

Влияние гиперпараметров нейронной сети на её численную обусловленность

С. В. Шолтанюк, ассистент кафедры компьютерных технологий
и систем ФПМИ

E-mail: SSholtanyuk@bsu.by

ORCID ID: 0000-0003-0266-7135

Белорусский государственный университет, пр. Независимости, д. 4,
220030, г. Минск, Республика Беларусь

Аннотация. В данной работе рассмотрена задача оценивания численной обусловленности многослойного персептрона, прогнозирующего временные ряды методом скользящего окна. Рассмотрена работа прогностического персептрона при различных наборах гиперпараметров, в частности, при различном количестве нейронов на разных слоях нейронной сети, а также при использовании тех или иных функций активации. Выявлены основные факторы, влияющие на обусловленность нейронной сети, а также особенности её работы при различных функциях активации. Предложены формулы для оценки чисел обусловленности отдельных компонентов прогностического персептрона и самой нейронной сети в целом. Проведён сравнительный анализ результатов обучения прогностического персептрона при различных гиперпараметрах на примере смоделированных временных рядов. Сформулированы условия, обеспечивающие лучшую устойчивость и обусловленность нейронной сети.

Ключевые слова: прогнозирование временных рядов, нейронные сети, персептрон, численная обусловленность, функция активации

Для цитирования: Шолтанюк, С. В. Влияние гиперпараметров нейронной сети на её численную обусловленность / С. В. Шолтанюк // Цифровая трансформация. – 2020. – № 1 (10). – С. 43–50. <https://doi.org/10.38086/2522-9613-2020-1-43-50>



© Цифровая трансформация, 2020

Influence of the Neural Network Hyperparameters on its Numerical Conditioning

S. V. Sholtanyuk, assistant of the Department of Computer Applications
and Systems, FAMCS

E-mail: SSholtanyuk@bsu.by

ORCID ID: 0000-0003-0266-7135

Belarusian State University, 4 Independence Ave., 220030 Minsk,
Republic of Belarus

Abstract. In this paper, the task of assessment of numerical conditioning of multilayer perceptron, forecasting time series with sliding window method, has been considered. Performance of the forecasting perceptron with various hyperparameters sets, with different amount of neurons and various activation functions in particular, has been considered. Main factors, influencing on the neural net conditioning, have been revealed, as well as performance features, when using various activation functions. Formulas for assessment of condition numbers of individual components of the forecasting perceptron and of the neural network itself have been proposed. Comparative analysis of results of training the forecasting perceptron with various hyperparameters on modeled time series has been performed. Conditions, providing the best stability and conditioning for the neural network, have been formulated.

Key words: time series forecasting, neural networks, perceptron, numerical conditioning, activation function

For citation: Sholtanyuk S. V. Influence of the Neural Network Hyperparameters on its Numerical Conditioning. *Cifrovaja transformacija* [Digital transformation], 2020, 1 (10), pp. 43–50 (in Russian). <https://doi.org/10.38086/2522-9613-2020-1-43-50>

© Digital Transformation, 2020

Введение. Искусственные нейронные сети получили широкое применение при решении различных задач, например, классификации образов, распознавании речи, принятии решений, прогнозировании временных рядов. Существует множество программных реализаций и библио-

тек на всевозможных языках программирования, позволяющих построить нейронные сети заданной архитектуры.

В центре внимания современных исследователей стоит такая проблема как устойчивость и робастность нейронных сетей. Существуют раз-

личные аспекты устойчивости нейронных сетей, например, различают устойчивость по входным данным (динамическую устойчивость), означающую малую чувствительность результата к возмущениям во входных данных, и структурную, которая подразумевает незначительность изменений результата в случае изменения каких-либо числовых характеристик самой нейронной сети [1]. Наиболее ярко проявления неустойчивости нейронных сетей возникает в задачах, связанных с распознаванием образов при помощи глубоких свёрточных сетей. Обучаясь на определённом наборе данных, они становятся неспособными к классификации объектов на зашумлённых изображениях, которые неотличимы от оригинала для человеческого глаза, но являются совершенно другими с точки зрения нейронной сети [2].

Целью данной статьи является выявление факторов, влияющих на динамическую устойчивость нейронных сетей, используемых в задачах прогнозирования временных рядов. Для этой цели рассмотрен вопрос обусловленности компонентов нейронных сетей, в частности, вычисление их чисел обусловленности и влияние этих показателей на работу нейронных сетей, а также проведён сравнительный анализ их работы при различных наборах гиперпараметров.

Архитектура и обучение прогностического персептрона. Объектом исследования данной статьи является трёхслойный персептрон следующей архитектуры (рис. 1):

- входной слой содержит произвольное число нейронов p , на которые подаются p последовательных значений нормированного временного ряда;
- скрытый слой содержит произвольное число нейронов n ;
- выходной слой содержит единственный нейрон, на выходе которого получается прогнозное значение для следующего, $p+1$ -ого элемента нормированного временного ряда;
- входной и скрытый слои содержат также нейроны смещения;
- нейронная сеть является полносвязной, т. е. все нейроны каждого слоя соединены синапсами со всеми нейронами следующего слоя.

