Doctor of Philosophy

Extension and optimization of the perceptron convergence algorithm

The Graduate School
of the University of Ulsan

Department of Mathematics
Laith M. Almomani

Extension and optimization of the perceptron convergence algorithm

Supervisor: Prof. Sang-Mok Choo

A Dissertation

Submitted to
the Graduate School of the University of Ulsan
In partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Laith M. Almomani

Department of Mathematics
University of Ulsan, Korea

August 2021

Extension and optimization of the perceptron convergence algorithm

This certifies that the dissertation/doctoral thesis of Laith M. Almomani is approved

Prof. Chang Rim Jang

Prof. Dougu Nam

Prof. Sang-Mok Choo

Prof. Yoora Kim

Prof. Jangwon Ju

Department of Mathematics
University of Ulsan, Korea

August 2021

Extension and optimization of the perceptron convergence algorithm

This certifies that the dissertation/doctoral thesis
of Laith M. Almomani is approved

Prof. Chang Rim Jang

Prof. Dougu Nam

Prof. Sang-Mok Choo

Prof. Yoora Kim

Prof. Jangwon Ju

Department of Mathematics
University of Ulsan, Korea

August 2021

국문요약

인공신경망은 다양한 분야에서 사용되고 중요한 역할을 하고 있다. 그러나 신경망에 대한 수학적 분석에 대한 결과는 매우 적다. 특히, 인공신경망 알고리즘의 수렴 속도를 가속화하기 위하여 훈련 데이터의 순서를 제시하는 이론적 접근법은 전무하다.

단일층의 페셉트론 수렴 알고리즘을 대상으로 이러한 순서를 이론적으로 제시하였고, 더 나아가 다양한 구조를 가지는 페셉트에 대하여 새로운 수렴 알고리즘을 만들었다.

이러한 수렴 알고리즘의 증명을 기반으로 수렴 속도를 가속화할 수 있는 훈련 데이터 순서를 제시하였고 수치적 예제를 통하여 확장된 수렴 정리와 훈련 데이터 순서의 효과를 제시하였다.

Acknowledgements

First of all, I thank God for giving me the strength, patience, and courage to complete this work.

I express my deep appreciation to my supervisor Prof. Sang-Mok Choo for his inspiring ideas, generous attitudes, helpful suggestions and valuable remarks, throughout every step of this thesis. I have learned a lot from our work together and appreciate that you took me on as your student.

Also, I am deeply grateful to the committee members for agreeing to read my thesis and taking a part in this process.

I would also like to thank the Korean Government for giving me the scholarship to finish my PhD degree and allow me to discover and know Korean culture.

Finally, I would sincerely like to thank all professors in the department of mathematics, friends, my parents and my family for helping me during working on this thesis.

Dedication

To my father and mother. Without your support, this wouldn't be possible.

Contents

1 Introduction

Neural networks originally mimic the way the human brain operates as a set of neurons joined together by a set of connections. Neurons here have a weighted sum of inputs followed by an activation function. The McCulloch [1] neuron is the simplest and earliest example of an artificial neuron. The perceptron was introduced first by Rosenblatt [2] which is a parameterized function that takes a real-valued vector as input and produces a Boolean output. And in machine learning, the perceptron is an algorithm for supervised learning of training data that should be classified into corresponding only two categories for classification called binary classifiers. Coolen [3] defined a student feed-forward network having its single binary neuron learning rule as: Let questions be the classification of data in $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\right\}$ for some positive integers N and r. Let teacher's answer T be a function from Ω to $\{0, 1\}$

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & if \quad (\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* > 0), \\ & \qquad\qquad\qquad\qquad \text{and } \min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* \mid > 0, \\ 0 & if \quad (\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* < 0), \end{cases}$$

for some $\boldsymbol{w}^* \in R^N$ and $\theta^* \in R$ and all $1 \leq i \leq r$. Let $\boldsymbol{w} \in R^N$ and $\theta \in R$. Student's answer $S_{\boldsymbol{w},\theta}$ is a function from Ω to $\{0, 1\}$ defined as

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & if \quad (\boldsymbol{w} \cdot \boldsymbol{x}^i - \theta > 0), \\ 0 & if \quad (\boldsymbol{w} \cdot \boldsymbol{x}^i - \theta \leq 0), \end{cases}$$

which can be represented as Figure 1.1

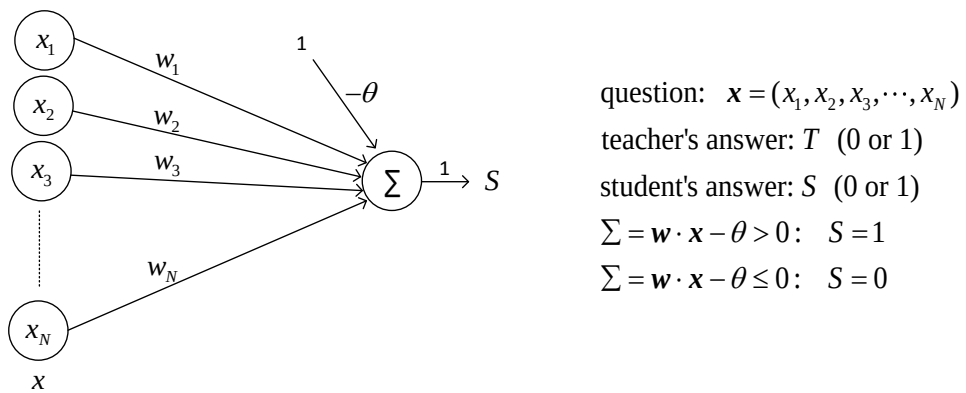


Figure 1.1: Student feed-forward network

Since the values of $\boldsymbol{w}$ and θ are unknown, it was a question to construct sequences $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ with the limits $\boldsymbol{w}$ and θ satisfying $S_{\boldsymbol{w},\theta} = T$. So Coolen [3] defined the sequences as follows:
Initial elements $\boldsymbol{w_0} \in R^N$ and $\theta_0 \in R$ are random.
Let $n \geq 0$, and $I = \{1, \cdots, r\}$.

Case 1. For all $i \in I$, $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T(\boldsymbol{x}^i)$.

$$\text{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n, \theta_n).$$

Case 2. For some $i \in I$, $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ and $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1, T(\boldsymbol{x}^i) = 0$.

$$\text{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n - \boldsymbol{x}^i, \theta_n + 1).$$

Case 3. For some $i \in I$, $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ and $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0, T(\boldsymbol{x}^i) = 1$.

$$\text{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1).$$

Our goal is to extend this theorem and to construct approaches to increase the convergence speed. In Chapter 2 a brief history of Neural Networks is given with supervised learning algorithms, linear classifiers, single-layer neural networks, and multi-layer neural network. In Chapter 3 we present extensions of the perceptron convergence algorithm. Since the perceptron algorithm can be convergent slowly if Ω is a big data set, we provide approches to make the algorithm converge faster in Chapter 4. Finally, in Chapter 5 we present numerical examples of our extended theorems and convergence speed.

2 Neural Networks

Neural networks are part of a broader family of machine learning, which is one of the most important topics in computer science and is a category of Artificial Intelligence (AI). Here AI provides systems with the ability to automatically learn and improve from experience without any intervention or assistance of human beings [4]. The aim of it is to make the computers modify their actions in order to adjust actions to get more accuracy, where accuracy is the number of correctly predicted data out of all the data.

2.1 Brief history

According to [5,6] artificial neural networks are techniques that mimic the mechanism of learning in biological organisms. The human nervous system contains cells, called nerve cells or neurons, which are specialized to carry "messages" through an electrochemical process. And they are connected to each other with axons and dendrites. The connecting regions between axons and dendrites are referred to as synapses. This biological mechanism is simulated in artificial neural networks as in Figure 2.1, which contains computation units that are referred to as neurons. Throughout this thesis, we use the term "neural networks" to refer to artificial neural networks rather than biological ones.

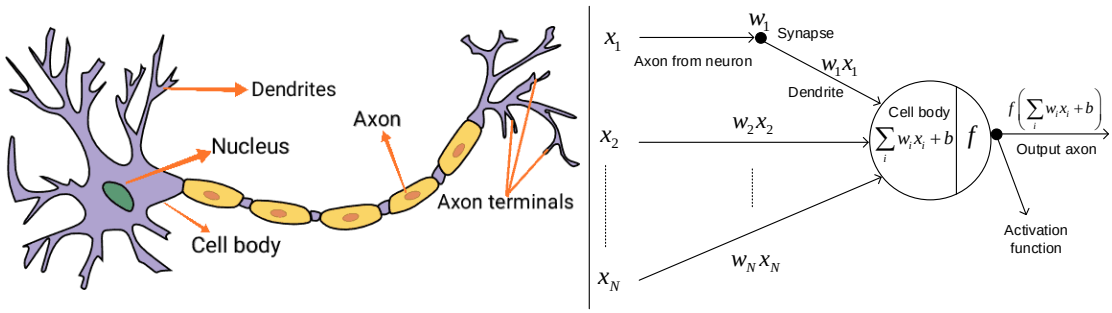


Figure 2.1: Biological neural network and artificial neural network

The earliest reported work in the field of neural networks began in the 1940s with McCulloch [1] neuron, which is the simplest example of an artificial neuron. In 1957, Frank Rosenblatt [2] applied McCulloch idea to early AI when he introduced the perceptron. Rosenblatt conceived of the percetron as a simplified mathematical

model of how the neurons in brains take an input real vector $\boldsymbol{x}$ (nearby neurons) and real vector of weights $\boldsymbol{w}$ (the synapse strength to each nearby neuron). The model returns 1 if the dot product of the two vectors is more than some threshold θ. Otherwise the model returns 0. In 1965 Ivakhnenko and Lapa [7] introduced the group method of data handling to learn multi-layered networks, which is perhaps the first deep learning systems of the feedforward multilayer perceptron(MLPs). In 1982 Fukushima [8] proposed a neocognitron model, which is the inspiration to the modern convolutional neural networks (CNNs), and at the same year Werbos [9] applied the chain rule and backpropagation to multilayer perceptrons. Yann LeCun [10], the first work on modern CNNs in the 1990s, combined backpropagation with CNNs to get a successful classification of hand writing digits. And also backpropagation was used in recurrent neural networks (RNNs) [11]. Hochreiter and Schmidhuber [12] made an extension of RNNs and designed long short-term memory (LSTMs). Until 2006 researchers were unsuccessful at training deep networks so they were not sure whether deep networks could be trained. The interest of deep feedforward networks was revived by two papers in 2006 [13] and 2010 [14], where they found that backpropagation indeed can work well with deep networks as long as appropriate activation functions are used and the weights are initialized in a clever way. And lately at 2014 Kyunghyun Cho [15] introduced gated recurrent unit (GRU) is like a LSTM with a forget gate, but has fewer parameters than LSTM.

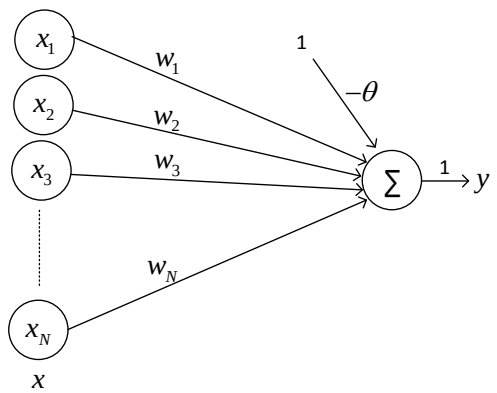
Nowadays neural networks are highly valuable because it is used in real life applications. In many fields they use collection of data that is huge in volume and grows exponentially with time. This type of data is called big data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. So AI here comes to solve storing and managing such big data using various methods like artificial neural networks. AI and big data complement each other: lots of data are needed for AI to be efficient. Neural networks play a very important rule in technology such as in robotics, social media applications, online shopping, medical robots and biology, where Sang-Mok Choo [16] proposed a boolean feedforward neural network modeling by combining neural network and boolean network modeling to identify control targets that can induce desired cellular state conversion.

2.2 Supervised machine learning algorithms

For a classification task, supervised machine learning algorithms take a training set of data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ to learn how to predict a random variable $y \in Y$, which is called a class label based on an input data $x \in X$ [17]. Some of these algorithms are linear classifier, single layer neural network the perceptron and multi layer neural network.

2.2.1 Linear classifier

According to [18] in machine learning, the goal of classification is to put items of same feature values into same groups. Timothy [19] stated that a linear classifier achieves this by making a classification decision based on the value of the linear combination of the features. A linear classifier is often used in situations where the speed of classification is an issue, since it is often the fastest classifier, especially when data is sparse, and linear classifiers often work very well when the number of dimensions is large. The most famous linear classifier is the perceptron as in Figure 2.2, where $\boldsymbol{x}$ is an input data and $\boldsymbol{w}$ are weights. The perceptron returns 1 if the dot product of the two vectors is more than threshold (θ) and otherwise returns 0.



$$y = 1 \text{ if } \boldsymbol{w} \cdot \boldsymbol{x} - \theta > 0,$$

$$y = 0 \text{ if } \boldsymbol{w} \cdot \boldsymbol{x} - \theta \leq 0.$$

Figure 2.2: Linear perceptron

2.2.2 Single and multi layer neural network

A single-layer neural network is the simplest form of neural networks, in which there is only one layer that sends the activation of weighted input values to an

output node as in Figure 2.3. This simple neural network is also referred to as the *perceptron* which is defined in the subsection 2.2.1.

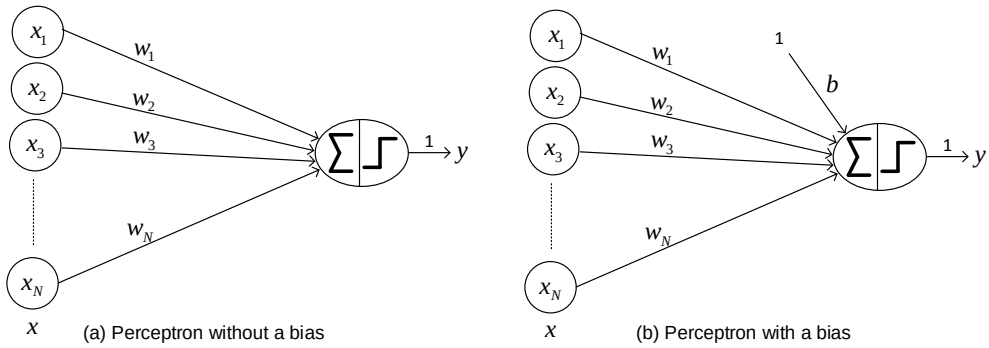


Figure 2.3: The basic architecture of the perceptron

From [20] the variables $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$ of neuron are called inputs of the neuron and the symbol $\sum$ is a weighted sum of the inputs. Sometimes a constant, called bias b, is added.

$$\sum = \boldsymbol{w} \cdot \boldsymbol{x} \quad \text{(without a bias)} \qquad \sum = \boldsymbol{w} \cdot \boldsymbol{x} + b \quad \text{(with a bias)}$$

Function f is termed an activation function and the output y of a neuron with input $\boldsymbol{x}$ is given by $y = f(\boldsymbol{w} \cdot \boldsymbol{x} + b)$.

Multi-layer neural networks [5] contain multiple computational layers and the additional intermediate layers (between input and output) are referred to as *hidden layers*. The basic multi-layer neural networks are called feed-forward networks, because each neuron in one layer has directed connections only to the neurons of the subsequent layer, where the direction is from input to output. This neural network is also referred to as the multilayer perceptron(MLP).

