

ВИЗНАЧЕННЯ ОПТИМАЛЬНИХ ПАРАМЕТРІВ НЕЙРОМЕРЕЖІ У ЗАДАЧАХ БІНАРНОЇ КЛАСИФІКАЦІЇ ПРИ ВИКОРИСТАННІ МАЛИХ ОБСЯГІВ ДАНИХ

DETERMINATION OF OPTIMAL NEURAL NETWORK PARAMETERS IN BINARY CLASSIFICATION PROBLEMS WHEN USING SMALL AMOUNTS OF DATA

Піскунова О. В.

Доктор економічних наук, професор,
професор кафедри математичного моделювання та статистики,
Київський національний економічний університет
імені Вадима Гетьмана,
м. Київ, Україна
ORCID: 0000-0002-0373-5358

Савіна С. С.

Кандидат економічних наук, доцент,
доцент кафедри математичного моделювання та статистики,
Київський національний економічний університет
імені Вадима Гетьмана,
м. Київ, Україна
ORCID: 0000-0003-0227-7081

Olena Piskunova

Doctor of Economics, Professor,
Professor of the Department of Mathematical Modeling and Statistics,
Kyiv National Economic University
named after Vadym Hetman,
Kyiv, Ukraine

Svitlana Savina

Candidate of Economics, Associate Professor,
Associate Professor of the Department of Mathematical Modeling and Statistics,
Kyiv National Economic University
named after Vadym Hetman,
Kyiv, Ukraine

*Досліджено практичні питання застосування нейромереж для розв'язування задач бінарної класифікації на малих обсягах вибірок. Розглянуто використання багатошарового перцептрону та проблему оптимального вибору такого основного параметру нейромережі, як кількість нейронів на проміжному шарі. Досліджувалась задача прогнозування ймовірності успішності стартапів. Встановлено, що на даних вибірки малого об'єму, при неможливості отримати результат таким класичним методом, як побудова логістичної регресії, нейромережа у вигляді багатошарового перцептрону дає високу точність класифікації. Для вибору оптимальних параметрів нейромережі потрібно проводити серії експериментів з метою встановлення взаємозв'язку між показниками ефективності нейромережі та зміною значення відповідного параметру, що надалі можливо апроксимувати функціональним зв'язком. Отримані залежності дають основу для вибору кращих параметрів нейромережі. **Ключові слова:** нейромережа; багатошаровий перцептрон; оптимальні параметри; нейрони проміжного шару; крива навчання.*

The practical issues of using neural networks for solving binary classification problems on small samples have been studied. It has been established that when the sample is artificially increased, some types of models lose their ability to perform qualitative classification. The use of a multilayer perceptron on a sample of a small volume and the problem of optimal selection of such a basic parameter of the neural network as the number of neurons in the intermediate layer are considered. The problem of forecasting the probability of success of startups was investigated. It was

found that on the data of a small sample, if it is impossible to obtain a result by such a classical method as building a logistic regression, a neural network in the form of a multilayer perceptron gives high accuracy of classification. Known classic approaches to optimizing parameter values based on learning curves did not yield the desired results. Therefore, a modified approach was used that retains the idea of the learning curve, but implements it in a different way. A learning curve is constructed for each value of the neural network parameter. The curve itself establishes the relationship between the number of runs of the neural network and the average value of classification efficiency. The described approach was implemented on the data of the training sample, which includes 376 values for each class. The test sample contains 95 values for each class. As a result of research, the best learning curve was chosen, which describes the results of a neural network containing 4 neurons on the intermediate layer, and it was established that the averaged result of no more than two or three series of experiments on the neural network of the selected architecture should be used. According to the described approach, it is possible to analyze the behavior of the effectiveness of the results of the neural network for each value of the studied parameter. If one better value of the parameter is not selected, then it is advisable to use the ensemble structure and combine the results of several better models.

Key words: *neural network; multilayer perceptron; optimal parameters; intermediate layer neurons; learning curve.*

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями. Задачі бінарної класифікації є одними з найбільш поширених у практичній діяльності економічних об'єктів. Найбільш універсальними класифікаторами у сучасному математичному моделюванні вважаються нейронні мережі. В останні десятиріччя був розроблений широкий спектр нейронних мереж, тому, в залежності від постановки задачі та наявної інформаційної бази, важливими є питання вибору архітектури та оптимізації параметрів нейромереж.

