1 Теорема Цыбенко об аппроксимации

Определение. $\sigma : \mathbb{R} \to \mathbb{R}$ *называется сигмоидной функцией, если пределы* $\lim_{x\to\infty} \sigma(x)$, $\lim_{x\to-\infty} \sigma(x)$ *существуют и* $\lim_{x\to\infty} \sigma(x) = 1, \lim_{x\to-\infty} \sigma(x) = 0$. σ *должна быть непрерывной.*

Например

$$\sigma(x) = \frac{1}{1 + \exp -x}$$

является сигмоидной функцией, также сигмоидой является

$$\begin{cases} 1, x \geqslant 0 \\ 0, x < 0 \end{cases}$$

Теорема доказана Джорджем Цыбенко в 1989 году и утверждает, что нейронная сеть с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью. Условиями являются: достаточное количество нейронов скрытого слоя и удачный подбор параметров. В [1] показано, что сети с активационной функцией ReLU с шириной n + 1 достаточны для аппроксимации любой непрерывной функции n-мерных входных переменных.

Теорема. *Если* $\sigma(x)$ *это ограниченнная сигмоидная функция, и* $f(x)$ *это непрерывная функция на* $(-\infty, \infty)$ *для которой* $\lim_{x\to-\infty} \grave{}(x) = A$ *и* $\lim_{x\to\infty} \grave{}(x) = B$, *где* A, B *- константы, то для любого* $\varepsilon > 0$ *существуют* N, c_i, y_i, θ_i *такие что*

$$|f(x) - \sum_{i=1}^{N} c_i \sigma(y_i x + \theta_i)| < \varepsilon$$

для всех $x \in (-\infty, \infty)$.

Доказательство. Доказательство является конструктивным. Из предположения, для любого $\varepsilon > 0$ мы можем найти $M > 0$ такое что $|f(x) - A| < \frac{\varepsilon}{4}$ если $x < -M$; $|f(x) - B| < \frac{\varepsilon}{4}$ если $x > M$; и $|f(x') - f(x")| < \frac{\varepsilon}{4}$ если $|x'| \leq M$, $|x''| \leq M$ и $|x' - x''| \leq \frac{1}{M}$.

Разделим $[-M; M]$ на $2M^2$ равных отрезков, каждый длины $\frac{1}{M}$, и положим

$$-M = x_0 < x_1 < \cdots < x_{M^2} = 0 < x_{M^2+1} < \cdots < x_{2M^2} = M$$

Пусть $t_i = \frac{1}{2}(x_i + x_{i+1})$ будет центром интервала $[x_i, x_{i+1}]$. Теперь построим функцию

$$g(x) = f(-M) + \sum_{i=1}^{N} [f(x_i) - f(x_{i-1})] \sigma(K(x - t_{i-1}))$$

где $N = 2M^2$.

Исходя из предположения, существует $W > 0$ такое что если $u > W$ тогда $|\sigma(u) - 1| < \frac{1}{M^2}$, если $u < -W$ тогда $|\sigma(u)| < \frac{1}{M^2}$. Положим $K > 0$ так чтобы $K\frac{1}{2M} > W$.

Мы утверждаем, что $|f(x) - g(x)| < \varepsilon$ выполняется для всех $x \in (-\infty, \infty)$.

(i) Если $x < -M$, тогда $|f(x) - f(-M)| < \frac{\varepsilon}{2}$,

$$|g(x) - f(-M)| \leq \sum_{i=1}^{2M^2} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{i-1}))| < \sum_{i=1}^{2M^2} \frac{\varepsilon}{4}\frac{1}{M^2} = \frac{\varepsilon}{2}$$

следовательно $|f(x) - g(x)| < \varepsilon$.

(ii) Если $x > M$, тогда $|f(x) - f(M)| < \frac{\varepsilon}{2}$,

$$|g(x) - f(M)| \leq \sum_{i=1}^{2M^2} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{i-1})) - 1| < \sum_{i=1}^{2M^2} \frac{\varepsilon}{4}\frac{1}{M^2} = \frac{\varepsilon}{2}$$

следовательно $|f(x) - g(x)| < \varepsilon$.

