Конспект по курсу

Дискретный анализ

Contributors:
Андрей Степанов
Анастасия Торунова

Лектор:
Райгородский А.М.

МФТИ

Последнее обновление: 21 апреля 2015 г.

Содержание

1 Числа Рамсея

Определение 1.1. Число Рамсея $R(s,t)$ для натуральных s и t – это минимальное натуральное число n, такое, что при любой реберной раскарске полного графа на n вершинах в два цвета, либо найдется полный подграф на s вершинах первого цвета, либо полный подграф на t вершинах второго цвета.

Определение 1.2 (Числа Рамсея, альтернативное определение)**.** $R(s,t)$ – минимальное такое n, что для любого графа на n вершинах в нем есть либо K_s клика, либо $\overline{K}_t$ антиклика

Пример.

1. $R(3,3) = 6$
2. $R(1,t) = 1$
3. $R(2,t) = t$

Утверждение 1.1.

$$(\frac{1}{4} + o(1))\frac{t^2}{\ln t} \le R(3,t) \le (1 + o(1))\frac{t^2}{\ln t}$$

Замечание. 1/4 была получена в 2013 году, а 1/162 – Кимом. Числа Рамсея были придуманы Рамсеем в 1930 году. В 1935 году Эрдеш и Секереш переоткрыли их в своей работе.

Утверждение 1.2.

$$R(3,t) \ge c\frac{t^2}{\ln^2 t}$$

Теорема 1.3. $R(s,t) \le R(s-1,t) + R(s,t-1)$

Доказательство. Обозначим $r_1 = R(s-1,t), r_2 = R(s,t-1), n = r_1 + r_2$. Положим также $\deg_+ v = \deg v$, $\deg_- v = n - 1 - \deg_+ v$.

Рассмотрим граф G на n вершинах и произвольную вершину v этого графа. Ясно, что либо $\deg_+ v \ge r_1$, либо $\deg_- v \ge r_2$. В первом случае вершина v смежна с подграфом на r_1 вершинах, в котором есть либо $\overline{K}_t$ (в этом случае все хорошо), либо K_{s-1}. Но тогда этот K_{s-1} вместе с вершиной v дает K_s и тоже все хорошо. Второй случай рассматривается аналогично. □

Следствие.

$$R(s,t) \le \binom{s-1}{s+t-2}$$

Доказательство. Индукция по $s+t$: применяем рекурсивную формулу из прошлой теоремы, а также рекурсивную формулу для треугольника Паскаля. □

Определение 1.3. Диагональные числа Рамсея – это числа $R(s,s)$.

Следствие (из следствия)**.**

$$R(s,s) \leq \binom{s-1}{s+s-2} \approx \frac{4^s}{\sqrt{\pi s}}$$

Замечание. Самая сильная верхняя оценка, которую людям удалось доказать – это

$$\exists \gamma > 0 : R(s,s) \leq 4^s \cdot e^{-\gamma \cdot \frac{\ln^2 s}{\ln \ln s}}$$

. Это сделал Конлон.

Теорема 1.4. *Пусть дано s – натуральное. Найдем такое n, что*

$$\binom{n}{s} \cdot 2^{1-\binom{s}{2}} < 1$$

Тогда $R(s,s) > n$

Доказательство. Докажем эту задачу вероятностным методом. То, что нужно доказать, равносильно тому, что $\exists$ раскраска ребер полного графа на n вершинах при которой нет одноцветной клики на s вершинах.

Рассмотрим вероятностное пространство $G(n, \frac{1}{2})$. Введем случайную величину ξ – количество одноцветных s-клик. Пусть ξ_S – другая индикаторная случайная величина, которая равна 1, если подграф S одноцветен. Тогда

$$P(\xi_S = 1) = 2 \cdot \left(\frac{1}{2}\right)^{\binom{s}{2}} = 2^{1-\binom{s}{2}}$$

Но вследствие линейности математического ожидания:

$$\xi = \sum_{S, |S|=s} \xi_S$$

А значит,

$$E\xi = \sum_{S, |S|=s} E\xi_S = \binom{n}{s} \cdot 2^{1-\binom{s}{2}} < 1$$

А значит существует такая раскраска, что $\xi = 0$ □

Следствие.

$$R(s,s) \geq (1+o(1)) \frac{s}{e\sqrt{2}} 2^{s/2}$$

Доказательство. Положим $n = (1+f(s))\frac{s}{e\sqrt{2}}2^{s/2}$, где $f(s) = o(1)$

$$C_n^s \cdot 2^{1-C_s^2} \leq \frac{n^s}{s!} \cdot 2^{1-\frac{s(s-1)}{2}} = (1+f(s))^s \frac{s^s}{e^s 2^{s/2} s!} \cdot 2^{s^2/2} \cdot 2^{1-s^2/2+s/2}$$
$$= \frac{(1+o(1))^s \cdot 2}{(1+o(1))\sqrt{2\pi s}} < 1$$