Аналіз останніх досліджень і публікацій, в яких започатковано розв'язання даної проблеми і на які спирається автор. Питання дослідження залежності ефективності класифікації на навчальних та тестових вибірках від обсягів вибірок досліджувались у роботах [1, 2]. В [3] показано, що можливо визначити найкращу функцію, якою апроксимується вище описана залежність, що підтверджує загальний характер такого підходу. Криві навчання застосовувались у дослідженні [4] при визначенні оптимальних значень параметрів нейромережі для забезпечення найкращих результатів процесу навчання моделі. В роботах останнього періоду розглядаються різні варіанти побудови кривих навчання на основі використання ваг синапсів [5] для визначення взаємозв'язку з похибками класифікації або генетичних алгоритмів [6] оптимального налаштування вагових коефіцієнтів для мінімізації сумарної квадратичної похибки класифікації.

Виділення невирішених раніше частин загальної проблеми, котрим присвячується означена стаття. У сучасній практичній діяльності економічних об'єктів доволі часто виникає необхідність проведення бінарної класифікації. Тому розроблено значну кількість видів математичних моделей, які дозволяють розв'язати таку проблему. Нейромережі вважаються одними з найкращих класифікаторів. Проте для досягнення якісних результатів при їх використанні необхідна наявність значного обсягу початкових даних, що не завжди можливо. Стандартною процедурою

для випадків баз даних незначної розмірності вважається процедура зашумлення даних або застосування спеціальних процедур, які змінюють початкові дані, наприклад, як це зроблено в [7]. Однак, не для всіх випадків це може привести до якісних результатів. У наведеному дослідженні розглядається підхід оптимального налаштування параметрів нейромереж для умов наявності вибірки незначної розмірності.

Формулювання цілей статті (постановка завдання). Метою даної роботи є поглиблення дослідження способів визначення оптимальних параметрів нейромереж на основі встановлення взаємозв'язків між ефективністю моделі та зміною шуканого параметру.

Виклад основного матеріалу дослідження. Найбільш поширеними типами моделей, які застосовуються при розв'язанні задач бінарної класифікації є дерева рішень, випадковий ліс, метод опорних векторів, логістична регресія, нейронні мережі. За співвідношенням точності та прозорості структури моделі нейронні мережі є складними з точки зору розуміння вибору параметрів архітектури моделі, проте дають найвищу точність результатів [8]. Тому для високої ефективності розв'язування задач класифікації доцільно застосовувати саме нейромережі.

В якості статистичного матеріалу використано інформацію стосовно стартапів, які діяли в Україні у період з 2017 по 2021 роки. Для побудови моделей розглядався ряд чинників, які мають найбільший вплив на успішність діяльності стартапів, а саме: місце головного офісу; час до отримання першого фінансування; приріст власних медіа каналів за даними платформи Alexa; кількість відвідувачів сайту за останні 12 місяців; сума першого інвестиційного раунду; наявність позитивних відгуків у Facebook та Twitter; кількість усіх інвестиційних раундів. Успішність стартапу відображається бінарною змінною, яка приймає значення 1, якщо стартап отримав фінансування та 0 – в протилежному випадку.

Загально відомо, що при побудові моделей класифікації найвища ефективність моделювання може бути досягнута за моделями, навчан-

ня яких здійснювалось на даних з однаковими пропорціями двох класів об'єктів. У нашому випадку суттєво більша кількість даних відповідає успішним стартапам, тому використовувалась методика штучного збільшення обсягу навчальних даних. В результаті використовувалась база даних з навчальною вибіркою, яка мстить по 376 об'єктів класу 0 та 1, у тестовій вибірці – по 95 об'єктів обох класів.

Для можливості порівняння, спершу класифікацію було проведено за допомогою логістичної регресії. Включити у модель всі вище вказані чинники виявилось не можливо, за рахунок близького до розрідженого характеру матриці початкових даних. В модель було введено лише 5 із 7 чинників. Результати моделювання вияви-

лись наступними: модель адекватна ($\chi^2 = 308,47$ при значенні $p < 0,05$), Матриця погано обумовлена, статистичну значущість оцінок параметрів розрахувати неможливо. В таблиці 1 наведено матрицю класифікації.