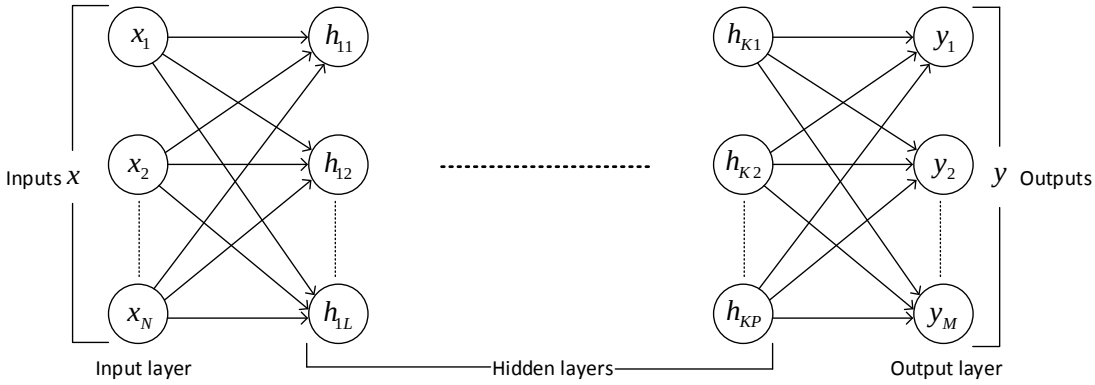


Figure 2.4: Multilayer perceptron without a bias

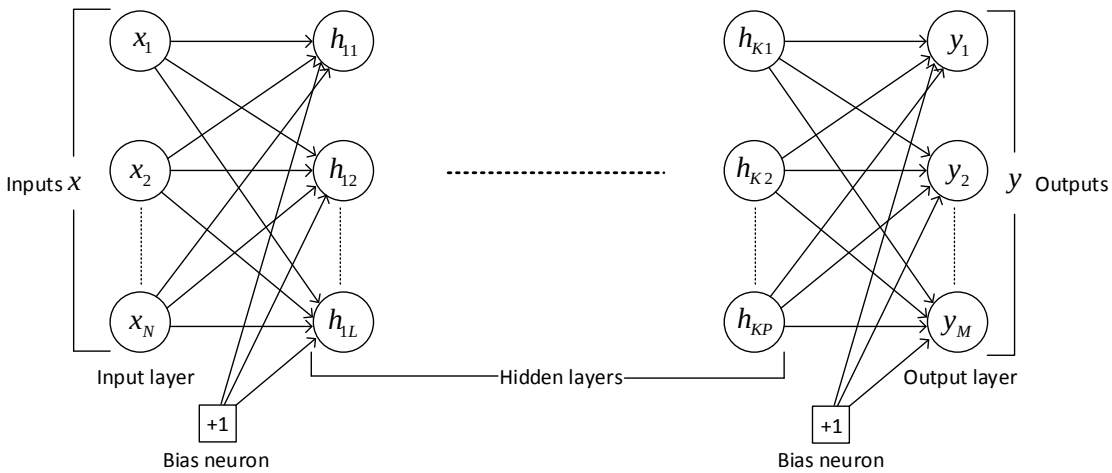


Figure 2.5: Multilayer perceptron with a bias

Figures 2.4 and 2.5 are the general forms of multi layer neural network without and with bias, where each has K hidden layers. Starting from the input layer, data is propagated forward to the output layer, which step is called the forward propagation. Then the parameters of the network is updated in the direction of decreasing the error (sum of squares of the difference between the predicted and known outcome) by using a backpropagate algorithm. Repeating such an update we finally get the weights that give us good predicted class labels.

3 Extensions of the perceptron convergence algorithm

Haykin [21] present the proof of the perceptron convergence algorithm. In this chapter we rewrite the proof the convergence of perceptron learning algorithm and make extension theorems of perceptron learning algorithm.

3.1 The perceptron convergence algorithm

Theorem 3.1. *Let* $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\right\}$ *for some positive integers* N *and* r*. Let* T *be a function from* Ω *to* $\{0,1\}$*. Assume that there exist* $\boldsymbol{w}^* \in R^N$ *and* $\theta^* \in R$ *such that for* $1 \leq i \leq r$

$$min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* \mid > 0$$

and

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* < 0. \end{cases}$$

Let $\boldsymbol{w} \in R^N$ *and* $\theta \in R$*. Function* $S_{\boldsymbol{w},\theta} : \Omega \to \{0,1\}$ *is defined such that for* $1 \leq i \leq r$

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{x}^i - \theta > 0, \\ 0 & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{x}^i - \theta \leq 0. \end{cases}$$

Sequences $\{\boldsymbol{w}_n\}$ *and* $\{\theta_n\}$ *are defined as follows:*
Initial elements $\boldsymbol{w}_0 \in R^N$ *and* $\theta_0 \in R$ *are random.*
Let $n \geq 0$ *and* $I = \{1, \cdots, r\}$*.*

Case 1. For all $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T(\boldsymbol{x}^i)$*.*

$$\text{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n, \theta_n).$$

Case 2. For some $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \leq l \leq i-1)$ *and* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$*,* $T(\boldsymbol{x}^i) = 0$*.*

$$\text{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n - \boldsymbol{x}^i, \theta_n + 1).$$

Case 3. For some $i \in I$, $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \le l \le i-1)$ and $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$, $T(\boldsymbol{x}^i) = 1$.

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1).$$

Then $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.

Proof. Using the definitions of the sequences, we have

$$(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) = \begin{cases} (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) & \text{(Case 1)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 3)} \end{cases} \tag{1}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are infinite sequences.
Then (Case1) in (1) is not possible, which give that for $n \ge 0$

$$(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) = \begin{cases} (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 3)} \end{cases} \tag{2}$$

Claim 1. $(\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) \ge (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid$ for all $n \ge 1$
Claim 2. $\mid (\boldsymbol{w}_n, \theta_n) \mid \le \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}$ for all $n \ge 1$

Using Claim 2, Cauchy-schwarz inequality and Claim 1, we have that for all n

$$\begin{aligned} \mid (\boldsymbol{w}^*, \theta^*) \mid \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)} &\ge \mid (\boldsymbol{w}_n, \theta_n) \mid\mid (\boldsymbol{w}^*, \theta^*) \mid \\ &\ge (\boldsymbol{w}_n, \theta_n)(\boldsymbol{w}^*, \theta^*) \\ &\ge (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) \\ &\quad + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid . \end{aligned}$$

Then

$$\mid (\boldsymbol{w}^*, \theta^*) \mid \ge \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid}{\sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}},$$

which gives

$$\begin{aligned} \mid (\boldsymbol{w}^*, \theta^*) \mid &= \lim_{n\to\infty} \mid (\boldsymbol{w}^*, \theta^*) \mid \\ &\ge \lim_{n\to\infty} \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid}{\sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}} = \infty. \end{aligned}$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.

Proof of Claim 1.

Case 2: $T(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* < 0$, which gives

$$\begin{aligned}(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \\ &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \mid \\ &\geq (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid.\end{aligned} \tag{3}$$

Case 3: $T(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* > 0$, which gives

$$\begin{aligned}(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \\ &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \mid \\ &\geq (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid.\end{aligned} \tag{4}$$

Using (2), (3), and (4), we obtain

$$(\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid .$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} (\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq \sum_{j=0}^{n-1} \left\{ (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid \right\}$$

and then

$$\sum_{j=0}^{n-1} (\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq \sum_{j=0}^{n-1} (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + n \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid .$$

This gives

$$\begin{aligned}&\sum_{j=1}^{n-1} (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) \\ &\qquad \geq (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + \sum_{j=1}^{n-1} (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + n \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid\end{aligned}$$

and hence

$$(\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid .$$

Proof of Claim 2.
Case 2: $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \theta_n > 0$, which gives

$$\begin{aligned}
|\ (\boldsymbol{w}_{n+1}, \theta_{n+1})\ |^2 &= |\ (\boldsymbol{w}_n - \boldsymbol{x}^i, \theta_n + 1)\ |^2 \\
&= (\boldsymbol{w}_n - \boldsymbol{x}^i, \theta_n + 1) \cdot (\boldsymbol{w}_n - \boldsymbol{x}^i, \theta_n + 1) \\
&= (\boldsymbol{w}_n - \boldsymbol{x}^i) \cdot (\boldsymbol{w}_n - \boldsymbol{x}^i) + (\theta_n + 1)^2 \\
&= |\ \boldsymbol{w}_n\ |^2 - 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i + |\ \boldsymbol{x}^i\ |^2 + {\theta_n}^2 + 2\theta_n + 1 \\
&= |\ \boldsymbol{w}_n\ |^2 + {\theta_n}^2 - 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i + 2\theta_n + |\ \boldsymbol{x}^i\ |^2 + 1 \\
&= |\ (\boldsymbol{w}_n, \theta_n)\ |^2 - 2(\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \theta_n) + |\ \boldsymbol{x}^i\ |^2 + 1 \\
&\leq |\ (\boldsymbol{w}_n, \theta_n)\ |^2 + \max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1.
\end{aligned}$$

Then

$$|\ (\boldsymbol{w}_{j+1}, \theta_{j+1})\ |^2 \leq |\ (\boldsymbol{w}_j, \theta_j)\ |^2 + \max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1 \text{ for all } j \geq 0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} |\ (\boldsymbol{w}_{j+1}, \theta_{j+1})\ |^2 \leq \sum_{j=0}^{n-1} \Big\{ |\ (\boldsymbol{w}_j, \theta_j)\ |^2 + \max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1 \Big\}$$

and then

$$\sum_{j=0}^{n} |\ (\boldsymbol{w}_{j+1}, \theta_{j+1})\ |^2 + |\ (\boldsymbol{w}_n, \theta_n)\ |^2 \leq \sum_{j=0}^{n-1} |\ (\boldsymbol{w}_j, \theta_j)\ |^2 + n(\max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1).$$

This gives

$$\sum_{j=1}^{n-1} |\ (\boldsymbol{w}_j, \theta_j)\ |^2 + |\ (\boldsymbol{w}_n, \theta_n)\ |^2 \leq \sum_{j=1}^{n-1} |\ (\boldsymbol{w}_j, \theta_j)\ |^2 + |\ (\boldsymbol{w}_0, \theta_0)\ |^2 + n(\max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1)$$

and hence

$$|\ (\boldsymbol{w}_n, \theta_n)\ |^2 \leq |\ \boldsymbol{w}_0, \theta_0)\ |^2 + n(\max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1). \tag{5}$$

Case 3: $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \theta_n < 0$, which gives

$$
\begin{aligned}
\mid (\boldsymbol{w}_{n+1}, \theta_{n+1}) \mid^2 &= \mid (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1) \mid^2 = (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1) \cdot (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1) \\
&= (\boldsymbol{w}_n + \boldsymbol{x}^i) \cdot (\boldsymbol{w}_n + \boldsymbol{x}^i) + (\theta_n - 1)^2 \\
&= \mid \boldsymbol{w}_n \mid^2 + 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i + \mid \boldsymbol{x}^i \mid^2 + {\theta_n}^2 - 2\theta_n + 1 \\
&= \mid \boldsymbol{w}_n \mid^2 + {\theta_n}^2 + 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i - 2\theta_n + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_n, \theta_n) \mid^2 + 2(\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \theta_n) + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_n, \theta_n) \mid^2 + \max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1.
\end{aligned}
$$

Similarly to case 2, we can obtain that

$$
\mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1). \tag{6}
$$

Using (5) and (6), we obtain the desired result

$$
\mid (\boldsymbol{w}_n, \theta_n) \mid \leq \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)}.
$$

□

Theorem 3.2. *Let $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\right\}$ for some positive integers N and r. Let T be a function from Ω to $\{0, 1\}$. Let ϕ be a function from Ω to R^N. Assume that there exist $\boldsymbol{w}^* \in R^N$ and $\theta^* \in R$ such that for $1 \leq i \leq r$*

$$
T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* < 0, \end{cases} \quad \text{and } min_\Omega \mid \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* \mid > 0.
$$

Let $\boldsymbol{w} \in R^N$ and $\theta \in R$. Function $S_{\boldsymbol{w},\theta} : \Omega \to \{0, 1\}$ is defined such that for $1 \leq i \leq r$

$$
S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w} \cdot \phi(\boldsymbol{x}^i) - \theta > 0, \\ 0 & \text{if} \quad \boldsymbol{w} \cdot \phi(\boldsymbol{x}^i) - \theta \leq 0. \end{cases}
$$

Sequences $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are defined as follows:
Initial elements $\boldsymbol{w}_0 \in R^N$ and $\theta_0 \in R$ are random.
Let $n \geq 0$ and $I = \{1, \cdots, r\}$.

Case 1. For all $i \in I$ $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T(\boldsymbol{x}^i)$.

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n, \theta_n).$$

Case 2. For some $i \in I$ $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \leq l \leq i-1)$ and $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$, $T(\boldsymbol{x}^i) = 0$.

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n - \phi(\boldsymbol{x}^i), \theta_n + 1).$$

Case 3. For some $i \in I$ $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \leq l \leq i-1)$ and $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$, $T(\boldsymbol{x}^i) = 1$.

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n + \phi(\boldsymbol{x}^i), \theta_n - 1).$$

Then $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.

Proof. Using the definitions of the sequences, we have

$$(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) = \begin{cases} (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) & \text{(Case 1)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) & \text{(Case 2).} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) & \text{(Case 3)} \end{cases} \tag{7}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are infinite sequences.
Then (Case1) in 7 is not possible, which give that for $n \geq 0$

$$(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) = \begin{cases} (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) & \text{(Case 3)} \end{cases} \tag{8}$$

Claim 1. $(\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* \mid$ for all $n \geq 1$
Claim 2. $\mid (\boldsymbol{w}_n, \theta_n) \mid \leq \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1)}$ for all $n \geq 1$

Using Claim 2, Cauchy-schwarz inequality and Claim1, we have that for all n

$$\begin{aligned} \mid (\boldsymbol{w}^*, \theta^*) \mid \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1)} &\geq \mid (\boldsymbol{w}_n, \theta_n) \mid\mid (\boldsymbol{w}^*, \theta^*) \mid \\ &\geq (\boldsymbol{w}_n, \theta_n)(\boldsymbol{w}^*, \theta^*) \\ &\geq (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) \\ &\quad + n\min_\Omega \mid \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* \mid . \end{aligned}$$

Then

$$| (\boldsymbol{w}^*, \theta^*) | \geq \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* |}{\sqrt{| (\boldsymbol{w}_0, \theta_0) |^2 + n \left\{\max_\Omega | \phi(\boldsymbol{x}^i) |^2 + 1\right\}}},$$

which gives

$$\begin{aligned} | (\boldsymbol{w}^*, \theta^*) | &= \lim_{n\to\infty} | (\boldsymbol{w}^*, \theta^*) | \\ &\geq \lim_{n\to\infty} \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* |}{\sqrt{| (\boldsymbol{w}_0, \theta_0) |^2 + n \left\{\max_\Omega | \phi(\boldsymbol{x}^i) |^2 + 1\right\}}} = \infty. \end{aligned}$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.

Proof of Claim 1.

