

# Асимптотика логарифма сложности перестройки нейронов

А. П. Соколов

## Введение.

Пороговые функции алгебры логики являются математической моделью нейронов. Они представляют интерес благодаря своим универсальным вычислительным возможностям, а также благодаря возможности их обучения. Последнее свойство с успехом применяется на практике при решении плохоформализуемых задач.

В качестве средства задания пороговых функций в работе рассматриваются линейные формы вида  $x_1w_1 + \dots + x_nw_n - \sigma$  с целочисленными коэффициентами и свободным членом.

Исследуется сложность преобразования одной пороговой функции, заданной линейной формой, к другой, путем пошагового изменения коэффициентов линейной формы. В качестве меры сложности данного процесса принимается изменение коэффициента или свободного члена линейной формы на единицу. Данный процесс может интерпретироваться как процесс обучения нейрона с пороговой функцией активации.

Ранее, в работе [3], для характеристики сложности обучения в худшем случае исследовалась шенноновская функция  $\rho(n)$ . Она говорит о том, сколько минимально достаточно выполнить единичных модификаций исходной линейной формы от  $n$  переменных для задания желаемой пороговой функции. Было показано, что при стремлении  $n$  к бесконечности величина  $\log_2 \rho(n)$  растет по порядку как  $n \log_2 n$ .

В настоящей работе найдена асимптотика логарифма  $\rho(n)$ . Показано, что при стремлении  $n$  к бесконечности имеет место

$$\log_2 \rho(n) \sim \frac{1}{2} n \log_2 n.$$

## 1. Основные понятия, постановки и результаты.

Пусть  $U = \{u_1, u_2, \dots\}$  — счетный алфавит переменных. Каждое из переменных  $u_i$  может принимать значения из множества  $E_2 = \{0, 1\}$ . В дальнейшем во избежание употребления сложных индексов мы будем использовать для обозначения букв алфавита  $U$  метасимволы  $x_i$  с индексами или без них.

Введем определения линейной формы и пороговой функции.

*Линейной формой* назовем функцию вида

$$l_{\vec{w}, \sigma}(x_1, \dots, x_n) = \sum_{i=1}^n x_i w_i - \sigma,$$

где  $w_i$  и  $\sigma$  суть целые числа при  $i = 1, \dots, n$ .

Вектор  $\vec{w} = (w_1, \dots, w_n)$  называют вектором весовых коэффициентов, а  $\sigma$  — порогом.

Функция  $f(x_1, \dots, x_n) : E_2^n \rightarrow E_2$ , называется *пороговой*, если существует линейная форма  $l_{\vec{w}, \sigma}(x_1, \dots, x_n) = x_1 w_1 + \dots + x_n w_n - \sigma$  такая, что

$$f(x_1, \dots, x_n) = \begin{cases} 1, & \text{если } \sum_{i=1}^n x_i w_i - \sigma \geq 0; \\ 0, & \text{иначе.} \end{cases}$$

В этом случае говорим, что *линейная форма*  $l_{\vec{w}, \sigma}$  *задает пороговую функцию*  $f(x_1, \dots, x_n)$ , и записывается это так:

$$l_{\vec{w}, \sigma} \rightarrow f(x_1, \dots, x_n),$$

или просто  $f_{\vec{w}, \sigma}$ .

Множество всех пороговых функций от  $n$  переменных  $x_1, \dots, x_n$  обозначим  $T^n$ .

В связи с тем, что линейные формы с целочисленными коэффициентами и порогом позволяют задать любую пороговую функцию, далее в работе рассматриваются только такие линейные формы.

Введем понятие расстояния между линейными формами и пороговыми функциями.

Пусть  $l_{\vec{w}', \sigma'}$  и  $l_{\vec{w}'', \sigma''}$  — линейные формы от  $n$  переменных. Расстоянием между линейными формами  $l_{\vec{w}', \sigma'}$  и  $l_{\vec{w}'', \sigma''}$  назовем следующую величину

$$\rho(l_{\vec{w}', \sigma'}; l_{\vec{w}'', \sigma''}) = |\sigma' - \sigma''| + \sum_{i=1}^n |w'_i - w''_i|.$$

Эту величину интерпретируем как необходимость сделать  $\rho$  последовательных единичных изменений компонент одной линейной формы, чтобы получить другую.