За даними таблиці 1 спостерігається значна перевага у розпізнаванні класу, який позначено 1. На тестових даних результати виявились наступними: 34 % правильно класифікованих об'єктів класу 0, та 97% правильно класифікованих об'єктів класу 1. Таким чином, можна дійти висновку, що штучно сформований масив даних, який був доповнений даними про об'єкти класу 0 не дає змоги логістичній регресії навчитись розпізнавати цей клас об'єктів. Застосування моделі такого типу неможливе.

Таблиця 1

Матриця класифікації за результатами логістичної регресії

Показники	Об'єкти, які віднесені моделлю до класу 0	Об'єкти, які віднесені моделлю до класу 1	Відсоток правильно класифікованих об'єктів
Кількість об'єктів класу 0	320	56	85,10638
Кількість об'єктів класу 1	3	373	99,20213

Джерело: розраховано авторами.

Подальші розрахунки проводились з використанням багатошарового перцептрон, реалізація моделей здійснювалась засобами ППП Statistica. Результати моделювання наступні: на навчальній вибірці правильно класифіковано від 90% до 100% об'єктів, на тестовій - від 68% до 99%.

При моделюванні за допомогою нейромереж, однією з важливих проблем є необхідність вибору значної кількості параметрів. Архітектура нейромережі у вигляді багатошарового перцептрон

потребує встановлення функцій активації, кількості проміжних шарів та кількість нейронів для кожного шару, тощо. Для вибору оптимальних параметрів нейромережі було обрано методику на основі застосування кривих навчання. В [9] дійшли висновку, що криві навчання чітко проявляють свої асимптотичні властивості. Однак, для досліджуваного прикладу, криві навчання не проявляють стандартної поведінки, що можна пояснити незначним обсягом масиву даних. Вигляд такої кривої наведено на рис.1.

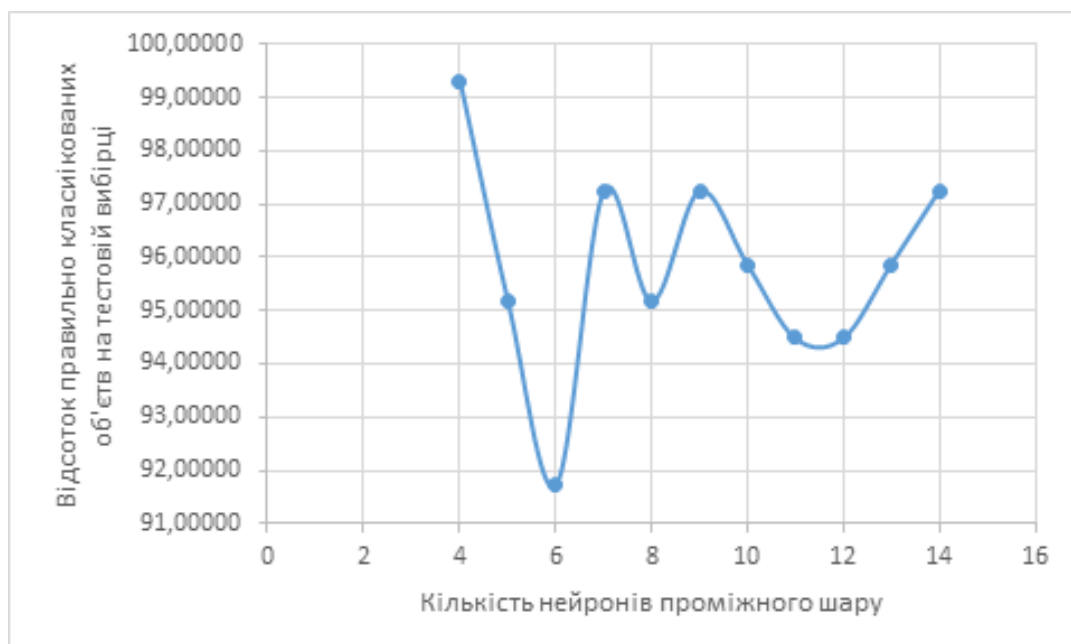


Рис. 1. Крива навчання при зміні кількості нейронів проміжного шару

Джерело: складено за розрахунками авторів.

