\phi(p'') = 1\varepsilon$, получаем, что $p' - p'' \leqslant c(\varepsilon)/\log n$. Это и есть sharp threshold.

Список литературы

[FK] E. Friedgut, G. Kalai, “Every monotone graph property has a sharp threshold”, *Proc.AMS*, **124**:10 (1996).