при правильном выборе $f(s)$

□

Теорема 1.5 (Эрдеша)**.**

$$R(s,s) \geq (1+o(1))\frac{s}{e} \cdot 2^{s/2}$$

Теорема 1.6 (Спенсера)**.**

$$R(s,s) \geq (1+o(1)) \cdot \frac{s\sqrt{2}}{e} 2^{s/2}$$

Определение 1.4. Событие B не зависит от совокупности событий $A_1, \dots, A_n$, если

$$\forall J \subset \{1, \dots, n\} : P(A| \cap_{j \in J} A_j) = P(B)$$

Лемма 1.7 (симметричная локальная лемма Ловаса)**.** *Пусть $A_1, \dots, A_n$ – события. Путь дополнительно $\exists p : \forall i : P(A_i) \leq p$. Дополнительно предположим, что $\forall i : A_i$ не зависит от совокупности всех остальных событий, кроме не более чем d штук. Пусть также $ep(d+1) \leq 1$. Тогда $P(\cap_{i=1}^n A_i^c) > 0$*

Доказательство теоремы Спенспера. Нам нужно доказать, что

$$P(\cap_{S, |S|=s} A_S) > 0$$

. $P(A_s) = 2^{1-C_s^2} = p$. Чему же равно d? A_S зависит от тех A_T, у которых $|S \cap T| \geq 2$. Тогда $d \leq C_s^2 \cdot C_{n-2}^{s-2}$. Осталось доказать, что $ep(d+1) < 1$.

$$ep(d+1) = e \cdot 2^{1-C_s^2} \cdot (C_s^2 C_{n-2}^{s-2} - 1) = e(1+o(1))2^{1-C_s^2} C_s^2 C_{n-2}^{s-2}$$
$$\leq e(1+o(1))2^{1-s^2/2+s/2} \cdot \frac{s^2}{2} \cdot \frac{n^{s-2}}{(s-2)!} = e(1+o(1))2^{1-s^2/2+s/2} \cdot \frac{s^4}{2} \cdot \frac{(1+o(1))^{s-2} 2^{s/2-1}}{e^{s-2}(s)!} s^{s-2} \cdot 2^{s^2/2-s}$$
$$\frac{(1+o(1))es^{s+2}(1+o(1))^{s-2}}{e^{s-2} 2(1+o(1))\sqrt{2\pi s}\left(\frac{s}{e}\right)^s}$$

Если взять $o(1) = -\frac{1}{\sqrt{s}}$, то все получится. □

2 Ещё более жаркая

Определение 2.1. Пусть $A_1, \dots, A_n$ – события на некотором вероятностном пространстве. Ориентированный граф $G = (\{A_1, \dots, A_n\}, E)$ является орграфом этих зависимостей, если $\forall i \in \{1, \dots, n\} : A_i$ не зависит от совокупности тех событий A_j для которых $(A_i, A_j) \notin E$

Пример. Полный граф является орграфом зависимостей. Но это не интересный пример. Возьмем события A_1, A_2, A_3 которые независимы попарно, но зависимы в совокупности. Тогда из любой вершины этого графа должно выходить хотя бы одно ребро, в противном случае этот граф не будет являтся орграфом зависимостей. Тогда понятно, что в минимальном орграфе зависимостей для этого набора должно быть хотя бы 3 ребра.

Замечание. Если A_i, A_j зависимы, то в графе зависимостей обязаны быть ребра $(A_i, A_j), (A_j, A_i)$

Теорема 2.1 (Локальная лемма Ловаса)**.** *Пусть $A_1, \dots, A_n$ – события на каком-то вероятностном пространстве. $G = (V, E)$ – такой орграф зависимостей, что*

$$\exists x_1, \dots, x_n \in [0, 1) : \forall i : P(A_i) \le x_i \prod_{j:(A_i, A_j) \in E} (1 - x_j)$$

Тогда

$$P(\cap_{i=1}^n \overline{A}_i) \ge \prod_{j=1}^n (1 - x_j) > 0$$

Следствие (симметричная локальная лемма Ловаса)**.** *Пусть $A_1, \dots, A_n$ – события. Путь дополнительно $\exists p : \forall i : P(A_i) \le p$. Дополнительно предположим, что $\forall i : A_i$ не зависит от совокупности всех остальных событий, кроме не более чем d штук. Пусть также $ep(d+1) \le 1$. Тогда $P(\cap_{i=1}^n \overline{A}_i^c) > 0$*

доказательство следствия. Рассмотрим $G = (V, E)$, у которого A_i соедина ровно с теми “паразитами” которые мешают независимости.

Рассмотрим сначала дурацкий случай: $d = 0$. В этом случае они независимы в совокупности. Тогда $P(\cap_{i=1}^n \overline{A}_i) = \prod_{i=1}^n (1 - P(A_i)) \ge (1-p)^n \ge (1 - \frac{1}{e})^n > 0$.