Расстоянием между пороговыми функциями  $f'(x_1, \dots, x_n)$  и  $f''(x_1, \dots, x_n)$  назовем величину

$$\rho(f'; f'') = \min_{\substack{l_{\vec{w}', \sigma'} \rightarrow f' \\ l_{\vec{w}'', \sigma''} \rightarrow f''}} \rho(l'; l'').$$

Здесь минимум берется по всем линейным формам  $l_{\vec{w}', \sigma'}$  и  $l_{\vec{w}'', \sigma''}$ , задающим функции  $f'$  и  $f''$ , соответственно.

Определим величину  $\rho(n)$  следующим образом

$$\rho(n) = \max_{f', f'' \in T^n} \rho(f'; f'').$$

Данная величина характеризует расстояние между наиболее удаленными пороговыми функциями от  $n$  переменных.

Сформулируем основной результат данной работы.

**Теорема 1.** При  $n \rightarrow \infty$  выполнено

$$\log_2 \rho(n) \sim \frac{1}{2} n \log_2 n.$$

## 2. Доказательство теоремы 1

Введенные ранее пороговые функции будем также называть  $(0, 1)$ -пороговыми функциями.

Введем понятие  $(-1, 1)$ -пороговой функции.

Пусть  $V = \{v_1, v_2, \dots\}$  — счетный алфавит переменных, каждое из которых может принимать значения из множества  $\bar{E}_2 = \{-1, 1\}$ . Для обозначения букв алфавита  $V$  будем использовать метасимволы  $y_i$  с индексами или без них.

Функция  $g(y_1, \dots, y_n) : \bar{E}_2^n \rightarrow \bar{E}_2$  называется  $(-1, 1)$ -пороговой, если существует линейная форма  $l_{\vec{w}, \sigma}(y_1, \dots, y_n) = y_1 w_1 + \dots + y_n w_n - \sigma$  такая, что

$$g(y_1, \dots, y_n) = \begin{cases} -1, & \text{если } \sum_{i=1}^n y_i w_i - \sigma \geq 0; \\ 1, & \text{иначе.} \end{cases}$$

Множество  $(-1, 1)$ -пороговых функций обозначим  $\bar{T}^n$ .

Для того, чтобы отличать  $(0, 1)$ -пороговые и  $(-1, 1)$ -пороговые функции, введем следующие обозначения:  $f^{0,1}$  и  $g^{-1,1}$ . Иногда для  $(0, 1)$ -пороговых функций верхний индекс будем опускать.

Сопоставим переменным алфавита  $U$  переменные алфавита  $V$  по следующему правилу:  $\varphi(u_i) = v_i$ , для всех  $i$ . Положим также

$$\begin{aligned} \varphi(0) &= -1; \\ \varphi(1) &= 1. \end{aligned}$$

Определим изоморфизм множеств  $T^n$  и  $\bar{T}^n$  следующим образом: каждой  $(0, 1)$ -пороговой функции  $f^{0,1}(x_1, \dots, x_n)$  поставим в соответствие  $(-1, 1)$ -пороговую функцию  $g^{-1,1}(y_1, \dots, y_n)$  следующим образом

$$\varphi(f(x_1, \dots, x_n)) = g(\varphi(x_1), \dots, \varphi(x_n)).$$

Очевидно, что определенное соответствие является взаимно-однозначным. Далее, для краткости, соответствующие друг другу в терминах описанного изоморфизма функции  $f$  и  $g$  будем называть *изоморфными* и обозначать их  $f^{0,1}$  и  $f^{-1,1}$ .

Следующее утверждение, доказанное в работе [4], позволяет строить линейные формы, задающие изоморфные пороговые функции.

**Теорема 2.** *Имеют место следующие утверждения:*

$$1) \text{ если } l_{\vec{w}, \sigma} \rightarrow f^{0,1} \text{ и } l_{\vec{w}, 2\sigma - \sum_{i=1}^n w_i} \rightarrow g^{-1,1}, \text{ то } f^{0,1} \sim g^{-1,1};$$

2) если  $l_{\vec{w}, \sigma} \rightarrow g^{-1,1}$  и  $l_{\vec{w}, \frac{1}{2} \left( \sum_{i=1}^n w_i + \sigma \right)} \rightarrow f^{0,1}$ , то  $g^{-1,1} \sim f^{0,1}$ .

В работе [3] было показано, что

$$\log_2 \rho(n) \leq \frac{1}{2} n \log_2 n + o(n \log_2 n).$$

Таким образом, для доказательства теоремы 1 достаточно доказать нижнюю оценку.

Докажем вспомогательный результат.