Прогностический персептрон работает с нормированным временным рядом, значения которого вычисляются по следующей формуле:

Прогностический персептрон работает с нормированным временным рядом, значения которого вычисляются по следующей формуле:

$$\tilde{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}, i = \overline{1, N}, \quad (1)$$

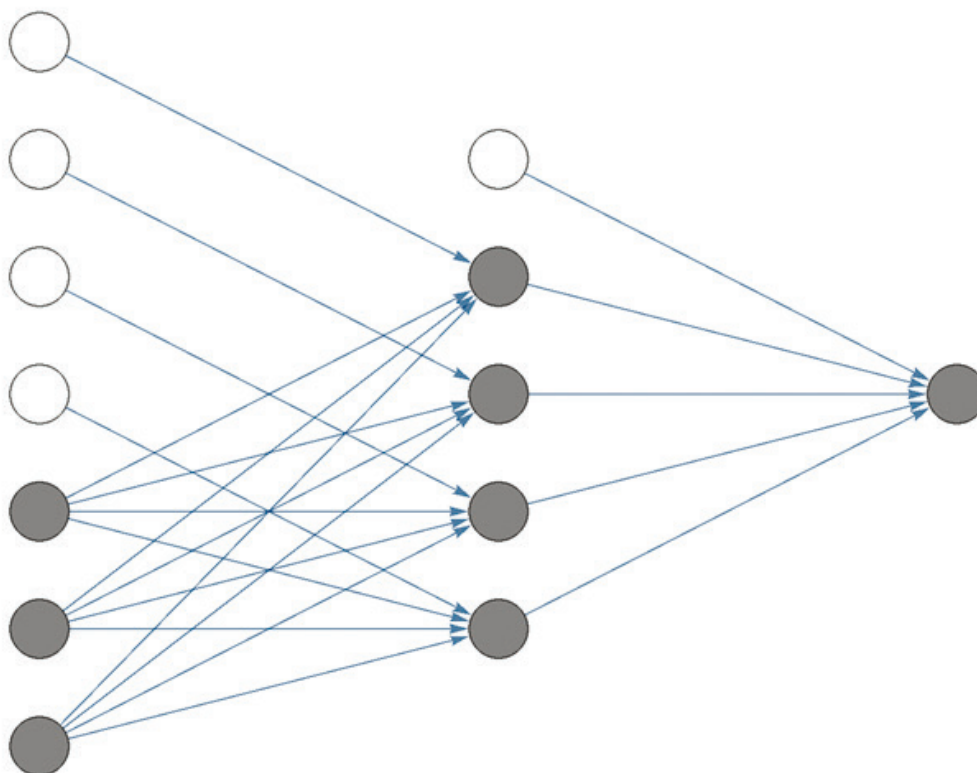


Рис. 1. Пример архитектуры прогностического персептрона при $p=3$, $n=4$. Серым цветом закрашены обычные нейроны, белым – нейроны смещения
 Fig. 1. An example of forecasting perceptron architecture for $p=3$, $n=4$. Regular neurons are painted in grey, and bias ones are painted in white

где x_i – i -ое значение первоначального временного ряда, $N \in \mathbb{N}$ – длина временного ряда, x_{min} и x_{max} – соответственно минимальное и максимальное значение исходного временного ряда. Значения нормированного ряда $\{\tilde{x}_j\}_{j=1}^N$ принадлежат отрезку $[0,1]$. Выходное значение нейронной сети подвергается обратному преобразованию, в результате чего получается окончательное прогнозируемое значение:

$$y = \tilde{y}(x_{max} - x_{min}) + x_{min}. \quad (2)$$

На каждом нейроне скрытого слоя прогностического перцептрона действует функция активации. В данной работе рассмотрены следующие функции активации:

- rectifier linear unit (ReLU):

$$\text{ReLU}(x) = \max(0, x); \quad (3)$$

- leaky rectifier linear unit (LeakyReLU):

$$\text{LeakyReLU}(x) = \begin{cases} \alpha x, & x < 0, \\ x, & x \geq 0, \end{cases} \quad (4)$$

где $\alpha \ll 1$ – параметр, предотвращающий проблему исчезновения градиента, присущую функции ReLU [3]. В данной работе использовалось значение $\alpha = 0,3$;

- softplus:

$$\text{softplus}(x) = \ln(1 + e^x). \quad (5)$$

Обучается перцептрон по выборке, которая формируется по принципу скользящего окна, т. е. в качестве обучающей выборки используется множество векторов $\vec{x}^j = (\tilde{x}_j, \tilde{x}_{j+1}, \dots, \tilde{x}_{j+p-1})$, каждому из которых ставится в соответствие ожидаемый результат прогнозирования, равный \tilde{x}_{j+p} , $j = \overline{1, N-p}$. После обучения прогностического перцептрона возможно предсказание будущих значений временного ряда, например, по входным значениям $(\tilde{x}_{N-p+1}, \tilde{x}_{N-p+2}, \dots, \tilde{x}_N)$ получить на выходе нейронной сети оценку для \tilde{x}_{N+1} .

Обусловленность нейронной сети. Прогностический перцептрон можно рассматривать как векторный оператор $N: R_{p,1} \rightarrow R$. Одной из мер устойчивости векторных и матричных операторов является число обусловленности, которое для невырожденного оператора A определяется по следующей формуле:

$$\text{cond } A = \|A\| \|A^{-1}\|, \quad (6)$$

где $\|\cdot\|$ – норма оператора, подчиненная евклидовой векторной нормой $\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^p x_i^2}$, т. е.

$$\|A\| = \sup_{\|\vec{x}\|_2 \neq 0} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2}. \quad (7)$$

Оператор N можно представить в виде композиции следующих операторов $N_j, j = \overline{1, 7}$:

- нормировка первоначального вектора $\vec{a} \in R_{p,1}$ по формуле (1):

$$N_1: R_{p,1} \rightarrow R_{p,1}, N_1(\vec{a}) = \frac{1}{x_{max} - x_{min}} \vec{a} - \frac{x_{min}}{x_{max} - x_{min}} \vec{e}, \quad (8)$$