Пусть теперь d > 0. Рассмотрим $x_1 = \dots = x_n = \frac{1}{d+1}$. Мы знаем, что $P(A_i) \le p \le \frac{1}{(d+1)e}$. Хочется доказать, что $P(A_i) \le \frac{1}{d+1} \prod_{k:(A_i, A_k) \in E} (1 - \frac{1}{d+1})$. Понятно, что $(1 - \frac{1}{d+1})^d \ge \frac{1}{e}$. Но тогда симметричный случай доказан. □

доказательство леммы Ловаса.

$$\begin{aligned} P(\cap_{i=1}^n \overline{A}_i) &= P(\overline{A}_1) \cdot P(\overline{A}_2 | \overline{A}_1) \cdot \ldots \cdot P(\overline{A}_n | \overline{A}_1 \cap \dots \cap \overline{A}_n) \\ &= (1 - P(A_1))(1 - P(A_1 | \overline{A}_2)) \cdots (1 - P(A_n | \overline{A}_1 \cap \dots \cap \overline{A}_n)) \end{aligned}$$

Лемма: $\forall i : \forall J \subset \{1, 2, \cdots, n\} \setminus \{i\} : P(A_i | \cap_{j \in J} \overline{A}_j) \le x_i$

База: $|J| = 0, P(A_i | \cap_{j \in \emptyset} \overline{A}_j) = P(A_i) \le x_i$

Переход: рассмотрим произвольное множество $J : |J| = k + 1$. Представим $J = J_1 \cup J_2$. Положим $J_1 = \{j \in J : (A_i, A_j) \in E\}$, $J_2 = J \setminus J_1$.

Рассмотрим случай, когда $J_1 = \emptyset$. Тогда $P(A_i | \cap_{j \in J} \overline{A}_j) = P(A_i | \cap_{j \in J_2} \overline{A}_j) = P(A_i) \le x_i$

Рассмотрим второй случай: $J_1 = \{j_1, \dots, j_r\}, r \ge 1$. Тогда

$$P(A_i | \cap_{j \in J} \overline{A}_j) = P(A_i | \cap_{j \in J_1} \overline{A}_j \cap \cap_{j \in J_2} \overline{A}_j) = \frac{P(A_i \cap \cap_{j \in J_1} \overline{A}_j | \cap_{j \in J_2} \overline{A}_j)}{P(\cap_{j \in J_1} \overline{A}_j | \cap_{j \in J_2} \overline{A}_j}$$

$$\le \frac{P(A_i | \cap_{j \in J_2} \overline{A}_j)}{P(\cap_{j \in J_1} \overline{A}_j | \cap_{j \in J_2} \overline{A}_j} = \frac{P(A_i)}{P(\cap_{j \in J_1} \overline{A}_j | \cap_{j \in J_2} \overline{A}_j} = \frac{P(A_i)}{P(A_{j_1} | \cap \cdots) \cdot P(A_{j_2} | \cap \cdots) \cdots} =$$

$$asasd$$

$\square$

Полезность несимметричного случая: $R(3,t) > n$. Берем случайную раскраску.

3 Конструктивные нижние оценки чисел Рамсея

Что значит снизу оценить число Рамсея?

Оценка $R(s,s) > n \Leftrightarrow$ существует граф $G = (V,E), |V| = n$, в котором нет K_s и $\overline{K}_s$, то есть $\omega(G) < s, \alpha(G) < s$.

Теорема 3.1 (Франкл, Уилсон, 1981). $\exists \varphi : \varphi(s) \to 0$, *при* $s \to \infty$ *причем,* $\forall s : \exists G = (V,E) : \omega(G) < s, \alpha(G) < s, |V| \geq (e^{1/4} + \varphi(s))^{\frac{\ln^2 s}{\ln \ln s}}$

Доказательство. Пусть p – простое. Положим $m = p^3, k = p^2$. Пусть множество вершин $V = \{(x_1, \cdots, x_m) : x_i \in \{0,1\}, x_1 + \cdots + x_m = k\}$. А множество ребер $E = \{\{x,y\} : (x,y) \equiv 0(\mod p)\}$. Отметим, что $n = |V| = \binom{p^3}{p^2}$

Лемма 3.2.

$$\alpha(G) \leq \sum_{i=0}^{p-1} \binom{m}{i}$$

Доказательство. Рассмотрим произвольное независимое множество $W = \{x_1, \cdots, x_t\}$ вершин нашего графа G, $\forall i,j, i \neq j : (x_i, x_j) \not\equiv 0(\mod p)$. Сопоставим каждому x_i многочлен $F_{x_i} \in \mathbb{Z}_p[y_1, \cdots, y_m]$.

Положим $F_{x_i}(y) := \prod_{j=1}^{p-1}(j - (x_i, y))$. $F'_{x_i}(y)$ – многочлен F_{x_i} со срезанными коэфициентами в каждом одночлене. Докажем, что эти многочлены линейно независимы в $\mathbb{Z}_p$.