**Лемма 1.** Если  $\vec{m}_1, \dots, \vec{m}_n$  — линейно-независимые вектора с координатами из  $\{-1, 1\}$  и  $\varphi(\vec{m}_i)$  — функция, сопоставляющая каждому вектору  $\vec{m}_i$  либо 1, либо  $-1$ , тогда найдется вектор  $\vec{w} = (w_1, \dots, w_n)$ , где  $w_i$  целые числа, такой что  $\vec{w} \cdot \vec{m} \neq 0$  для всех  $\vec{m} \in \{-1, 1\}^n$  и

$$\begin{cases} \vec{w} \cdot \vec{m}_i > 0, & \text{если } \varphi(\vec{m}_i) = 1; \\ \vec{w} \cdot \vec{m}_i < 0, & \text{если } \varphi(\vec{m}_i) = -1 \end{cases}$$

для всех  $i = 1, \dots, n$ .

**Доказательство.** Пусть  $\dot{m}_1, \dots, \dot{m}_n$  концы векторов  $\vec{m}_1, \dots, \vec{m}_n$ . Так как вектора  $\vec{m}_i$  линейно-независимы, то найдется гиперплоскость  $l$  размерности  $n - 1$ , которая содержит точки  $\dot{m}_1, \dots, \dot{m}_n$ . Отметим также, что  $l$  не проходит через начало координат, иначе вектора  $\vec{m}_1, \dots, \vec{m}_n$  не были бы линейно-независимыми.

Рассмотрим следующие множества

$$\begin{aligned} A &= \{\dot{m}_i, \varphi(i) = 1\}; \\ B &= \{\dot{m}_i, \varphi(i) = -1\}. \end{aligned}$$

Обозначим  $\text{conv}(X)$  выпуклую оболочку множества  $X$ . Заметим, что

$$\text{conv}(A) \cap \text{conv}(B) = \emptyset.$$

Иначе некоторый вектор  $\vec{m}_i$  линейно выражался бы через вектора  $\vec{m}_1, \dots, \vec{m}_{i-1}, \vec{m}_{i+1}, \dots, \vec{m}_n$ .

Так как все точки  $\dot{m}_1, \dots, \dot{m}_n$  лежат на плоскости  $l$ , то найдется прямая  $\gamma$ , также лежащая на  $l$ , которая разделяет множества  $A$  и  $B$  и, при этом, элементы  $A$  и  $B$  не лежат на этой прямой.

Все, что осталось сделать, это провести гиперплоскость  $\lambda$  через начало координат и прямую  $\gamma$ . Искомый вектор  $\vec{w}$  выбирается ортогональным плоскости  $\lambda$ . Его направление определяется исходя из условия

$$\begin{cases} \vec{w} \cdot \vec{m}_i > 0, & \text{если } \varphi(\vec{m}_i) = 1; \\ \vec{w} \cdot \vec{m}_i < 0, & \text{если } \varphi(\vec{m}_i) = -1. \end{cases}$$

Лемма доказана.

Содержательно лемма 1 говорит о том, что если на  $n$  линейно-независимых наборах из множества  $\{-1, 1\}^n$  произвольным образом заданы значения из множества  $\{-1, 1\}$ , то найдется самодвойственная  $(-1, 1)$ -пороговая функция  $f^{-1,1}$ , которая принимает указанные значения на данных наборах.

Самодвойственность  $f^{-1,1}$  обосновывается следующим образом. Заметим, что значение функции  $f^{-1,1}$  на наборе  $\vec{m}$  определяется знаком скалярного произведения  $\vec{w} \cdot \vec{m}$ . Так как вектор  $\vec{w}$  задает  $f^{-1,1}$  строгим образом, то на противоположном наборе  $-\vec{m}$  функция принимает противоположное значение. Следовательно,  $f^{-1,1}$  — самодвойственная.

При задании самодвойственных пороговых функций имеет место следующая особенность.

**Лемма 2.** *Если  $l_{\vec{w},\sigma} \rightarrow f^{-1,1}$  и  $f^{-1,1}$  — самодвойственная  $(-1, 1)$ -пороговая функция, тогда  $l_{\vec{w},0} \rightarrow f^{-1,1}$ .*

**Доказательство.** Возможны два случая. Если  $\sigma = 0$ , то утверждение очевидно.

Пусть  $\sigma \neq 0$ . Рассмотрим значение  $f^{-1,1}$  на наборе  $\alpha$ . Если  $f^{-1,1}(\alpha) = 1$ , тогда, так как  $f^{-1,1}$  самодвойственная, то  $f^{-1,1}(-\alpha) = -1$ . Иными словами

$$\begin{cases} \vec{w} \cdot \alpha - \sigma \geq 0, \\ -\vec{w} \cdot \alpha - \sigma < 0. \end{cases}$$

Следовательно,  $\vec{w} \cdot \alpha \geq |\sigma| > 0$ .