Очевидно, що апроксимація отриманої залежності будь-якою асимптотичною кривою неможлива. Подальше штучне збільшення початкового масиву даних є недоцільним, оскільки вже проведене збільшення призводить до неефективного навчання моделей. Тому, орієнтуючись на ідею побудови кривих навчання, було реалізовано специфічний підхід, який дає змогу встановлювати оптимальні параметри нейромереж навіть для незначних об'ємів початкових даних. Підхід було доопрацьовано наступним чином. Проводились кілька се-

рій експериментів по 15 запусків нейромереж в кожній серії з поступовою зміною значення потрібного параметру та фіксувався результат ефективності моделі на навчальній вибірці. Для підвищення стійкості даних обчислень і визначення тенденції, починаючи з другої серії експериментів значення ефективностей моделей усереднювалось. Фрагмент результатів (середні значення перших та останніх запусків нейромереж для кількості нейронів проміжного шару від 4 до 9 та 12, 13, 14) розрахунків наведено в таблиці 2.

Таблиця 2

Усереднені значення ефективності моделі на тестових даних

Усереднені значення ефективності моделі	Кількість нейронів проміжного шару									
	4	5	6	7	8	9	10	11	12	13
Значення ефективності на 1 запуску нейромережі	99,3	95,2	91,7	97,2	95,2	97,2	95,9	94,5	94,5	95,9
Усереднені значення з 1 по 14 запуски нейромережі	96,9	96,3	95,9	96,1	96,3	96,2	95,9	96,2	96,2	95,5
Усереднені значення з 1 по 15 запуски нейромережі	97,0	96,4	96,0	96,2	96,1	96,1	95,9	96,3	96,0	95,4

Джерело: розраховано авторами.

Запропонований підхід дозволив стабілізувати результати розрахунків. Стандартні відхилення усереднених даних зменшились порівняно по початкових значень в 2-7 разів. Використаємо отримані усереднені результати для побудови кривих навчання. Така крива навчання, на відміну від класичного підходу, будується для

кожного значення розглянутого параметру нейромережі, однак, одночасний аналіз всіх отриманих кривих дасть змогу виявити оптимальне значення досліджуваного параметру нейромережі.

На рис.2 наведено зображення всіх кривих навчання.