$c_1 F'_{x_1} + \cdots + c_t F'_{x_t} = 0$. То есть $\forall y \in W : c_1 F'_{x_1} + \cdots + c_t F'_{x_t} = 0 \mod p$. Возьмем, например $y = x_1$. Тогда $c_1 F_{x_1}(x_1)' + \cdots + c_t F'_{x_t}(x_1) = F_{x_1}(x_1) + \cdots + F_{x_t}(x_1)$. Причем $F_{x_1}(x_1) \equiv 0 \mod p$, а для $k \neq 1$ $F_{x_k}(x_1) \not\equiv 0 \mod p$. Тогда $c_1 \equiv 0 \mod p$

Значит, все многочлены независимы. Но их не может быть больше $\sum_{i=0}^{p-1} \binom{m}{i}$ $\square$

Лемма 3.3.

$$\omega(G) \leq \sum_{i=0}^{p} \binom{m}{i}$$

Доказательство. Рассмотрим произвольное множество вершин $W = \{x_1, \cdots, x_t\}$ в графе G, которое образует клику. То есть $\forall i,j, i \neq j : (x_i, x_j) \equiv 0 \mod p \Leftrightarrow (x_i, x_j) \in \{0, p, 2p, \cdots, p^2 - p\}$

Сопоставим каждому x_i многочлен $F_{x_i} \in \mathbb{Q}[y_1, \cdots, y_m]$ по следующему правилу $F_{x_i} = (x_i, x_j)((x_i, x_j) - p)((x_i, x_j) - 2p) \cdots ((x_i, x_j) - (p^2 - p))$. Опять же срежем степени всех одночленов, получим F'_{x_i}. Докажем их линейную независимость

$\forall y \in W : c_1 F_{x1}(y) + \cdots c_t F_{x_t} = 0$

$F_{x_1}(x_1) \neq 0$, для $i > 1 : F_{x_i}(x_1) = 0$ Значит, $c_1 = 0$. Аналогично для остальных c_i. Получили, что многочлены независимы. Значит, их не больше, чем $\sum_{i=0}^{p} \binom{m}{i}$

□

Обозначим $s = \sum_{i=0}^{p} \binom{m}{i} + 1$. Тогда из лемм следует, что $\alpha(G) < s, \omega(G) < s$. Докажем что n как функция от s имеет вид $(e^{1/4} + o(1))^{\frac{\ln^2 s}{\ln \ln s}}$.

$n = \binom{p^3}{p^2} = \frac{p^3(p^3-1)\cdots(p^3-p^2+1)}{(p^2)!}$ Понятно, что $p^3 - i = p^{3(1+o(1))}$ Тогда $n = \frac{p^{3p^2(1+o(1))}}{(p^2)!}$, $(p^2)! = p\sqrt{2\pi}\left(\frac{p^2}{e}\right)^{p^2} = p^{2p^2(1+o(1))}$, $n = p^{p^2(1+o(1))}$, $\binom{m}{p} = \frac{p^{3p(1+o(1))}}{p^{p(1+o(1)}}$, $s \leq (p+1)p^{2p(1+o(1))}+1$, $s \geq p^{2p(1+o(1))}$, короче говря $s = p^{2p(1+o(1))}$.

$\ln s = 2p(1 + o(1)) \ln p$, $\ln^2 s = 4p^2(1 + o(1)) \ln^2 p$, $\ln \ln s = \ln 2p + \ln(1 + o(1)) + \ln \ln p = (1 + o(1)) \ln p$

$\frac{\ln^2 s}{\ln \ln s} = 4p^2 \ln p(1+o(1))$, $(e^{1/4}+o(1))^{4p^2 \ln p(1+o(1))} = e^{1/4 \cdot 4p^2 \ln p(1+o(1))(1+o(1))}$, $n = e^{p^2 \ln p(1+o(1))}$ Подбираем правильно $o(1)$ которое в нашей власти и все получилось. Что делать для произвольного s: находим максимальное простое $p : s > s_0 := \sum_{i=0}^{p} \binom{m}{i} + 1$. Ясно, что $R(s, s) \geq R(s_0, s_0) \geq (e^{1/4} + o(1))^{\frac{\ln^2 s_0}{\ln \ln s_0}} \sim (e^{1/4} + o(1))^{\frac{\ln^2 s}{\ln \ln s}}$. □

4 Двудольные числа Рамсея

Определение 4.1. $b(k, k)$ – это минимальное такое l, что при любой раскраске ребёр $K_{l,l}$ в красный и синий цвета, найдется одноцветный $K_{k,k}$

Теорема 4.1.

$$b(k, k) \geq (1 + o(1))\frac{2}{e}k2^{k/2}$$

Замечание. Берём случайную раскраску ребёр полного двудольного графа $K_{l,l}$. Рассматриваем случайную величину ξ = число одноцветных $K_{k,k}$. $E\xi = \binom{l}{k}^2 \cdot 2^{1-k^2}$. Это матожидание отличается от аналогичного для чисел Рамсея совсем чуть-чуть.

Теорема 4.2 (Конлон)**.**

$$b(k, k) \leq (1 + o(1)) \log_2 k \cdot 2^{k+1}$$

Лемма 4.3. *Пусть числа $m, n, r, s \in \mathbb{N}$ и $p \in [0, 1]$ таковы, что $(s-1)\binom{m}{r} < n\binom{mp}{r}$. Пусть $G_{m,n}$ – любой подграф $K_{m,n} : \frac{|E(G_{m,n})|}{mn} \geq p$. Тогда в $G_{m,n}$ есть $K_{r,s}$*

Доказательство. Предположим противное. Пусть в $G_{m,n}$ нет $K_{r,s}$.

Подсчитаем двумя разными способами число подграфов $K_{r,1}$ в графе $G_{m,n}$.

