УДК 519.633

Теорема обучения для алгоритма конкуренции

В.С. Антюфеев

Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, просп. Акад. М.А. Лаврентьева, 6, Новосибирск, 630090

E-mail: ant@osmf.sscc.ru

Антюфеев В.С. Теорема обучения для алгоритма конкуренции // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2013. — Т. 16, № 1. — С. 1–9.

Настоящая работа является продолжением [1], где был предложен новый решающий алгоритм. Его функционирование напоминает действия искусственных нейронных сетей. Однако функционирование этого алгоритма основано на других принципах, в определении алгоритма не используются понятия сети, нейрона. Здесь доказана теорема обучения для нового алгоритма.

Ключевые слова: *теорема обучения, сходимость по вероятности, искусственная нейронная сеть.*

Antyufeev V.S. Theorem of training for a competition algorithm // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2013. — Vol. 16, № 1. — P. 1–9.

This paper is an extension of [1], where a new decision algorithm was proposed. In its operation, the unit resembles artificial neural networks. However the functioning of the algorithm proposed is based on the different concepts. It does not use the concept of a net, a neuron. The theorem of training for the new competition algorithm is proved.

Key words: *theorem of training, probabilistic convergence, artificial neural network.*

1. Введение

Создание искусственных систем распознавания образов остается сложной теоретической и технической проблемой. Необходимость в таком распознавании возникает в различных областях жизни — от военного дела и систем безопасности до оцифровки аналоговых сигналов.

Задачи распознавания часто решают с помощью искусственных нейронных сетей (ИНС) [2, 3]. Имеется много впечатляющих демонстраций возможностей ИНС: сеть научили превращать текст в фонетическое представление, которое затем с помощью уже иных методов превращалось в речь [4]; другая сеть может распознавать рукописные буквы [5]; сконструирована система сжатия изображений, основанная на нейронной сети [6].

ИНС были созданы в 50-х и 60-х годах XX века. Их сразу начали активно использовать для решения разнообразных задач, например для предсказания погоды, анализа электрокардиограмм. Действительно, сравнительно простые задачи могут быть решены с помощью "однослойных сетей" [7]. Причем для простейших однослойных сетей — персептронов — была доказана теорема об обучении персептронов, т. е. утверждение, что такие сети способны решать задачи определенного вида [8].

Однако вскоре было обнаружено, что "сети не могли решать задачи, внешне весьма сходные с теми, которые они успешно решали" [7]. Исследования, проведенные Минским [9], показали, что используемые в то время персептроны неспособны решить многие простые задачи, и интерес к исследованиям ИНС снизился.

Исследования ИНС возобновились после того, как были разработаны так называемые многослойные ИНС [7] и методы их обучения, которые позволили решать задачи, неразрешимые для персептронов. В настоящее время все ИНС для решения практических задач являются многослойными.

И все же, несмотря на огромное количество исследований в этой области, до сих пор имеются нерешенные теоретические и практические вопросы, связанные с ИНС.

ИНС — это логические алгоритмы, функционирование которых связывают с биологическими представлениями о работе мозга [10, 11]. Сложность этих алгоритмов затрудняет теоретическое исследование ИНС. Исследователям трудно понять, что на самом деле происходит "внутри" сети. Приведем цитаты из учебников и статей по ИНС:

"... нет гарантии, что сеть может быть обучена за конечное время. Много усилий, израсходованных на обучение, пропадает напрасно после затрат большого количества машинного времени. Когда это происходит, попытка обучения повторяется — без всякой уверенности, что результат окажется лучше [7]."

"... никакая из сегодняшних сетей не является панацеей, все они страдают от ограничений в своих возможностях обучаться и вспоминать [7]."

"Возможно, будет открыта какая-то мощная теорема о сходимости или найдена глубокая причина неудач доказать интересную теорему обучения для многослойных машин [7]."

"... трудность заключается в неспособности традиционных искусственных нейронных сетей объяснить, как они решают задачу. Внутреннее представление, получающееся в результате обучения, часто настолько сложно, что его невозможно проанализировать, за исключением самых простых случаев [7]."