Case 2: $T(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* < 0$, which gives

$$\begin{aligned} (\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) \\ &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + | (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) | \\ &\geq (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* |. \end{aligned} \tag{9}$$

Case 3: $T(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* > 0$, which gives

$$\begin{aligned} (\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) \\ &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + | (\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^*) | \\ &\geq (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* |. \end{aligned} \tag{10}$$

Using (8), (9), and (10), we obtain

$$(\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* | .$$

Taking a summation, we have that

$$\sum_{j=0}^{n-1} (\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq \sum_{j=0}^{n-1} \left\{ (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* | \right\}$$

and then

$$\sum_{j=0}^{n-1} (\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq \sum_{j=0}^{n-1} (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega | \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* | .$$

This gives

$$\sum_{j=1}^{n-1}(\boldsymbol{w}_j,\theta_j)\cdot(\boldsymbol{w}^*,\theta^*)+(\boldsymbol{w}_n,\theta_n)\cdot(\boldsymbol{w}^*,\theta^*)$$
$$\geq(\boldsymbol{w}_0,\theta_0)\cdot(\boldsymbol{w}^*,\theta^*)+\sum_{j=1}^{n-1}(\boldsymbol{w}_j,\theta_j)\cdot(\boldsymbol{w}^*,\theta^*)+n\min_\Omega\mid\boldsymbol{w}^*\cdot\phi(\boldsymbol{x}^i)-\theta^*\mid$$

and hence

$$(\boldsymbol{w}_n,\theta_n)\cdot(\boldsymbol{w}^*,\theta^*)\geq(\boldsymbol{w}_0,\theta_0)\cdot(\boldsymbol{w}^*,\theta^*)+n\min_\Omega\mid\boldsymbol{w}^*\cdot\phi(\boldsymbol{x}^i)-\theta^*\mid.$$

Proof of Claim 2.
Case 2: $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)=1$. Then $\boldsymbol{w}_n\cdot\phi(\boldsymbol{x}^i)-\theta_n>0$, which gives

$$\begin{aligned}
\mid(\boldsymbol{w}_{n+1},\theta_{n+1})\mid^2 &= \mid(\boldsymbol{w}_n-\phi(\boldsymbol{x}^i),\theta_n+1)\mid^2\\
&=(\boldsymbol{w}_n-\phi(\boldsymbol{x}^i),\theta_n+1)\cdot(\boldsymbol{w}_n-\phi(\boldsymbol{x}^i),\theta_n+1)\\
&=(\boldsymbol{w}_n-\phi(\boldsymbol{x}^i))\cdot(\boldsymbol{w}_n-\phi(\boldsymbol{x}^i))+(\theta_n+1)^2\\
&=\mid\boldsymbol{w}_n\mid^2-2\boldsymbol{w}_n\cdot\phi(\boldsymbol{x}^i)+\mid\phi(\boldsymbol{x}^i)\mid^2+{\theta_n}^2+2\theta_n+1\\
&=\mid\boldsymbol{w}_n\mid^2+{\theta_n}^2-2\boldsymbol{w}_n\cdot\phi(\boldsymbol{x}^i)+2\theta_n+\mid\phi(\boldsymbol{x}^i)\mid^2+1\\
&=\mid(\boldsymbol{w}_n,\theta_n)\mid^2-2(\boldsymbol{w}_n\cdot\phi(\boldsymbol{x}^i)-\theta_n)+\mid\phi(\boldsymbol{x}^i)\mid^2+1\\
&\leq\mid(\boldsymbol{w}_n,\theta_n)\mid^2+\max_\Omega\mid\phi(\boldsymbol{x}^i)\mid^2+1.
\end{aligned}$$

Then

$$\mid(\boldsymbol{w}_{j+1},\theta_{j+1})\mid^2\leq\mid(\boldsymbol{w}_j,\theta_j)\mid^2+\max_\Omega\mid\phi(\boldsymbol{x}^i)\mid^2+1\text{ for all }j\geq0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1}\mid(\boldsymbol{w}_{j+1},\theta_{j+1})\mid^2\leq\sum_{j=0}^{n-1}\Big\{\mid(\boldsymbol{w}_j,\theta_j)\mid^2+\max_\Omega\mid\phi(\boldsymbol{x}^i)\mid^2+1\Big\}$$

and then

$$\sum_{j=0}^{n}\mid(\boldsymbol{w}_{j+1},\theta_{j+1})\mid^2+\mid(\boldsymbol{w}_n,\theta_n)\mid^2\leq\sum_{j=0}^{n-1}\mid(\boldsymbol{w}_j,\theta_j)\mid^2+n(\max_\Omega\mid\phi(\boldsymbol{x}^i)\mid^2+1).$$

This gives

$$\sum_{j=1}^{n-1} \mid (\boldsymbol{w}_j, \theta_j) \mid^2 + \mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \sum_{j=1}^{n-1} \mid (\boldsymbol{w}_j, \theta_j) \mid^2 + \mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1)$$

and hence

$$\mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \mid \boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1). \tag{11}$$

Case 3: $S_{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_n \cdot \phi(\boldsymbol{x}^i) - \theta_n < 0$, which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_{n+1}, \theta_{n+1}) \mid^2 &= \mid (\boldsymbol{w}_n + \phi(\boldsymbol{x}^i), \theta_n - 1) \mid^2 \\
&= (\boldsymbol{w}_n + \phi(\boldsymbol{x}^i), \theta_n - 1) \cdot (\boldsymbol{w}_n + \phi(\boldsymbol{x}^i), \theta_n - 1) \\
&= (\boldsymbol{w}_n + \phi(\boldsymbol{x}^i)) \cdot (\boldsymbol{w}_n + \phi(\boldsymbol{x}^i)) + (\theta_n - 1)^2 \\
&= \mid \boldsymbol{w}_n \mid^2 + 2\boldsymbol{w}_n \cdot \phi(\boldsymbol{x}^i) + \mid \phi(\boldsymbol{x}^i) \mid^2 + {\theta_n}^2 - 2\theta_n + 1 \\
&= \mid \boldsymbol{w}_n \mid^2 + {\theta_n}^2 + 2\boldsymbol{w}_n \cdot \phi(\boldsymbol{x}^i) - 2\theta_n + \mid \phi(\boldsymbol{x}^i) \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_n, \theta_n) \mid^2 + 2(\boldsymbol{w}_n \cdot \phi(\boldsymbol{x}^i) - \theta_n) + \mid \phi(\boldsymbol{x}^i) \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_n, \theta_n) \mid^2 + \max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1.
\end{aligned}$$

Similarly to Case 2, we can obtain

$$\mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1). \tag{12}$$

Using (11) and (12), we obtain the desired result

$$\mid (\boldsymbol{w}_n, \theta_n) \mid \leq \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1)}.$$

□

Theorem 3.3. *Let $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\right\}$ for some positive integers N and r. Let T be a function from Ω to $\{0, 1\}$. Let ψ be an increasing function from R to R. Assume that there exist $\boldsymbol{w}^* \in R^N$ and $\theta^* \in R$ such that for $1 \leq i \leq r$*

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i) - \theta^* > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i) - \theta^* < 0, \end{cases} \quad \text{and } min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) \mid > 0.$$

Let $\boldsymbol{w} \in R^N$ *and* $\theta \in R$*. Function* $S_{\boldsymbol{w},\theta} : \Omega \to \{0,1\}$ *is defined such that for* $1 \le i \le r$

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \textit{if} \quad \psi(\boldsymbol{w} \cdot \boldsymbol{x}^i) - \theta > 0, \\ 0 & \textit{if} \quad \psi(\boldsymbol{w} \cdot \boldsymbol{x}^i) - \theta \le 0. \end{cases}$$

Sequences $\{\boldsymbol{w}_n\}$ *and* $\{\theta_n\}$ *are defined as follows:*
Initial elements $\boldsymbol{w}_0 \in R^N$ *and* $\theta_0 \in R$ *are random.*
Let $n \ge 0$*, and* $I = \{1, \cdots, r\}$*.*

Case 1. For all $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T(\boldsymbol{x}^i)$*.*

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n, \theta_n).$$

Case 2. For some $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \le l \le i-1)$ *and* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$*,* $T(\boldsymbol{x}^i) = 0$*.*

$$\textit{Define } (\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1})) = (\boldsymbol{w}_n - \boldsymbol{x}^i, \psi^{-1}(\theta_n) + 1).$$

Case 3. For some $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \le l \le i-1)$ *and* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$*,* $T(\boldsymbol{x}^i) = 1$*.*

$$\textit{Define } (\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1})) = (\boldsymbol{w}_n + \boldsymbol{x}^i, \psi^{-1}(\theta_n) - 1).$$

Then $\{\boldsymbol{w}_n\}$ *and* $\{\theta_n\}$ *are finite sequences.*

Proof. Note that

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i) - \theta^* > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i) - \theta^* < 0, \end{cases}$$

which is equivalent to

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) < 0. \end{cases}$$

Similarly,

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}\cdot\boldsymbol{x}^i) - \theta > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w}\cdot\boldsymbol{x}^i) - \theta \leq 0, \end{cases}$$

which is equivalent to

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}\cdot\boldsymbol{x}^i - \psi^{-1}(\theta) > 0, \\ 0 & \text{if} \quad \boldsymbol{w}\cdot\boldsymbol{x}^i - \psi^{-1}(\theta) \leq 0. \end{cases}$$

Using the definitions of the sequences, we have
$(\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1}))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*))$

$$= \begin{cases} (\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*)) & \text{(Case 1)} \\ (\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*)) - (\boldsymbol{w}^*\cdot\boldsymbol{x}^i - \psi^{-1}(\theta^*)) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + (\boldsymbol{w}^*\cdot\boldsymbol{x}^i - \psi^{-1}(\theta^*)) & \text{(Case 3)} \end{cases} \tag{13}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are infinite sequences. Then (Case1) in 13 is not possible, which give that for $n \geq 0$

$$(\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1}))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*))$$
$$= \begin{cases} (\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*)) - (\boldsymbol{w}^*\cdot\boldsymbol{x}^i - \psi^{-1}(\theta^*)) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + (\boldsymbol{w}^*\cdot\boldsymbol{x}^i - \psi^{-1}(\theta^*)) & \text{(Case 3)} \end{cases} \tag{14}$$

Claim 1. $(\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_0, \psi^{-1}(\theta_0))\cdot(\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{w}^*\cdot\boldsymbol{x}^i - \psi^{-1}(\theta^*) \mid$ for all $n \geq 1$

Claim 2. $\mid (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \mid \leq \sqrt{\mid (\boldsymbol{w}_0, \psi^{-1}(\theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)}$ for all $n \geq 1$
Using Claim 2, Cauchy-schwarz inequality and Claim1, we have that for all n

$$\begin{aligned} &\mid (\boldsymbol{w}^*, \theta^*) \mid \sqrt{\mid (\boldsymbol{w}_0, \psi^{-1}(\theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}} \\ &\qquad \geq \mid (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \mid\mid (\boldsymbol{w}^*, \theta^*) \mid \\ &\qquad \geq (\boldsymbol{w}_n, \psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*, \theta^*) \\ &\qquad \geq (\boldsymbol{w}_0, \theta_0)\cdot(\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{w}^*\cdot\boldsymbol{x}^i - \psi^{-1}(\theta^*) \mid . \end{aligned}$$

Then

$$| (\boldsymbol{w}^*, \theta^*) | \geq \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega | \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) |}{\sqrt{| (\boldsymbol{w}_0, \psi^{-1}(\theta_0)) |^2 + n \left\{\max_\Omega | \boldsymbol{x}^i |^2 + 1\right\}}},$$

which gives

$$\begin{aligned} | (\boldsymbol{w}^*, \theta^*) | &= \lim_{n\to\infty} | (\boldsymbol{w}^*, \theta^*) | \\ &\geq \lim_{n\to\infty} \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega | \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) |}{\sqrt{| (\boldsymbol{w}_0, \psi^{-1}(\theta_0)) |^2 + n \left\{\max_\Omega | \boldsymbol{x}^i |^2 + 1\right\}}} = \infty. \end{aligned}$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.
Proof of Claim 1.
Case 2: $T(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) < 0$, which gives

$$\begin{aligned} &(\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1})) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) \\ &\quad = (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) - (\boldsymbol{w}^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta^*)) \\ &\quad = (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + | (\boldsymbol{w}^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta^*)) | \\ &\quad \geq (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + \min_\Omega | \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) |. \end{aligned} \tag{15}$$

Case 3: $T(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) > 0$, which gives

$$\begin{aligned} &(\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1})) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) \\ &\quad = (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + (\boldsymbol{w}^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta^*)) \\ &\quad = (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + | (\boldsymbol{w}^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta^*)) | \\ &\quad \geq (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + \min_\Omega | \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) |. \end{aligned} \tag{16}$$

Using (14), (15), and (16), we obtain

$$\begin{aligned} &(\boldsymbol{w}_{k+1}, \psi^{-1}(\theta_{k+1})) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) \\ &\qquad \geq (\boldsymbol{w}_k, \psi^{-1}(\theta_k)) \cdot (\boldsymbol{w}^*, \psi^{-1}(\theta^*)) + \min_\Omega | \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) |. \end{aligned}$$

Taking a summation, we have that

$$\sum_{j=0}^{n-1}(\boldsymbol{w}_{j+1},\psi^{-1}(\theta_{j+1}))\cdot(\boldsymbol{w}^*,\psi^{-1}(\theta^*))$$
$$\geq \sum_{j=0}^{n-1}\left\{(\boldsymbol{w}_k,\psi^{-1}(\theta_k))\cdot(\boldsymbol{w}^*,\psi^{-1}(\theta^*))+\min_\Omega \mid \boldsymbol{w}^*\cdot\boldsymbol{x}^i-\psi^{-1}(\theta^*)\mid\right\}$$

and then

$$\sum_{j=0}^{n-1}(\boldsymbol{w}_{j+1},\psi^{-1}(\theta_{j+1}))\cdot(\boldsymbol{w}^*,\psi^{-1}(\theta^*))$$
$$\geq \sum_{j=0}^{n-1}(\boldsymbol{w}_k,\psi^{-1}(\theta_k))\cdot(\boldsymbol{w}^*,\psi^{-1}(\theta^*))+n\min_\Omega \mid \boldsymbol{w}^*\cdot\boldsymbol{x}^i-\psi^{-1}(\theta^*)\mid.$$

This gives

$$\sum_{j=1}^{n-1}(\boldsymbol{w}_j,\psi^{-1}(\theta_j))\cdot(\boldsymbol{w}^*,\psi^{-1}(\theta^*))+(\boldsymbol{w}_n,\psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*,\theta^*)$$
$$\geq(\boldsymbol{w}_0,\psi^{-1}(\theta_0))\cdot(\boldsymbol{w}^*,\theta^*)+\sum_{j=1}^{n-1}(\boldsymbol{w}_j,\psi^{-1}(\theta_j))\cdot(\boldsymbol{w}^*,\psi^{-1}(\theta^*))$$
$$+n\min_\Omega \mid \boldsymbol{w}^*\cdot\boldsymbol{x}^i-(\psi^{-1}(\theta^*))\mid$$

and hence

$$(\boldsymbol{w}_n,\psi^{-1}(\theta_n))\cdot(\boldsymbol{w}^*,\theta^*)\geq(\boldsymbol{w}_0,\psi^{-1}(\theta_0))\cdot(\boldsymbol{w}^*,\theta^*)+n\min_\Omega \mid \boldsymbol{w}^*\cdot\boldsymbol{x}^i-\psi^{-1}(\theta^*)\mid$$

Proof of Claim 2.

