

<https://doi.org/10.15407/dopovidi2022.01.034>

UDC 519.7

V.I. Norkin, <https://orcid.org/0000-0003-3255-0405>

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”
E-mail: Vladimir.norkin@gmail.com

On the finite convergence of the NN classification learning on mistakes

Presented by Corresponding Member of the NAS of Ukraine P.S. Knopov

The paper establishes an analog of well-known Novikoff's theorem on the perceptron learning algorithm's finite convergence in linearly separated classes. We obtain a similar result concerning the nearest neighbor classification algorithm in the case of compact classes in a general metric space for the case of non-intersecting classes. The learning process consists of gradual modification of the algorithm in misclassification cases. The process is studied in the deterministic setting. Classes are understood as compacts in complete metric space, and class separation is defined as the non-intersection of compacts. The number of learning steps is bounded by the number of elements in some ε -net for the considered classes.

Keywords: *classification learning, finite convergence, nearest neighbor method, learning on mistakes.*

F. Rosenblatt [1] developed one of the first learning machines (perceptron) for pattern recognition and proposed an error correction method for its learning. When analyzing the perceptron's work, V.M. Glushkov [2] emphasized the importance of obtaining rigorous mathematical results regarding the learning process. One of such results is the well-known Novikoff theorem [3] on the finite convergence of the learning process of a perceptron for linearly separable classes. A generalization of this theorem to the nonlinear case was given in [4, Ch. V, § 5, Theorem II]. There is a similar result for the Kozinets algorithm for separating the convex hulls of two sets [5]. A detailed presentation of these and related results is available in books [6, 7].

In this paper, we present similar learning finite convergence results concerning the nearest neighbor (NN) classification algorithm [8, 9] in the case of compact non-intersecting classes in general metric space. Our exploration is deterministic and is based on functional analysis arguments. We focus on the iterative learning of the NN algorithm. Unlike [10], we consider classes as compact sets in complete metric space. The probabilistic asymptotic analysis of the

Цитування: Norkin V.I. On the finite convergence of the NN classification learning on mistakes. *Допов. Нац. акад. наук Укр.* 2022. № 1. С. 34–38. <https://doi.org/10.15407/dopovidi2022.01.034>

NN classification rule can be found in [8, 9]. In the latter and related works, the learning problem, i.e., the gradual formation of the class of representatives, was not studied.

Consider the following iterative algorithm for a machine classification learning on the nearest neighbor principle. Suppose there is a flow of objects from some finite set of classes. From this flow, the representative groups of objects for each class are gradually formed. The initial representative groups are formed arbitrarily. For each new object with a known affiliation to a particular class, its proximity to each existing representative group is assessed. When the object is the closest to another representative group, i.e., a classification error occurs, it replenishes its representative group. When the object is the closest to its representative group, the groups do not change, the object is ignored, and the next object in the stream is considered. Under certain conditions, over a finite number of replenishment steps, the learning process ends, i.e., the representative groups cease to change. After that, each new object is correctly classified, i.e., it is the closest to its representative group.

This classification method can be interpreted as an iterative version of the nearest neighbor method. It is a non-parametric method because it does not use parametric functions to separate classes. Next, we consider strict statements about the convergence of this method of machine classification learning.

Consider the case where the classes do not intersect.

Assumption 1.

The objects are points in complete metric space C with distance $dist(\cdot, \cdot)$.

The classes, $C_n \subset C$, $n = 1, 2, \dots, N < \infty$, are compacts in this metric space.

Different classes do not intersect, $C_i \cap C_j = \emptyset$, $i \neq j$.

Algorithm 1 (iterative classification learning for non-intersecting classes).

Step 0 (Initialization). The initial representations of classes, compact sets $C_n^0 \subseteq C_n$, $C_n^0 \neq \emptyset$, $n = 1, \dots, N$, are created. It is set $t = 0$.

Step 1 (Presentation of an object). At iteration t an arbitrary object $c^t \in C_{n_t}$ from some class n_t is selected.

