

Спецкурс “Многомерная аппроксимация и анализ булевых функций”. Лекция 1: Функции нескольких переменных

4 октября 2026 г.

1 Анонс курса

Вначале мы кратко рассмотрим функции *нескольких переменных* $f(x_1, \dots, x_d)$, считая что d фиксировано. Классический подход состоит в определении пространств дифференцируемых в том или ином смысле функций, скажем,

$$C^r[0, 1]^d = \{f: [0, 1]^d \rightarrow \mathbb{R} \mid \partial^\alpha f \in C, |\alpha| \leq r\}.$$

Сформулируем без доказательства классические теоремы аппроксимации в одномерном и многомерном вариантах. Посмотрим на специфику многомерного случая.

Переходя к случаю *многих переменных*, мы увидим так называемое “проклятие размерности”: сложность задач (аппроксимации, оптимизации, интегрирования) может расти экспоненциально по d . Рассмотрим сейчас простейший такой пример. Если функция $f: [0, 1] \rightarrow \mathbb{R}$ липшицева, $|f(x) - f(y)| \leq |x - y|$, то для нахождения любого значения $f(x)$ с точностью порядка ε достаточно знать значения в точках $f(x_k)$, $x_k = \lfloor k\varepsilon \rfloor$, их порядка $(1/\varepsilon)$. Рассмотрим теперь куб $[0, 1]^d$ и пусть функция $f: [0, 1]^d \rightarrow \mathbb{R}$ липшицева в следующем смысле:

$$|f(x) - f(y)| \leq \|x - y\|_\infty.$$

Тогда для того, чтобы уметь находить значения $f(x)$ с точностью ε , понадобится уже сеть размера $(1/\varepsilon)^d$. При фиксированном ε эта величина

растёт экспоненциально по d . Скажем, при $d = 1000$ она невообразимо велика. Это называется “проклятием размерности”.

Мы рассмотрим и другие, с виду более простые задачи, которые также страдают от “проклятия размерности”.

Однако, в реальности удаётся работать с функциями многих переменных. Причина в том, что классические множества функций слишком широкие, реальные функции многих переменных имеют дополнительную структуру.

Пример функции многих переменных. MatrixNet Яндекса (поиск 2010-х). Поиск в ответ на *запрос* подбирает *документы* (скачанные страницы сайтов), после чего их *ранжирует* по релевантности и отдаёт пользователю. Релевантность можно считать некоей функцией $\text{Rel}(\text{Query}, \text{Doc})$. Как с ней работают на практике? Определяются *факторы*. Каждый фактор это отображение $X_i: \{(Q, D)\} \rightarrow \mathbb{R}$. Например, количество слов запроса Q , найденных в документе D . Или “авторитетность” сайта, с которого получен D . Если факторы достаточно полны, то можно считать, что

$$\text{Rel}(Q, D) = F_{\text{rel}}(X_1(Q, D), \dots, X_n(Q, D))$$

для некоторой функции F . Факторов больше тысячи, над отдельными наиболее важными работают целые группы разработчиков-исследователей.

Ассессоры размечают данные по парам запрос-документ $(Q^{(i)}, D^{(i)})$, размеры выборок — десятки и сотни тысяч и более. Получаются значения релевантности $Y^{(i)} = \text{Rel}(Q^{(i)}, D^{(i)})$. Далее для $F_{\text{rel}}(X_1, \dots, X_n)$ строится аппроксимация на основе размеченных данных $(X^{(i)}, Y^{(i)})$:

$$F_{\text{approx}}(X^{(i)}) \approx Y^{(i)}.$$

Здесь

$$X^{(i)} = (X_1(Q^{(i)}, D^{(i)}), \dots, X_n(Q^{(i)}, D^{(i)})),$$

т.е. это факторное описание. Используются специальные деревья решений. При этом каждая переменная X_k дискретизируется (≈ 16 значений). Многие переменные (факторы) уже изначально булевы (да/нет).

В этом примере функция F_{rel} больше похожа на булеву функцию, нежели на какую-то гладкую. Основным предметом изучения в курсе будут булевы функции

$$f: \{0, 1\}^n \rightarrow \mathbb{R}.$$

Они будут для нас модельным примером функций большого числа переменных.