где $\vec{e} = (\underbrace{1, 1, \dots, 1}_p)^T$;

- умножение полученного вектора слева на матрицу весов $A_1 \in R_{n,1}$, у которой элемент i -ой строки, j -ого столбца равен весу синапса, соединяющего i -ый нейрон входного слоя с j -ым нейроном скрытого слоя:

$$N_2: R_{p,1} \rightarrow R_{n,1}, N_2(\vec{a}) = A_1 \vec{a};$$

- сложение полученного вектора с вектором смещения $\vec{b}^1 \in R_{n,1}$, у которого i -ая компонента равна весу синапса между i -ым нейроном смещения входного слоя с i -ым обычным нейроном скрытого слоя:

$$N_3: R_{n,1} \rightarrow R_{n,1}, N_3(\vec{a}) = \vec{a} + \vec{b}^1;$$

- функция активации, действующая на каждом нейроне скрытого слоя:

$$N_4: R_{n,1} \rightarrow R_{n,1}, N_4(\vec{a}) = \text{act}(a_1), \dots, \text{act}(a_p)^T,$$

где act – одна из функций активаций (3-5);

- умножение слева на матрицу весов $A_2 \in R_{1,n'}$, у которой i -ый элемент равен весу синапса, соединяющего i -ый нейрон скрытого слоя с нейроном выходного слоя:

$$N_5: R_{n,1} \rightarrow R, N_5(\vec{a}) = A_2 \vec{a};$$

- сложение с весом b синапса между нейроном смещения скрытого слоя и нейроном выходного слоя:

$$N_6: R \rightarrow R, N_6(a) = a + b;$$

- преобразование, обратное нормировке, которое осуществляется по формуле (2):

$$N_7: R \rightarrow R, N_7(a) = a(x_{max} - x_{min}) + x_{min}.$$

Число обусловленности оператора N можно оценить сверху произведением чисел обуслов-

ленности операторов $N_i, i = \overline{1,7}$ согласно свойству субмультипликативности нормы (7):

$$\text{cond}N = \prod_{i=1}^7 \text{cond}N_i.$$

Оценим числа обусловленности операторов $N_i, i = \overline{1,7}$. Вначале покажем, что $\text{cond}N_1 = 1$. Для этого воспользуемся следующей формулой [4]:

$$\text{cond}f = \sup_{|\bar{a}| \neq 0} \frac{\|J\|}{\|f(\bar{a})\|/\|\bar{a}\|}, \quad (9)$$

где J – матрица Якоби функции f , а матричная норма $\|J\|$ подчинена векторной норме, используемой в этой же формуле. Для функции, определённой формулой (8), матрица J равна

$$\begin{pmatrix} \frac{1}{x_{\max} - x_{\min}} & 0 & \dots & 0 \\ 0 & \frac{1}{x_{\max} - x_{\min}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{x_{\max} - x_{\min}} \end{pmatrix},$$

а её норма – $\frac{1}{x_{\max} - x_{\min}}$. После подстановки в (9) имеем

$$\begin{aligned} \text{cond}N_1 &= \sup_{|\bar{a}| \neq 0} \frac{\|J\|}{\|N_1(\bar{a})\|/\|\bar{a}\|} = \frac{1}{x_{\max} - x_{\min}} \sup_{|\bar{a}| \neq 0} \frac{\|\bar{a}\|}{\|N_1(\bar{a})\|} = \\ &= \frac{1}{x_{\max} - x_{\min}} \|N_1^{-1}\|. \end{aligned} \quad (10)$$

Для оператора $N_1^{-1}(\bar{a}) = (x_{\max} - x_{\min})\bar{a} + x_{\min}$ аналогично получаем

$$\text{cond}N_1^{-1} = (x_{\max} - x_{\min})\|N_1\|. \quad (11)$$

Перемножив равенства (10) и (11) и учитывая формулу (6), окончательно имеем $\|N_1\|\|N_1^{-1}\| = 1$, откуда и следует, что число обусловленности оператора N_1 равно 1. Таким же образом доказывается, что $\text{cond}N_i = 1$ при $i = 3, 6, 7$.

Для операторов N_2, N_5 числа обусловленности равны числам обусловленности матриц A_1 и A_2 соответственно, причём $\text{cond}A_2 = 1$, так как матрица $A_2 \in R_{1,n}$ имеет единственное сингулярное число. Оценивание числа обусловленности матрицы A_1 проводилось после обучения прогностического персептрона при различных гиперпараметрах и обучающих данных.

Рассмотрим оператор N_4 . Матрица Якоби для него имеет диагональный вид, следовательно, норма этой матрицы, подчинённая евклидовой векторной норме, равна наибольшему из

абсолютных значений диагональных элементов. Для функций ReLU и LeakyReLU имеем $\|J\| = 1$, т. к. на множестве положительных чисел $\text{ReLU}'(x) = \text{LeakyReLU}'(x) = 1$, а для отрицательных чисел $\text{ReLU}'(x) = 0$, $\text{LeakyReLU}'(x) = \alpha < 1$. В случае использования функции softplus в качестве функции активации получаем следующую оценку:

$$\|J\| \leq \sup_x \text{softplus}'(x) = \sup_x (\ln(1+e^x))' = \sup_x \frac{e^x}{1+e^x} = 1.$$

