DOI: 10.34185/1991-7848.itmm.2023.01.037

МОДЕЛІ ТА МЕТОДИ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ З РОЗШИРЕНИМ ВЕКТОРОМ ВАРІЙОВАНИХ ПАРАМЕТРІВ

Зеленцов Д.Г., Шаптала Т.М.
Український державний хіміко-технологічний університет, Дніпро

Вступ. Оскільки головною ідеєю нейронної мережі (НМ) є відтворення принципу функціонування біологічного нейрону мозку, то заздалегідь НМ не знає як вирішувати поставлені перед собою задачі. Таким чином, одним із ключових етапів роботи НМ – це процес її навчання [1]. Для НМ з неперервними та диференційованими функціями активації задача навчання формулюється як задача нелінійного математичного програмування. Якість навчання НМ залежіть, у тому числі, від кількості та конкретного змісту варійованих параметрів та методу розв'язання задачі багатовимірної оптимізації. Слід зазначити, що вибір методу розв'язання задачі оптимізації в деяких випадках також залежить від змісту варійованих параметрів. В переважній більшості постановок задач навчання НМ в якості варійованих параметрів приймаються вагові коефіцієнти, а я к метод навчання – метод зворотного поширення похибки (МЗПП) [2].

Відомо, що в загальному випадку поверхня функції похибки НМ складається з пагорбів, долин, складок та ярів у просторі високої розмірності. Це викликає проблему локальних мінімумів. Частково її вдається вирішити за допомогою інерційних складових, які додаються в формули перерахунку вагових коефіцієнтів за ітераціями процесу навчання [3]. Інший підхід полягає в збільшенні розмірності оптимізаційної задачі шляхом розширення вектору варійованих параметрів.

Основний матеріал. Розширення вектору варійованих параметрів пропонується за рахунок параметрів функцій активації прихованих та вихідного шарів НМ. У цьому випадку архітектура НМ залишається незмінною, але використання МЗПП в класичному виді для розв'язання задачі навчання стає неможливим. В даній роботі використовуються метод градієнтного спуску, метод найскорішого спуску та методи спряжених градієнтів (алгоритми Флетчера-Рівза та Полака-Ріб'єра).

Для ілюстрації запропонованого підходу розглядається задача апроксимації функції

$$F(x,y)=\left(y-x^2\right)^2+\left(1-x\right)^2 \tag{1}$$

в області: $x \in [0;2]$, $y \in [0;3]$.

Нейронна мережа, що апроксимує функцію (1), наведена на рис. 1.

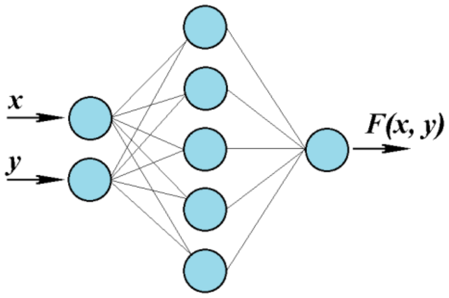

Рисунок 1

Функції активації прихованого та вихідного шарів – логістичні:

$$f_i(z) = \frac{1}{1 + \exp(-\alpha_i \cdot z)}; \; i = 1,2. \tag{2}$$

З урахуванням елементів зсуву загальна кількість вагових коефіцієнтів складає 21. Розглянемо два додаткових варійованих параметра – коефіцієнти функцій активації α_1 та α_2. Отже, маємо дві постановки задачі оптимізації, які відрізняються кількістю варійованих параметрів:

$$E(\overline{W_1}) = \frac{1}{N}\sum_{s=1}^{N}\left(t^s - y^s(\overline{W_1})\right)^2 \to \min$$
$$\overline{W_1} = [w_1, w_2, \ldots, w_{21}]^T \tag{3}$$

та

$$E(\overline{W_2}) = \frac{1}{N}\sum_{s=1}^{N}\left(t^s - y^s(\overline{W_2})\right)^2 \to \min$$
$$\overline{W_2} = [w_1, w_2, \ldots, w_{21}, \alpha_1, \alpha_2]^T. \tag{4}$$

Для постановки (3) значення параметрів $\alpha_1 = \alpha_2 = 1$.