В дискретной математике изучаются обычно $\{0, 1\}$ -значные функции. Для вещественнозначных функций можно построить свой анализ и получать не только результаты “комбинаторного” типа. Для них есть естественный гармонический анализ, определены производные. Справедливы некоторые аналоги классических неравенств.

Для булевых функций определены различные меры сложности. Одна из наиболее интересных для нас: минимальная степень полинома (многих переменных), аппроксимирующего функцию. Почти все меры сложности оказываются эквивалентны, последний нерешенный случай был закрыт в 2019 году Нео Хуанг (sensitivity theorem). Приведём формулировку. Для функции $f: \{0, 1\}^n \rightarrow \{0, 1\}$ чувствительность в точке x равна количеству i , для которых изменение бита x_i в точке x меняет значение функции. Чувствительность f это максимальная чувствительность по всем x . Степень функции $\deg f$ это степень полинома, равного f . Хуанг доказал $s(f) \geq \sqrt{\deg f}$.

2 Функции нескольких переменных

Начнём с классики. Теорема Вейерштрасса: $f \in C[-1, 1]$, $\mathcal{P}_n = \{a_0 + a_1x + \dots + a_nx^n\}$, тогда

$$\lim_{n \rightarrow \infty} E(f, \mathcal{P}_n)_C = 0.$$

Здесь $E(f, V)_C = \inf_{g \in V} \|f - g\|_{C[-1, 1]}$ это расстояние до подпространства V , или *наилучшее приближение*.

Структурные свойства функции соответствуют аппроксимативным. Наиболее законченная и стройная теория получается для периодических функций, или, эквивалентно, функций на $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. Гармонический анализ, $f \sim \sum \hat{f}(k)e^{ikx}$, тригонометрические полиномы $\mathcal{T}_n = \{\sum_{|k| \leq n} c_k e^{ikx}\}$.

Теорема (прямая теорема приближения, Jackson). Если $f \in C^r(\mathbb{T})$, то $E(f, \mathcal{T}_n) = O(n^{-r})$.

Теорема (обратная теорема приближения, Бернштейн). Если $E(f, \mathcal{T}_n) = O(n^{-r-\varepsilon})$ для некоторого $\varepsilon > 0$, то $f \in C^r$.

На самом деле есть точное описание класса функций, для которых $E(f, \mathcal{T}_n) = O(n^{-\alpha})$, для любого $\alpha > 0$.

Функции нескольких переменных: $f(x_1, \dots, x_d)$. Считаем, что d фиксировано. В функциональном анализе основным объектом является то

или иное *функциональное пространство* — линейное нормированное пространство функций. Класс функций — обычно шар в подходящем пространстве.

1. Пространство Лебега $L_p[0, 1]^d$. Очень широкие пространства функций, из принадлежности $f \in L_p$ нельзя сделать каких-либо количественных выводов о возможности аппроксимации полиномами, например.

2. Пространство гладких функций

$$C^r[0, 1]^d := \{f: \partial^\alpha f \in C[0, 1]^d, |\alpha| \leq r\}.$$

Норму можно определять по-разному, обычно это максимум норм производных.

3. Липшицевы функции

$$\text{Lip}[0, 1]^d = \{f: |f(x) - f(y)| \leq C\|x - y\|\}.$$

При фиксированном d не важно, в какой метрике $x - y$.

4. Пространства Соболева:

$$W_p^r(\Omega) := \{f: \partial^\alpha f \in L_p(\Omega), |\alpha| \leq r\}.$$

Производные понимаются в слабом смысле (интегрирования по частям). Пространства нужны для PDE и уравнений математической физики. Требование обычной C^r -гладкости для PDE излишне ограничительное, часто в теоремах существования доказывается именно наличие решения из пространства Соболева. При $d > 1$ непрерывность функций из W_p^r можно гарантировать при $r/d > 1/p$. Пример: функция $\log \log(c/|x|)$ не ограничена, но лежит в W_2^1 (где $|x| = \sqrt{x_1^2 + x_2^2}$). Здесь проявляется многомерная специфика; в одномерном случае всё “хорошо”, класс Соболева состоит из функций, абсолютно непрерывных вплоть до $(r - 1)$ -й производной, для которых $f^{(r-1)} \in L_p$.