В связи с этим в [1] предложен новый решающий алгоритм — *алгоритм конкуренции*, в котором не используются понятия нейрона, сети. Этот алгоритм позволяет решать примерно те же задачи, что и ИНС. Однако обучение и работа этого алгоритма основаны на других принципах. Эти новые принципы позволили существенно упростить алгоритм, сделать его работу понятной.

В частности, оказалось возможным применить к исследованию нового решающего алгоритма методы математического анализа, теории вероятностей. Это позволило доказать соответствующую теорему обучения, которую не удалось доказать для многослойных ИНС.

Для произвольного решающего алгоритма способность обучаться означает, что он способен давать правильные ответы после достаточно длительной процедуры обучения [7]. На математическом языке обучаемость алгоритма конкуренции означает, что его приближенные решающие функции сходятся к идеальной решающей функции [1], если обучающие сигналы выбраны специальным образом. В доказательстве теоремы указаны вид сходимости функций и способ моделирования обучающих сигналов, который обеспечивает эту сходимость.

2. Обозначения, определения, вспомогательные утверждения

В этом пункте мы напомним некоторые понятия, относящиеся к ИНС и алгоритму конкуренции (АК), а также введем несколько новых определений, которые будут нужны для доказательства теоремы.

ИНС функционирует в двух режимах: рабочий режим и режим обучения. На математическом языке упрощенную модель ИНС в рабочем режиме можно отождествить с функцией ИНС : $R^n \to \{0; 1\}$ или ИНС : $R^n \to [0, 1]$. Обратим внимание на области значений этих отображений. Область значений для отображения слева — числа 0 и 1.

Это так называемая *четкая* нейронная сеть. Областью значений для отображения справа является промежуток $[0,1]$. Отображение с такой областью значений соответствует *нечеткой* нейронной сети [12–15].

Функция ИНС — это логический алгоритм, сопоставляющий *сигналу* (точке из R^n) *ответ* (число из соответствующей области значений). Для четкой сети ответ на сигнал может быть только *верным* или *неверным* (относительно ответов нечеткой сети см., например, [15]). Верный ответ неизвестен ИНС до того, как проведено *обучение.*

Способность к обучению — самое важное свойство ИНС. Во время процедуры обучения в ИНС изменяются, корректируются специальные параметры, веса сети. После обучения ИНС должна давать верный ответ для каждой точки $z \in Z \subset R^n$.

Новый алгоритм АК является нечетким. Его также можно отождествить с функцией АК : $R^n \to [0,1]$. Он, как и ИНС, проходит процедуру обучения.

Перечислим некоторые определения.

- Сигнал z — это числовой кортеж, точка в R^n. Координаты сигнала z_k, $k=1,\ldots,n$, — параметры исследуемого объекта; n — фиксированная размерность пространства сигналов.
- Обозначим Z ($Z \subset R^n$) множество всех допустимых сигналов. Пусть Z — множество, измеримое относительно меры Лебега. Множество Z будет служить основным пространством для сигналов. Поэтому будем далее рассматривать лишь точки-сигналы $z \in Z$ и измеримые множества $U, V, \ldots \subset Z$. Их меру будем обозначать $\mu(U)$, $\mu(V), \ldots$ соответственно. Пусть по определению $0 < \mu(Z) < \infty$.
- Разобьем множество Z на два подмножества: X и Y “положительных” и “отрицательных” сигналов:

  $$X \cup Y = Z, \quad X \cap Y = \emptyset, \quad \mu(X) > 0.$$

  Для точек из множеств X и Y ответы на основной вопрос будут соответственно положительными и отрицательными. Множество X — это *решающее множество.* Задача распознавания состоит в том, чтобы для любого сигнала $z \in R^n$ дать ответ на *основной вопрос*: верно ли, что z принадлежит X?
- Вспомогательная функция $\varepsilon(z)$ — это *индикатор* или *знак* сигнала z:

  $$\varepsilon(z) = \begin{cases} +1, & z \in X, \\ -1, & z \in Y. \end{cases}$$

Каждый обучающий сигнал z^k создает вокруг себя скалярное поле влияния [1]. Знак этого поля совпадает со знаком $\varepsilon(z^k)$ сигнала z^k.