Step 2 (Calculation of distances). The distances, $dist(c^t, C_n^t) = \min_{c \in C_n^t} dist(c^t, c)$, $n = 1, \dots, N$, are calculated, and the set of indices $v_t = \arg \min_{n \in \{1, \dots, N\}} dist(c^t, C_n^t)$ is found.

Step 3 (Replenishment of the representative classes in the case of classification error, the so-called productive step).

$$C_n^{t+1} = \begin{cases} C_n^t, n = 1, \dots, N, & n_t = v_t \text{ (no changes);} \\ C_n^t, n \in \{1, \dots, N\} \setminus n_t, & n_t \neq v_t, \\ C_n^t \cup c^t, n = n_t, & n_t \neq v_t \text{ (productive step).} \end{cases}$$

Step 4 (Transition to the next iteration). It is set $t := t + 1$, and the transition to Step 1 is done.

Theorem 1. *In assumption 1, the learning process ends after a finite number of steps with the replenishment of some representative class, i.e., the sequence of sets $\{C_n^t, n = 1, \dots, N\}$, $t = 1, \dots$, is stabilized after a finite number of productive steps of algorithm 1. If the classes are separated*

by a distance $\varepsilon > 0$, then the number of productive steps of the classification learning algorithm is bounded by the total number of elements in the ε' -nets for these classes, $\varepsilon' < \varepsilon/2$.

Remark 1. The algorithm does not use the actual value of the separation distance between classes. The first statement of the theorem only assumes that the separation distance is positive. In the second statement of the theorem, the unknown value of the separation distance is used to bound the total number of the productive (replenishment) steps of the algorithm.

Remark 2. The number of different elements in a class can be finite or infinite. The total number of iterations of the algorithm until the stabilization is almost surely bounded, if, at each iteration, each class appears in the flow with fixed positive probability.

Consequence 1. If the learning process is stopped, i.e., the class of representatives is not further changed, then each new object is correctly classified according to the nearest neighbor principle.

Indeed, if the class of representatives is not changed, each new presented object will be the closest to its class of representatives.

Proof. Let us define the following distance between classes C_i , and let $C_j: \text{dist}(C_i, C_j) = \inf\{\text{dist}(c_i, c_j) : c_i \in C_i, c_j \in C_j\}$. By assumption, there is $\varepsilon > 0$ such that $\text{dist}(C_i, C_j) \geq \varepsilon$ for any $i \neq j$. Let the representation of a class, say n , be replenished at moments $\{t_k, k = 1, 2, \dots\}$, with elements $c_n^{t_k} \in C_n$. At these moments, it is either $\min_{n' \in \{1, \dots, N\}} \text{dist}(c_n^{t_k}, C_n^{t_k}) = \text{dist}(c_n^{t_k}, c_{n'}^{t_k}) < \text{dist}(c_n^{t_k}, C_n^{t_k})$ holds for some n , or, $n \subset v_t = \arg \min_{n' \in \{1, \dots, N\}} \text{dist}(c_n^{t_k}, C_n^{t_k})$, $n \neq v_t$.

In the first case, due to ε -separation of classes, $\text{dist}(c_n^{t_k}, c_{n'}^{t_k}) \geq \varepsilon$, then $\text{dist}(c_n^{t_k}, C_n^{t_k}) > \varepsilon$ for all $k = 1, 2, \dots$, i.e., each new element $c_n^{t_k}$ is separated from all previous representatives $C_n^{t_k}$ of the class C_n by a distance not less than ε .

Consider the second case where $n \subset v_t$, $n \neq v_t$, i.e., there exists $n'' \neq n$, and $n'' \in v_t$.