Используя формулу (9), имеем

$$\text{cond}N_4 = \sup_{|\bar{a}| \neq 0} \frac{\|J\|}{\|N_4(\bar{a})\|/\|\bar{a}\|} \leq \sup_{|\bar{a}| \neq 0} \frac{\|\bar{a}\|}{\|N_4(\bar{a})\|} \quad (12)$$

Так как на вход нейронной сети подаётся нормированный вектор с компонентами из отрезка $[0,1]$, для нормы вектора \bar{a} справедливо следующее неравенство:

$$\|\bar{a}\| = \|A\bar{x} + \bar{b}^{-1}\| \leq \|A\|\|\bar{x}\| + \|\bar{b}^{-1}\| \leq \sqrt{n}\|A\| + \|\bar{b}^{-1}\|. \quad (13)$$

Поэтому для оценки числа обусловленности оператора N_4 по формуле (12) можно использовать значение $\sup_{0 < |\bar{a}| \leq k} \frac{\|\bar{a}\|}{\|N_4(\bar{a})\|}$ при заданном $k > 0$.

При использовании функции активации ReLU это значение может принимать сколь угодно большие значения. Действительно, полагая

$$\bar{a} = \left(-\frac{k}{\sqrt{n}}, -\frac{k}{\sqrt{n}}, \dots, -\frac{k}{\sqrt{n}}, -\sqrt{\frac{2}{n}k^2 - \varepsilon^2}, \varepsilon\right), \varepsilon > 0, \text{ имеем } \|\bar{a}\| = \sqrt{\frac{n-2}{n}k^2 + \frac{2}{n}k^2 - \varepsilon^2 + \varepsilon^2} = k^2, \|N_4(\bar{a})\| = \|(0, 0, \dots, 0, \varepsilon)\| = \varepsilon,$$

а отношение норм этих векторов, равное k^2/ε при $\varepsilon \rightarrow 0$ стремится к бесконечности.

Для функции активации LeakyReLU это отношение можно оценить сверху нормой обратного оператора, и тогда из формулы (12) следует $\text{cond}N_4 \leq \|N_4^{-1}\|$. Сдругой стороны, $\text{cond}N_4^{-1} \leq \frac{1}{\alpha}\|N_4\|$, т. к. для оператора N_4^{-1} матрица Якоби J содержит диагональные элементы, равные 1 и $1/\alpha$, а $\alpha < 1$. Перемножая два неравенства, имеем

$$\text{cond}N_4 \text{cond}N_4^{-1} = (\|N_4\|\|N_4^{-1}\|)^2 \leq \frac{1}{\alpha}\|N_4\|\|N_4^{-1}\|,$$

откуда следует $\text{cond}N_4^{-1} \leq \frac{1}{\alpha}$. Значение $1/\alpha$ достигается для векторов \bar{a} , у которых все компоненты отрицательны:

$$\begin{aligned} \frac{\|\bar{a}\|}{\|N_4(\bar{a})\|} &= \frac{\|(a_1, \dots, a_n)\|}{\|(\text{LeakyReLU}(a_1), \dots, \text{LeakyReLU}(a_n))\|} = \\ &= \frac{\|(a_1, \dots, a_n)\|}{\|(aa_1, \dots, aa_n)\|} = \frac{1}{\alpha} \end{aligned}$$

Для функции активации softplus вначале оценим $\inf_{\|\vec{a}\|=k_1} \|N_4(\vec{a})\|$ при заданном значении нормы $0 < k_1 \leq k$:

$$\inf_{\|\vec{a}\|=k_1} \|N_4(\vec{a})\| = \inf_{\|\vec{a}\|=k_1} (\ln(1 + e^{a_1}), \ln(1 + e^{a_2}), \dots, \ln(1 + e^{a_n})) = \inf_{\|\vec{a}\|=k_1} \sqrt{\sum_{i=1}^n \ln^2(1 + e^{a_i})}. \quad (14)$$

Следовательно, необходимо решить следующую задачу нахождения условного минимума:

$$\begin{cases} \text{vf}(\vec{a}) = \sqrt{\sum_{i=1}^n \ln^2(1 + e^{a_i})} \rightarrow \min, \\ \sum_{i=1}^n a_i^2 = k_1^2. \end{cases}$$

Так как квадратный корень – это возрастающая функция на всей своей области определения, то минимум функции $\text{vf}(\vec{a})$ будет достигаться в той же точке, что и минимум функции $f(\vec{a}) = \sum_{i=1}^n \ln^2(1 + e^{a_i})$. Также отметим, что условный минимум этой функции будет достигаться при $a_i < 0, i = \overline{1, n}$, так как функция $f(\vec{a})$ является возрастающей по каждому аргументу на всей своей области определения.

Функция Лагранжа для функции f имеет

$$\text{вид } L(\vec{a}, \lambda) = \sum_{i=1}^n \ln^2(1 + e^{a_i}) + \lambda (\sum_{i=1}^n a_i^2 - k_1^2) = \sum_{i=1}^n (\ln^2(1 + e^{a_i}) + \lambda a_i^2) - \lambda k_1^2. \text{ Её частные производные } \frac{\partial L}{\partial a_i} = 2 \ln(1 + e^{a_i}) \frac{e^{a_i}}{1 + e^{a_i}} + 2\lambda a_i \text{ равны нулю при}$$

$$\lambda = -\frac{\ln(1 + e^{a_i}) e^{a_i}}{(1 + e^{a_i}) a_i}, i = \overline{1, n}. \quad (15)$$

$$\text{Докажем, что функция } \lambda(\alpha) = -\frac{\ln(1 + e^\alpha) e^\alpha}{(1 + e^\alpha) \alpha}$$

возрастает на интервале $(-\infty, 0)$. Действительно,

$$\text{если } \alpha < \beta < 0, \text{ то отношение } \frac{\lambda(\alpha)}{\lambda(\beta)} = \frac{\ln(1 + e^\alpha) e^\alpha}{(1 + e^\alpha) \alpha} \times \frac{(1 + e^\beta) \beta}{\ln(1 + e^\beta) e^\beta} = \frac{\beta}{\alpha} \frac{e^\beta + 1}{e^\alpha + 1} \frac{\ln(1 + e^\alpha)}{\ln(1 + e^\beta)} < 1. \text{ С учётом}$$

того, что $\lambda(\beta) > 0$, окончательно имеем $\lambda(\alpha) < \lambda(\beta)$.