Case 2: $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)=1$. Then $\boldsymbol{w}_n\cdot\boldsymbol{x}^i-\psi^{-1}(\theta_n)>0$, which gives

$$\begin{aligned}
\mid(\boldsymbol{w}_{n+1},\psi^{-1}(\theta_{n+1}))\mid^2 &= \mid(\boldsymbol{w}_n-\boldsymbol{x}^i,\psi^{-1}(\theta_n)+1)\mid^2\\
&=(\boldsymbol{w}_n-\boldsymbol{x}^i,\psi^{-1}(\theta_n)+1)\cdot(\boldsymbol{w}_n-\boldsymbol{x}^i,\psi^{-1}(\theta_n)+1)\\
&=(\boldsymbol{w}_n-\boldsymbol{x}^i)\cdot(\boldsymbol{w}_n-\boldsymbol{x}^i)+(\psi^{-1}(\theta_n)+1)^2\\
&=\mid\boldsymbol{w}_n\mid^2-2\boldsymbol{w}_n\cdot\boldsymbol{x}^i+\mid\boldsymbol{x}^i\mid^2+(\psi^{-1}(\theta_n))^2+2\psi^{-1}(\theta_n)+1\\
&=\mid\boldsymbol{w}_n\mid^2+(\psi^{-1}(\theta_n))^2-2\boldsymbol{w}_n\cdot\boldsymbol{x}^i+2\psi^{-1}(\theta_n)+\mid\boldsymbol{x}^i\mid^2+1\\
&=\mid(\boldsymbol{w}_n,\psi^{-1}(\theta_n))\mid^2-2(\boldsymbol{w}_n\cdot\boldsymbol{x}^i-\psi^{-1}(\theta_n))+\mid\boldsymbol{x}^i\mid^2+1\\
&\leq\mid(\boldsymbol{w}_n,\psi^{-1}(\theta_n))\mid^2+\max_\Omega\mid\boldsymbol{x}^i\mid^2+1.
\end{aligned}$$

Then

$$| (\boldsymbol{w}_{j+1}, \psi^{-1}(\theta_{j+1})) |^2 \leq | (\boldsymbol{w}_j, \psi^{-1}(\theta_j)) |^2 + \max_\Omega | \boldsymbol{x}^i |^2 + 1 \text{ for all } j \geq 0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} | (\boldsymbol{w}_{j+1}, \psi^{-1}(\theta_{j+1})) |^2 \leq \sum_{j=0}^{n-1} \Big\{ | (\boldsymbol{w}_j, \psi^{-1}(\theta_j)) |^2 + \max_\Omega | \boldsymbol{x}^i |^2 + 1 \Big\}$$

and then

$$\sum_{j=0}^{n-1} | (\boldsymbol{w}_{j+1}, \psi^{-1}(\theta_{j+1})) |^2 \leq \sum_{j=0}^{n-1} | (\boldsymbol{w}_j, \psi^{-1}(\theta_j)) |^2 + n\max_\Omega | \boldsymbol{x}^i |^2 + 1.$$

This gives

$$\begin{aligned} &\sum_{j=1}^{n-1} | (\boldsymbol{w}_j, \psi^{-1}(\theta_j)) |^2 + | (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) |^2 \\ &\qquad \leq \sum_{j=1}^{n-1} | (\boldsymbol{w}_j, \psi^{-1}(\theta_j)) |^2 + | (\boldsymbol{w}_0, \psi^{-1}(\theta_0)) |^2 + n(\max_\Omega | \boldsymbol{x}^i |^2 + 1) \end{aligned}$$

and hence

$$| (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) |^2 \leq | (\boldsymbol{w}_0, \psi^{-1}(\theta_0)) |^2 + n(\max_\Omega | \boldsymbol{x}^i |^2 + 1). \tag{17}$$

Case 3: $S_{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_n) < 0$, which gives

$$\begin{aligned} | (\boldsymbol{w}_{n+1}, \psi^{-1}(\theta_{n+1})) |^2 &= | (\boldsymbol{w}_n - \boldsymbol{x}^i, \psi^{-1}(\theta_n) - 1) |^2 \\ &= (\boldsymbol{w}_n - \boldsymbol{x}^i, \psi^{-1}(\theta_n) - 1) \cdot (\boldsymbol{w}_n - \boldsymbol{x}^i, \psi^{-1}(\theta_n) - 1) \\ &= (\boldsymbol{w}_n + \boldsymbol{x}^i) \cdot (\boldsymbol{w}_n + \boldsymbol{x}^i) + (\psi^{-1}(\theta_n) - 1)^2 \\ &= | \boldsymbol{w}_n |^2 - 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i + | \boldsymbol{x}^i |^2 + (\psi^{-1}(\theta_n))^2 - 2\psi^{-1}(\theta_n) + 1 \\ &= | \boldsymbol{w}_n |^2 + (\psi^{-1}(\theta_n))^2 + 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i + 2\psi^{-1}(\theta_n) - | \boldsymbol{x}^i |^2 + 1 \\ &= | (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) |^2 + 2(\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_n)) + | \boldsymbol{x}^i |^2 + 1 \\ &\leq | (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) |^2 + \max_\Omega | \boldsymbol{x}^i |^2 + 1. \end{aligned}$$

Similarly to case 2 we can obtain

$$| (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) |^2 \leq | (\boldsymbol{w}_0, \psi^{-1}(\theta_0)) |^2 + n(\max_\Omega | \boldsymbol{x}^i |^2 + 1). \tag{18}$$

Using (17) and (18), we obtain the desired result

$$| (\boldsymbol{w}_n, \psi^{-1}(\theta_n)) | \leq \sqrt{| (\boldsymbol{w}_0, \psi^{-1}(\theta_0)) |^2 + n(\max_\Omega | \boldsymbol{x}^i |^2 + 1)}.$$

□

Theorem 3.4. *Let* $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\right\}$ *for some positive integers* N *and* r*. Let* T *be a function from* Ω *to* $\{0, 1\}$*. Let* ψ *be an increasing function from* R *to* R*. Assume that there exist* $\boldsymbol{w}^* \in R^N$ *and* $\theta^* \in R$ *such that for* $1 \leq i \leq r$

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) < 0, \end{cases} \quad \text{and } min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* \mid > 0.$$

Let $\boldsymbol{w} \in R^N$ *and* $\theta \in R$*. Function* $S_{\boldsymbol{w},\theta} : \Omega \to \{0, 1\}$ *is defined such that for* $1 \leq i \leq r$

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w} \cdot \boldsymbol{x}^i - \theta) > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w} \cdot \boldsymbol{x}^i - \theta) \leq 0. \end{cases}$$

Sequences $\{\boldsymbol{w}_n\}$ *and* $\{\theta_n\}$ *are defined as follows:*
Initial elements $\boldsymbol{w}_0 \in R^N$ *and* $\theta_0 \in R$ *are random.*
Let $n \geq 0$ *and* $I = \{1, \cdots, r\}$*.*

Case 1. For all $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T(\boldsymbol{x}^i)$*.*

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n, \theta_n).$$

Case 2. For some $i \in I$*,* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \leq l \leq i-1)$ *and* $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$*,* $T(\boldsymbol{x}^i) = 0$*.*

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n - \boldsymbol{x}^i, \theta_n + 1).$$

Case 3. For some $i \in I$, $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T(\boldsymbol{x}^l)$ $(1 \le l \le i-1)$ and $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$, $T(\boldsymbol{x}^i) = 1$.

$$\textit{Define } (\boldsymbol{w}_{n+1}, \theta_{n+1}) = (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1).$$

Then $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.

Proof. Note that the assumption of Theorem 3.4 is the existence of a hyper-surface $\psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) = 0 (x \in R)$ that separates the given set $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \le i \le r\right\}$, where $\boldsymbol{w}^* \in R^N$ and increasing function ψ from R to R. Then

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) < 0, \end{cases}$$

which is equivalent to

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* < 0. \end{cases}$$

Similarly,

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w} \cdot \boldsymbol{x}^i - \theta) > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w} \cdot \boldsymbol{x}^i - \theta) \le 0, \end{cases}$$

which is equivalent to

$$S_{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{x}^i - \theta > 0, \\ 0 & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{x}^i - \theta \le 0. \end{cases}$$

Using the definitions of the sequences, we have

$$(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) = \begin{cases} (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) & \text{(Case 1)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 3)} \end{cases} \tag{19}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are infinite sequences. Then (Case1) in (19) is not possible, which give that for $n \ge 0$

$$(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) = \begin{cases} (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 2)} \\ (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) & \text{(Case 3)} \end{cases} \tag{20}$$

Claim 1. $(\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid$ for all $n \geq 1$

Claim 2. $\mid (\boldsymbol{w}_n, \theta_n) \mid \leq \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}$ for all $n \geq 1$
Using Claim 2, Cauchy-schwarz inequality and Claim1, we have

$$\begin{aligned}
\mid (\boldsymbol{w}^*, \theta^*) \mid \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)} \geq & \mid (\boldsymbol{w}_n, \theta_n) \mid\mid (\boldsymbol{w}^*, \theta^*) \mid \\
\geq & (\boldsymbol{w}_n, \theta_n)(\boldsymbol{w}^*, \theta^*) \\
\geq & (\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) \\
& + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid
\end{aligned}$$

for all n. Then

$$\mid (\boldsymbol{w}^*, \theta^*) \mid \geq \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid}{\sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}},$$

which gives

$$\mid (\boldsymbol{w}^*, \theta^*) \mid = \lim_{n\to\infty} \mid (\boldsymbol{w}^*, \theta^*) \mid \geq \lim_{n\to\infty} \frac{(\boldsymbol{w}_0, \theta_0) \cdot (\boldsymbol{w}^*, \theta^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid}{\sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}} = \infty.$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ are finite sequences.

Proof of Claim 1.
Case 2: $T(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* < 0$, which gives

$$\begin{aligned}
(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) \quad &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \\
&= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \mid \\
&\geq (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid.
\end{aligned} \tag{21}$$

Case 3: $T(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* > 0$, which gives

$$\begin{aligned}
(\boldsymbol{w}_{n+1}, \theta_{n+1}) \cdot (\boldsymbol{w}^*, \theta^*) \quad &= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \\
&= (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^*) \mid \\
&\geq (\boldsymbol{w}_n, \theta_n) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid.
\end{aligned} \tag{22}$$

Using (20), (21), and (22), we obtain

$$(\boldsymbol{w}_{j+1}, \theta_{j+1}) \cdot (\boldsymbol{w}^*, \theta^*) \geq (\boldsymbol{w}_j, \theta_j) \cdot (\boldsymbol{w}^*, \theta^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* - \theta^* \mid .$$

Taking a summation, we have

$$\sum_{j=0}^{n-1}(\boldsymbol{w}_{j+1},\theta_{j+1})\cdot(\boldsymbol{w}^*,\theta^*) \geq \sum_{j=0}^{n-1}\left\{(\boldsymbol{w}_j,\theta_j)\cdot(\boldsymbol{w}^*,\theta^*)+\min_\Omega \mid \boldsymbol{x}^i\cdot\boldsymbol{w}^*-\theta^* \mid\right\}$$

and then

$$\sum_{j=0}^{n-1}(\boldsymbol{w}_{j+1},\theta_{j+1})\cdot(\boldsymbol{w}^*,\theta^*) \geq \sum_{j=0}^{n-1}(\boldsymbol{w}_j,\theta_j)\cdot(\boldsymbol{w}^*,\theta^*)+n\min_\Omega \mid \boldsymbol{x}^i\cdot\boldsymbol{w}^*-\theta^* \mid .$$

This gives

$$\begin{aligned}&\sum_{j=1}^{n-1}(\boldsymbol{w}_j,\theta_j)\cdot(\boldsymbol{w}^*,\theta^*)+(\boldsymbol{w}_n,\theta_n)\cdot(\boldsymbol{w}^*,\theta^*)\\ &\qquad\geq(\boldsymbol{w}_0,\theta_0)\cdot(\boldsymbol{w}^*,\theta^*)+\sum_{j=1}^{n-1}(\boldsymbol{w}_j,\theta_j)\cdot(\boldsymbol{w}^*,\theta^*)+n\min_\Omega \mid \boldsymbol{x}^i\cdot\boldsymbol{w}^*-\theta^* \mid\end{aligned}$$

and hence

$$(\boldsymbol{w}_n,\theta_n)\cdot(\boldsymbol{w}^*,\theta^*)\geq(\boldsymbol{w}_0,\theta_0)\cdot(\boldsymbol{w}^*,\theta^*)+n\min_\Omega \mid \boldsymbol{x}^i\cdot\boldsymbol{w}^*-\theta^* \mid .$$

Proof of Claim 2.
Case 2: $S_{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)=1$. Then $\boldsymbol{w}_n\cdot\boldsymbol{x}^i-\theta_n>0$, which gives

$$\begin{aligned}\mid(\boldsymbol{w}_{n+1},\theta_{n+1})\mid^2 &= \mid(\boldsymbol{w}_n-\boldsymbol{x}^i,\theta_n+1)\mid^2=(\boldsymbol{w}_n-\boldsymbol{x}^i,\theta_n+1)\cdot(\boldsymbol{w}_n-\boldsymbol{x}^i,\theta_n+1)\\ &=(\boldsymbol{w}_n-\boldsymbol{x}^i)\cdot(\boldsymbol{w}_n-\boldsymbol{x}^i)+(\theta_n+1)^2\\ &=\mid\boldsymbol{w}_n\mid^2-2\boldsymbol{w}_n\cdot\boldsymbol{x}^i+\mid\boldsymbol{x}^i\mid^2+{\theta_n}^2+2\theta_n+1\\ &=\mid\boldsymbol{w}_n\mid^2+{\theta_n}^2-2\boldsymbol{w}_n\cdot\boldsymbol{x}^i+2\theta_n+\mid\boldsymbol{x}^i\mid^2+1\\ &=\mid(\boldsymbol{w}_n,\theta_n)\mid^2-2(\boldsymbol{w}_n\cdot\boldsymbol{x}^i-\theta_n)+\mid\boldsymbol{x}^i\mid^2+1\\ &\leq\mid(\boldsymbol{w}_n,\theta_n)\mid^2+\max_\Omega\mid\boldsymbol{x}^i\mid^2+1.\end{aligned}$$

Then

$$\mid(\boldsymbol{w}_{j+1},\theta_{j+1})\mid^2\leq\mid(\boldsymbol{w}_j,\theta_j)\mid^2+\max_\Omega\mid\boldsymbol{x}^i\mid^2+1 \text{ for all } j\geq 0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} \mid (\boldsymbol{w}_{j+1}, \theta_{j+1}) \mid^2 \leq \sum_{j=0}^{n-1} \Big\{ \mid (\boldsymbol{w}_j, \theta_j) \mid^2 + \max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1 \Big\}$$

and then

$$\sum_{j=0}^{n} \mid (\boldsymbol{w}_{j+1}, \theta_{j+1}) \mid^2 + \mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \sum_{j=0}^{n-1} \mid (\boldsymbol{w}_j, \theta_j) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1).$$

This gives

$$\sum_{j=1}^{n-1} \mid (\boldsymbol{w}_j, \theta_j) \mid^2 + \mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \sum_{j=1}^{n-1} \mid (\boldsymbol{w}_j, \theta_j) \mid^2 + \mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)$$

and hence

$$\mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \mid \boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1). \tag{23}$$

Case 3: $S_{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \theta_n < 0$, which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_{n+1}, \theta_{n+1}) \mid^2 &= \mid (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1) \mid^2 = (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1) \cdot (\boldsymbol{w}_n + \boldsymbol{x}^i, \theta_n - 1) \\
&= (\boldsymbol{w}_n + \boldsymbol{x}^i) \cdot (\boldsymbol{w}_n + \boldsymbol{x}^i) + (\theta_n - 1)^2 \\
&= \mid \boldsymbol{w}_n \mid^2 + 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i + \mid \boldsymbol{x}^i \mid^2 + {\theta_n}^2 - 2\theta_n + 1 \\
&= \mid \boldsymbol{w}_n \mid^2 + {\theta_n}^2 + 2\boldsymbol{w}_n \cdot \boldsymbol{x}^i - 2\theta_n + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_n, \theta_n) \mid^2 + 2(\boldsymbol{w}_n \cdot \boldsymbol{x}^i - \theta_n) + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_n, \theta_n) \mid^2 + \max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1.
\end{aligned}$$

Similarly to case 2, we can obtain

$$\mid (\boldsymbol{w}_n, \theta_n) \mid^2 \leq \mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1). \tag{24}$$

Using (23) and (24), we obtain the desired result

$$\mid (\boldsymbol{w}_n, \theta_n) \mid \leq \sqrt{\mid (\boldsymbol{w}_0, \theta_0) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)}.$$

□

3.2 Extensions of the algorithm

In this section we present extensions of the perceptron convergence algorithm to networks of two or three layers with m output nodes.