Первый способ соответствует предположению противного. $\binom{m}{r} \cdot (s-1)$ – максимальное количество $K_{r,1}$ в $G_{m,n}$ в виду сделанного нами предположения противного.

С другой стороны, обозначим $d_1, \cdots, d_n$ – степени вершин графа $G_{m,n}$ в правой доле. Тогда количество таких $K_{r,1}$ – это $\binom{d_1}{r} + \cdots + \binom{d_n}{r} \geq \binom{\frac{d_1+\cdots+d_n}{n}}{r}$ По условию это больше либо равно $n\binom{pm}{r}$. Пришли к противоречию. □

Мы знаем, что если $r^2 = o(m)$, то $\binom{m}{r} \sim \frac{m^r}{r!}$, $\binom{mp}{r} \sim \frac{(mp)^r}{r!}$

Лемма 4.4 (Та же самая, только в асимптотическом виде)**.** *Пусть* $m = m(k), n = n(k), r = r(k), s = s(k)$. $p \in [0,1]$. *Предположим, что* $r^2 = o(m)$, $n > (s-1) \cdot p^{-r}(1+o(1))$. *Пусть для каждого* k $G_{m,n}$ *– любой произвольный подграф* $K_{m,n}$ *такой, что* $|E(G_{m,n})| \geq pmn$. *Тогда в* $G_{m,n}$ *есть* $K_{r,s}$

Докажем теорему Колона. Мы хотим доказать, что для $\varepsilon > 0$ $b(k,k) \leq (1+\varepsilon)(\log_2 k) \cdot 2^{k+1}, k \geq k_0$ Обозначим $l = (1+\varepsilon) \cdot (\log_2 k) \cdot 2^{k+1}$. Это равносильно тому, что при любой раскраске рёбер графа $K_{l,l}$ в красный и синий цвета найдется одноцветный $K_{k,k}$. Зафиксируем произвольную раскраску. Назовём вершину красной, если её красная степень не меньше, чем синяя. В противном случая назовём её синей. Б.о.о считаем, что в правой доле красных хотя бы $\frac{l}{2}$. Возьмём из них первые $\frac{l}{2}$. Пусть $m(k) = l(k), n(k) = \frac{l(k)}{2}$. $G(m,n)$ – это граф из красных рёбер. $p = \frac{1}{2}$. Положим $s(k) = k^2 \log_2 k$, $r(k) = k - 2\log_2 k$

$\frac{l}{2} = (1+\varepsilon)(\log_2 k) \cdot 2^k > (k^2 \log_2 k - 1)2^{k-2\log_2 k}$

Тогда из леммы следует, что при каждом $k \geq k_0$ в $G_{m,n}$ есть $K_{r,s}$ Рассмотрим $m = k^2 \log_2 k$, $n = l - (k - 2\log_2 k)$ Возьмём $G_{m,n}$ – из красных рёбер. $r = k, s = 2\log_2 k$, $p = (\frac{l}{2} - k)/l = \frac{1}{2} - \frac{k}{l}$ После второго применения леммы победа □

5 Гиперграфовые числа Рамсея

Определение 5.1. $R_k(l_1, \cdots, l_r)$ – минимальное такое n, что при любой раскраске рёбер полного k-однородного гиперграфа на n вершинах в r цветов найдется такое i и найдется такое l_i-элементное подмножество множества вершин, такое что все рёбра которые целиком содержатся в этом подмножетсве покрешены в i-цвет.

Несложно доказать, что $R_k(l_1, \cdots, l_r) \leq R_{k-1}(R_k(l_1-1, \cdots, l_r), \cdots, R_k(l_1, \cdots, l_r-1)$.

Теорема 5.1. $R_3(s,t) \leq 4^{4^{4^{4}}}$ $s+t$ *раз*

Доказательство. Доказываем по индукции: $R_3(s,t) \leq R_2(R_3(s-1,t), R_3(s,t-1)) \leq R_2(4^{4^{4^{4}}}, 4^{4^{4^{4}}}) \leq (1+o(1))4^{4^{4^{4}}}$ □

Если использовать вероятностный метод, то: $\binom{n}{3}\cdot(\frac{1}{2})^{\binom{s}{3}}\cdot 2 \geq 2^{s^2/6}(1+o(1))$ Короче, все плохо.

6 Система представителей

Определение 6.1. Рассмотрим k-однородный гиперграф $H = (\mathcal{R}_n, \mathcal{M})$, где $\mathcal{R}_n = \{1, \cdots, n\}$. Обозначим $|\mathcal{M}| = s$ Назовём системой общих представителей (СОП) для $\mathcal{M}$ произвольное подмножество $S \subset \mathcal{R}_n$: $\forall M \in \mathcal{M} : M \cap S \neq \emptyset$

$\tau(\mathcal{M}) = \min\{|S| : S$ – СОП для $\mathcal{M}\}$

Утверждение 6.1. $\forall n, k, s : \forall \mathcal{M} : \tau(\mathcal{M}) \leq \min\{s, n-k+1\}$

Утверждение 6.2. $\forall n, k, s : \exists \mathcal{M} : \tau(\mathcal{M}) \geq \min\{s, \lceil \frac{n}{k} \rceil\}$