Что получается для аппроксимации? Для $f \in C^r$ аналогично

$$E(f, \mathcal{T}_{n,\dots,n}) = O(n^{-r}).$$

Здесь

$$\mathcal{T}_{n,\dots,n} = \sum_{\substack{|k_j| \leq n \\ j=1,\dots,d}} c_{k_1,\dots,k_d} e^{i(k_1 x_1 + \dots + k_d x_d)}.$$

Однако, чтобы получить правильное представление об этой величине, нужно учесть количество параметров. Здесь используется порядка n^d

коэффициентов. Удобно использовать определение колмогоровского перечника

$$d_n(W, X) := \inf_{V_n} \sup_{f \in W} E(f, V_n)_X,$$

где инфимум по всем n -мерным линейным подпространствам. Поскольку $\dim \mathcal{T}_{n,n,\dots,n} = (2n+1)^d \asymp n^d$, для единичного шара рассмотренного пространства получается

$$d_n(B(C^r), C(\mathbb{T}^d)) \lesssim n^{-r/d}$$

(на самом деле, эта оценка точна по порядку).

Пример другой задачи — квадратуры с n узлами (заменяем интеграл на сумму по n узлам). Тоже ошибка порядка $n^{-r/d}$.

Проблема с $n^{-r/d}$ в том, что при больших d знаменатель портит скорость аппроксимации.

Смешанная гладкость. Коробов обнаружил, что смешанная гладкость позволяет получать лучшие результаты:

$$W_p^{r,r,\dots,r} := \{f : \partial^\alpha f \in L_p, \max \alpha_j \leq r\}.$$

Тут погрешность квадратур $n^{-r} \log^c(n)$ для некоторого $c = c(d)$, а класс гораздо шире класса Соболева W_p^{rd} . Аппроксимация тоже хорошая. Единичный шар пространства смешанной гладкости приближается тригонометрическими полиномами со спектром в гиперболическом кресте:

$$\mathcal{T}(N, d) = \left\{ \sum_{k \in \Gamma(N, d)} c_k e^{i\langle k, x \rangle} \right\}, \quad \Gamma(N, d) = \left\{ k \in \mathbb{Z}^d : \prod_{j=1}^d (|k_j| + 1) \leq N \right\}.$$

Получаются примерно такие же порядки:

$$d_n(B(W_p^{r,r,\dots,r}), L_p) \lesssim E(B(W_p^{r,r,\dots,r}), \mathcal{T}(N, d))_{L_p} \lesssim n^{-r} \log^c n,$$

при подходящем N .

Поэтому в многомерной аппроксимации обычно изучают именно классы смешанной гладкости. См. книгу [1].

Дискрепанс. Приведём ещё один пример проблемы, связанной с многомерностью. Рассмотрим куб $[0, 1]^d$ и в нём множество точек $X_1, \dots, X_N$. Для параллелепипеда $\Pi_{a_1, \dots, a_N} = [0, a_1] \times \dots \times [0, a_N]$ сравним его объём и долю точек, попавших в Π . Дискрепанс определяется как модуль разности этих величин в худшем случае:

$$D(X_1, \dots, X_N) = \max_{a_1, \dots, a_N} \left| a_1 \cdots a_N - \frac{1}{N} \#\{k: a_k \in \Pi_a\} \right|.$$

Неравенство Koksma–Hlawka связывает дискрепанс с погрешностью кубатурной формулы для интеграла, а именно,

$$\left| \int_{[0,1]^d} f(x) dx - \frac{1}{N} \sum_{k=1}^N f(X_k) \right| \leq D(X_1, \dots, X_N) \cdot V(f).$$

Насколько мал может быть дискрепанс? Положим

$$D_d(N) := \min_{X_1, \dots, X_N \in [0,1]^d} D(X_1, \dots, X_N).$$

Известно, что $D_2(N) \asymp (\log N)/N$, а вот D_3 остаётся открытой проблемой — степень логарифма там неизвестна,

$$\frac{\log^{1+c} N}{N} \lesssim D_3(N) \lesssim \frac{\log^2 N}{N}.$$

Список литературы

- [1] D. Dung, V. Temlyakov, T. Ullrich, *Hyperbolic Cross Approximation*. Birkhäuser, Cham. 2018.