Theorem 3.5 (Extension of Theorem 3.1). *Let I be the set of positive integers less than or equal to r and $\Omega = \left\{\boldsymbol{x}^i \in R^N | i \in I\right\}$ for some positive integers N and r. Assume that there exist $(\boldsymbol{w}_m^*, \theta_m^*) \in R^N \times R$ satisfying $\min_{\boldsymbol{x}^i \in \Omega} |\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*| > 0$ for a positive integer M and all $1 \leq m \leq M$. Let T_m be a function from Ω to $\{a_m, b_m\}$ for real numbers a_m and b_m $(a_m > b_m)$ such that for all $i \in I$*

$$T_m(x^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^* > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^* < 0. \end{cases}$$

Let $(\boldsymbol{w}_m, \theta_m) \in R^N \times R$ and define $S_m^{\boldsymbol{w},\theta} : R^N \to \{a_m, b_m\}$ as

$$S_m^{\boldsymbol{w},\theta}(\boldsymbol{x}) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x} - \theta_m > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x} - \theta_m < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x} - \theta_m = 0, \end{cases}$$

where $randr(\{a_m, b_m\})$ is a number randomly chosen from $\{a_m, b_m\}$. Sequences $\{\boldsymbol{w}_{m,n}\}_{n=0}^{\infty}$ and $\{\theta_{m,n}\}_{n=0}^{\infty}$ are defined recursively as follows:

i) Initial elements $\boldsymbol{w}_{m,0} \in R^N$ and $\theta_{m,0} \in R$ are randomly sampled.

ii) $\boldsymbol{w}_{m,n+1}$ and $\theta_{m,n+1}$ $(n \geq 0)$ are defined depending on the values $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)$ and $T_m(\boldsymbol{x}^i)$.

a) Case 1: $(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n}, \theta_{m,n})$ if $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T_m(\boldsymbol{x}^i)$ for all $i \in I$.

b) Case 2: $(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \theta_{m,n} + 1)$ if $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^\ell) = T_m(\boldsymbol{x}^\ell)$ and $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = a_m, T_m(\boldsymbol{x}^i) = b_m$ for some $i \in I$ and all $1 \leq \ell \leq i-1$.

c) Case 3: $(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1)$ if $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^\ell) = T_m(\boldsymbol{x}^\ell)$ and $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = b_m, T_m(\boldsymbol{x}^i) = a_m$ for some $i \in I$ and all $1 \leq \ell \leq i-1$.

Then $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are finite sequences for any fixed m.

Proof. Using the definitions of the sequences, we have

$$(\boldsymbol{w}_{m,n+1},\theta_{m,n+1})\cdot(\boldsymbol{w}_m^*,\theta_m^*)=\begin{cases}(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*) & \text{(Case 1)}\\ (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)-(\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*) & \text{(Case 2)}\\ (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+(\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*) & \text{(Case 3)}\end{cases} \tag{25}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are infinite sequences. Then Case1 in (25) is not possible, which give that for $n\geq 0$

$$(\boldsymbol{w}_{m,n+1},\theta_{m,n+1})\cdot(\boldsymbol{w}_m^*,\theta_m^*)=\begin{cases}(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)-(\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*) & \text{(Case 2)}\\ (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+(\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*) & \text{(Case 3)}\end{cases} \tag{26}$$

So, we can obtain two claims:
Claim 1. $(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)\geq(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\delta$ for all $n\geq 1$

Claim 2. $\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid\leq\sqrt{\mid(\boldsymbol{w}_{m,0},\theta_{m,0})\mid^2+n(\chi+1)}$ for all $n\geq 1$
Here

$$\delta=\min_{\Omega}\mid\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*\mid,\ \chi=\max_{\Omega}\mid\boldsymbol{x}^i\mid^2.$$

Using Claim 2, Cauchy-schwarz inequality and Claim1, we have

$$\begin{aligned}\mid(\boldsymbol{w}_m^*,\theta_m^*)\mid\sqrt{\mid(\boldsymbol{w}_{m,0},\theta_{m,0})\mid^2+n(\chi+1)}&\geq\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid\mid(\boldsymbol{w}_m^*,\theta_m^*)\mid\\&\geq(\boldsymbol{w}_{m,n},\theta_{m,n})(\boldsymbol{w}_m^*,\theta_m^*)\\&\geq(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\delta\end{aligned}$$

for all n. Then

$$\mid(\boldsymbol{w}_m^*,\theta_m^*)\mid\geq\frac{(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\delta}{\sqrt{\mid(\boldsymbol{w}_{m,0},\theta_{m,0})\mid^2+n(\chi+1)}},$$

which gives

$$\mid(\boldsymbol{w}_m^*,\theta_m^*)\mid=\lim_{n\to\infty}\mid(\boldsymbol{w}_m^*,\theta_m^*)\mid\geq\lim_{n\to\infty}\frac{(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\delta}{\sqrt{\mid(\boldsymbol{w}_{m,0},\theta_{m,0})\mid^2+n(\chi+1)}}=\infty.$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are finite sequences.

Proof of Claim 1.
Case 2: $T_m(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* < 0$, which gives

$$\begin{aligned}(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \\ &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \mid \quad (27) \\ &\geq (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \delta.\end{aligned}$$

Case 3: $T(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* > 0$, which gives

$$\begin{aligned}(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \\ &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \mid \quad (28) \\ &\geq (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \delta.\end{aligned}$$

Using (26), (27), and (28), we obtain

$$(\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \delta.$$

Taking a summation, we have that

$$\sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq \sum_{j=0}^{n-1} \{(\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \delta\}$$

and then

$$\sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq \sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\delta.$$

This gives

$$\begin{aligned}&\sum_{j=1}^{n-1} (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\boldsymbol{w}_{m,n}, \theta_n) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \\ &\qquad \geq (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \sum_{j=1}^{n-1} (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\delta\end{aligned}$$

and hence

$$(\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\delta.$$

Proof of Claim 2.
Case 2: $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \theta_{m,n} > 0$, which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \mid^2 &= \mid (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \theta_{m,n} + 1) \mid^2 \\
&= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \theta_{m,n} + 1) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \theta_{m,n} + 1) \\
&= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i) + (\theta_{m,n} + 1)^2 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + \mid \boldsymbol{x}^i \mid^2 + {\theta_{m,n}}^2 + 2\theta_{m,n} + 1 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + {\theta_{m,n}}^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + 2\theta_{m,n} + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 - 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \theta_{m,n}) + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + (\chi + 1).
\end{aligned}$$

Then

$$\mid (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \mid^2 \leq \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + (\chi + 1) \text{ for all } j \geq 0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} \mid (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \mid^2 \leq \sum_{j=0}^{n-1} \left\{ \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + (\chi + 1) \right\}$$

and then

$$\sum_{j=0}^{n} \mid (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \mid^2 + \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \sum_{j=0}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + n\,(\chi + 1).$$

This gives

$$\sum_{j=1}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \sum_{j=1}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\,(\chi + 1)$$

and hence

$$\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \mid \boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\,(\chi + 1). \tag{29}$$

Case 3: $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \theta_{m,n} < 0$, which gives

$$
\begin{aligned}
\mid (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \mid^2 &= \mid (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1) \mid^2 \\
&= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1) \\
&= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i) + (\theta_{m,n} - 1)^2 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + \mid \boldsymbol{x}^i \mid^2 + {\theta_{m,n}}^2 - 2\theta_{m,n} + 1 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + {\theta_{m,n}}^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - 2\theta_{m,n} + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \theta_{m,n}) + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + (\chi + 1).
\end{aligned}
$$

Similarly to case 2, we can obtain

$$
\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\,(\chi + 1). \tag{30}
$$

Using (29) and (30), we obtain the desired result

$$
\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid \leq \sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\,(\chi + 1)}.
$$

□

Theorem 3.6 (Extension of Theorem 3.2). *Let* $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\right\}$ *for some positive integers* N *and* r*. Assume that there exist* $(\boldsymbol{w}_m^*, \theta_m^*) \in R^N \times R (1 \leq m \leq M)$ *and function* ϕ *from* Ω *to* R^N *satisfying* $\min_{\boldsymbol{x}^i \in \Omega} |\boldsymbol{w}_m^* \cdot \phi(\boldsymbol{x}^i) - \theta_m^*| > 0$ *for all* $1 \leq m \leq M$*. Let* $T_{\boldsymbol{w}_m^*, \theta_m^*}$ *be a function from* Ω *to* $\{a_m, b_m\}$ *for any fixed real numbers* a_m *and* b_m $(a_m > b_m)$ *such that for all* $1 \leq i \leq r$

$$
T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \phi(\boldsymbol{x}^i) - \theta_m^* > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \phi(\boldsymbol{x}^i) - \theta_m^* < 0. \end{cases}
$$

Let $(\boldsymbol{w}_m, \theta_m) \in R^N \times R$ *and define* $S_{\boldsymbol{w}_m, \theta_m} : R^N \to \{a_m, b_m\}$ *as*

$$
S_m^{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m \cdot \phi(\boldsymbol{x}^i) - \theta_m > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m \cdot \phi(\boldsymbol{x}^i) - \theta_m < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \boldsymbol{w}_m \cdot \phi(\boldsymbol{x}^i) - \theta_m = 0, \end{cases}
$$

where $randr(\{a_m, b_m\})$ *is a number randomly chosen from the set* $\{a_m, b_m\}$*. Sequences* $\{\boldsymbol{w}_{m,n}\}$ *and* $\{\theta_{m,n}\}$ *are defined as follows.*

Initial elements $\boldsymbol{w}_{m,0} \in R^N$ *and* $\theta_{m,0} \in R$ *are random.* $\boldsymbol{w}_{m,n+1}$ *and* $\theta_{m,n+1}$ $(n \geq 0)$ *are defined depending on the values of* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)$ *and* $T_m(\phi(\boldsymbol{x}^i))$.
Case 1. For all $i \in I$, $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i)$.

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n}, \theta_{m,n}).$$

Case 2. For some $i \in I$, $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ *and* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = a_m, T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = b_m$.

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} - \phi(\boldsymbol{x}^i), \theta_{m,n} + 1).$$

Case 3. For some $i \in I$, $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ *and* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = b_m, T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = a_m$.

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} + \phi(\boldsymbol{x}^i), \theta_{m,n} - 1).$$

Then $\{\boldsymbol{w}_{m,n}\}$ *and* $\{\theta_{m,n}\}$ *are finite sequences.*

Proof. Using the definitions of the sequences, we have

$$(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) = \begin{cases} (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) & \text{(Case 1)} \\ (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 2)} \\ (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 3)} \end{cases} \tag{31}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are infinite sequences.
Then (Case1) in 31 is not possible, which give that for $n \geq 0$

$$(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) = \begin{cases} (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 2)} \\ (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 3)} \end{cases} \tag{32}$$

Claim 1. $(\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n \min_\Omega \mid \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* \mid$ for all $n \geq 1$

Claim 2. $\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid \leq \sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1)}$ for all $n \geq 1$

Using Claim 2, Cauchy-schwarz inequality and Claim1, we have

$$\begin{aligned}
| (\boldsymbol{w}_m^*, \theta_m^*) | \sqrt{| (\boldsymbol{w}_{m,0}, \theta_{m,0}) |^2 + n(\max_\Omega | \phi(\boldsymbol{x}^i) |^2 + 1)} \geq & | (\boldsymbol{w}_{m,n}, \theta_{m,n}) || (\boldsymbol{w}_m^*, \theta_m^*) | \\
\geq & (\boldsymbol{w}_{m,n}, \theta_{m,n})(\boldsymbol{w}_m^*, \theta_m^*) \\
\geq & (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \\
& + n\min_\Omega | \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* |
\end{aligned}$$

for all n. Then

$$| (\boldsymbol{w}_m^*, \theta_m^*) | \geq \frac{(\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega | \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* |}{\sqrt{| (\boldsymbol{w}_{m,0}, \theta_{m,0}) |^2 + n \left\{\max_\Omega | \phi(\boldsymbol{x}^i) |^2 + 1\right\}}},$$

which gives

$$\begin{aligned}
| (\boldsymbol{w}_m^*, \theta_m^*) | &= \lim_{n\to\infty} | (\boldsymbol{w}_m^*, \theta_m^*) | \\
&\geq \lim_{n\to\infty} \frac{(\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega | \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* |}{\sqrt{| (\boldsymbol{w}_{m,0}, \theta_{m,0}) |^2 + n \left\{\max_\Omega | \phi(\boldsymbol{x}^i) |^2 + 1\right\}}} = \infty.
\end{aligned}$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are finite sequences.