Если  $f^{-1,1}(\alpha) = -1$ , тогда

$$\begin{cases} \vec{w} \cdot \alpha - \sigma < 0, \\ -\vec{w} \cdot \alpha - \sigma \geq 0. \end{cases}$$

Следовательно,  $\vec{w} \cdot \alpha \leq -|\sigma| < 0$ . Лемма доказана.

Пусть  $M = \|m_{ij}\|_{n \times n}$  — обратимая матрица и  $M^{-1} = \|m'_{ij}\|_{n \times n}$ .  
Обозначим

$$L(M) = \max_{i,j} |m'_{ij}|.$$

Следующее утверждение позволяет по обратимой матрице  $M$  построить  $(-1, 1)$ -пороговые функции, удаленные друг от друга на расстоянии не менее  $L(M)$ .

**Лемма 3.** Если  $M = \|m_{ij}\|_{n \times n}$  — обратимая матрица и  $m_{ij} \in \{-1, 1\}$ , тогда существуют  $(-1, 1)$ -пороговые функции  $f_1^{-1,1}$  и  $f_2^{-1,1}$ , такие что

$$\rho(f_1^{-1,1}, f_2^{-1,1}) \geq L(M).$$

**Доказательство.** Пусть  $M^{-1} = \|m'_{ij}\|_{n \times n}$ . Без ограничения общности будем полагать, что

$$L(M) = m'_{11}.$$

Иными словами

$$L(M) = \frac{\det M_{1,1}}{\det M}.$$

Здесь  $M_{1,1}$  — матрица, полученная из  $M$  удалением 1-го столбца и 1-й строки.

Рассмотрим следующее представление матрицы  $M$

$$M = \begin{pmatrix} \vec{m}_1 \\ \dots \\ \vec{m}_n \end{pmatrix}.$$

Здесь вектор  $\vec{m}_i$  соответствует  $i$ -й строке матрицы  $M$ .

Зададим функцию  $f_1^{-1,1}$  на наборах  $\vec{m}_1, \dots, \vec{m}_n$  следующим образом

$$f_1^{-1,1}(\vec{m}_i) = \begin{cases} sg\left((-1)^{i+1} \frac{\det M_{i,1}}{\det M}\right), & \text{если } \det M_{i,1} \neq 0; \\ 1, & \text{иначе.} \end{cases}$$

Так как матрица  $M$  обратимая, то вектора  $\vec{m}_1, \dots, \vec{m}_n$  линейно-независимы. В таком случае, по лемме 1, найдется вектор  $\vec{w}$  такой, что  $\vec{w} \cdot \vec{m} \neq 0$  для всех  $\vec{m} \in \{-1, 1\}^n$  и

$$sg(\vec{w} \cdot \vec{m}_i) = f_1^{-1,1}(\vec{m}_i), \quad i = 1, \dots, n.$$

Таким образом, мы определили значение функции  $f_1^{-1,1}$  на наборах из множества  $\vec{m}_1, \dots, \vec{m}_n$ . На остальных наборах  $\vec{m}$  из множества  $\{-1, 1\}^n$  доопределим функцию  $f_1^{-1,1}$  следующим образом

$$f_1^{-1,1}(\vec{m}) = sg(\vec{w} \cdot \vec{m}).$$

Полученная функция  $f_1^{-1,1}$ , очевидно, является  $(-1, 1)$ -пороговой.

Пусть линейная форма  $l_{\vec{w}', \sigma'}$  задает  $f_1^{-1,1}$ . Покажем, что  $w'_1 \geq L(M)$ .

Сначала устраним из рассмотрения порог  $\sigma'$ . Так как  $f_1^{-1,1}$  самодвойственная, то по лемме 2 линейная форма  $l_{\vec{w}', 0}$  также задает  $f_1^{-1,1}$ . Поэтому в дальнейшем порог  $\sigma'$  можно полагать равным нулю.

Рассмотрим систему уравнений

$$M \cdot \vec{w}'^T = \vec{s}^T.$$

Очевидно, что  $sg(s_i) = f_1^{-1,1}(\vec{m}_i)$  при  $i = 1, \dots, n$ .

