

# Нова Модель Штучного Нейрону для Побудови Нейронних Мереж у Задачі Апроксимації

Олена Чумаченко  
кафедра технічної кібернетики  
НТУУ «КПІ імені Ігоря Сікорського»  
Київ, Україна  
lobach21@mail.ru

Владислав Горбатюк  
кафедра технічної кібернетики  
НТУУ «КПІ імені Ігоря Сікорського»  
Київ, Україна  
vladislav.horbatiuk@gmail.com

## New Artificial Neuron Model for Building Neural Networks in Approximation Tasks

Olena Chumachenko  
dept. of technical cybernetics  
NTUU «Igor Sikorsky KPI»  
Kyiv, Ukraine  
lobach21@mail.ru

Vladyslav Gorbatiuk  
dept. of technical cybernetics  
NTUU «Igor Sikorsky KPI»  
Kyiv, Ukraine  
vladislav.horbatiuk@gmail.com

**Анотація**—Запропоновано нову модель штучного нейрону, що певним чином узагальнює та вдосконалює відомі існуючі моделі сімейства ReLU: ReLU, PReLU та maxout моделі. Проаналізовані теоретичні переваги нової моделі по відношенню до вказаних існуючих. Проведено експериментальні порівняльні тести на реальних вибірках, які підтвердили апроксимуючу ефективність нової моделі.

**Abstract**—A new artificial neuron model is presented, which generalizes and in certain way improves known models that belong to ReLU “family”: ReLU, PReLU and maxout models. Its theoretical advantages in regards to mentioned known models are analyzed. Experimental comparison tests on real data sets were conducted, confirming the approximation effectiveness of the suggested model.

**Ключові слова**—апроксимація; нейронні мережі; штучний нейрон; ReLU нейрон

**Keywords**—approximation; neural networks; artificial neuron; ReLU neuron

### I. ВСТУП

Математична задача апроксимації невідомої функції за наявною вибіркою потенційно зашумлених спостережень виникає при вирішенні великої кількості практичних задач – прогнозування часових рядів, класифікація зображень, розпізнавання мови – усі ці задачі в тому чи іншому вигляді зводяться до задачі апроксимації.

Останнім часом найкращих результатів у цій області досягають нейромережеві підходи. При цьому, саме

використання ReLU нейронів [1] дозволило значно поліпшити найкращу якість класифікації мережі у змаганні ImageNet. У даній роботі пропонується нова модель нейрону Sigmoid Piecewise (SP), яка певним чином узагальнює та вдосконалює моделі нейронів «сімейства» ReLU (ReLU, leaky ReLU [2], PReLU [3], maxout [4]).

### II. ОПИС МОДЕЛІ НОВОГО НЕЙРОНУ

Неперервну функцію багатьох змінних  $f: R^n \rightarrow R$  можна з будь-якою заданою точністю апроксимувати кусочно-лінійною функцією, що розбиває простір  $R^n$  на опуклі множини  $S_1, \dots, S_N \subseteq R^n$ , які не перетинаються, і на кожній множині задає деяку лінійну функцію  $f_i(\vec{x}) = \vec{w}_i^T \vec{x}, i = 1, \dots, N, \vec{w}_i \in R^n$ .

Розглянемо модель простого кусочно-лінійного нейрону як базового елемента для побудови таких складних кусочно-лінійних функцій у задачі апроксимації:

$$PL(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h}) = \begin{cases} \vec{w}_+^T \vec{x}, \vec{h}^T \vec{x} \geq 0 \\ \vec{w}_-^T \vec{x}, \vec{h}^T \vec{x} < 0 \end{cases}, \vec{w}_+, \vec{w}_-, \vec{h} \in R^n.$$

Вектори параметрів  $\vec{w}_+, \vec{w}_-, \vec{h}$  мають наступний зміст:

- вектор  $\vec{h}$  задає гіперплощину, що розділяє простір  $R^n$  на 2 півпростори;

- вектор  $\vec{w}_+$  задає ваги кусочно–лінійної функції PL на півпросторі, де  $\vec{h}^T \vec{x} \geq 0$ ;
- вектор  $\vec{w}_-$  задає ваги кусочно–лінійної функції PL на півпросторі, де  $\vec{h}^T \vec{x} < 0$ .