Из монотонности функции $\lambda(\alpha)$ и формулы (15) следует, что все a_i должны быть равны. Значит, минимум функции $\text{vf}(\vec{a})$ достигается при $\vec{a} = (-\frac{k_1}{\sqrt{n}}, \dots, -\frac{k_1}{\sqrt{n}})$ и равен $\sqrt{n \ln^2(1 + e^{-k_1/\sqrt{n}})} = \sqrt{n} \ln(1 + e^{-k_1/\sqrt{n}})$. Таким образом, $\inf_{\|\vec{a}\|=k_1} \|N_4(\vec{a})\| = \sqrt{n} \ln(1 + e^{-k_1/\sqrt{n}})$.

Вычисляя значение отношения $\sup_{\|\vec{a}\|=k} \frac{\|\vec{a}\|}{\|N_4(\vec{a})\|}$ получим $\frac{k_1}{\sqrt{n} \ln(1 + e^{-k_1/\sqrt{n}})}$. Это выражение является

возрастающей по k_1 функцией, следовательно,

$$\text{cond} N_4 \leq \sup_{0 < \|\vec{a}\| \leq k} \frac{\|\vec{a}\|}{\|N_4(\vec{a})\|} = \frac{k}{\sqrt{n} \ln(1 + e^{-k_1/\sqrt{n}})}. \quad (16)$$

Таким образом, удалось получить оценки сверху для числа обусловленности оператора N_4 в случае использования функций активации LeakyReLU и softplus. Для функции LeakyReLU эта оценка равна $1/\alpha$ и не зависит от других гиперпараметров нейронной сети. Для функции softplus оценка представляет собой возрастающую функцию от числа k/\sqrt{n} , которое, в свою очередь, согласно формуле (13) можно оценить сверху числом, равным $\|A\| + \|\vec{b}^1\|/\sqrt{n}$, то есть оценка числа обусловленности оператора N_4 зависит от норм весовых матриц, а также от количества нейронов на скрытом слое. Более точную оценку числа обусловленности оператора N_4 при использовании любой из функций (3-5) можно получить, используя формулу (12).

Вычислительный эксперимент и его результаты. Тестирование прогностического перцептрона осуществлялось на примере временного ряда $\vec{x} = (x_1, x_2, \dots, x_N)$, смоделированного по следующей формуле:

$$x_t = t^\alpha + \beta \sin \gamma t, t = \overline{1, N}.$$

Также рассматривались временные ряды $\vec{x}^j = \vec{x} + \xi^j$, где $\xi^j = (\xi_1^j, \xi_2^j, \dots, \xi_N^j)$, $\xi_t^j \in N(0, \sigma_j^2)$ – случайные величины, распределённые по нормальному распределению с нулевым математическим ожиданием и заданными значениями дисперсии σ_j^2 . Обучение прогностического перцептрона проводилось при различных значениях p, n , а также при функциях активации, заданных формулами (3-5), в течение 7 эпох. В качестве функции оптимизации во время обучения использовался алгоритм AdaGrad [5], который показал лучшие результаты по устойчивости нейронной сети [6]. Числа обусловленности функций активации оценивались согласно формуле (12), где компоненты вектора \vec{a} равны значениям, поступившим на вход нейронов скрытого слоя, а $N_4(\vec{a})$ – вектор из значений на выходе скрытого слоя. Оценка числа обусловленности нейронной сети вычислялась по формуле $\text{cond} N = \text{cond} A_1 \cdot \text{cond} N_4$.

Результаты по обусловленности прогностического перцептрона и его компонент для $N = 1000, \alpha = 1, 2, \beta = 100, \gamma = 0,05$ при различных значениях параметров p, n, σ_j приведены на рисунках 2-4 (в подписи горизонтальной оси каждой гистограммы сверху вниз указаны значения пара-

метров σ_j , n , p , на рисунках 3-4 используется логарифмическая шкала для вертикальной оси).

Как видно на рисунке 2, обусловленность матрицы A_1 зависит преимущественно от количества нейронов на входном и скрытом слоях, причём при увеличении числа нейронов на входном слое матрица A_1 становится хуже обусловленной, при увеличении числа нейронов на скрытом слое – лучше обусловленной. Влияние функции активации скрытого слоя на матрицу A_1 незначительно.