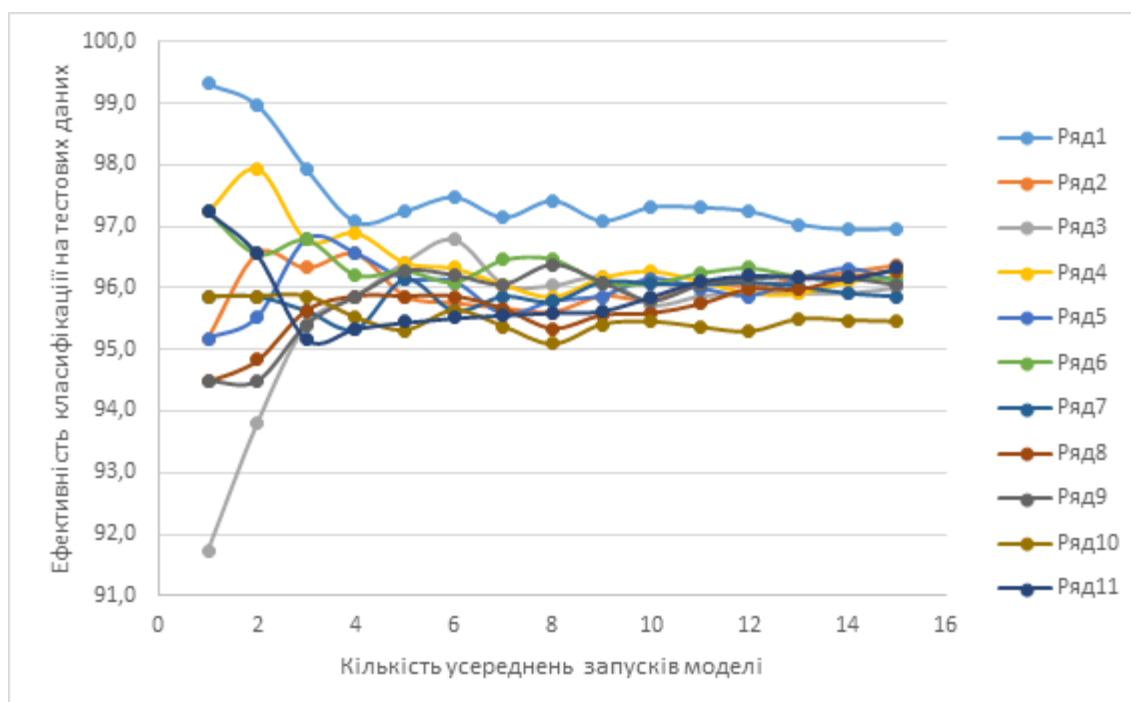


Рис. 2. Криві навчання на усереднених значеннях ефективності моделі

Джерело: складено за розрахунками авторів.

Ряд 1 відповідає даним нейромережі з 4 нейронами проміжного шару, ряд 2 – 5 нейронів проміжного шару і т.д., ряд 11 – 14 нейронів проміжного шару. З рис. 2 очевидно, що зі збільшенням кількості усереднень результатів моделювання можна отримати все більш стійкі результати. Коливання значень ефективності суттєво зменшуються і виходять на певний асимптотичний рівень.

Тоді як за результатами перших кількох серій експериментів значення ефективності мають значний розкид, в межах від 91-99%, то для усереднень за всіма серіями експериментів така розбіжність складає 95,4-96,9%. Разом з тим на рис.2 чітко виділяється крива, яка протягом всіх серій експериментів дає вищі показники ефективності, ніж інші. Ця крива описує результати нейромережі, яка містить 4 нейрони на проміжному шарі і використовувати усереднений результат не більше ніж двох-трьох серій експериментів. Таким чином встановлено, що для розрахунків на даному масиві даних слід використовувати багатoshаровий персептрон з 4 нейронами на проміжному шарі.

Висновки з цього дослідження і перспективи подальших розвідок у даному напрямку. На масивах даних незнаного обсягу не завжди

спрацьовують класичні технології застосування математичних моделей. Штучне збільшення початкових даних привело до неможливості застосування такого класифікатора, як логістична регресія. Кращі результати демонструє нейромережа у вигляді багатoshарового персептрону. Для такої архітектури нейромереж необхідно перш за все встановити значення такого параметру, як кількість нейронів на проміжному шарі. Відомі класичні підходи оптимізації значення параметрів на основі кривих навчання не дали потрібних результатів. Тому було застосовано модифікований підхід, який зберігає ідею кривої навчання, але реалізує її по іншому. Крива навчання будується для кожного значення параметру нейромережі. Самою кривою встановлюється взаємозв'язок між кількістю проведених запусків нейромережі та середнім значенням ефективності класифікації. Встановлено, що за таким підходом можна проаналізувати поведінку ефективності результатів нейромережі для кожного значення досліджуваного параметру. Якщо, на відміну, від розглянутого прикладу, не буде виділено одне краще значення параметру, то тоді доцільно використати ансамблеву структуру та поєднати результати кількох кращих моделей.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Cortes C., Jackel L.D., Solla S.A., Vapnik V., Denker J.S. Learning curves: Asymptotic values and rate of convergence *Advances in Neural Information Processing Systems* 6. Morgan Kaufmann, 1994. pp.327-334.
2. Cortes C., Jackel L.D., Chiang W.-P. Limits on learning machine accuracy imposed by data quality *Advances in Neural Information Processing Systems* 7. MIT Press, 1995. pp.239-246.
3. Gu B., Hu F., Liu H. Modelling classification performance for large data sets: An empirical study. *Lecture Notes in Computer Science*, 2001. Vol.2118. pp.317-328.
4. Watanabe E., Shimizu H. Relationships between internal representation and generalization ability in multi layered neural network for binary pattern classification problem *Proc. IJCNN*. 1993, Nagoya, Japan, 1993. Vol.2, pp.1736-1739.
5. Царегородцев В.Г. Определение оптимального размера нейросети обратного распространения через сопоставление средних значений модулей весов синапсов. *Материалы 14 международной конференции по нейрокибернетике*, Ростов-на-Дону, 2005. Т.2, С. 60-64.
6. Гридин В.Н., Солодовников В.И., Карнаков В.В. Выбор начальных значений и оптимизация параметров нейронной сети. *Новые информационные технологии в автоматизированных системах*. 2017. URL: <https://cyberleninka.ru/article/n/vybor-nachalnyh-znacheniy-i-optimizatsiya-parametrov-neyronnoy-seti> (дата звернення: 01.12.2022).
7. Drucker H., Schapire R., Simard P. Improving performance in neural networks using a boosting algorithm *Advances in Neural Information Processing Systems*. 5, Morgan Kaufmann, 1993. pp.42-49.
8. Константинов М. Краткий курс машинного обучения или как создать нейронную сеть для решения скоринг задачи. *Хабр: вебсайт*. 2017. URL: <https://habr.com/ru/post/340792/> (дата звернення: 01.11.2022).
9. Царегородцев В.Г. Оптимизация экспертов busting-коллектива по их кривым обучения. *Материалы XIII Всеросс. семинара "Нейроинформатика и её приложения"*, Красноярск, 2004. С.152-157.