Очевидно, що функція PL не є диференційовною в точках  $x : \vec{h}^T \vec{x} = 0$ . Якщо використовувати нейрони з такою моделлю у задачах апроксимації за навчальною вибіркою, застосування градієнтних методів для пошуку оптимальних значень параметрів буде дещо ускладненим. Для уникнення цієї проблеми пропонується наступна функція–модель нейрону:

$$SP(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h}) = \frac{\vec{w}_+^T \vec{x}}{1 + e^{-k\vec{h}^T \vec{x}}} + \frac{\vec{w}_-^T \vec{x}}{1 + e^{k\vec{h}^T \vec{x}}}; k > 0,$$

де, для усіх векторів  $\vec{x}$ , таких, що  $\vec{h}^T \vec{x} \neq 0$ , маємо:

$$k \rightarrow \infty : SP(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h}) \rightarrow PL(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h}).$$

Хоча функція SP «стає» функцією PL лише у ліміті, уже навіть при невеликих значеннях  $k$  вона дуже добре апроксимує цю функцію. У цьому можна впевнитися, записавши функцію PL наступним чином:

$$PL(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h}) = \vec{w}_+^T \vec{x} * step'(\vec{h}^T \vec{x}) + \vec{w}_-^T \vec{x} * step''(-\vec{h}^T \vec{x}),$$

де:

$$step'(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}, step''(x) = \begin{cases} 1, x > 0 \\ 0, x \leq 0 \end{cases}.$$

Власне, функція SP відрізняється від функції PL тим, що замість  $step'(\vec{h}^T \vec{x})$  та  $step''(\vec{h}^T \vec{x})$  використовуються множники  $\frac{1}{1 + e^{-k\vec{h}^T \vec{x}}}$  та  $\frac{1}{1 + e^{k\vec{h}^T \vec{x}}}$  відповідно. Порівняємо графіки функцій  $step'(x)$  та  $sigm(x; k)$  при  $k = 10$ :

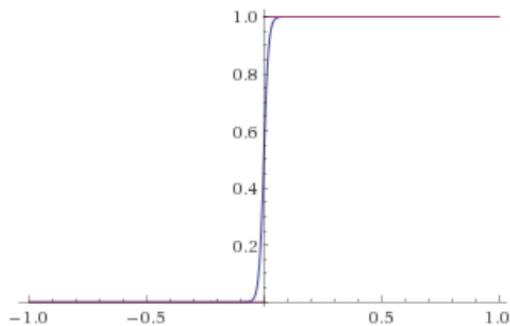


Рис. 1. Графіки функцій  $step'(x)$  та  $sigm(x, k=10)$

Як бачимо, апроксимація справді доволі точна навіть при відносно невеликому значенні параметру  $k$ .

Структурно нейрон з моделлю SP можна представити наступним чином:

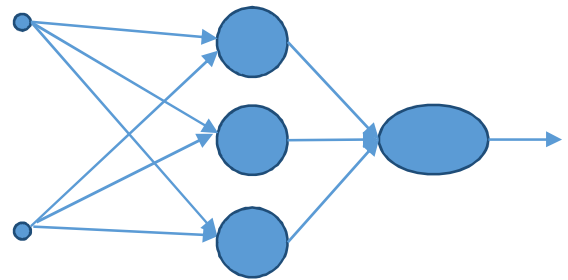


Рис. 2. Структура нейрону SP

Як відомо, класичний штучний нейрон [6] складається з 2 частин: блоку–зваженого суматору  $S(\vec{x}; \vec{w}) = \vec{w}^T \vec{x}$  та функції активації  $AF(x; \vec{\theta})$ , на вхід якої подається вихід суматору, і разом ці 2 частини описують модель нейрону вигляду  $f(\vec{x}; \vec{w}, \vec{\theta}) = AF(S(\vec{x}; \vec{w}); \vec{\theta})$ . У випадку нейрону з математичною моделлю SP структура дещо складніша: маємо 3 блоки суматори –  $S_+(\vec{x}; \vec{w}_+)$ ,  $S_-(\vec{x}; \vec{w}_-)$ ,  $S_h(\vec{x}; \vec{h})$  та, відповідно, функцію активації, що залежить від трьох змінних –  $AF(S_+, S_-, S_h; k) = \frac{S_+}{1 + e^{-kS_h}} + \frac{S_-}{1 + e^{kS_h}}$ , і використовуючи ці частини повна модель нейрону може бути записана як:

$$SP(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h}) = AF(S_+(\vec{x}; \vec{w}_+), S_-(\vec{x}; \vec{w}_-), S_h(\vec{x}; \vec{h}); k).$$