Proof of Claim 1.
Case 2: $T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = 0$. Then $\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* < 0$, which gives

$$\begin{aligned}
(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) \\
&= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + | (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) | \\
&\geq (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega | \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* |.
\end{aligned} \tag{33}$$

Case 3: $T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = 1$. Then $\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* > 0$, which gives

$$\begin{aligned}
(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) \\
&= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + | (\phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^*) | \\
&\geq (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega | \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* |.
\end{aligned} \tag{34}$$

Using (32), (33) and (34), we obtain

$$(\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega | \phi(\boldsymbol{x}^i) \cdot \boldsymbol{w}_m^* - \theta_m^* | .$$

Taking a summation, we have that

$$\sum_{j=0}^{n-1}(\boldsymbol{w}_{m,j+1},\theta_{m,j+1})\cdot(\boldsymbol{w}_m^*,\theta_m^*)\geq\sum_{j=0}^{n-1}\left\{(\boldsymbol{w}_{m,j},\theta_{m,j})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+\min_\Omega\mid\phi(\boldsymbol{x}^i)\cdot\boldsymbol{w}_m^*-\theta_m^*\mid\right\}$$

and then

$$\sum_{j=0}^{n-1}(\boldsymbol{w}_{m,j+1},\theta_{m,j+1})\cdot(\boldsymbol{w}_m^*,\theta_m^*)\geq\sum_{j=0}^{n-1}(\boldsymbol{w}_{m,j},\theta_{m,j})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\min_\Omega\mid\phi(\boldsymbol{x}^i)\cdot\boldsymbol{w}_m^*-\theta_m^*\mid.$$

This gives

$$\begin{aligned}&\sum_{j=1}^{n-1}(\boldsymbol{w}_{m,j},\theta_j)\cdot(\boldsymbol{w}_m^*,\theta_m^*)+(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)\\&\qquad\geq(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+\sum_{j=1}^{n-1}(\boldsymbol{w}_{m,j},\theta_{m,j})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\min_\Omega\mid\phi(\boldsymbol{x}^i)\cdot\boldsymbol{w}_m^*-\theta_m^*\mid\end{aligned}$$

and hence

$$(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)\geq(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\min_\Omega\mid\phi(\boldsymbol{x}^i)\cdot\boldsymbol{w}_m^*-\theta_m^*\mid.$$

Proof of Claim 2.
Case 2: $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)=1$. Then $\boldsymbol{w}_{m,n}\cdot\phi(\boldsymbol{x}^i)-\theta_{m,n}>0$, which gives

$$\begin{aligned}\mid(\boldsymbol{w}_{m,n+1},\theta_{m,n+1})\mid^2&=\mid(\boldsymbol{w}_{m,n}-\phi(\boldsymbol{x}^i),\theta_{m,n}+1)\mid^2\\&=(\boldsymbol{w}_{m,n}-\phi(\boldsymbol{x}^i),\theta_{m,n}+1)\cdot(\boldsymbol{w}_{m,n}-\phi(\boldsymbol{x}^i),\theta_{m,n}+1)\\&=(\boldsymbol{w}_{m,n}-\phi(\boldsymbol{x}^i))\cdot(\boldsymbol{w}_{m,n}-\phi(\boldsymbol{x}^i))+(\theta_{m,n}+1)^2\\&=\mid\boldsymbol{w}_{m,n}\mid^2-2\boldsymbol{w}_{m,n}\cdot\phi(\boldsymbol{x}^i)+\mid\phi(\boldsymbol{x}^i)\mid^2+{\theta_{m,n}}^2+2\theta_{m,n}+1\\&=\mid\boldsymbol{w}_{m,n}\mid^2+{\theta_{m,n}}^2-2\boldsymbol{w}_{m,n}\cdot\phi(\boldsymbol{x}^i)+2\theta_{m,n}+\mid\phi(\boldsymbol{x}^i)\mid^2+1\\&=\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid^2-2(\boldsymbol{w}_{m,n}\cdot\phi(\boldsymbol{x}^i)-\theta_{m,n})+\mid\phi(\boldsymbol{x}^i)\mid^2+1\\&\leq\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid^2+\max_\Omega\mid\phi(\boldsymbol{x}^i)\mid^2+1.\end{aligned}$$

Then

$$\mid(\boldsymbol{w}_{m,j+1},\theta_{m,j+1})\mid^2\leq\mid(\boldsymbol{w}_{m,j},\theta_{m,j})\mid^2+\max_\Omega\mid\phi(\boldsymbol{x}^i)\mid^2+1\text{ for all }j\geq0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} \mid (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \mid^2 \leq \sum_{j=0}^{n-1} \Big\{ \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + \max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1 \Big\}$$

and then

$$\sum_{j=0}^{n} \mid (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \mid^2 + \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \sum_{j=0}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1).$$

This gives

$$\begin{aligned}
&\sum_{j=1}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_j) \mid^2 + \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \\
&\qquad\qquad \leq \sum_{j=1}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1)
\end{aligned}$$

and hence

$$\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \mid \boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1). \tag{35}$$

Case 3: $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_{m,n} \cdot \phi(\boldsymbol{x}^i) - \theta_{m,n} < 0$, which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \mid^2 &= \mid (\boldsymbol{w}_{m,n} + \phi(\boldsymbol{x}^i), \theta_{m,n} - 1) \mid^2 \\
&= (\boldsymbol{w}_{m,n} + \phi(\boldsymbol{x}^i), \theta_{m,n} - 1) \cdot (\boldsymbol{w}_{m,n} + \phi(\boldsymbol{x}^i), \theta_{m,n} - 1) \\
&= (\boldsymbol{w}_{m,n} + \phi(\boldsymbol{x}^i)) \cdot (\boldsymbol{w}_{m,n} + \phi(\boldsymbol{x}^i)) + (\theta_{m,n} - 1)^2 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + 2\boldsymbol{w}_{m,n} \cdot \phi(\boldsymbol{x}^i) + \mid \phi(\boldsymbol{x}^i) \mid^2 + {\theta_{m,n}}^2 - 2\theta_{m,n} + 1 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + {\theta_{m,n}}^2 + 2\boldsymbol{w}_{m,n} \cdot \phi(\boldsymbol{x}^i) - 2\theta_{m,n} + \mid \phi(\boldsymbol{x}^i) \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + 2(\boldsymbol{w}_{m,n} \cdot \phi(\boldsymbol{x}^i) - \theta_{m,n}) + \mid \phi(\boldsymbol{x}^i) \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + \max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1.
\end{aligned}$$

Similarly to case 2, we can obtain that

$$\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \phi(\boldsymbol{x}^i) \mid^2 + 1). \tag{36}$$

Using (35) and (36), we obtain the desired result

$$| (\boldsymbol{w}_{m,n}, \theta_{m,n}) | \leq \sqrt{| (\boldsymbol{w}_{m,0}, \theta_{m,0}) |^2 + n(\max_\Omega | \phi(\boldsymbol{x}^i) |^2 + 1)}.$$

□

Theorem 3.7 (Extension of Theorem 3.3). *Let* $\Omega = \{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\}$ *for some positive integers* N *and* r*. Assume that there exist* $(\boldsymbol{w}_m^*, \theta_m^*) \in R^N \times R (1 \leq m \leq M)$ *and increasing function* ψ *from* R *to* R *with* $\psi(0) = 0$ *satisfying* $\min_{\boldsymbol{x}^i \in \Omega} |\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*)| > 0$ *for all* $1 \leq m \leq M$*. Let* $T_{\boldsymbol{w}_m^*, \theta_m^*}$ *be a function from* Ω *to* $\{a_m, b_m\}$ *for any fixed real numbers* a_m *and* b_m $(a_m > b_m)$ *such that for all* $1 \leq i \leq r$

$$T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i) - \theta_m^* > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i) - \theta_m^* < 0. \end{cases}$$

Let $(\boldsymbol{w}_m, \theta_m) \in R^N \times R$ *and define* $S_{\boldsymbol{w}_m, \theta_m} : R^N \to \{a_m, b_m\}$ *as*

$$S_m^{\boldsymbol{w}, \theta}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i) - \theta_m > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i) - \theta_m < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i) - \theta_m = 0, \end{cases}$$

where $randr(\{a_m, b_m\})$ *is a number randomly chosen from the set* $\{a_m, b_m\}$*. Sequences* $\{\boldsymbol{w}_{m,n}\}$ *and* $\{\theta_{m,n}\}$ *are defined as follows.*

Initial elements $\boldsymbol{w}_{m,0} \in R^N$ *and* $\theta_{m,0} \in R$ *are random.*

$\boldsymbol{w}_{m,n+1}$ *and* $\theta_{m,n+1}$ $(n \geq 0)$ *are defined depending on the values of* $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i)$ *and* $T_m(\boldsymbol{x}^i)$*.*

Case 1. For all $i \in I$*,* $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i)$*.*

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n}, \theta_{m,n}).$$

Case 2. For some $i \in I$*,* $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^l) = T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ *and* $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = a_m, T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = b_m$*.*

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi(\psi^{-1}(\theta_{m,n}) + 1)).$$

Case 3. For some $i \in I$*,* $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^l) = T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ *and* $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = b_m, T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = a_m$*.*

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \psi(\psi^{-1}(\theta_{m,n}) - 1)).$$

Then $\{\boldsymbol{w}_{m,n}\}$ *and* $\{\theta_{m,n}\}$ *are finite sequences.*

Proof. Note that

$$T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i) - \theta_m^* > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i) - \theta_m^* < 0, \end{cases}$$

which is equivalent to

$$T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) < 0. \end{cases}$$

Similarly,

$$S_m^{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i) - \theta_m > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i) - \theta_m < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i) - \theta_m = 0, \end{cases}$$

which is equivalent to

$$S_m^{\boldsymbol{w},\theta}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m) > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m) < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m) = 0. \end{cases}$$

Using the definitions of the sequences, we have

$$\begin{aligned} &(\boldsymbol{w}_{m,n+1}, \psi^{-1}(\theta_{m,n+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\ &= \begin{cases} (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) & \text{(Case 1)} \\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) - (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*)) & \text{(Case 2)} \\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*)) & \text{(Case 3)} \end{cases} \end{aligned} \tag{37}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are infinite sequences. Then (Case1) in 37 is not possible, which give that for $n \geq 0$

$$\begin{aligned} &(\boldsymbol{w}_{m,n+1}, \psi^{-1}(\theta_{m,n+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\ &= \begin{cases} (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) - (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*)) & \text{(Case 2)} \\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*)) & \text{(Case 3)} \end{cases} \end{aligned} \tag{38}$$

Claim 1. $(\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid$ for all $n \geq 1$

Claim 2. $\mid (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \mid \leq \sqrt{\mid (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}$ for all $n \geq 1$

Using Claim 2, Cauchy-schwarz inequality and Claim1, we have

$$\begin{aligned}
&\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \sqrt{\mid (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}} \\
&\qquad \geq \mid (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \mid\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \\
&\qquad \geq (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \\
&\qquad \geq (\boldsymbol{w}_{m,0}, \theta_0) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid
\end{aligned}$$

for all n. Then

$$\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \geq \frac{(\boldsymbol{w}_{m,0}, \theta_0) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid}{\sqrt{\mid (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}},$$

which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid &= \lim_{n\to\infty} \mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \\
&\geq \lim_{n\to\infty} \frac{(\boldsymbol{w}_{m,0}, \theta_0) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid}{\sqrt{\mid (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}} = \infty.
\end{aligned}$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are finite sequences.

Proof of Claim 1.

Case 2: $T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \psi^{-1}(\theta_m^*) < 0$, which gives

$$\begin{aligned}
&(\boldsymbol{w}_{m,n+1}, \psi^{-1}(\theta_{m,n+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\
&\quad = (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) - (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta_m^*)) \\
&\quad = (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + \mid (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta_m^*)) \mid \qquad (39) \\
&\quad \geq (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + \min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid.
\end{aligned}$$

Case 3: $T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \psi^{-1}(\theta_m^*) > 0$, which gives

$$
\begin{aligned}
&(\boldsymbol{w}_{m,n+1}, \psi^{-1}(\theta_{m,n+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\
&\quad = (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta_m^*)) \\
&\quad = (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + \mid (\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta_m^*)) \mid \qquad (40) \\
&\quad \geq (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + \min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid.
\end{aligned}
$$

Using (38), (39), and (40), we obtain

$$
\begin{aligned}
&(w_{k+1}, \psi^{-1}(\theta_{k+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\
&\qquad \geq (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_{m,j})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + \min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid.
\end{aligned}
$$

Taking a summation, we have

$$
\begin{aligned}
&\sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j+1}, \psi^{-1}(\theta_{m,j+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\
&\qquad \geq \sum_{j=0}^{n-1} \left\{ (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_{m,j})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + \min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid \right\}
\end{aligned}
$$

and then

$$
\begin{aligned}
&\sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j+1}, \psi^{-1}(\theta_{m,j+1})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\
&\qquad \geq \sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_{m,j})) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid.
\end{aligned}
$$

This gives

$$
\begin{aligned}
&\sum_{j=1}^{n-1} (w_j, \psi^{-1}(\theta_j)) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) + (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \\
&\qquad \geq (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \sum_{j=1}^{n-1} (w_j, \psi^{-1}(\theta_j)) \cdot (\boldsymbol{w}_m^*, \psi^{-1}(\theta_m^*)) \\
&\qquad\quad + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - (\psi^{-1}(\theta_m^*)) \mid
\end{aligned}
$$

and hence

$$
(\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_m^*) \mid .
$$

Proof of Claim 2.
Case 2: $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_n) > 0$, which gives

$$\begin{aligned}
|\ (\boldsymbol{w}_{m,n+1}, \psi^{-1}(\theta_{m,n+1}))\ |^2 &= |\ (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi^{-1}(\theta_{m,n}) + 1)\ |^2 \\
&= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi^{-1}(\theta_{m,n}) + 1) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi^{-1}(\theta_{m,n}) + 1) \\
&= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i) + (\psi^{-1}(\theta_n) + 1)^2 \\
&= |\ \boldsymbol{w}_{m,n}\ |^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + |\ \boldsymbol{x}^i\ |^2 + (\psi^{-1}(\theta_n))^2 + 2\psi^{-1}(\theta_n) + 1 \\
&= |\ \boldsymbol{w}_{m,n}\ |^2 + (\psi^{-1}(\theta_n))^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + 2\psi^{-1}(\theta_n) + |\ \boldsymbol{x}^i\ |^2 + 1 \\
&= |\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_n))\ |^2 - 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_n)) + |\ \boldsymbol{x}^i\ |^2 + 1 \\
&\leq |\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_n))\ |^2 + \max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1.
\end{aligned}$$

Then

$$|\ (\boldsymbol{w}_{m,j+1}, \psi^{-1}(\theta_{m,j+1}))\ |^2 \leq |\ (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_j))\ |^2 + \max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1 \text{ for all } j \geq 0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1} |\ (\boldsymbol{w}_{m,j+1}, \psi^{-1}(\theta_{m,j+1}))\ |^2 \leq \sum_{j=0}^{n-1} \Big\{ |\ (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_j))\ |^2 + \max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1 \Big\}$$

and then

$$\sum_{j=0}^{n-1} |\ (\boldsymbol{w}_{m,j+1}, \psi^{-1}(\theta_{m,j+1}))\ |^2 \leq \sum_{j=0}^{n-1} |\ (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_j))\ |^2 + n\max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1.$$

This gives

$$\begin{aligned}
&\sum_{j=1}^{n-1} |\ (\boldsymbol{w}_j, \psi^{-1}(\theta_j))\ |^2 + |\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n}))\ |^2 \\
&\quad \leq \sum_{j=1}^{n-1} |\ (\boldsymbol{w}_{m,j}, \psi^{-1}(\theta_j))\ |^2 + |\ (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0))\ |^2 + n(\max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1)
\end{aligned}$$

and hence

$$|\ (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_n))\ |^2 \leq |\ (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0))\ |^2 + n(\max_\Omega |\ \boldsymbol{x}^i\ |^2 + 1). \tag{41}$$

Case 3: $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_n) < 0$, which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_{m,n+1}, \psi^{-1}(\theta_{m,n+1})) \mid^2 &= \mid (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi^{-1}(\theta_{m,n}) - 1) \mid^2 \\
&= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi^{-1}(\theta_{m,n}) - 1) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \psi^{-1}(\theta_{m,n}) - 1) \\
&= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i) + (\psi^{-1}(\theta_n) - 1)^2 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + \mid \boldsymbol{x}^i \mid^2 + (\psi^{-1}(\theta_n))^2 - 2\psi^{-1}(\theta_n) + 1 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + (\psi^{-1}(\theta_n))^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + 2\psi^{-1}(\theta_n) - \mid \boldsymbol{x}^i \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_n)) \mid^2 + 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \psi^{-1}(\theta_n)) + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_n)) \mid^2 + \max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1.
\end{aligned}$$

Similarly to case 2, we can obtain

$$\mid (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_n)) \mid^2 \leq \mid (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1). \tag{42}$$

Using (41) and (42), we obtain the desired result

$$\mid (\boldsymbol{w}_{m,n}, \psi^{-1}(\theta_{m,n})) \mid \leq \sqrt{\mid (\boldsymbol{w}_{m,0}, \psi^{-1}(\theta_0)) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)}.$$