- Зададим функцию

  $$h(z) = 1/|z|^m. \tag{1}$$

  Здесь $m \geq n$ — произвольное число; n — размерность пространства сигналов. Смысл числа m выясним в доказательстве теоремы обучения. Величину интенсивности поля, индуцируемого обучающим сигналом z^k в произвольной точке $z \in Z$, задает *функция влияния сигнала* [1]:

  $$h_k(z) = \varepsilon(z^k)\ h(z - z^k). \tag{2}$$

- Пусть $z^1, z^2, \ldots, z^N$ — конечная последовательность обучающих сигналов. Каждый из сигналов z^k задает свое поле влияния с интенсивностью $h_k(z)$ в точке z. Суперпозиция этих полей определяется сложением. Величину интенсивности суперпозиции полей в точке z обозначим $H_N(z)$:

$$H_N(z) = \sum_{k=1}^{N} h_k(z), \tag{3}$$

$H_N(z)$ — функция влияния конечной совокупности обучающих сигналов $\{z^1, z^2, \ldots, z^N\}$.

- $F(t)$ — вспомогательная функция одной переменной: $F(t) = \frac{1}{2}\left[\frac{t}{\sqrt{t^2+1}} + 1\right]$ с характерным s-образным графиком (рис. 1).

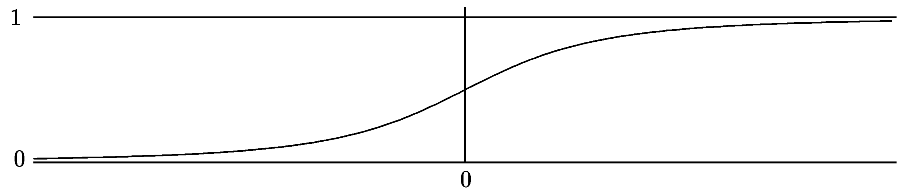

Рис. 1. График функции $F(t)$

- $f_N(z)$ — приближенная решающая функция, соответствующая обучающим сигналам z^1, z^2, …, z^N:

$$f_N(z) = \begin{cases} F(H_N(z)), & z \neq z^k, \quad k = 1, \ldots, N, \\ 1, & z = z^k \in X, \quad k = 1, \ldots, N, \\ 0, & z = z^k \in Y, \quad k = 1, \ldots, N. \end{cases} \tag{4}$$

- Идеальная решающая функция $\chi_X(z) = \begin{cases} 1, & z \in X, \\ 0, & z \notin X \end{cases}$ является характеристической функцией [16] решающего множества X.

3. Теорема обучения

В этом пункте доказано, что если выбирать обучающие сигналы в Z специальным образом, то алгоритм конкуренции может быть обучен "сколь угодно хорошо". Другими словами, правильное обучение позволяет сколь угодно приблизить к единице долю верных ответов.

По ходу доказательства мы дважды воспользуемся следующим утверждением из математического анализа [17].

Пусть $\lim_{k\to\infty} a_k = 0$ и $a_k \geq b_k \geq 0$ для всех k. Тогда $\lim_{k\to\infty} b_k = 0$. Будем ссылаться на утверждение о последовательности.

Теорема. *Пусть $\{z^1, z^2, \ \ldots\}$ — бесконечная последовательность случайных взаимно независимых точек, равномерно распределенных на множестве $Z \subset R^n$, $0 < \mu(Z) < \infty$. Тогда последовательность приближенных решающих функций $f_k(z)$ сходится по вероятности к идеальной решающей функции $\chi_X(z)$ во всех точках $z \in Z$ за исключением, может быть, точек границы $\Gamma(X)$ искомого множества X.*

Другими словами, если обучающие точки-сигналы выбраны указанным выше образом, то выполняются следующие предельные соотношения:

$$f_k(z) \xrightarrow{p} 1 \quad \forall\, z \in \mathrm{Int}(X), \qquad f_k(z) \xrightarrow{p} 0 \quad \forall\, z \in \mathrm{Int}(Y). \tag{5}$$

Здесь $\mathrm{Int}(W)$ — совокупность внутренних точек множества W, т. е. точек, принадлежащих W вместе с некоторой окрестностью.

Доказательство. Пусть $x^* \in X$ — произвольная точка, которая не лежит на границе множества X. Мы собираемся доказать, что $f_k(x^*) \xrightarrow{p} 1$. Если $y^* \in Y$ — точка, не лежащая на границе X, то предельное соотношение $f_k(y^*) \xrightarrow{p} 0$ можно доказать аналогичным образом.