(iii) Рассмотрим $x \in [x_{k-1}, x_k]$, тогда $|x - t_{i-1}| \leq \frac{1}{2M}$ если $i = k$; $|x - t_{i-1}| > \frac{1}{2M}$ если $i \neq k$. Кроме того, если $i < k$ тогда $K(x - t_{i-1}) > W$, и следовательно $|\sigma(K(x - t_{i-1})) - 1| < \frac{1}{M^2}$; если $i > k$, тогда $K(x - t_{i-1}) < -W$, и следовательно $|\sigma(K(x - t_{i-1}))| < \frac{1}{M^2}$.

В результате мы имеем

$$|g(x) - f(-M) - [f(x_k) - f(x_{k-1})]\sigma(K(x - t_{k-1})) - \sum_{i=1}^{k-1}[f(x_i) - f(x_{i-1})]|$$

$\leq \sum_{i=1}^{k-1} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{k-1})) - 1| + \sum_{i=k+1}^{2M^2} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{i-1}))|$
$\leq \sum_{i=1}^{k-1} \frac{\varepsilon}{4}\frac{1}{M^2} + \sum_{i=k+1}^{2M^2} \frac{\varepsilon}{4}\frac{1}{M^2} < \frac{\varepsilon}{2}$

Ясно, что

$$f(-M) + [f(x_k) - f(x_{k-1})]\sigma(K(x - t_{k-1})) + \sum_{i=1}^{k-1}[f(x_i) - f(x_{i-1})]| = f(x_{k-1}) + [f(x_k) - f(x_{k-1})]\sigma(K(x - t_{k-1}))$$

и объединив это с предыдущим неравенством, получаем

$$|g(x) - f(x)| < \frac{\varepsilon}{2} + |f(x) - f(x_{k-1})| + |f(x_k) - f(x_{k-1})||\sigma(K(x - t_{k-1}))| < \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon.$$

Здесь мы предположили, без потери общности, что $\sigma(x) \leq 1$ для всех $x \in (-\infty, \infty)$.

Для того, чтобы закончить доказательство теоремы, нам необходимо только позаботиться о $f(-M)$, или дополнительно показать что консанта 1 может быть аппроксимирована с помощью линейной комбинации $\sum c_i\sigma(y_i x + \theta_i)$. Пусть N, M, x_i, t_i, K будут теми же константами, что и ранее, тогда это легко подтвердить что

$$\frac{1}{N}\sum_{i=1}^{N}[\sigma(K(x-t_{i-1}))+\sigma(-K(x-t_{i-1}))]$$

$\square$

Таким образом, любую непрерывную функцию нескольких переменных можно с любой точностью реализовать с помощью обычного трехслойного перцептрона с достаточным количеством нейронов в скрытом слое.

2 Теорема Колмогорова

Теорема. *Каждая непрерывная функция n переменных, заданная на единичном кубе n-мерного пространства E_n, представима в виде*

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} h_q[\sum_{p=1}^{n}\varphi_p^q(x_p)], \tag{1}$$

где функции $h_q(u)$ — непрерывны, а функции $\varphi_p^q(x_p)$, кроме того, еще стандартны, т.е. не зависят от выбора функции f; в частности, каждая непрерывная функция двух переменных представима в виде

$$f(x,y) = \sum_{q=1}^{5} h_q[\varphi_q(x) + \psi_q(y)]. \tag{2}$$

При $n = 3$ из (1) следует

$$f(x,y,z) = \sum_{q=1}^{7} h_q[\varphi_q(x) + \psi_q(y) + \chi_q(z)] = \sum_{q=1}^{7} F_q[g_q(x,y), z]$$

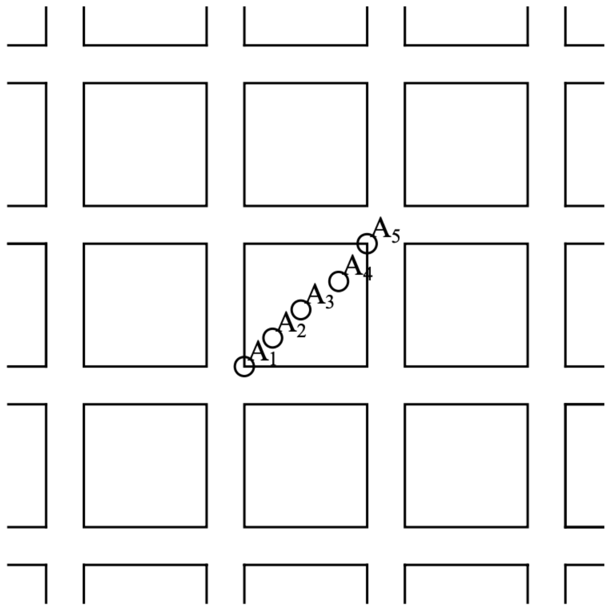
где положено

$$F_q(u,z) = h_q[u + \chi_q(z)], g_q = \varphi_q(x) + \psi_q(y).$$

Так как все идеи доказательства отчетливо проявляются уже в случае $n = 2$, то мы будем говорить лишь о представлении (2) произвольной непрерывной функции $f(x,y)$