Нажаль, через те, що функція активації  $AF(S_+, S_-, S_h; k)$  є функцією від 3 змінних, неможливо у загальному випадку зобразити її у вигляді двовимірного або тривимірного графіку. Однак можна це зробити для певних часткових випадків, наприклад якщо умовитися, що  $S_- = CS_+$ , де  $C$  – деяка константа. Ось як виглядає графік функції активації при  $C = 5$  та  $k = 2$ :

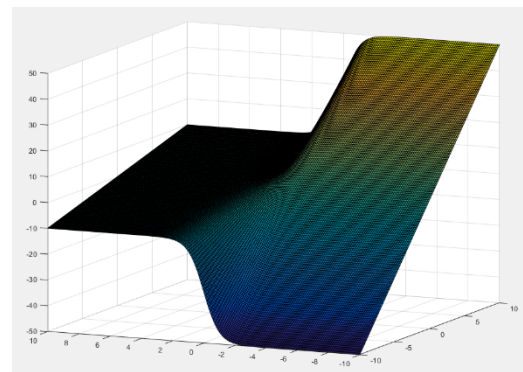


Рис. 3. Поверхня функції активації SP нейрону при  $C = 5$  та  $k = 2$

### III. ПОРІВНЯННЯ З «РОДИНОЮ» ReLU НЕЙРОНІВ

У роботі [7] був запропонований нейрон з моделлю вигляду  $ReLU(\vec{x}; \vec{w}) = \max(0, \vec{w}^T \vec{x})$ , тобто нейрону, що

складається зі зваженого суматора та функції активації, графік якої має наступний вигляд:

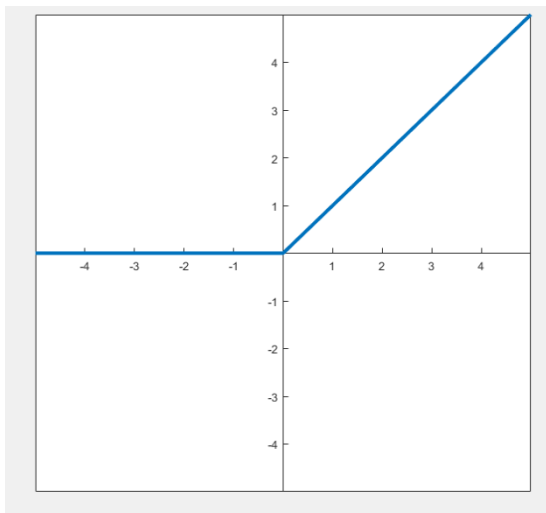


Рис. 4. Графік функції активації ReLU нейрону

Модель нейрону ReLU і деяких його модифікацій – Parametric ReLU (PReLU) нейрону, тахout нейрону є частковими випадками більш загальної моделі PL :

- $\text{ReLU}(\vec{x}; \vec{w}) = \text{PL}(\vec{x}; \vec{w}_+ = \vec{w}, \vec{w}_- = \vec{0}, \vec{h} = \vec{w})$ ,
- $\text{PReLU}(\vec{x}; \vec{w}, a) = \text{PL}(\vec{x}; \vec{w}_+ = \vec{w}, \vec{w}_- = a * \vec{w}, \vec{h} = \vec{w})$ ,
- $\text{taхout}(\vec{x}; \vec{w}_1, \vec{w}_2) = \text{PL}(\vec{x}; \vec{w}_+ = \vec{w}_1, \vec{w}_- = \vec{w}_2, \vec{h} = \vec{w}_1 - \vec{w}_2)$

У свою чергу, функція  $\text{SP}(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h})$  при достатньо великому значенні  $k$  може точно апроксимувати функцію  $\text{piecewise\_linear}(\vec{x}; \vec{w}_+, \vec{w}_-, \vec{h})$  – отже модель SP є більш загальною, ніж ReLU, PReLU, та тахout, однак при цьому вона має більше параметрів, які потрібно налаштувати.

Іншою перевагою нейрону SP є можливість окремо налаштувати вектор параметрів, що задає роздільну гіперплощину і окремо – лінійні трансформації для кожного підпростору; в усіх нейронах «сімейства» ReLU є певний зв'язок між цими 3 векторами.