Очевидно теорема справедлива, если точка x^* совпадает с одним из обучающих сигналов $z^k \in X$, так как в этом случае $f_k(z^k) \equiv 1$ для всех $k = 1, 2, \dots$ согласно определению (4).

Пусть точка $x^* \in X$ не совпадает ни с одним из обучающих сигналов z^k. Согласно определению (4) решающей функции f_k (см. также [1]), утверждение (5) эквивалентно следующему предельному соотношению:

$$H_N(x^*) \xrightarrow{p} +\infty, \quad N \to \infty. \tag{6}$$

Сходимость по вероятности к бесконечному пределу определим в ходе доказательства по аналогии со сходимостью к конечному пределу [19].

Итак, доказательство теоремы сходимости сводится к доказательству предельного соотношения (6). Не ограничивая общности, будем считать для удобства, что

$$x^* = \mathbf{0}. \tag{7}$$

Если $x^* \neq \mathbf{0}$, то выполним параллельный перенос системы координат в R^n.

Разделим конечную совокупность последовательных обучающих сигналов $z^1, z^2, \dots, z^N$ на два не пересекающихся подмножества:

$$\{z^1, z^2, \dots, z^N\} = \{x^1, x^2, \dots, x^{N_X}\} \cup \{y^1, y^2, \dots, y^{N_Y}\}, \quad N_X + N_Y = N. \tag{8}$$

В первую группу отнесем те сигналы z^k, которые попали в X, вторая группа состоит из сигналов z^k, попавших в Y.

Запишем и преобразуем выражение для $H_N(x^*)$. Согласно определениям (2), (3),

$$H_N(x^*) = \sum_{k=1}^{N} h_k(x^*) = \sum_{k=1}^{N} \frac{\varepsilon(z^k)}{|z^k - x^*|^m}.$$

Так как $x^* = \mathbf{0}$ (см. (7)), это выражение можно упростить: $\sum_{k=1}^{N} \frac{\varepsilon(z^k)}{|z^k - x^*|^m} = \sum_{k=1}^{N} \frac{\varepsilon(z^k)}{|z^k|^m}$. Учитывая соотношения $\varepsilon(x^i) = +1$, $\varepsilon(y^j) = -1$ и разбиение множества $\{z^1, z^2, \dots, z^N\}$, имеем $\sum_{k=1}^{N} \frac{\varepsilon(z^k)}{|z^k|^m} = \sum_{i=1}^{N_X} \frac{1}{|x^i|^m} - \sum_{j=1}^{N_Y} \frac{1}{|y^j|^m}$. Для краткости обозначим: $H_N^X = \sum_{i=1}^{N_X} \frac{1}{|x^i|^m}$, $H_N^Y = \sum_{j=1}^{N_Y} \frac{1}{|y^j|^m}$ $(N_X + N_Y = N)$. Итак,

$$H_N(x^*) = H_N^X - H_N^Y. \tag{9}$$

Напомним, что, согласно условиям теоремы, обучающие сигналы x^i, y^j — это случайные точки и, следовательно, H_N^X, H_N^Y — случайные величины. Найдем пределы по вероятности выражений $\frac{1}{N} H_N^X$ и $\frac{1}{N} H_N^Y$.

1. Сначала рассмотрим выражение

$$\frac{1}{N_X} H_N^X = \frac{1}{N_X} (h(x^1) + \dots + h(x^{N_X})) \tag{10}$$

(см. определение (1) функции h) и найдем его предел.

Пусть $\boldsymbol{x}$ — случайная точка, равномерно распределенная на множестве X. Тогда $\boldsymbol{h} = h(\boldsymbol{x})$ — случайная величина. Точки x^i можно считать выборочными значениями $\boldsymbol{x}$. Выражения вида (10) используют в качестве статистических оценок для вычисления многомерных интегралов методом Монте-Карло [18] на основании закона больших чисел.