двух переменных x и y. Возможность такого представления доказывается в несколько этапов.

(i) «Внутренние» функции $\varphi_q(x)$ и $\psi_q(y)$ представления (2) совершенно не зависят от разлагаемой функции $f(x,y)$. Для определения этих функций нам понадобятся неко-



торые предварительные построения. Рассмотрим изображенный на рис. 1 «город» — систему одинаковых «кварталов» (непересекающихся замкнутых квадратов) разделенных узкими «улицами» одной и той же ширины. Уменьшим гомотетично наш «город» в N раз; за центр гомотетии можно принять, например, точку A_1 - мы получим новый «город», который будем называть «городом ранга 2». «Город ранга 3» точно также получается из «города ранга 2» гомотетичным уменьшением с коэффициентом гомотетии $\frac{1}{N}$; «город ранга 4» получается гомотетичным уменьшением в N раз раз «города ранга 3» и т.д. Вообще «город ранга k» получается из исходного «города» (который мы будем называть «городом первого ранга») гомотетичным уменьшением в N^k раз (с центром гомотетии в A_1; впрочем, выбор центра гомотетии не существенен для дальнейшего).

Построенную систему «городов» мы назовем 1-й системой. «Город первого ранга q-й системы» ($q = 2, \ldots, 5$) получается из изображенного на рис. 1 «города» при помощи параллельного переноса, совмещающего точку A_1 с точкой A_q. Нетрудно понять, что «улицы» «города» можно выбрать настолько узкими, что каждая точка плоскости будет покрыта по крайней мере тремя кварталами наших пяти «городов первого ранга». Точно так же «город k-го ранга» q-й системы $(k = 2, 3, \ldots; q = 2, \ldots, 5)$ получается из «города k-го ранга 1-й системы» параллельным переносом, переводящим точку A_1^k в точку A_q^k, где A_1^k и A_q^k получаются из точек A_1 и A_q гомотетией, переводящей «город первого ранга» 1-й системы (т.е. наш исходный «город») в «город k-го ранга» той же системы; при этом каждая точка плоскости будет принадлежать кварталам по крайней мере трех из пяти "городов"любого фиксированного ранга k.

Функцию

$$\Phi_q(x, y) = \varphi_q(x) + \psi_q(y)(q = 1, 2, \ldots, 5)$$

мы определим теперь так, чтобы она разделяла любые два «квартала» каждого «города» системы q, т.е. чтобы множество значений, принимаемых $\Phi_q(x, y)$ на определенном «квартале» «города k-го ранга» (здесь k - произвольное фиксированное число) q-й системы, не пересекалось с множеством значений, принимаемых $\Phi_q(x, y)$ на любом другом «квартале» того же «города». При этом нам, разумеется, будет достаточно рассматривать функцию $\Phi_q(x, y)$ на единичном квадрате (а не на всей плоскости).

Для того, чтобы функция $\Phi_q(x, y) = \varphi_q(x) + \psi_q(y)$ разделяла «кварталы» «города первого ранга», можно потребовать, например, чтобы $\varphi_q(x)$ на проекциях «кварталов» «города» на ось x весьма мало отличалась от различных целых чисел, а $\psi_q(y)$ на проекциях «кварталов» на ось y весьма мало отличалась от различных кратных $\sqrt{2}$ (ибо $m + n\sqrt{2} = m' + n'\sqrt{2}$ при целых m, n, m', n', лишь если $m = m', n = n'$). При этом наложенные условия не определяют пока еще, разумеется, функций $\varphi_q(x)$ и $\psi_q(y)$ (на «улицах» функция $\Phi_q(x, y) = \varphi_q(x) + \psi_q(y)$ вообще пока может задаваться совершенно произвольно); используя это, можно подобрать границы значений $\varphi_q(x)$ и $\psi_q(y)$ на «кварталах» «города второго ранга» так, чтобы функция $\Phi_q(x, y) = \varphi_q(x) + \psi_q(y)$ разделяла не только «кварталы» «города 1 -го ранга», но и «кварталы» «города 2 -го ранга». Намеченную программу можно осуществить, если N достаточно велико (так что кварталы последующих рангов не соединяют кварталы предыдущих). А.Н. Колмогоров выбрал $N = 18$. Привлекая подобным же образом к рассмотрению "города"последующих рангов и уточняя каждый раз значения функций $\varphi_q(x)$ и $\psi_q(y)$, мы в пределе получим непрерывные функции $\varphi_q(x)$ и $\psi_q(y)$ (можно даже потребовать, чтобы они были монотонными), удовлетворяющие поставленным условиям