#### IV. ОЦІНКА ЕФЕКТИВНОСТІ НОВОГО НЕЙРОНУ НА РЕАЛЬНИХ ВИБІРКАХ

Виконаємо порівняльний тест ефективності нового нейрону у задачі апроксимації на реальних вибірках. Порівнювати будемо з нейроном ReLU. При заданій вибірці  $\langle X, y : X \rightarrow R \rangle$  порівняльний тест складається з наступних кроків:

1. На першій ітерації тесту навчається 2 повнозв'язні мережі прямого поширення з одним прихованим шаром з 1 нейроном, та 1 вихідним лінійним нейроном: мережа з нейроном ReLU у прихованому шарі, та мережа з нейроном SP у прихованому шарі.

2. Якщо помилка обох мереж  $< MSE_{target}$  – виконання тесту закінчується.

3. Після  $i$  ітерацій маємо мережу з  $i$  нейронами у прихованому шарі.

4. На ітерації  $i+1$  до прихованого шару кожної мережі додається ще один нейрон відповідного типу, після чого повторно виконується навчання обох мереж, але усі параметри мереж, крім параметрів нових нейронів та їх ваг у вихідному лінійному нейроні, фіксуються.

5. Для кожної мережі розраховується  $i$  запам'ятовується поточна помилка  $MSE_{i+1}$  – тобто помилка відповідної мережі, досягнута при використанні  $i+1$  нейронів.

6. Якщо на ітерації  $i+1$  помилка обох мереж стала  $< MSE_{target}$  – виконання тесту зупиняється.

7. Після зупинки тесту маємо дані вигляду «кількість нейронів у прихованому шарі»  $\rightarrow$  «досягнуте значення помилки апроксимації».

Перший порівняльний тест було виконано на часовому ряді денних значень курсу USD до EUR:

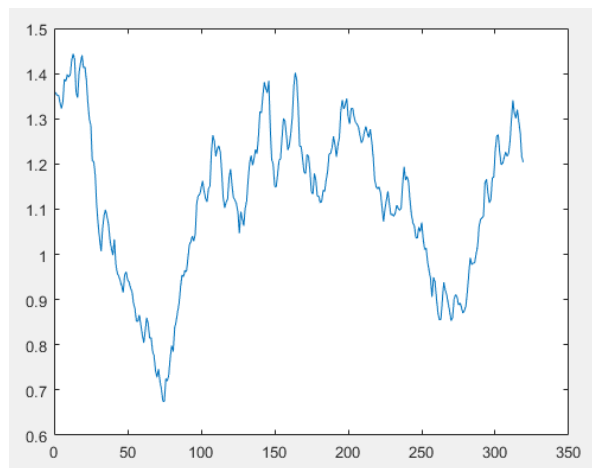


Рис. 5. Графік даних, що використовувалися для першого порівняльного тесту

Навчальна вибірка  $\langle X, y : X \rightarrow \mathfrak{R} \rangle$  була побудована з вихідного часового ряду методом вкладення часових рядів з розмірністю вкладення  $m = 5$  та горизонтом прогнозування  $k = 4$ . Середньоквадратична помилка наївної моделі вигляду  $\hat{g}_{i+4} = g_i$  дорівнює  $MSE_{naive} = 7.16 * 10^{-2}$ .

Задамо рівень середньоквадратичної помилки, який потрібно досягнути  $MSE_{target} = 5 * 10^{-2}$ , виконаємо описаний тест і побудуємо графік залежності кількості нейронів від середньоквадратичної помилки мережі:

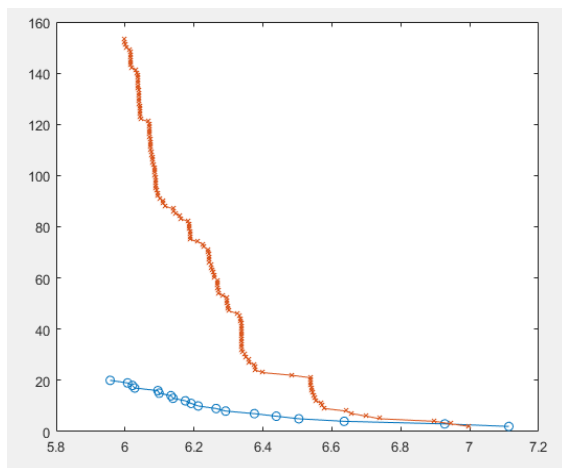


Рис. 6. Графік залежності кількості нейронів від досягнутого рівня помилки апроксимації у першому порівняльному тесті. Помаранчевий графік – для мережі з ReLU нейронами, синій – для мережі з SP нейронами

Починаючи приблизно зі значення помилки  $MSE_{threshold} \approx 6.5 * 10^{-2}$  для досягнення значення помилки було потрібно в 3 рази більше нейронів типу ReLU, ніж нейронів типу SP, після чого розрив у потрібній кількості нейронів тільки збільшувався, що свідчить про більшу ефективність нейрону типу SP з точки зору критерію «кількості параметрів, потрібних для досягнення заданого рівня помилки», якщо рівень помилки достатньо малий.