Согласно теореме Колмогорова [19], случайная величина вида (10) сходится по вероятности к соответствующему интегралу от h в том и только в том случае, если существует конечное математическое ожидание $\mathbf{E}\boldsymbol{h}$. Но этот интеграл от h расходится в окрестности особой точки $\mathbf{0}$ при $m \geq n$ [17], следовательно, конечное математическое ожидание $\mathbf{E}\boldsymbol{h}$ не существует.

Вычислим предел выражения (10), используя лишь случайные величины с конечным математическим ожиданием.

Обозначим $S(\varepsilon)$ шар радиуса $\varepsilon > 0$ в R^n с центром в точке $\mathbf{0}$. Удалим из X окрестность $S(\varepsilon)$ особой точки $\mathbf{0}$ и обозначим $X_\varepsilon = X - S(\varepsilon)$ оставшееся множество (рис. 2). Рассмотрим регуляризованную функцию

$$h_\varepsilon = \begin{cases} h(x), & x \in X_\varepsilon, \\ 0, & x \in S_\varepsilon. \end{cases} \tag{11}$$

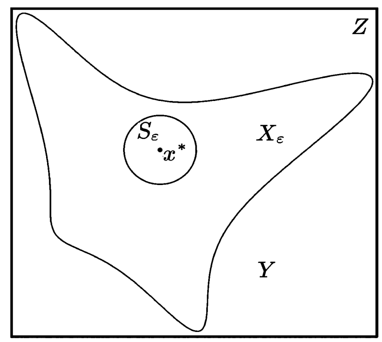

Рис. 2. Разбиение множества X

Функция h_ε ограничена, поэтому интеграл по области X от h_ε сходится. Обозначим I_ε его значение: $I_\varepsilon = \int_X h_\varepsilon(x)\,dx < \infty$. Поскольку интеграл от функции h расходится в окрестности точки $\mathbf{0}$, выполняется предельное соотношение

$$\lim_{\varepsilon\to 0} I_\varepsilon = \lim_{\varepsilon\to 0} \int\limits_{X\setminus S_\varepsilon} h(x)\,dx = +\infty. \tag{12}$$

Выберем произвольное число $A > 0$. В силу (12) найдется такое $\varepsilon_A > 0$, что при $\varepsilon < \varepsilon_A$ выполняется неравенство

$$I_\varepsilon > A. \tag{13}$$

Функции h_ε соответствует регуляризованная случайная величина $\boldsymbol{h}_\varepsilon = h_\varepsilon(\boldsymbol{x})$ с конечным математическим ожиданием $\mathbf{E}\boldsymbol{h}_\varepsilon$. Для $\boldsymbol{h}_\varepsilon$ выполняются условия упомянутой теоремы Колмогорова. Поэтому

$$\frac{1}{N_X}(h_\varepsilon(x^1) + \dots + h_\varepsilon(x^{N_X})) \xrightarrow{p} \mathbf{E}\boldsymbol{h}_\varepsilon = I_\varepsilon < \infty$$

при $N_X \to \infty$. Другими словами, для любого фиксированного $\delta > 0$ выполняется предельное соотношение

$$P\left(\left|\frac{1}{N_X}(h_\varepsilon(x^1) + \dots + h_\varepsilon(x^{N_X})) - I_\varepsilon\right| > \delta\right) \to 0 \tag{14}$$

при $N_X \to \infty$. Выберем здесь положительное $\varepsilon < \varepsilon_A$ и $\delta = I_\varepsilon - A$. Тогда, согласно (13), $\delta > 0$.

Воспользуемся простым замечанием: для любой случайной величины $\boldsymbol{a}$ выполняется неравенство

$$P(|\boldsymbol{a}| > \delta) = P(\boldsymbol{a} > \delta) + P(\boldsymbol{a} < -\delta),$$

откуда

$$P(|\boldsymbol{a}| > \delta) \geq P(\boldsymbol{a} < -\delta).$$

Из (14) и из последнего замечания следует, что выполняется неравенство

$$P\left(\left|\frac{1}{N_X}(h_\varepsilon(x^1)+\cdots+h_\varepsilon(x^{N_X}))-I_\varepsilon\right|>\delta\right)$$
$$\geq P\left(\frac{1}{N_X}(h_\varepsilon(x^1)+\cdots+h_\varepsilon(x^{N_X}))-I_\varepsilon<-\delta\right)\geq 0.$$