С другой стороны, по методу Крамера имеем

$$w'_1 = \frac{\det M_1}{\det M},$$

где  $M_1$  — матрица, полученная из  $M$  заменой первого столбца на  $\vec{s}^T$ . Разложим определитель  $\det M_1$  по первому столбцу

$$w'_1 = \sum_{i=1}^n (-1)^{i+1} \frac{\det M_{i,1}}{\det M} \cdot s_i.$$

Все слагаемые данной суммы неотрицательные так как знак  $s_i$  совпадает со знаком выражения

$$(-1)^{i+1} \frac{\det M_{i,1}}{\det M}.$$

Следовательно,

$$w'_1 \geq \frac{\det M_{1,1}}{\det M} = L(M).$$

Выберем в качестве функции  $f_2^{-1,1}$  функцию  $f_1^{-1,1}$ , у которой все переменные домножены на  $-1$ . В таком случае, по теореме «о сигналах», доказанной в работе [3], выполнено

$$\rho(f_1^{-1,1}, f_2^{-1,1}) \geq L(M).$$

Лемма доказана.

Аналогичный результат верен и для  $(0, 1)$ -пороговых функций.

Обозначим  $L_i(l_{\vec{w}, \sigma}) = |w_i|$ , где  $i \in \{1, \dots, n\}$ . Введем следующую характеристику сложности задания пороговой функции линейными формами

$$L_i(f^{0,1}) = \min_{l_{\vec{w}, \sigma} \rightarrow f^{0,1}} L_i(l_{\vec{w}, \sigma}).$$

Здесь минимум берется по всем линейным формам, задающим  $f^{0,1}$ .

Аналогичная характеристика вводится для  $(-1, 1)$ -пороговых функций

$$L_i(f^{-1,1}) = \min_{l_{\vec{w}, \sigma} \rightarrow f^{-1,1}} L_i(l_{\vec{w}, \sigma}).$$

Здесь минимум берется по всем линейным формам, задающим  $f^{-1,1}$ .

Имеет место следующее утверждение.

**Лемма 4.** *Если  $f^{0,1}$  пороговая функция, тогда для всех  $i \in \{1, \dots, n\}$  выполнено*

$$L_i(f^{0,1}) \geq L_i(f^{-1,1}).$$

**Доказательство.** Предположим противное, то есть для некоторого  $i$

$$L_i(f^{0,1}) < L_i(f^{-1,1}).$$

Пусть  $L_i(f^{0,1})$  достигается на линейной форме  $l_{\vec{w}, \sigma}$ . В таком случае по теореме 2 имеем

$$l_{\vec{w}, 2\sigma - \sum_i w_i} \rightarrow f^{-1,1}.$$

Следовательно,

$$L_i(f^{0,1}) = L_i(l_{\vec{w}, \sigma}) = L_i\left(l_{\vec{w}, 2\sigma - \sum_i w_i}\right) \geq L_i(f^{-1,1}).$$

Противоречие, лемма доказана.

Из леммы 4 вытекает аналог леммы 3 для  $(0, 1)$ -пороговых функций.

**Лемма 5.** Если  $M = \|m_{ij}\|_{n \times n}$  — обратимая матрица и  $m_{ij} \in \{-1, 1\}$ , тогда существуют  $(0, 1)$ -пороговые функции  $f_1^{0,1}$  и  $f_2^{0,1}$ , такие что

$$\rho(f_1^{0,1}, f_2^{0,1}) \geq L(M).$$

Сформулируем один известный результат.

**Теорема 3.** ([1]) При  $n \rightarrow \infty$  выполнено

$$2^{\frac{1}{2}n \log_2 n - 2n - o(n)} \leq L(M) \leq 2^{\frac{1}{2}n \log_2 n - n - o(n)}.$$

Теорема 1 очевидным образом следует из леммы 5 и теоремы 3.

## Благодарности

Я благодарю Валерия Борисовича Кудрявцева и других участников семинара «Кибернетика и информатика» за ценные консультации и обсуждения, возникавшие по ходу работы. Также хочу поблагодарить свою жену Анну за ту поддержку, которую она мне всегда оказывала.

## Список литературы

- [1] Graham R. L., Sloane N. J. A. Anti-Hadamard matrices // Linear algebra and its applications. 62. 1984.
- [2] Кострикин А. И. Введение в алгебру. Линейная алгебра. М.: Физматлит, 2000.
- [3] Соколов А. П. О конструктивной характеристике пороговых функций // Интеллектуальные системы. Т. 12, вып. 1–4. 2008. С. 363–388.
- [4] Соколов А. П. Об одном семействе нейронов с ограниченной сложностью взаимной перестройки // Интеллектуальные системы. Т. 13, вып. 1–4. 2009. С. 475–488.