На рисунке 3 видно, что наибольшее число обусловленности оператора N_4 даёт использование функции активации ReLU, наименьшее – softplus. Влияние выбора функции активации особенно велико при малых значениях гиперпараметра p и временных рядах с малой дисперсией, а при $p = 15$ функции ReLU и LeakyReLU дают почти равные числа обусловленности оператора N_4 вне зависимости от дисперсии σ_j обучающего временного ряда. Таким образом, проблема исчезающего градиента оказывает сильное воздействие на численную обусловленность нейронной сети. Следует отметить, что полученные на практике оценки числа обусловленности, представленные на рисунке 3, в некоторых случаях заметно меньше ранее полученных теоретических оценок. Действительно, для функции LeakyReLU эта оценка составила $1/\alpha \approx 3,33$, для функции softplus, согласно формулам

(13) и (16), число обусловленности оператора N_4 можно оценить выражением

$$\frac{\|A\| + \|b^{-1}\| / \sqrt{n}}{\ln(1 + e^{-|A| - \|b^{-1}\| / \sqrt{n}})},$$

которое для рассмотренных значений гиперпараметров лежит в промежутке от 7,7 до 15,3. На практике же числа обусловленности для функции LeakyReLU составили от 1,5 до 3,1, для функции softplus – от 0,15 до 0,5.

Из гистограммы на рисунке 4 можно сделать вывод о том, что значительное влияние на обусловленность прогностического персептрона оказывает как количество нейронов на входном и скрытом слоях, так и функция активации. Так, лучшую обусловленность нейронной сети обеспечивает функция активации softplus по сравнению с функциями ReLU и LeakyReLU, а также большее число нейронов на скрытом слое. Что касается числа нейронов на входном слое, то его увеличение улучшает обусловленность прогностического персептрона, если используется функция ReLU либо LeakyReLU, и ухудшает, если функция активации персептрона – softplus. При $p = 10$ и $p = 15$ функции ReLU и LeakyReLU дают примерно равную обусловленность нейронной сети. Дисперсия исходных временных рядов оказывает заметное влияние на обусловленность только при использовании функций ReLU и LeakyReLU, а также при

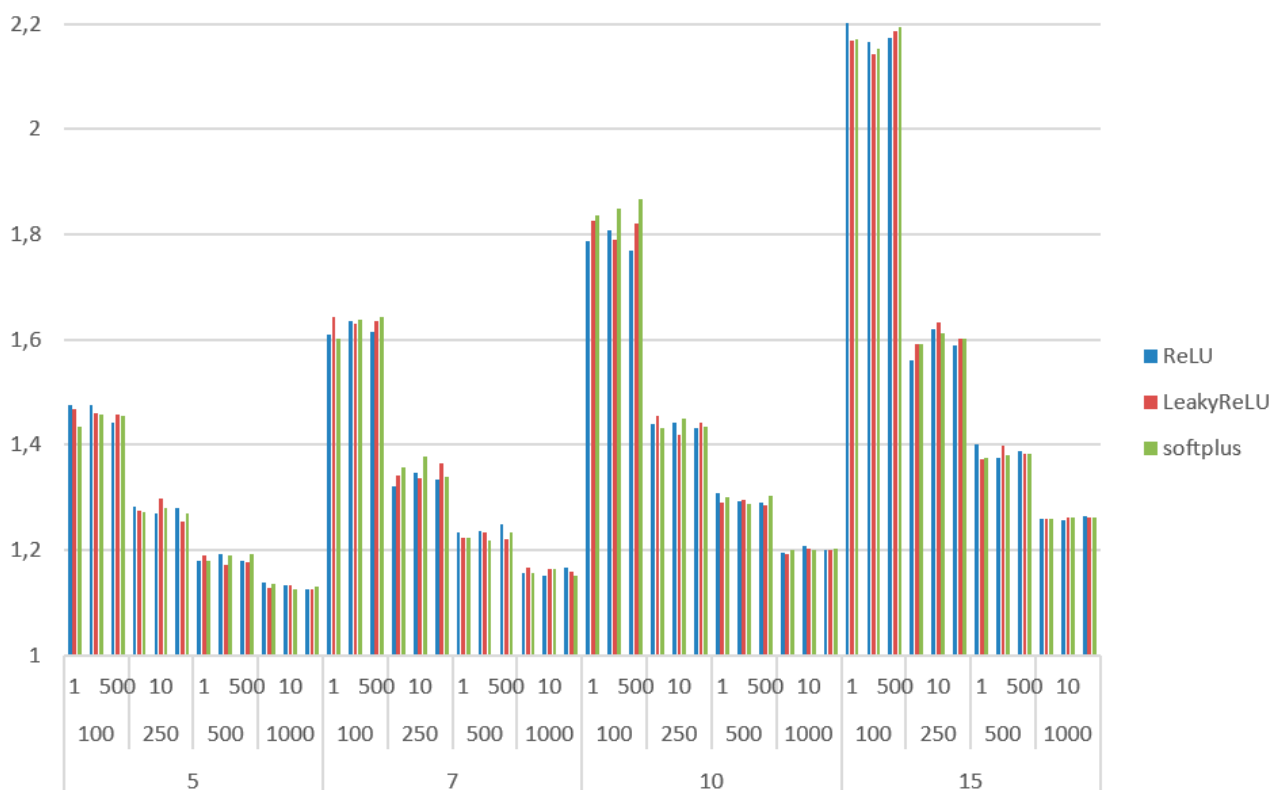


Рис. 2. Зависимость числа обусловленности матрицы A_1 от гиперпараметров прогностического персептрона
 Fig. 2. Dependence of condition number of the matrix A_1 on forecasting perceptron hyperparameters