Учитывая предельное соотношение (14) и $\delta=I_\varepsilon-A>0$, получаем из утверждения о последовательности, что

$$P\left(\frac{1}{N_X}(h_\varepsilon(x^1)+\cdots+h_\varepsilon(x^{N_X}))-I_\varepsilon<-\delta\right)$$
$$=P\left(\frac{1}{N_X}(h_\varepsilon(x^1)+\cdots+h_\varepsilon(x^{N_X}))<I_\varepsilon-\delta\right)=P\left(\frac{1}{N_X}(h_\varepsilon(x^1)+\cdots+h_\varepsilon(x^{N_X}))<A\right)\to 0$$

при $N_X\to\infty$.

Еще одно замечание: если для случайных величин $\boldsymbol{a}$, $\boldsymbol{b}$ выполняется неравенство $\boldsymbol{a}\geq\boldsymbol{b}$ и C — произвольное число, то $P(\boldsymbol{b}<C)\geq P(\boldsymbol{a}<C)$. Поскольку $h(z)\geq h_\varepsilon(z)$ (см. (11)), из этого замечания следует, что

$$P\left(\frac{1}{N_X}(h_\varepsilon(x^1)+\cdots+h_\varepsilon(x^{N_X}))<A\right)\geq P\left(\frac{1}{N_X}(h(x^1)+\cdots+h(x^{N_X}))<A\right).$$

Следовательно, по утверждению о последовательности

$$P\left(\frac{1}{N_X}(h(x^1)+\cdots+h(x^{N_X}))<A\right)\equiv P\left(\frac{1}{N_X}H_N^X<A\right)\to 0.$$

Напомним, что число A здесь произвольное, поэтому его можно выбрать сколь угодно большим. По аналогии с определением сходимости по вероятности к конечному пределу можно считать, что последнее предельное соотношение означает сходимость последовательности случайных величин $\frac{1}{N_X}H_N^X$ к бесконечности:

$$\frac{1}{N_X}H_N^X\xrightarrow{p}+\infty. \tag{15}$$

Осталось найти предел по вероятности случайной величины $\frac{1}{N}H_N^X$ ($N=N_X+N_Y>N_X$). Перепишем выражение для $\frac{1}{N}H_N^X$:

$$\frac{1}{N}H_N^X=\left(\frac{N_X}{N}\right)\left(\frac{1}{N_X}H_N^X\right).$$

Предел сомножителя $\left(\frac{1}{N_X}H_N^X\right)$ уже найден (см. (15)). Перепишем сомножитель $\left(\frac{N_X}{N}\right)$ в следующем виде:

$$\frac{N_X}{N}=\frac{\chi_X(z^1)+\cdots+\chi_X(z^N)}{N}.$$

В числителе последней дроби слагаемые $\chi_X(z^k)$ являются случайными величинами Бернулли. Те N_X из них, для которых $z^k=x^i\in X$, равны 1, а остальные равны 0. Согласно закону больших чисел,

$$\frac{\chi_X(z^1)+\cdots+\chi_X(z^N)}{N}=\frac{\chi_X(x^1)+\cdots+\chi_X(x^N_X)}{N}\xrightarrow{p}E(\chi_X(\boldsymbol{x})=\int\limits_Z\chi_X(x)\,dx=\frac{\mu(X)}{\mu(Z)}>0, \tag{16}$$

так как $\mu(X) > 0$ и $\mu(Z) < \infty$ по определению. Из (15), (16) следует искомое предельное соотношение (в том смысле, как определено выше):

$$\frac{1}{N} H_N^X \xrightarrow{p} +\infty. \tag{17}$$

2. Вернемся к (9) и оценим сверху выражение $\frac{1}{N} H_N^Y$. Эту оценку можно получить, не используя теории вероятности.