REFERENCES:

1. Cortes, C., Jackel, L.D., Solla, S.A., Vapnik, V. & Denker, J.S. (1994). Learning curves: Asymptotic values and rate of convergence. *Advances in Neural Information Processing Systems* 6. Morgan Kaufmann. 327-334.
2. Cortes, C., Jackel, L.D. & Chiang, W.-P. (1995). Limits on learning machine accuracy imposed by data quality. *Advances in Neural Information Processing Systems* 7. MIT Press, 239-246.
3. Gu, B., Hu, F. & Liu, H. (2001). Modelling classification performance for large data sets: An empirical study *Lecture Notes in Computer Science*, 2118, 317-328.
4. Watanabe, E. & Shimizu, H. (1993). Relationships between internal representation and generalization ability in

multi layered neural network for binary pattern classification problem Proc. *IJCNN*. 1993, Nagoya, Japan, 2, 1736-1739.

5. Caregorodcev, V.G. (2005). Opredelenie optimal'nogo razmera nejroseti obratnogo rasprostraneniya cherez sopostavlenie srednih znachenij modulej vesov sinapsov [Determining the optimal size of the backpropagation neural network by comparing the average values of the modules of synapse weights]. *Materialy 14 mezhdunarodnoj konferencii po nejrokibernetike - Proceedings of the 14th International Conference on Neurocybernetics*, Rostov-na-Donu, T.2, 60-64 [in Russian].

6. Gridin, V.N., Solodovnikov, V.I. & Karnakov, V.V. (2016). Vybory nachal'nykh znacheniy i optimizatsiya parametrov neyronnoy seti [Selection of initial values and optimization of neural network parameters]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh – New information technologies in automated systems*. Retrieved from: <https://cyberleninka.ru/article/n/vybor-nachalnyh-znacheniy-i-optimizatsiya-parametrov-neyronnoy-seti> [in Russian].

7. Drucker, H., Schapire, R., & Simard, P. (1993). Improving performance in neural networks using a boosting algorithm. *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann.

8. Konstantinov, M. (2017). Kratkiy kurs mashinnogo obucheniya ili kak sozdat' neyronnuyu set' dlya resheniya skoring zadachi [A short course in machine learning or how to create a neural network to solve a scoring problem]. *Khabr: vebsite - Habr: website*, Retrieved from: <https://habr.com/ru/post/340792/> [in Russian].

9. Caregorodcev, V.G. (2004). Optimizatsiya jekspertov busting-kollektiva po ih krivym obuchenija [Optimization of busting team experts according to their learning curves]. *Materialy XIII Vseross. seminara "Nejroinformatika i ejo prilozheniya" – Proceedings of XIII All-russian seminar "Neuroinformatics and its applications"*, Krasnoyarsk [in Russian].