(ii) Функции $h_q(u)$ разложения (2), напротив того, существенно зависят от исходной функции $f(x,y)$. Для построения этих функций докажем прежде всего, что *любую непрерывную функцию $f(x,y)$ двух переменных x и y, заданную на единичном квадрате, можно представить в виде*

$$f(x,y)=\sum_{q=1}^{5}h_q^{(1)}(\Phi_q(x,y))+f_1(x,y), \tag{3}$$

где $\Phi_q(x,y)=\varphi_q(x)+\psi_q y$ - функции, построенные выше и

$$M_1=\max|f_1(x,y)|\leq\frac{5}{6}\max|f(x,y)|=\frac{5}{6}M, \tag{4}$$

$$\max|h_q^{(1)}(\Phi_q(x,y))|\leq\frac{1}{3}M, q=1,\ldots,5. \tag{5}$$

Выберем ранг k столь большим, чтобы колебание (т.е. разность наибольшего и наименьшего значений) функции $f(x,y)$ на каждом «квартале» любого из «городов ранга k» не превосходило $\frac{1}{6}M$; это, разумеется, возможно, так как с ростом ранга k размеры «кварталов» уменьшаются неограниченно. Далее, пусть $p_1^{(ij)}$ - определенный «квартал» «города 1 -й системы» (и выбранного ранга k); в таком случае (непрерывная) функция $\Phi_1(x,y)$ принимает на этом «квартале» значения, принадлежащие определенному сегменту $\Delta_1^{(ij)}$ числовой оси (причем в силу определения функции Φ_1 этот сегмент не пересекается с сегментами значений, принимаемых Φ_1 на всех других «кварталах»). Положим теперь функцию $h_1^{(1)}$ на сегменте $\Delta_1^{(ij)}$ постоянной, равной $\frac{1}{3}$ значения, принимаемого функцией $f(x,y)$ в какой-либо (безразлично какой) внутренней точке $M_1^{(ij)}$ квартала $p_1^{(ij)}$ (эту точку можно назвать «центром квартала»). Таким же образом мы определим функцию $h_1^{(1)}$ на любом другом из сегментов, задаваемых значениями функции $\Phi_1(x,y)$ на «кварталах» «города k-го ранга» 1 -й системы; при этом все значения $h_1^{(1)}$ будут по модулю не превосходить $\frac{1}{3}M$ (ибо значение $f(x,y)$ в «центре» любого «квартала» по модулю не превосходит M). Доопределим теперь функцию $h_1^{(1)}(u)$ при тех значениях аргумента u, при каких она еще не определена, произвольно, с тем лишь, чтобы она была непрерывна и чтобы выполнялось неравенство (5); совершенно аналогично определим и все остальные функции $h_q^{(1)}(u)(q=2,\ldots,5)$.

Докажем теперь, что разность

$$f_1(x,y)=f(x,y)-\sum_{q=1}^{5}h_q^{(1)}(\Phi_q(x,y))$$

довлетворяет условию (4), т.е. что

$$|f_1(x_0,y_0)| \leq \frac{5}{6}M,$$

где (x_0,y_0) - произвольная точка единичного квадрата. Эта точка (как и все точки плоскости) принадлежит по крайней мере трем кварталам «городов ранга k»; поэтому заведомо найдутся такие три из пяти функций $h_1^{(1)}(\Phi_q(x,y))$, которые принимают в точке (x_0,y_0) значение, равное $\frac{1}{3}$ значения $f(x,y)$ в «центре» соответствующего «квартала», т.е. отличающееся от $\frac{1}{3}f(x_0,y_0)$ не более чем на $\frac{1}{18}$M (ибо колебание $f(x,y)$ на каждом квартале не превосходит $\frac{1}{6}$M); сумма этих трех значений $h_q^{(1)}(\Phi_q(x_0,y_0))$ будет отличаться от $f(x_0,y_0)$ по модулю не более чем на $\frac{1}{6}$M. А так как каждое из оставшихся двух чисел $h_q^{(1)}(\Phi_q(x_0,y_0))$ в силу (3) по модулю не превосходит $\frac{1}{3}$M то мы получаем:

$$|f_1(x_0,y_0)| = \left| f(x_0,y_0) - \sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x_0,y_0)) \right| \leq \frac{1}{6}M + \frac{2}{3}M = \frac{5}{6}M,$$

что и доказывает (4). Применим теперь то же разложение (3) к входящей в (3) функции $f_1(x,y)$; мы получим:

$$f_1(x,y) = \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x,y)) + f_2(x,y)$$

или

$$f(x,y) = \sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x,y)) + \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x,y)) + f_2(x,y),$$

где

$$M_2 = \max|f_2(x,y)| \leq \frac{5}{6}M_1 = \left(\frac{5}{6}\right)^2 M,$$

и

$$\max\left|h_q^{(2)}(\Phi_q(x,y))\right| \leq \frac{1}{3}M_1 \leq \frac{1}{3}\cdot\frac{5}{6}M (q=1,2,\ldots,5).$$

Затем мы применим разложение (3) к полученной функции $f_2(x,y)$ и т.д.; после n - кратного применения этого разложения мы будем иметь:

$$f(x,y) = \sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x,y)) + \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x,y)) + f_2(x,y) + \ldots\cdots + \sum_{q=1}^{5} h_q^{(n-1)}(\Phi_q(x,y)) + f_n(x,y)$$

где

$$M_n = \max|f_n(x,y)| \leq \left(\frac{5}{6}\right)^n M$$

и

$$\max\left|h_q^{(s)}(\Phi_q(x,y))\right| \leq \frac{1}{3}\cdot\left(\frac{5}{6}\right)^{s-1} M(q=1,2,...,5; s=1,2,...,n-1).$$

Последние оценки показывают, что при $n\to\infty$ получим:

$$f(x,y)=\sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x,y))+\sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x,y))+f_2(x,y)+......+\sum_{q=1}^{5} h_q^{(n)}(\Phi_q(x,y))+...$$

где стоящий справа бесконечный ряд сходится равномерно; также и каждый из пяти рядов

$$h_q^{(1)}(\Phi_q(x,y))+h_q^{(2)}(\Phi_q(x,y))+h_q^{(n)}(\Phi_q(x,y))+...(q=1,2,...,5)$$

сходится равномерно, что позволяет ввести обозначения

$$h_q(u)=h_q^{(1)}+h_q^{(2)}+...+h_q^{(n)}+...(q=1,2,...,5).$$

Итак, окончательно получаем:

$$f(x,y)=\sum_{q=1}^{5} h_q(\Phi_q(x,y))=\sum_{q=1}^{5} h_q\left[\varphi_q(x)+\psi_q(y)\right],$$

то есть требуемое разложение (2).

Таким образом, любую непрерывную функцию n переменных можно получить с помощью операций сложения, умножения и суперпозиции из непрерывных функций одного переменного. Так существуют ли функции многих переменных? В каком-то смысле - да, в каком-то - нет. Все непрерывные функции многих переменных могут быть получены из непрерывных функций одного переменного с помощью линейных операций и суперпозиции.

Выражение (1) имеет структуру нейронной сети с одним скрытым слоем из 2n+1 нейронов. Теорема Колмогорова показала принципиальную возможность реализации сколь угодно сложных зависимостей с помощью относительно простой нейронной сети, называемой многослойным персептроном.

3 Теорема о глубоких нейросетях

Эта теорема об аппроксимирующей силе глубины нейронных сетей с активационной функцией ReLU и ограниченной шириной. Особенно интересеы следующие вопросы:

какова минимальная ширина $\omega_{\min(d)}$, чтобы сети ReLU ширины $\omega_{\min(d)}$ (и произвольной глубины) могли сколь угодно хорошо приближать любую непрерывную функцию на единичном кубе $[0,1]^d$? Что можно сказать о глубине сети, необходимой для аппроксимации заданной функции, для сетей ReLU вблизи этой минимальной ширины?