Since the distance between different classes is not less than ε , we have $\min_{n' \in \{1, \dots, N\}} \text{dist}(c_n^{t_k}, C_n^{t_k}) = \text{dist}(c_n^{t_k}, c_{n''}^{t_k}) \geq \varepsilon$. In addition, the following relation holds:

$$\text{dist}(c_n^{t_k}, C_n^{t_k}) = \min_{n' \in \{1, \dots, N\}} \text{dist}(c_n^{t_k}, C_n^{t_k}) = \text{dist}(c_n^{t_k}, c_{n''}^{t_k}) \geq \varepsilon.$$

Due to the compactness of sets C_n for $\varepsilon' < \varepsilon/2$, there is a finite ε' -net $M_n(\varepsilon')$ with the number of elements $m_n(\varepsilon')$ that approximates the set C_n and its subsets $C_n^{t_k} \subset C_n$ with accuracy ε' . In other words, for each element $a \in C_n$, there is an element $b \in M_n(\varepsilon')$ such that $\text{dist}(a, b) \leq \varepsilon'$. Let $M_n^{t_k}(\varepsilon')$ be a part of this network with the number of elements $m_n^{t_k}(\varepsilon')$ that approximates the set $C_n^{t_k}$ with accuracy ε' , i.e., for each element $a \in C_n^{t_k}$, there is an element $b \in M_n^{t_k}(\varepsilon')$ such that $\text{dist}(a, b) \leq \varepsilon'$.

Since $\text{dist}(c_n^{t_k}, C_n^{t_k}) \geq \varepsilon$, the element $c_n^{t_k}$ is not approximated by the net $M_n^{t_k}(\varepsilon')$. Therefore, the number of elements $m_n^{t_{k+1}}(\varepsilon')$ in the net $M_n^{t_{k+1}}(\varepsilon')$ is greater than the number $m_n^{t_k}(\varepsilon')$ of elements in the net $M_n^{t_k}(\varepsilon')$, $m_n^{t_{k+1}}(\varepsilon') > m_n^{t_k}(\varepsilon')$. But since the total number of elements $m(\varepsilon')$ of the net is bounded, i.e., $m_n^{t_k}(\varepsilon') < m_n^{t_{k+1}}(\varepsilon') \leq m_n(\varepsilon')$, the number of replenishments K_n of the class of representatives $C_n^{t_k}$ is limited by the value $m_n(\varepsilon')$.

Then the total number of productive steps of the learning algorithm is finite and does not exceed the value $m_1(\varepsilon') + \dots + m_N(\varepsilon')$. The theorem is proved.

The case of convex classes.

Assumption 2.

The objects are points in the l -dimensional Euclidean space \mathbf{R}^l with distance $dist(x, y) = \|x - y\|$ for $x, y \in \mathbf{R}^l$.

The classes $C_n \subset C$, $n = 1, 2, \dots$, are compact sets (bounded closed sets) in this space \mathbf{R}^l .

The number of classes N is finite, $n \leq N < \infty$.

The classes $C_n \subset C$, $n = 1, 2, \dots$ are convexly separated, i.e., the convex hulls chC_n of the classes do not intersect, $chC_i \cap chC_j = \emptyset$, $i \neq j$.

Algorithm 2 (iterative classification learning one for convex classes).

Step 0 (Initialization). Initial representations of the classes $C_n^0 \subseteq C_n$, $C_n^0 \neq \emptyset$, $n = 1, \dots, N$, are created. It is set $t = 0$.

Step 1 (Presentation of an object). An arbitrary object $c^t \in C_{n^t}$ from some class n^t is selected.

Step 2 (Calculation of distances). The distances of the element c^t to the convex hulls chC_n^t of the sets C_n^t are calculated, and the set of indices $v_t = \arg \min_{n \in \{1, \dots, N\}} dist(c^t, C_n^t)$ is found.

Step 3 (Replenishment of the class of representatives, a productive step).

$$C_n^{t+1} = \begin{cases} C_n^t, n = 1, \dots, N, & n_t = v_t \text{ (no changes);} \\ C_n^t, n \in \{1, \dots, N\} \setminus n_t, & n_t \neq v_t, \\ C_n^t \cup c^t, n = n_t, & n_t \neq v_t \text{ (productive step).} \end{cases}$$

Step 4 (Transition to the next iteration). It is set $t := t + 1$, and the transition to Step 1 is fulfilled.