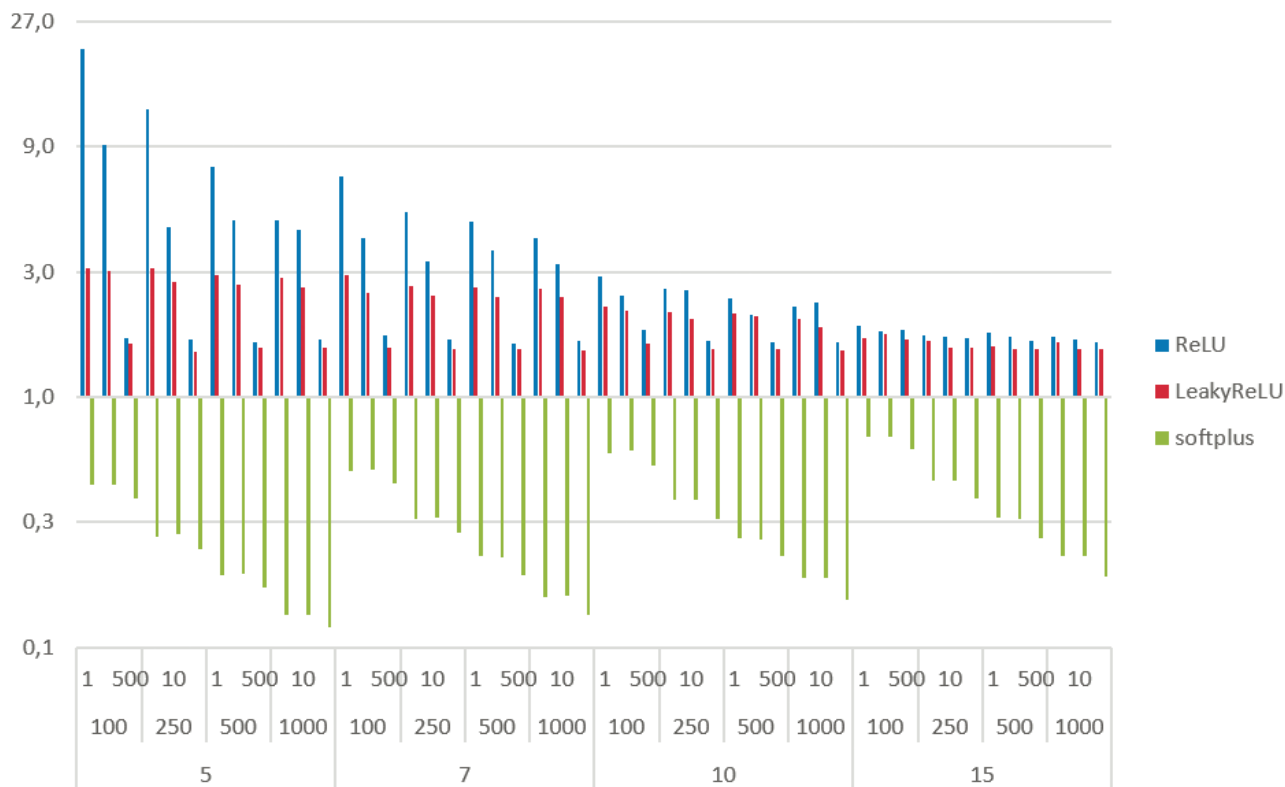


Рис. 3. Зависимость числа обусловленности оператора N_4 от гиперпараметров прогностического персептрона
 Fig. 3. Dependence of condition number of the operator N_4 on forecasting perceptron hyperparameters