Для начала введем несколько обозначений

$$||f||_{C_0} := \sup_{x\in[0,1]^d} |f(x)|$$

Далее обозначим через

$$\omega_f(\varepsilon) := \sup\{|f(x) - f(y)|, |x - y| \leq \varepsilon\}$$

модуль непрерывности функции f. Определим нейронную сеть сеть с активационной функцией ReLU, размерностью входа d_in, шириной скрытого слоя ω, глубиной n и размером выхода d_out.

$$ReLU \circ A_n \circ \cdots \circ ReLU \circ A_1$$

где

$$A_j : \mathbb{R}^\omega \to \mathbb{R}^\omega, j = 2, \ldots, n-1, A_1 : \mathbb{R}^{d_in} \to \mathbb{R}^\omega, A_n : \mathbb{R}^\omega \to \mathbb{R}^{d_out}$$

$$ReLU(x_1, \ldots, x_m) = (\max\{0, x_1\}, \ldots, \max\{0, x_m\})$$

Обозначим такую сеть за $\mathbb{N}$ и запишем

$$f_{\mathbb{N}}(x) = ReLU \circ A_n \circ \cdots \circ ReLU \circ A_1 \circ ReLU \circ A_1(x)$$

Теорема. *Пусть $n \geq 1$ и $f : [0,1]^n \to \mathbb{R}_+$ — положительная функция с нормой $||f||_{C_0} = 1$. Тогда:*

1. *Если f — непрерывная, то существует последовательность нейронных сетей прямого распространения $\mathcal{N}_k$ с функциями активации ReLU, размер входного слоя — n, скрытого — $n+2$, выходного слоя — 1 такие, что*

$$\lim_{k\to\infty} ||f - f_{\mathcal{N}_k}||_{C^0} = 0. \tag{6}$$

В частности, $\omega_{\min}(n) \leq n + 2$ Кроме того, если зафиксировать $\varepsilon > 0$ и взять модуль непрерывности функции f $\omega_f(\varepsilon)$, то найдется нейронная сеть прямого *распространения* $\mathcal{N}_\varepsilon$ *с функциями активации ReLU , размерами входного слоя* n*, скрытого слоя —* $n+3$*, выходного слоя —* 1*, и*

$$\operatorname{depth}(\mathcal{N}_\varepsilon) = \frac{2 \cdot n!}{\omega_f(\varepsilon)^n}.$$

такая, что

$$||f - f_{\mathcal{N}_\varepsilon}||_{C_0} \leq \varepsilon.$$

2. *Если* f *— выпуклая, то существует последовательность нейронных сетей прямого распространения* $\mathcal{N}_m$ *с функциями активации ReLU, размерами входного слоя —* n*, скрытого слоя —* $n+1$*, и выходного слоя —* 1*, такие, что*

$$\lim_{m\to\infty} ||f - f_{\mathcal{N}_m}||_{C_0} = 0. \tag{7}$$

В частности, верна такая оценка $\omega_{\min}^{\text{conv}}(n) \leq n+1$*. Более того, найдется* $C > 0$ *такая, что если* f *одновременно и выпуклая, и липшицева с константой Липшица* L*, тогда нейронные сети* $\mathcal{N}_m$ *в (7) могут быть выбраны такими, что они удовлетворяют*

$$\operatorname{depth}(\mathcal{N}_m) = m,\ ||f - f_{\mathcal{N}_m}||_{C_0} \leq CLn^{\frac{3}{2}} m^{-\frac{2}{n}}.$$

3. *Если* f *— гладкая, то найдутся константа* K*, зависящая только от* n *и константа* B*, зависящая только от максимума первых* K *производных от* f *такие, что для каждого* $m \geq 3$ *нейронные сети* $\mathcal{N}_m$ *с размерностью* $n+2$ *в (6) могут быть выбраны такими что*

$$\operatorname{depth}(\mathcal{N}_m) = m,\ ||f - f_{\mathcal{N}_m}||_{C_0} \leq B(m-2)^{-\frac{1}{n}}.$$

Список литературы

[1] Lu Z. et al. The expressive power of neural networks: A view from the width //Advances in neural information processing systems. – 2017. – C. 6231-6239.