Теорема 6.3 (Эрдеш)**.** $\forall n, k, s : \forall \mathcal{M} : \tau(\mathcal{M}) \leq \max\{\frac{n}{k}, \frac{n}{n} \ln \frac{sk}{n}\} + \frac{n}{k} + 1$

Доказательство. Пусть $s >> \frac{n}{k}$. В противном случае если, скажем $sleq\frac{n}{k}$, то $\tau(\mathcal{M}) = s \leq \frac{n}{k}$. Другой плохой случай – это когда $\frac{n}{k} \ln \frac{sk}{n} \leq n$. Тогда $\tau(n) \leq n \leq \frac{n}{k} \ln \frac{sk}{n}$

Теперь у нас $s > \frac{n}{k}$ и кроме того $\frac{n}{k} \ln \frac{sk}{n} < n$. Теперь зафиксируем $\mathcal{M} = \{M_1, \cdots, M_s\}$. Пусть ν_1 – вершина, содержащаяся в самом большом количестве M_i. Пусть $\rho_i = \{j : \nu_i \in M_j\}$. Тогда $|\rho_1| \geq \frac{sk}{n}$. Удалим элемент ν_1 из рассмотрения (выкенем его из $\mathcal{R}_n$, также выкинем из $\mathcal{M}$ все ρ_1). Пусть $s_1 = |\mathcal{M} \setminus \rho_1|$, тогда $|\rho_2| \geq \frac{s_1 k}{n-1} \leq \frac{s_1 k}{n}$. Сделаем N шагов, чтобы $N = \rceil \frac{n}{k} \ln \frac{sk}{n} \rceil + 1$. Осталось s_N ребер, причем $s_N = s_{N-1} - \rho_N \leq S_{N-1} - \frac{S_{N-1}k}{n} = S_{N-1}(1 - \frac{k}{n}) \leq \cdots \leq s(1-\frac{k}{n})^N \leq s(1-\frac{k}{n})^{\frac{n}{k}\ln\frac{sk}{n}} \leq se^{-\frac{k}{n}\cdot\frac{n}{k}\ln\frac{sk}{n}} = s \cdot \frac{n}{sk} = \frac{n}{k}$

Получили, что $\tau(\mathcal{M}) \leq N + \frac{n}{k} \leq \frac{n}{k} \ln \frac{sk}{n} + 1 + \frac{n}{k}$ □

7 Нижняя оценка для СОП

Теорема 7.1. *Пусть* $n \geq 16$. *Пусть* $k \leq \frac{n}{16}$. *Пусть* $s : 4 \leq \ln \frac{sk}{n} \leq k$. *Тогда* $\exists M : \tau(M) \geq \frac{n}{32k} \ln \frac{sk}{n}$

Доказательство. Обозначим $m := \lceil \frac{1}{2} \ln \frac{sk}{n} \rceil$. Ясно, что $m \geq 2$. Пусть $N_1^1, \cdots N_{\binom{2m}{m}}^1$ – все m-элементные подможества $\{1, \cdots, 2m\}$. $\tau(\{N_1, \cdots\}) = m+1$. Пусть $q := \lceil \frac{2k}{m} \rceil$. Пусть теперь $N_1^2, \cdots, N_{\binom{2m}{m}}^2$ – все m-элементные подмножества $\{2m+1, \cdots, 4m\}$. Аналогично определим $N^3, \cdots, N^q$. Пусть $M_1 = N_1^1 \cup N_1^2 \cup \cdots \cup N_1^q$. Аналогично определим $M_2, \cdots, M_{\binom{2m}{m}}$. Пусть $\mathcal{M}_1 = \{M_1, \cdots, M_{\binom{2m}{m}}\}$. Тогда $|M_i| = qm$. Заметим, что $\frac{2k}{m} \geq \frac{4k}{\frac{1}{2}\ln\frac{sk}{n}} \geq 4$. Кроме того, $q \geq \frac{k}{m}$. Тогда $qm \geq k$. $\tau(\mathcal{M}_1) = m+1 > m \geq \frac{1}{4} \ln \frac{sk}{n}$. Пусть $t := \lceil \frac{n}{2qm} \rceil$. Прододжжим эту конструкцию t раз. Получим множества $\mathcal{M}_1, \cdots, \mathcal{M}_t$. Теперь рассмотрим $\overline{\mathcal{M}} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_t$. Тогда $\tau(\overline{\mathcal{M}}) = t\tau(\mathcal{M}_1) \geq \frac{1}{4} t \ln \frac{sk}{n} \geq \frac{1}{4} \frac{n}{4qm} \ln \frac{sk}{n} \geq$

$\frac{1}{16}\frac{n}{2k}\ln\frac{sk}{n} = \frac{1}{32}\frac{n}{k}\ln\frac{sk}{n}$. Посчитаем $|\overline{\mathcal{M}}| = t\binom{2m}{m} \leq t2^{2m} \leq t\cdot 2^{2\frac{1}{2}\ln\frac{sk}{n}} \leq t\frac{sk}{n} \leq \frac{n}{2qm}\frac{sk}{n} = \frac{sk}{2qm} \leq \frac{sk}{2k} = \frac{s}{2}$. Каждое множество $M \in \overline{\mathcal{M}}$ при необходимости обрежем. К полученной совокупности добавим любые k-элементные множества так, чтобы итоговая совокупность $\mathcal{M}$, состояла ровно из s множеств. Понятно, что $\tau(\mathcal{M}) \geq \tau(\overline{\mathcal{M}})$. Конец. □