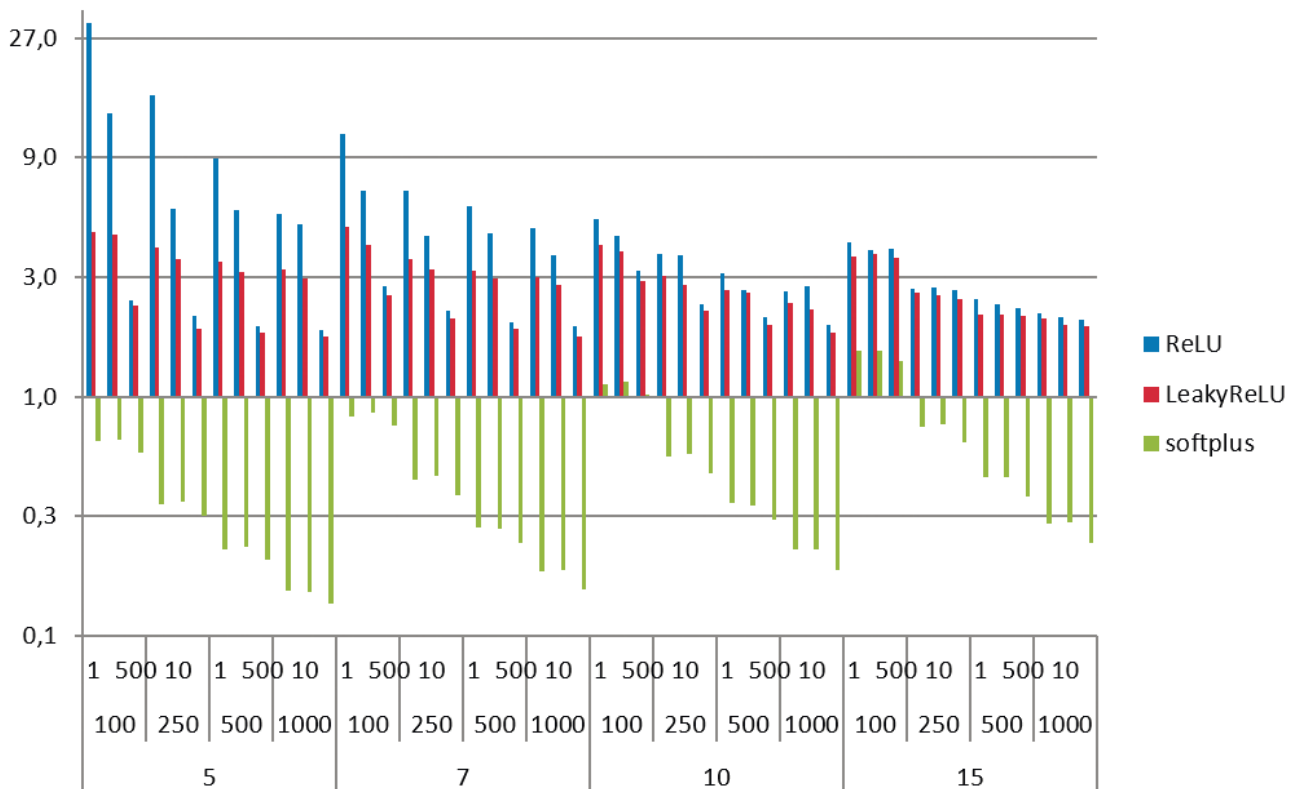


Рис. 4. Зависимость числа обусловленности прогностического персептрона от его гиперпараметров
 Fig. 4. Dependence of condition number of the forecasting perceptron on its hyperparameters

малых значениях p и n . Таким образом, можно судить о большей устойчивости прогностического персептрона при большем количестве нейронов на входном и скрытом слоях, а также при использовании функции активации softplus.

Заключение. Результаты обучения прогностического персептрона свидетельствуют о том, что численная обусловленность прогностического персептрона в первую очередь зависит от функции активации, действующей на скрытом слое. Так, функция softplus даёт лучшую обусловленность по сравнению с функциями ReLU и LeakyReLU. Кроме того, нейронная сеть с функцией активации softplus в меньшей степени подвержена влиянию погрешностей в исходном временном ряде. Это влияние можно минимизировать также путём увеличения количества нейронов на входном и скрытом слое.

Также значительным является количество нейронов прогностического персептрона. Так, увеличение числа нейронов на скрытом слое

уменьшает число обусловленности нейронной сети, а увеличение числа нейронов на входном слое, в зависимости от используемой функции активации, может как увеличить, так и уменьшить число обусловленности.

Получены теоретические и практические оценки числа обусловленности функций активации при различных значениях гиперпараметров. Функция softplus оказалась лучше обусловленной по сравнению с функциями ReLU и LeakyReLU. Значения оценок числа обусловленности, полученные на практике, оказались значительно меньше теоретических оценок. Это можно объяснить тем, что теоретические оценки достигаются только для векторов специального вида, которые могут получиться для ограниченного множества данных, подаваемых на вход нейронной сети. Тем не менее, теоретические оценки могут дать некоторое первоначальное представление об обусловленности рассматриваемой нейронной степени, для которого не требуется проводить её обучение.

Список литературы

1. Sengupta, B. How Robust are Deep Neural Networks [Electronic resource] / B. Sengupta, K.J. Friston // arXiv.org e-Print archive – Mode of access: <https://arxiv.org/abs/1804.11313>. – Date of access: 02.02.2020. – (Preprint / arXiv:1804.11313).
2. Goodfellow, I.J. Explaining and Harnessing Adversarial Examples [Electronic resource] / I.J. Goodfellow, J. Shlens, C. Szegedy // International Conference on Learning Representations: proceedings of 3rd International Conference, San Diego, 7-9 May 2015 // arXiv.org e-Print archive – Mode of access: <https://arxiv.org/abs/1412.6572>. – Date of access: 02.02.2020. – (Preprint / arXiv:1412.6572v3).
3. Maas, A.L. Rectifier Nonlinearities Improve Neural Network Acoustic Models // A.L. Maas, A.Y. Hannun, A.Y. Ng // International Conference on Machine Learning: proceedings of 30th International Conference, Atlanta, 16-21 June 2013 // Stanford Artificial Intelligence Laboratory – Mode of access: https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf. – Date of access: 02.02.2020.
4. Trefethen, L. N. Numerical Linear Algebra / L.N. Trefethen, D. Bau. – Philadelphia : Society for Industrial and Applied Mathematics, 1997. – 390 p.
5. Duchi, J. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization / J. Duchi, E. Hazan, Y. Singer // Journal of Machine Learning Research – 2011. – Vol. 12 – P. 2121–2159.
6. Шолтанюк, С. В. Сравнительный анализ нейросетевой и регрессионных моделей прогнозирования временных рядов / С. В. Шолтанюк // Цифровая трансформация. – 2019. – № 2 (7). – С. 60–68.

References

1. B. Sengupta, K.J. Friston. How Robust are Deep Neural Networks? arXiv preprint arXiv:1804.11313, 2018.
2. I.J. Goodfellow, J. Shlens, C. Szegedy. Explaining and Harnessing Adversarial Examples. International Conference on Learning Representations, arXiv:1412.6572, 2015.
3. A.L. Maas, A.Y. Hannun, A.Y. Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. International Conference on Machine Learning, 2013.
4. L. N. Trefethen, D. Bau. Numerical Linear Algebra, Philadelphia, Society for Industrial and Applied Mathematics, 1997, 390 p.
5. J. Duchi, E. Hazan, Y. Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research, 2011, vol. 12, pp. 2121-2159.
6. Sholtanyuk S.V. Comparative Analysis of Neural Networking and Regression Models for Time Series Forecasting. Цифровая трансформация [Digital transformation], 2019, 2 (7), pp. 60–68 (in Russian).

Received: 02.02.2020

Поступила: 02.02.2020