Обозначим d расстояние от точки x^* до множества Y:

$$d = \inf_{y \in Y} |y - x^*| = \inf_{y \in Y} |y - \mathbf{0}| = \inf_{y \in Y} |y|. \tag{18}$$

Согласно условиям теоремы, $x^* \in \mathrm{Int}(X)$, следовательно, $d > 0$. Очевидно, выполняется следующее неравенство:

$$H_N^Y = \sum_{j=1}^{N_Y} \frac{1}{|y^j|^m} \le N_Y \cdot \max_{y^j \in Y} \frac{1}{|y^j|^m}. \tag{19}$$

Оценим оба сомножителя в правой части последнего неравенства. Поскольку $N_X + N_Y = N$ (см. (8)), то

$$N_Y \le N. \tag{20}$$

Поскольку $\inf_{y \in Y} |y| = d$ (см. (18)), то для всех точек $y^j \in Y$ выполняется неравенство $|y^j| \ge d$. Следовательно,

$$\max_{y^j \in Y} \frac{1}{|y^j|^m} = \frac{1}{\min\limits_{y^j \in Y} |y^j|^m} \le \frac{1}{d^m}. \tag{21}$$

Из (19)–(21) получаем оценку сверху для $\frac{1}{N} H_N^Y$: $\frac{1}{N} H_N^Y \le \frac{1}{d^m} = C_1 < +\infty$. Здесь константа C_1 не зависит от количества N обучающих точек.

Итак, мы доказали (17), что $\frac{H_N^X}{N} \xrightarrow{p} +\infty$, и для всех N выполняется неравенство $\frac{H_N^Y}{N} \le C_1 < +\infty$. Отсюда следует, что $\frac{H_N^X - H_N^Y}{N} \xrightarrow{p} +\infty$, и тем более $[H_N^X - H_N^Y] \xrightarrow{p} +\infty$. $\square$

Литература

1. **Антюфеев В.С.** Решения задач распознавания методом Монте-Карло // RJNAMM. — 2012. — Т. 27, № 2. — P. 113–130.
2. **Барский А.Б.** Нейронные сети: распознавание, управление, принятие решений. — М.: Финансы и статистика, 2004.
3. **Осовский С.** Нейронные сети для обработки информации. — М.: Финансы и статистика, 2004.
4. **Sejnowski T.J., Rosenberg C.R.** Parallel networks that learn to pronounce English text // Complex Systems. — 1987. — № 3. — P. 145–168.
5. **Burr D.J.** Experiments with a connectionist text reader // Proc. of the First International on Neural Networks / Caudill M., Butler C. — San Diego, CA: SOS. — 1987. — Vol. 4. — P. 717–724.
6. **Cottrell G.W., Munro P., and Zipser D.** Learning internal representation from gray-scale images: An example of extensional programming // Proc. of the 9-th Annual cognitive sci.soc.conf., Seatle (USA), July16-18. — Norwood, NJ: Ablex, 1987. — Vol. 3. — P. 461–473.
7. **Уоссермен Ф.** Нейрокомпьютерная техника: Теория и практика. — М.: Мир, 1992.

8. **Розенблатт Ф.** Принципы нейродинамики. — М.: Мир, 1965.
9. **Минский М.Л., Пейперт С.** Персептроны. — М: Мир, 1971.
10. **Grossberg S.** Studies of Mind and Brain. — Boston: Reidel, 1982.
11. **McCulloch V.S., Pitts W.H.** A logical calculus of ideas immanent in nervous activity // Bull. Math. Biophysics. — 1943. — Vol. 2. — P. 548–558.
12. **Кофман А.** Введение в теорию нечетких множеств. — М.: Радио и связь, 1982.
13. **Орловский С.А.** Проблемы принятия решений при нечеткой исходной информации. — М.: Наука, 1981.
14. **Аверкин А.Н.** Нечеткие множества в моделях управления и искусственного интеллекта / Д.А. Поспелова. — М.: Наука, 1986.
15. **Рутковская Д., Пилиньский М., Рутковский Л.** Нейронные сети, генетические алгоритмы и нечеткие системы. — М.: Горячая линия–Телеком, 2004.
16. **Куратовский А.** Топология. Т. I. — М.: Мир, 1966.
17. **Фихтенгольц Г.М.** Курс дифференциального и интегрального исчисления. — М.: Наука, 1969.
18. **Ермаков С.М., Михайлов Г.А.** Статистическое моделирование.— М.: Наука, 1982.
19. **Феллер В.** Введение в теорию вероятностей и ее приложения. — М.: Мир, 1984.

Поступила в редакцию 17 ноября 2011 г.,
в окончательном варианте 16 декабря 2011 г.