Теорема 7.2. *Пусть n, k, s, l таковы, что $\binom{n}{l}\cdot\binom{\binom{n}{k}-\binom{n-l}{k}}{s}\frac{1}{\binom{\binom{n}{k}}{s}} < 1$. Тогда $\exists M : \tau(M) > l$*

Доказательство. Возьмём случайную M совокупность мощности s состоящую из k-элементных подмножеств $\{1,\cdots,n\}$. Всего таких совокупностей $\binom{\binom{n}{k}}{s}$. Рассмотрим $L_1,\cdots,L_{\binom{n}{l}} \subset \{1,\cdots,n\}$, причем $|L_i| = l$. Для каждого $i = \{1,\cdots,\binom{n}{l}\}$ определим события $A_1,\cdots,A_i$, заключающиеся в том, что L_i является СОП для M. Тогда $P(A_i) = \frac{\binom{\binom{n}{k}-\binom{n-l}{k}}{s}}{\binom{\binom{n}{k}}{s}}$. Тогда $P(\cup A_i) \leq \binom{n}{l}P(A_1) \leq 1$. Тогда существует такая совокупномть M, у которой можность СОП $> l$. □

Следствие. *Пусть при $n\to\infty$, $k = k(n)\to\infty, s = s(n)\to\infty, \frac{sk}{n}\to\infty$. Пусть $k^2 = o(n), \ln\ln k = o(\ln\frac{sk}{n}), \ln^2\frac{sk}{n} = o(k)$. Тогда $\exists n_0 : \forall n \geq n_0 : \exists M : \tau(M) \geq \frac{n}{k}\ln\frac{sk}{n} - \frac{n}{k}\ln\ln\frac{sk}{n} - \frac{n}{k}\ln\ln k - \frac{n}{k} = (1+o(1))\frac{n}{k}\ln\frac{sk}{n}$*

Доказательство. Проведем неформальное доказательство. $\binom{\binom{n}{k}-\binom{n-l}{k}}{s}$ □

<++>

$R_n = \{1,2,\cdots,n\}$. Нужно построить совокупность M, которая будет состоять из s k-элементных подмножеств, так, что $\tau(M) > l$. Допустим мы построили такую совокупность M: для любого множества L_j в R_n имеющего можность $n-l$. Найдется $M_i \in M$: $M_i \subset L_j$. Это очень похоже на СОП, только для множеств. Тогда конечно же, $\tau(M) > l$. Формализуем: пусть $L_1,\cdots,L_{\binom{n}{l}}$ – все $(n-l)$-элементные подмножества R_n. Нам нужна такая M, что $\forall j : \exists i : M_i \subset L_j$. Пусть $K_1,\cdots,K_{\binom{n}{k}}$ – все k-элементные подмножества R_n. Рассмотрим $R_{\binom{n}{k}} = \{1,\cdots,\binom{n}{l}\}$. Сопоставим $L_j \mapsto \Lambda_j = \{\nu : K_\nu \subset L_j\}$. $LL = \{\Lambda_1,\cdots,\Lambda_n\}$. $\tau := \tau(LL)$. Рассмотрим любую минимальную СОП $\sigma_1,\cdots,\sigma_\tau$ для LL. Рассмотрим $\overline{M} := \{K_{\sigma_1},\cdots,K_{\sigma_\tau}\}$. Утверждение $\forall j : \exists i : K_{\sigma_i} \subset L_j$. Если $\tau(LL) \leq s$, тогда $\exists M : \tau(M) > l$

Теорема 7.3. *Пусть $\max\{\frac{\binom{n}{k}}{\binom{n-l}{k}}, \frac{\binom{n}{k}}{\binom{n-l}{k}}\ln(\cdots)\} - \cdots \leq s$. Тогда $\exists M : \tau(M) > l$.*

8 Размерность Вапника-Червоненкиса

Пример. Задача. Пусть $S \subset \mathbb{R}^2$ – конечное множество, $|S| = n$. Будем пересекать множество со всевозможными треугольниками. $\mathcal{M}_S := \{M \subset$

$S : \exists \Delta \subset \mathbb{R}^2 : \Delta \cap S = M\}$. Возьмём $\varepsilon \in (0,1)$. Определим $\mathcal{M}_{S,\varepsilon} := \{M \subset S : \varepsilon\Delta \subset \mathbb{R}^2 : \Delta \cap S = M, |M| \geq \varepsilon n\}$ Имеет место следующая теорема: $\forall n : \forall S : \forall \varepsilon :\subset \mathbb{R}^2, |S| = n : \tau(\mathcal{M}_S) \leq \frac{500}{\varepsilon} \log_2 \frac{500}{\varepsilon}$.

Рассмотрим обобщение. Рассмотрим пару (X, R), где X – какое-то множество, а R – совокупность каких-то подмножеств.