Theorem 2. *In assumption 2, the learning process ends in a finite number of productive steps, i.e., the sequence of sets $\{C_n^t, n = 1, \dots, N\}, t = 1, \dots$, is stabilized in a finite number of productive steps of algorithm 2.*

Proof. The proof is completely analogous to the proof of Theorem 1, if we replace the sets C_n and C_n^t with their convex hulls chC_n and chC_n^t .

Thus, the finite convergence of the method of teaching classification according to the nearest neighbor principle under the condition of non-intersection of compact classes in a metric space is proved.

The work was supported by the grant of the National Research Foundation of Ukraine No. 2020.02/0121 "Analytical methods and machine learning in control theory and decision-making in conditions of conflict and uncertainty".

REFERENCES

1. Rosenblatt, F. (1962). Principles of Neurodynamics: Perceptron and Theory of Brain Mechanisms. Washington: Spartan Books. 616 p.
2. Glushkov, V. M. (1963). The theory of instruction for a class of discrete perceptrons. USSR. Comput. Math. and Math. Phys., 2, Iss. 2, pp. 338-355. [https://doi.org/10.1016/0041-5553\(63\)90410-5](https://doi.org/10.1016/0041-5553(63)90410-5)
3. Novikoff, A. B. J. (1962). On convergence proofs on perceptrons. Proceedings of the Symposium on the Mathematical Theory of Automata, 12, pp. 615-622. New York: Polytechnic Institute of Brooklyn.

4. Kozinets, V. N. (1973). Recurrent algorithm for separating convex hulls of two sets. Pattern recognition learning algorithms. Moscow: Sovetskoe Radio, pp. 43-50 (in Russian).
5. Aizerman, M. A., Braverman, E. M., Rozonoer, L. I. (1970). Method of Potential Functions in the Theory of Pattern Recognition. Moscow: Nauka (in Russian).
6. Schlesinger, M. I. & Hlavác, V. (2002). Ten lectures on statistical and structural pattern recognition. Dordrecht: Kluwer. <https://doi.org/10.1007/978-94-017-3217-8>
7. Vapnik, V. N. (1998). Statistical learning theory. New York: Wiley.
8. Cover, T. M. & Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 13, Iss. 1, pp. 21-27. <https://doi.org/10.1109/tit.1967.1053964>
9. Devroye, L., Györfi, L. & Lugosi, G. (1996). A Probabilistic Theory of Pattern Recognition. New York: Springer.
10. Vorontsov, K. V. (2010). Metric classification algorithms. Access. <http://machinelearning.ru/wiki/images/8/8f/Voron-ML-Metric1.pdf>

Received 16.09.2021

V.I. Норкін, <https://orcid.org/0000-0003-3255-0405>

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ,
Національний технічний університет України
“Київський політехнічний інститут ім. Ігоря Сікорського”
E-mail: Vladimir.norkin@gmail.com

ПРО СКІНЧЕННУ ЗБІЖНІСТЬ ПРОЦЕСУ НАВЧАННЯ NN КЛАСИФІКАЦІЇ НА ПОМИЛКАХ

Встановлено аналог відомої теореми Новікова про скінченну збіжність алгоритму навчання перцептрона у випадку лінійно розділених класів. Ми отримуємо аналогічний результат щодо алгоритму класифікації за принципом найближчого сусіда у випадку компактних класів у загальному метричному просторі для класів, що не перетинаються. Процес навчання полягає у поступовій модифікації алгоритму у випадках помилкової класифікації. Процес вивчається в детермінованій постановці. Класи розуміються як компакти в повному метричному просторі. Розділення класів визначається як неперетин компактів. Кількість кроків навчання обмежена числом елементів в деякій ε -сітці для розглянутих класів.

Ключові слова: класифікаційне навчання, кінцева збіжність, метод найближчого сусіда, навчання на помилках.