Другий тест було виконано на вибірці середніх процентних ставок за державними облігаціями в Австралії, 1969–1994 роки:

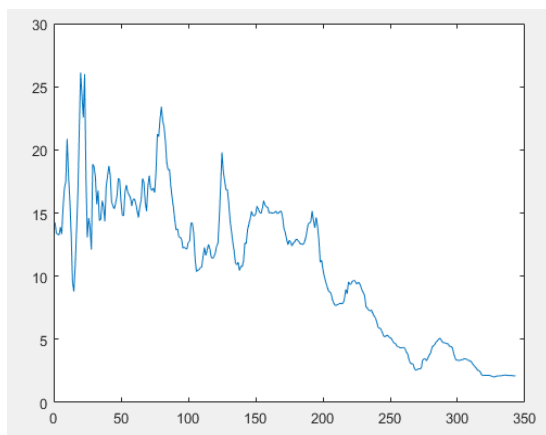


Рис. 7. Графік даних, що використовувалися для другого порівняльного тесту

Після виконання аналогічних кроків для отримання навчальної вибірки  $\langle X, y: X \rightarrow R \rangle$  помилка наївної моделі склала  $MSE_{naive} = 2.9157$ . Задамо помилку, яку потрібно досягнути  $MSE_{target} = 1.5$ , побудуємо аналогічні графіки:

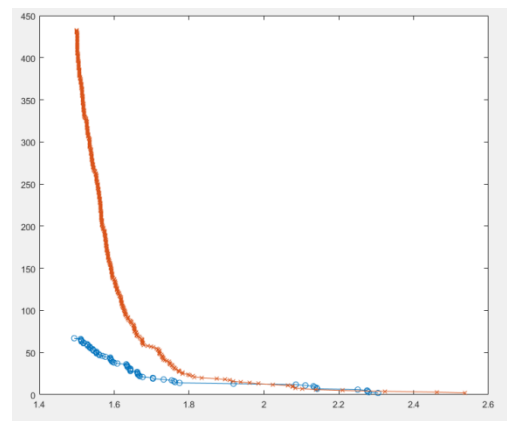


Рис. 8. Графік залежності кількості нейронів від досягнутого рівня помилки апроксимації у першому порівняльному тесті. Помаранчевий графік – для мережі з ReLU нейронами, синій – для мережі з SP нейронами

## ВИСНОВКИ

Нейронні мережі показують найкращі результати у задачі апроксимації функції за вибіркою спостережень. Великий вплив на апроксимуючі здібності мережі мають математична моделі нейронів, з яких вона складається. У роботі запропоновано принципово нову модель нейрону Sigmoid Piecewise (SP), яка є узагальненням існуючих моделей сімейства ReLU нейронів – ReLU, PReLU, leaky ReLU та maxout. Нейрон з моделлю SP має перевагу через можливість незалежно налаштувати вектор–нормаль до роздільної площини нейрону і вектори–параметри лінійних функцій, які будуть застосовуватися до векторів входів з різних напівпросторів, утворених роздільною площиною. Ефективність нової моделі було перевірено на реальних вибірках, де він показав кращі апроксимуючі властивості по відношенню до ReLU нейрону при достатньо низькому значенні помилки апроксимації, яку потрібно досягнути.

## ЛІТЕРАТУРА REFERENCES

- [1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097–1105. 2012.
- [2] Maas, Andrew L., Awni Y. Hannun, and Andrew Y. Ng. "Rectifier nonlinearities improve neural network acoustic models." In Proc. icml, vol. 30, no. 1, p. 3. 2013.
- [3] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." In Proceedings of the IEEE international conference on computer vision, pp. 1026–1034. 2015.
- [4] Goodfellow, Ian J., David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. "Maxout networks." arXiv preprint arXiv:1302.4389 (2013).
- [5] Hochreiter, Sepp. "The vanishing gradient problem during learning recurrent neural nets and problem solutions." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6, no. 02 (1998): 107–116.
- [6] McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." The bulletin of mathematical biophysics 5, no. 4 (1943): 115–133.
- [7] Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." In Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814. 2010.