Пример. $(X, R) = (\mathbb{R}^n, H)$, где H – все открытые полупространства $\mathbb{R}^n$. В ML это часто называют ранжированным пространством.

Пусть $A \subset X$. Введем обозначение $Pr_A R := \{r \cap A : r \in R\}$ – проекция R на A. $(A, Pr_A R)$ – ранжированное подпространство. Скажем, что A дробится областями из R, если $Pr_A R = 2^A$. $VC(X, R) := \max\{m : \exists A \subset X : |A| = m, A$ дробится областями из $R\}$ – размерность Вапника-Червоненкиса. $VC(\mathbb{R}^n, H) = n+1$. Для начала $n = 1$. Понятно, что любые 3 точки не дробятся. А 2 различные дробятся. Рассмотрим $n = 2$. Любой невырожденный треугольник дробится. И треугольник с точкой внутри тоже дробится. В более общём случае множество не будет дробится, если существует два его подмножества, у которых линейные оболочки пересекаются.

Теорема 8.1 (Радона). *Пусть $S \subset \mathbb{R}^n$: $|S| \geq n+2$. Тогда $\exists S_1 \cap S_2 = \emptyset : S = S_1 \cup S_2 : conv(S_1) \cap conv(S_2) \neq \emptyset$*

Лемма 8.2. *Пусть $S = (X, R)$ – ранжированное пространство, причем $|X| = n \in \mathbb{N}$. $VC(X, R) = d$. Тогда $R \leq g(n, d) := \sum_{i=0}^{d} \binom{n}{i}$*

Доказательство. Докажем по индукции по (n, d). База: $n = 0$. Тогда $d = 0$. $|R| \leq 1 = g(0, 0)$. Пусть $d = 0$. Тогда n – любое, а $|R| \leq 1 = g(n, 0)$. Шаг индукции. $S = (X, R), VC(X, R) = d$. Рассмотрим в S два подпространства. Возьмем $x \in X$. $S_1 := (X \setminus \{x\}, R_1), S_2 := (X \setminus \{x\}, R_2)$, где $R_1 := \{r \setminus \{x\}, r \in R\}$, $R_2 := \{r \in R : x \notin r, r \cup \{x\} \in R\}$. Тогда $|R| = |R_1| + |R_2|$. Ясно, что $|R_1| \leq g(n-1, d)$. Докажем, что $|R_2| \leq g(n-1, d-1)$. Предположим, что $\exists A \subset X \setminus \{x\}, |A| = d, A$ дробится R_2. Если мы возьмём $A \cup \{x\}$, то его мощность – это $d+1$, причем A дробится R. Завершаем доказательство применением формулы господина Паскаля. □

Следствие. *Скажем, что $S = (X, R)$. $VC(S) = d$. $A \subset X : |A| = n$. Тогда $|Pr_A R| \leq g(n, d)$.*

Доказательство. $VC(A, Pr_A R) \leq VC(X, R) \leq d$. Применяем предыдущую лемму. □

Определение 8.1. Возьмём $h \geq 2$, (X, R) – ранжированное пространство. h-измельчением системы R назовём $R_h := \{r : \exists r_1, \cdots, r_h \in R : r = r_1 \cap r_2 \cap \cdots \cap r_h\}$. Например, H_3 содержит в себе все треугольники.

Лемма 8.3. *Пусть $VC(X, R) = d \geq 2$. $h \geq 2$. Тогда $VC(X, R_h) \leq 2dh \log_2(dh)$*

Доказательство. Пусть $A \subset X$, $|A| = n$, A дробится с помощью R_h. Тогда $|Pr_A R| = 2^n$. С другой стороны, $|Pr_A R| \leq g(n, d) \leq n^d$. Но $|Pr_A R_h| \leq n^{dh}$. То есть $2^n \leq n^{dh}$. То есть если $n^{dh} < 2^n$, то A не может дробится. Но есть в качестве n взять $2dh \log_2(dh)$, то это неравенство будем выполнено, а значит $VC(X, R_h) \leq 2dh \log_2(dh)$ □

Пример. $VC(\mathbb{R}^2, H) = 3$. Тогда $VC(\mathbb{R}^2, T_3) \leq VC(\mathbb{R}^2, H_3) \leq 18 \log_2 9 \leq 60$. То есть в нашем первом примере размерность Вапника-Червоненкиса ≤ 60.

Определение 8.2. Пусть (X, R) – ранжированное пространство. $S \subset X$, $\varepsilon \in (0, 1)$. Положим $M_{S,\varepsilon} := \{M \subset S : \exists r \in R : r \cap S = M, |M| \geq \varepsilon |S|\}$.

Теорема 8.4. *Пусть* $VC(X, R) = d$. *Тогда* $\forall n : \forall S \subset X, |S| = n : \forall \varepsilon \in (0, 1) : \tau(M_{S,\varepsilon}) \leq \frac{8d}{\varepsilon} \log_2(\frac{8d}{\varepsilon})$

Замечание. Если $VC(X, R) = \infty$, то $\forall m : \exists S \subset X, |S| = m : S$ дробится, то $\tau(\mathcal{M}_{S,\varepsilon}) \sim m(1 - \varepsilon)$