□

Theorem 3.8 (Extension of Theorem 3.4). *Let $\Omega = \{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\}$ for some positive integers N and r. Assume that there exist $(\boldsymbol{w}_m^*, \theta_m^*) \in R^N \times R (1 \leq m \leq M)$ satisfying $\min_{\boldsymbol{x}^i \in \Omega} |\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*| > 0$ for all $1 \leq m \leq M$. Let ψ be an increasing function from R to R with $\psi(0) = 0$. Let $T_{\boldsymbol{w}_m^*,\theta_m^*}$ be a function from Ω to $\{a_m, b_m\}$ for any fixed real numbers a_m and b_m $(a_m > b_m)$ such that for all $1 \leq i \leq r$*

$$T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*) > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*) < 0. \end{cases}$$

Let $(\boldsymbol{w}_m, \theta_m) \in R^N \times R$ and define $S_{\boldsymbol{w}_m,\theta_m} : R^N \to \{a_m, b_m\}$ as

$$S_m^{\boldsymbol{w},\theta}(x) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m) > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m) < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m) = 0, \end{cases}$$

where $randr(\{a_m, b_m\})$ *is a number randomly chosen from the set* $\{a_m, b_m\}$*. Sequences* $\{\boldsymbol{w}_{m,n}\}$ *and* $\{\theta_{m,n}\}$ *are defined as follows.*

Initial elements $\boldsymbol{w}_{m,0} \in R^N$ *and* $\theta_{m,0} \in R$ *are random.*

$\boldsymbol{w}_{m,n+1}$ *and* $\theta_{m,n+1}$ $(n \geq 0)$ *are defined depending on the values of* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)$ *and* $T_m(\boldsymbol{x}^i)$*.*

Case 1. For all $i \in I$*,* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i)$ *.*

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n}, \theta_{m,n}).$$

Case 2. For some $i \in I$*,* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ *and* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = a_m, T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = b_m$*.*

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \theta_{m,n} + 1).$$

Case 3. For some $i \in I$*,* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^l) = T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^l) \quad (1 \leq l \leq i-1)$ *and* $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = b_m, T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = a_m$*.*

$$\textit{Define } (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1).$$

Then $\{\boldsymbol{w}_{m,n}\}$ *and* $\{\theta_{m,n}\}$ *are finite sequences.*

Proof. Note that

$$T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*) > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*) < 0, \end{cases}$$

which is equivalent to

$$T_{\boldsymbol{w}_m^*,\theta_m^*}(\boldsymbol{x}^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^* > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^* < 0. \end{cases}$$

Similarly,

$$S_m^{\boldsymbol{w},\theta}(x) = \begin{cases} a_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m) > 0, \\ b_m & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m) < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \psi(\boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m) = 0, \end{cases}$$

which is equivalent to

$$S_m^{\boldsymbol{w},\theta}(x) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x}^i - \theta_m = 0. \end{cases}$$

Using the definitions of the sequences, we have

$$(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) = \begin{cases} (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) & \text{(Case 1)} \\ (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 2)} \\ (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 3)} \end{cases} \tag{43}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are infinite sequences.
Then (Case1) in (43) is not possible, which give that for $n \geq 0$.

$$(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) = \begin{cases} (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 2)} \\ (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) & \text{(Case 3)} \end{cases} \tag{44}$$

Claim 1. $(\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_0, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid$ for all $n \geq 1$

Claim 2. $\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid \leq \sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}$ for all $n \geq 1$
Using Claim 2, Cauchy-schwarz inequality and Claim1, we have

$$\begin{aligned} \mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)} &\geq \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \\ &\geq (\boldsymbol{w}_{m,n}, \theta_{m,n})(\boldsymbol{w}_m^*, \theta_m^*) \\ &\geq (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \\ &\quad + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid \end{aligned}$$

for all n. Then

$$\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \geq \frac{(\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid}{\sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}},$$

which gives

$$
\begin{aligned}
\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid &= \lim_{n\to\infty} \mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \\
&\geq \lim_{n\to\infty} \frac{(\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid}{\sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n\left\{\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1\right\}}} = \infty.
\end{aligned}
$$

This is a contradiction. Therefore, $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ are finite sequences.

Proof of Claim 1.

Case 2: $T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* < 0$, which gives

$$
\begin{aligned}
(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) - (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \\
&= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \mid \\
&\geq (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid.
\end{aligned} \tag{45}
$$

Case 3: $T_{\boldsymbol{w}_m^*, \theta_m^*}(\boldsymbol{x}^i) = 1$. Then $\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* > 0$, which gives

$$
\begin{aligned}
(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \\
&= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \mid (\boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^*) \mid \\
&\geq (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid.
\end{aligned} \tag{46}
$$

Using (44), (45), and (46), we obtain

$$
(\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid .
$$

Taking a summation, we have that

$$
\sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq \sum_{j=0}^{n-1} \left\{ (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid \right\}
$$

and then

$$
\sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j+1}, \theta_{m,j+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \geq \sum_{j=0}^{n-1} (\boldsymbol{w}_{m,j}, \theta_{m,j}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n\min_\Omega \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid .
$$

This gives

$$\sum_{j=1}^{n-1}(\boldsymbol{w}_{m,j},\theta_j)\cdot(\boldsymbol{w}_m^*,\theta_m^*)+(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)$$
$$\geq(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+\sum_{j=1}^{n-1}(\boldsymbol{w}_{m,j},\theta_{m,j})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\min_\Omega\mid\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*\mid$$

and hence

$$(\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)\geq(\boldsymbol{w}_{m,0},\theta_{m,0})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+n\min_\Omega\mid\boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*\mid.$$

Proof of Claim 2.
Case 2: $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)=1$. Then $\boldsymbol{w}_{m,n}\cdot\boldsymbol{x}^i-\theta_{m,n}>0$, which gives

$$\begin{aligned}
\mid(\boldsymbol{w}_{m,n+1},\theta_{m,n+1})\mid^2 &= \mid(\boldsymbol{w}_{m,n}-\boldsymbol{x}^i,\theta_{m,n}+1)\mid^2\\
&=(\boldsymbol{w}_{m,n}-\boldsymbol{x}^i,\theta_{m,n}+1)\cdot(\boldsymbol{w}_{m,n}-\boldsymbol{x}^i,\theta_{m,n}+1)\\
&=(\boldsymbol{w}_{m,n}-\boldsymbol{x}^i)\cdot(\boldsymbol{w}_{m,n}-\boldsymbol{x}^i)+(\theta_{m,n}+1)^2\\
&=\mid\boldsymbol{w}_{m,n}\mid^2-2\boldsymbol{w}_{m,n}\cdot\boldsymbol{x}^i+\mid\boldsymbol{x}^i\mid^2+{\theta_{m,n}}^2+2\theta_{m,n}+1\\
&=\mid\boldsymbol{w}_{m,n}\mid^2+{\theta_{m,n}}^2-2\boldsymbol{w}_{m,n}\cdot\boldsymbol{x}^i+2\theta_{m,n}+\mid\boldsymbol{x}^i\mid^2+1\\
&=\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid^2-2(\boldsymbol{w}_{m,n}\cdot\boldsymbol{x}^i-\theta_{m,n})+\mid\boldsymbol{x}^i\mid^2+1\\
&\leq\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid^2+\max_\Omega\mid\boldsymbol{x}^i\mid^2+1.
\end{aligned}$$

Then

$$\mid(\boldsymbol{w}_{m,j+1},\theta_{m,j+1})\mid^2\leq\mid(\boldsymbol{w}_{m,j},\theta_{m,j})\mid^2+\max_\Omega\mid\boldsymbol{x}^i\mid^2+1\text{ for all }j\geq0.$$

Taking a summation, we have

$$\sum_{j=0}^{n-1}\mid(\boldsymbol{w}_{m,j+1},\theta_{m,j+1})\mid^2\leq\sum_{j=0}^{n-1}\Big\{\mid(\boldsymbol{w}_{m,j},\theta_{m,j})\mid^2+\max_\Omega\mid\boldsymbol{x}^i\mid^2+1\Big\}$$

and then

$$\sum_{j=0}^{n}\mid(\boldsymbol{w}_{m,j+1},\theta_{m,j+1})\mid^2+\mid(\boldsymbol{w}_{m,n},\theta_{m,n})\mid^2\leq\sum_{j=0}^{n-1}\mid(\boldsymbol{w}_{m,j},\theta_{m,j})\mid^2+n(\max_\Omega\mid\boldsymbol{x}^i\mid^2+1).$$

This gives

$$\sum_{j=1}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_j) \mid^2 + \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2$$
$$\leq \sum_{j=1}^{n-1} \mid (\boldsymbol{w}_{m,j}, \theta_{m,j}) \mid^2 + \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)$$

and hence

$$\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \mid \boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1). \tag{47}$$

Case 3: $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^i) = 0$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \theta_{m,n} < 0$, which gives

$$\begin{aligned}
\mid (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \mid^2 &= \mid (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1) \mid^2 \\
&= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1) \\
&= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i) + (\theta_{m,n} - 1)^2 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i + \mid \boldsymbol{x}^i \mid^2 + {\theta_{m,n}}^2 - 2\theta_{m,n} + 1 \\
&= \mid \boldsymbol{w}_{m,n} \mid^2 + {\theta_{m,n}}^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - 2\theta_{m,n} + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^i - \theta_{m,n}) + \mid \boldsymbol{x}^i \mid^2 + 1 \\
&\leq \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + \max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1.
\end{aligned}$$

Similarly to case 2 we can obtain that

$$\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 \leq \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1). \tag{48}$$

Using (47) and (48), we obtain the desired result

$$\mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid \leq \sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n(\max_\Omega \mid \boldsymbol{x}^i \mid^2 + 1)}.$$

□

4 Convergence speed of extended algorithms

In this chapter we present a generalized upper bound of the extended algorithm and construct the order of training data for accelerating convergence speed.

Theorem 4.1. *Let I be the set of positive integers less than $r+1$ and $\Omega = \{\boldsymbol{x}^i \in R^N | i \in I\}$ for some positive integers N and r. Assume that there exist $(\boldsymbol{w}_m^*, \theta_m^*) \in R^N \times R$ satisfying $\min_{\boldsymbol{x}^i \in \Omega} |\boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^*| > 0$ for a positive integer M and all $1 \leq m \leq M$. Let T_m be a function from Ω to $\{a_m, b_m\}$ for real numbers a_m and b_m $(a_m > b_m)$ such that for all $i \in I$*

$$T_m(x^i) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^* > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{x}^i - \theta_m^* < 0. \end{cases}$$

Let $(\boldsymbol{w}_m, \theta_m) \in R^N \times R$ and define $S_m^{\boldsymbol{w},\theta} : R^N \to \{a_m, b_m\}$ as

$$S_m^{\boldsymbol{w},\theta}(\boldsymbol{x}) = \begin{cases} a_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x} - \theta_m > 0, \\ b_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x} - \theta_m < 0, \\ randr(\{a_m, b_m\}) & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{x} - \theta_m = 0, \end{cases}$$

where $randr(\{a_m, b_m\})$ is a number randomly chosen from $\{a_m, b_m\}$.
Sequences $\{\boldsymbol{w}_{m,n}\}_{n=0}^{\infty}$ and $\{\theta_{m,n}\}_{n=0}^{\infty}$ are defined recursively as follows:

i) Initial elements $\boldsymbol{w}_{m,0} \in R^N$ and $\theta_{m,0} \in R$ are randomly sampled.

ii) $\boldsymbol{w}_{m,n+1}$ and $\theta_{m,n+1}$ $(n \geq 0)$ are defined depending on the values $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i)$ and $T_m(\boldsymbol{x}^i)$.

a) Case 1: $(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n}, \theta_{m,n})$ if $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = T_m(\boldsymbol{x}^i)$ for all $i \in I$.

b) Case 2: $(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} - \boldsymbol{x}^i, \theta_{m,n} + 1)$ if $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^\ell) = T_m(\boldsymbol{x}^\ell)$ and $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = a_m, T_m(\boldsymbol{x}^i) = b_m$ for some $i \in I$ and all $1 \leq \ell \leq i-1$.

c) Case 3: $(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) = (\boldsymbol{w}_{m,n} + \boldsymbol{x}^i, \theta_{m,n} - 1)$ if $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^\ell) = T_m(\boldsymbol{x}^\ell)$ and $S_m^{\boldsymbol{w}_n,\theta_n}(\boldsymbol{x}^i) = b_m, T_m(\boldsymbol{x}^i) = a_m$ for some $i \in I$ and all $1 \leq \ell \leq i-1$.

Then for any fixed m, $\{\boldsymbol{w}_{m,n}\}$ and $\{\theta_{m,n}\}$ reach to Case 1 (the update stops) at iteration number n if the following inequality satisfies

$$n > -\frac{DE - \frac{1}{2}AC}{E^2} + \sqrt{\left(\frac{DE - \frac{1}{2}AC}{E^2}\right)^2 - \frac{D^2 - AB}{E^2}},$$

where $A = \mid (\boldsymbol{w}_m^*, \theta_m^*) \mid^2$, $B = \mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2$, $C = \max_{\Omega} \mid \boldsymbol{x}^i \mid^2 + 1$, $D = (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*)$, $E = \min_{\Omega} \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid$ *and* $\left(\frac{DE - \frac{1}{2}AC}{E^2}\right)^2 - \frac{D^2 - AB}{E^2} \geq 0$.

Proof. Using the proof of Theorem 3.5 we have that when $(\boldsymbol{w}_{m,n}, \theta_{m,n})$ is updated, n must satisfies the following inequality

$$
\begin{aligned}
&\mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \geq \frac{(\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n \min_{\Omega} \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid}{\sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n \left(\max_{\Omega} \mid \boldsymbol{x}^i \mid^2 + 1\right)}} \\
&\iff \mid (\boldsymbol{w}_m^*, \theta_m^*) \mid \sqrt{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n \left(\max_{\Omega} \mid \boldsymbol{x}^i \mid^2 + 1\right)} \\
&\qquad \geq (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n \min_{\Omega} \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid \\
&\iff \mid (\boldsymbol{w}_m^*, \theta_m^*) \mid^2 \left\{\mid (\boldsymbol{w}_{m,0}, \theta_{m,0}) \mid^2 + n \left(\max_{\Omega} \mid \boldsymbol{x}^i \mid^2 + 1\right)\right\} \\
&\qquad \geq \left\{(\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + n \min_{\Omega} \mid \boldsymbol{x}^i \cdot \boldsymbol{w}_m^* - \theta_m^* \mid\right\}^2 \\
&\iff A\{B + nC\} \geq \{D + nE\}^2 = D^2 + 2nDE + n^2E^2 \\
&\iff 0 \geq \left(D^2 - AB\right) + 2n\left(DE - \frac{1}{2}AC\right) + n^2E^2 \\
&\iff 0 \geq E^2\left(n^2 + \frac{D^2 - AB}{E^2} + 2n\frac{DE - \frac{1}{2}AC}{E^2}\right) \\
&\iff 0 \geq n^2 + 2n\frac{DE - \frac{1}{2}AC}{E^2} + \frac{D^2 - AB}{E^2} \\
&\iff 0 \geq \left[n - \left\{-\frac{DE - \frac{1}{2}AC}{E^2} - \sqrt{\left(\frac{DE - \frac{1}{2}AC}{E^2}\right)^2 - \frac{D^2 - AB}{E^2}}\right\}\right] \\
&\qquad \times \left[n - \left\{-\frac{DE - \frac{1}{2}AC}{E^2} + \sqrt{\left(\frac{DE - \frac{1}{2}AC}{E^2}\right)^2 - \frac{D^2 - AB}{E^2}}\right\}\right]
\end{aligned}
$$

$$\iff -\frac{DE-\frac{1}{2}AC}{E^2}-\sqrt{\left(\frac{DE-\frac{1}{2}AC}{E^2}\right)^2-\frac{D^2-AB}{E^2}}$$
$$\leq n \leq -\frac{DE-\frac{1}{2}AC}{E^2}+\sqrt{\left(\frac{DE-\frac{1}{2}AC}{E^2}\right)^2-\frac{D^2-AB}{E^2}}.$$

Note that

$$\left(\frac{DE-\frac{1}{2}AC}{E^2}\right)^2-\frac{D^2-AB}{E^2}\geq 0$$

since n satisfies the inequality $0 \geq n^2+2n\frac{DE-\frac{1}{2}AC}{E^2}+\frac{D^2-AB}{E^2}$. Therefore if

$$n > -\frac{DE-\frac{1}{2}AC}{E^2}+\sqrt{\left(\frac{DE-\frac{1}{2}AC}{E^2}\right)^2-\frac{D^2-AB}{E^2}},$$

then the update stops. □

Remark 4.1. Assume $(\boldsymbol{w}_{m,0},\theta_{m,0})=\mathbf{0}$. Then we have $B=D=0$ in Theorem 4.1, which gives that if

$$n > \mid (\boldsymbol{w}_m^*,\theta_m^*)\mid^2 \frac{\max\limits_{\Omega}\mid \boldsymbol{x}^i\mid^2+1}{\min\limits_{\Omega}\mid \boldsymbol{x}^i\cdot\boldsymbol{w}_m^*-\theta_m^*\mid^2},$$

then the update stops.