Виходячи з рекомендацій, наведених в [3], кількість навчальних зразків – 25000; контрольних зразків – 5000; тестових зразків – 10000. Результати навчання для задачі (3) наведені в таблиці 1. Для всіх використаних методів градієнт функції похибки визначався чисельно, тому як критерій ефективності приймалась загальна кількість його обчислень.

Таблиця 1

Метод	Кількість обчислень градієнту функції похибки	Функція похибки
Метод градієнтного спуску	3000	0,000298
Метод найскорішого	535	0,000273
Метод Флетчера-Рівза	529	0,000124
Метод Полака-Ріб'єра	494	0,000153

На підставі даних таблиці 1 зроблено висновок про доцільність використання методів спряжених градієнтів для розв'язання задачі навчання НМ.

В таблиці 2 наведені результати порівняння розв'язків задачі навчання в постановках (3) та (4) з припустимою похибкою $\varepsilon^*=0{,}0001$. Для коректного зіставлення результатів розглядались шість стартових точок, перші 21 координати яких співпадали для обох постановок. Як метод навчання використовувався метод Флетчера-Рівза. Інформація про розв'язки задачі (4) наведена в дужках.

Таблиця 2

№ стартової точки	Кількість обчислень градієнту функції похибки	Кількість обчислень функції похибки
1	624 (340)	275184 (179860)
2	1571 (732)	692811 (387228)
3	839 (911)	401751 (443831)
4	1324 (488)	583884 (258152)
5	501 (565)	249165 (265029)
6	1357 (852)	598437 (450708)

Висновки. Аналіз результатів чисельних експериментів дозволив зробити висновок про доцільність розширення вектору варійованих параметрів в задачах навчання НМ з неперервними та диференційованими функціями активації. Незважаючи на збільшення розмірності задачі оптимізації, в цілому ефективність нової постановки більш висока, ніж традиційної. Пояснюється це, на думку авторів, тим, що значна частка обчислювальних витрат в традиційній постановці припадає на спроби покинути околиці локальних мінімумів, в той час, коли збільшення розмірності простору розв'язків дозволяє здійснити це зі значно меншими витратами.

Література

1. Рассел С., Норвіг П. Artificial Intelligence: A Modern Approach, четверте видання. Лондон: Pearson, 2020. 1136 с.
2. Фройнд Й., Хауслер Д. Unsupervised learning of distributions on binary vectors using two layer networks. *In Advances in Neural Information Processing Systems 4*: Конференція та семінар з систем обробки нейронної інформації, Денвер, 1992. С. 912–919.
3. Хайкін С. Neural Networks: A comprehensive foundation. Prentice Hall, 1999. 842 с.

MODELS AND METHODS FOR TRAINING NEURAL NETWORKS WITH AN EXTENDED VECTOR OF VARYING PARAMETERS

Dmytro Zelentsov, Taras Shaptala

Abstracts. A studied of models and methods for training neural networks using an extended vector of varying parameters is conducted. The training problem is formulated as a continuous multidimensional unconditional optimization problem. The extended vector of varying parameters implies that it includes some parameters of activation functions in addition to weight coefficients. The introduction of additional varying parameters does not change the architecture of a neural network, but makes it impossible to use the back propagation method. A number of gradient methods have been used to solve optimization problems. Different formulations of optimization problems and methods for their solution have been investigated according to accuracy and efficiency criteria.

Keywords: neural networks, neural network training task, multidimensional optimization, vector of variable parameters, gradient methods.

Reference

1. Russell S., Norvig P. Artificial Intelligence: A Modern Approach, fourth edition. London: Pearson, 2020. 1136 p.
2. Freund Y., Hausler D. Unsupervised learning of distributions on binary vectors using two layer networks. *In Advances in Neural Information Processing Systems 4*: Conference and Workshop on Neural Information Processing Systems, Denver, 1992. P. 912–919.
3. Haykin S. Neural Networks: A comprehensive foundation. Prentice Hall, 1999. 842 p.