Remark 4.2. Assume that $(\boldsymbol{w}_{m,n},\theta_{m,n})$ is updated. Define

$$a_n = \operatorname*{argmax}_{i\in I, S_m^{\boldsymbol{w}_n}(\boldsymbol{x}^i)\neq T_m(\boldsymbol{x}^i)} \mid \boldsymbol{x}^i\cdot\boldsymbol{w}_m^*\mid .$$

We consider Case 2 and Case 3.
Case 2: $T(\boldsymbol{x}^{a_n})=0$. Then $\boldsymbol{x}^{a_n}\cdot\boldsymbol{w}_m^*-\theta_m^*<0$, which gives

$$\begin{aligned}(\boldsymbol{w}_{m,n+1},\theta_{m,n+1})\cdot(\boldsymbol{w}_m^*,\theta_m^*) &= (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)-(\boldsymbol{x}^{a_n}\cdot\boldsymbol{w}_m^*-\theta_m^*)\\ &= (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+\mid \boldsymbol{x}^{a_n}\cdot\boldsymbol{w}_m^*-\theta_m^*\mid.\end{aligned}$$

Case 3: $T(\boldsymbol{x}^{a_n})=1$. Then $\boldsymbol{x}^{a_n}\cdot\boldsymbol{w}_m^*-\theta_m^*>0$, which gives

$$\begin{aligned}(\boldsymbol{w}_{m,n+1},\theta_{m,n+1})\cdot(\boldsymbol{w}_m^*,\theta_m^*) &= (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+(\boldsymbol{x}^{a_n}\cdot\boldsymbol{w}_m^*-\theta_m^*)\\ &= (\boldsymbol{w}_{m,n},\theta_{m,n})\cdot(\boldsymbol{w}_m^*,\theta_m^*)+\mid \boldsymbol{x}^{a_n}\cdot\boldsymbol{w}_m^*-\theta_m^*\mid.\end{aligned}$$

Then

$$(\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) = (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \mid \boldsymbol{x}^{a_n} \cdot \boldsymbol{w}_m^* - \theta_m^* \mid,$$

which gives

$$\begin{aligned}(\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) &= (\boldsymbol{w}_{m,n-1}, \theta_{m,n-1}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \mid \boldsymbol{x}^{a_{n-1}} \cdot \boldsymbol{w}_m^* - \theta_m^* \mid \\ &= (\boldsymbol{w}_{m,0}, \theta_{m,0}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) + \sum_{0 \le k \le n-1} \mid \boldsymbol{x}^{a_k} \cdot \boldsymbol{w}_m^* - \theta_m^* \mid.\end{aligned}$$

If $(\boldsymbol{w}_{m,0}, \theta_{m,0}) = \mathbf{0}$, then

$$(\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) = \sum_{0 \le k \le n-1} \mid \boldsymbol{x}^{a_k} \cdot \boldsymbol{w}_m^* - \theta_m^* \mid . \tag{49}$$

We again consider the two cases:
Case 2: $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^{a_n}) = 1$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n} > 0$, which gives

$$\begin{aligned}\mid (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \mid^2 &= \mid (\boldsymbol{w}_{m,n} - \boldsymbol{x}^{a_n}, \theta_{m,n} + 1) \mid^2 \\ &= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^{a_n}, \theta_{m,n} + 1) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^{a_n}, \theta_{m,n} + 1) \\ &= (\boldsymbol{w}_{m,n} - \boldsymbol{x}^{a_n}) \cdot (\boldsymbol{w}_{m,n} - \boldsymbol{x}^{a_n}) + (\theta_{m,n} + 1)^2 \\ &= \mid \boldsymbol{w}_{m,n} \mid^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} + \mid \boldsymbol{x}^{a_n} \mid^2 + {\theta_{m,n}}^2 + 2\theta_{m,n} + 1 \\ &= \mid \boldsymbol{w}_{m,n} \mid^2 + {\theta_{m,n}}^2 - 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} + 2\theta_{m,n} + \mid \boldsymbol{x}^{a_n} \mid^2 + 1 \\ &= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 - 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n}) + \mid \boldsymbol{x}^{a_n} \mid^2 + 1 \\ &= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 - 2|\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n}| + \mid \boldsymbol{x}^{a_n} \mid^2 + 1.\end{aligned}$$

Case 3: $S_m^{\boldsymbol{w}_n, \theta_n}(\boldsymbol{x}^{a_n}) = 0$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n} < 0$, which gives

$$\begin{aligned}\mid (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) \mid^2 &= \mid (\boldsymbol{w}_{m,n} + \boldsymbol{x}^{a_n}, \theta_{m,n} - 1) \mid^2 \\ &= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^{a_n}, \theta_{m,n} - 1) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^{a_n}, \theta_{m,n} - 1) \\ &= (\boldsymbol{w}_{m,n} + \boldsymbol{x}^{a_n}) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{x}^{a_n}) + (\theta_{m,n} - 1)^2 \\ &= \mid \boldsymbol{w}_{m,n} \mid^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} + \mid \boldsymbol{x}^{a_n} \mid^2 + {\theta_{m,n}}^2 - 2\theta_{m,n} + 1 \\ &= \mid \boldsymbol{w}_{m,n} \mid^2 + {\theta_{m,n}}^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - 2\theta_{m,n} + \mid \boldsymbol{x}^{a_n} \mid^2 + 1 \\ &= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 + 2(\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n}) + \mid \boldsymbol{x}^{a_n} \mid^2 + 1 \\ &= \mid (\boldsymbol{w}_{m,n}, \theta_{m,n}) \mid^2 - 2|\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n}| + \mid \boldsymbol{x}^{a_n} \mid^2 + 1.\end{aligned}$$

Then we have

$$| (\boldsymbol{w}_{m,n+1}, \theta_{m,n+1}) |^2 = | (\boldsymbol{w}_{m,n}, \theta_{m,n}) |^2 - 2|\boldsymbol{w}_{m,n} \cdot \boldsymbol{x}^{a_n} - \theta_{m,n}| + | \boldsymbol{x}^{a_n} |^2 + 1,$$

which gives

$$\begin{aligned} | (\boldsymbol{w}_{m,n}, \theta_{m,n}) |^2 &= | (\boldsymbol{w}_{m,n-1}, \theta_{m,n-1}) |^2 - 2|\boldsymbol{w}_{m,n-1} \cdot \boldsymbol{x}^{a_{n-1}} - \theta_{m,n-1}| + | \boldsymbol{x}^{a_{n-1}} |^2 + 1 \\ &= | (\boldsymbol{w}_{m,0}, \theta_{m,0}) |^2 + \sum_{0 \le k \le n-1} \left\{ -2|\boldsymbol{w}_{m,k} \cdot \boldsymbol{x}^{a_k} - \theta_{m,k}| + | \boldsymbol{x}^{a_k} |^2 + 1 \right\}. \end{aligned}$$

If $(\boldsymbol{w}_{m,0}, \theta_{m,0}) = \mathbf{0}$, then

$$| (\boldsymbol{w}_{m,n}, \theta_{m,n}) | = \sqrt{\sum_{0 \le k \le n-1} \left\{ -2|\boldsymbol{w}_{m,k} \cdot \boldsymbol{x}^{a_k} - \theta_{m,k}| + | \boldsymbol{x}^{a_k} |^2 + 1 \right\}}. \tag{50}$$

Using (49) and (50), we obtain

$$\begin{aligned} & | (\boldsymbol{w}_m^*, \theta_m^*) | \sqrt{\sum_{0 \le k \le n-1} \left\{ -2|\boldsymbol{w}_{m,k} \cdot \boldsymbol{x}^{a_k} - \theta_{m,k}| + | \boldsymbol{x}^{a_k} |^2 + 1 \right\}} \cos b_n \\ &= | (\boldsymbol{w}_m^*, \theta_m^*) | \, | (\boldsymbol{w}_{m,n}, \theta_{m,n}) | \cos b_n \\ &= (\boldsymbol{w}_{m,n}, \theta_{m,n}) \cdot (\boldsymbol{w}_m^*, \theta_m^*) \\ &= \sum_{0 \le k \le n-1} | \boldsymbol{x}^{a_k} \cdot \boldsymbol{w}_m^* - \theta_m^* |, \end{aligned}$$

where b_n is the angle between $(\boldsymbol{w}_m^*, \theta_m^*)$ and $(\boldsymbol{w}_{m,n}, \theta_{m,n})$. So, we have

$$| (\boldsymbol{w}_m^*, \theta_m^*) | \sqrt{\sum_{0 \le k \le n-1} \left(-2|\boldsymbol{w}_{m,k} \cdot \boldsymbol{x}^{a_k} - \theta_{m,k}| + | \boldsymbol{x}^{a_k} |^2 + 1 \right)} \ge \sum_{0 \le k \le n-1} | \boldsymbol{x}^{a_k} \cdot \boldsymbol{w}_m^* - \theta_m^* | .$$

If $\theta_m^* = 0$, then

$$| \boldsymbol{w}_m^* | \sqrt{\sum_{0 \le k \le n-1} \left(-2|\boldsymbol{w}_{m,k} \cdot \boldsymbol{x}^{a_k}| + | \boldsymbol{x}^{a_k} |^2 \right)} \ge \sum_{0 \le k \le n-1} | \boldsymbol{x}^{a_k} \cdot \boldsymbol{w}_m^* | .$$

Thus when updating $(\boldsymbol{w}_{m,n-1}, \theta_{m,n-1})$, the number n must satisfy the following inequality

$$\sqrt{\sum_{0 \le k \le n-1} \left(-2|\boldsymbol{w}_{m,k} \cdot \boldsymbol{x}^{a_k}| + | \boldsymbol{x}^{a_k} |^2 \right)} \ge \sum_{0 \le k \le n-1} \left| \boldsymbol{x}^{a_k} \cdot \frac{\boldsymbol{w}_m^*}{|\boldsymbol{w}_m^*|} \right| .$$

If the just above inequality fails to hold, the update stops. So, in order to make the update stop faster we reorder $(\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \cdots, \boldsymbol{x}^r)$ as $\left(\boldsymbol{x}^{\xi_1}, \boldsymbol{x}^{\xi_2}, \boldsymbol{x}^{\xi_3}, \cdots, \boldsymbol{x}^{\xi_r}\right)$ such that

$$\xi_k = \underset{i \in I, S_m^{\boldsymbol{w}_k}(\boldsymbol{x}^i) \neq T_m(\boldsymbol{x}^i)}{\operatorname{argmax}} \left| \boldsymbol{x}^i \cdot \frac{\boldsymbol{w}_m^*}{|\boldsymbol{w}_m^*|} \right|.$$

Theorem 4.2. *Let I be the set of positive integers less than $r+1$ and $\Omega = \left\{\boldsymbol{d}^i \in R^N | i \in I\right\}$ for some positive integers N and r. Let $\boldsymbol{w}_m^* \in R^N$ satisfying $\min_{\boldsymbol{d}^i \in \Omega} |\boldsymbol{w}_m^* \cdot \boldsymbol{d}^i| > 0$ for a positive integer M and all $1 \leq m \leq M$ and let T_m be a function from Ω to $\{\alpha_m, \beta_m\}$ for real numbers α_m and β_m $(\alpha_m > \beta_m)$ such that for all $i \in I$*

$$T_m(\boldsymbol{d}^i) = \begin{cases} \alpha_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{d}^i > 0, \\ \beta_m & \text{if} \quad \boldsymbol{w}_m^* \cdot \boldsymbol{d}^i < 0. \end{cases}$$

Let $\boldsymbol{w}_m \in R^N$ and define $S_m^{\boldsymbol{w}} : R^N \to \{\alpha_m, \beta_m\}$ as

$$S_m^{\boldsymbol{w}}(\boldsymbol{d}) = \begin{cases} \alpha_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{d} > 0, \\ \beta_m & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{d} < 0, \\ randr(\{\alpha_m, \beta_m\}) & \text{if} \quad \boldsymbol{w}_m \cdot \boldsymbol{d} = 0, \end{cases}$$

where $randr(\{\alpha_m, \beta_m\})$ is a number randomly chosen from $\{\alpha_m, \beta_m\}$. Sequence $\{\boldsymbol{w}_{m,n}\}_{n=0}^{\infty}$ is defined recursively as follows:

i) Initial elements $\boldsymbol{w}_{m,0} \in R^N$ is randomly sampled.

ii) $\boldsymbol{w}_{m,n+1}$ $(n \geq 0)$ is defined depending on the values $S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^i)$ and $T_m(\boldsymbol{d}^i)$.

a) Case 1: $\boldsymbol{w}_{m,n+1} = \boldsymbol{w}_{m,n}$ if $S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^i) = T_m(\boldsymbol{d}^i)$ for all $i \in I$.

b) If Case 1 is not true, define the sequence ξ_n as

$$\xi_n = \underset{i \in I, S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^i) \neq T_m(\boldsymbol{d}^i)}{\operatorname{argmax}} \left| \frac{\boldsymbol{w}_m^*}{|\boldsymbol{w}_m^*|} \cdot \boldsymbol{d}^i \right|.$$

1) Case 2: $\boldsymbol{w}_{m,n+1} = \boldsymbol{w}_{m,n} - \boldsymbol{d}^{\xi_n}$ if $S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^{\xi_n}) = \alpha_m$ and $T_m(\boldsymbol{d}^{\xi_n}) = \beta_m$.

2) Case 3: $\boldsymbol{w}_{m,n+1} = \boldsymbol{w}_{m,n} + \boldsymbol{d}^{\xi_n}$ if $S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^{\xi_n}) = \beta_m$ and $T_m(\boldsymbol{d}^{\xi_n}) = \alpha_m$.

Then $\{\boldsymbol{w}_{m,n}\}$ is a finite sequence for any fixed m.

Proof. Using the definitions of the sequences, we have

$$\boldsymbol{w}_{m,n+1} \cdot \boldsymbol{w}_m^* = \begin{cases} \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* & \text{(Case 1)} \\ \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* - \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} & \text{(Case 2)} \\ \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* + \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} & \text{(Case 3)} \end{cases} \tag{51}$$

Suppose, on the contrary, that $\{\boldsymbol{w}_{m,n}\}$ is an infinite sequence. Then Case1 in (51) is not possible, which give that for $n \geq 0$

$$\boldsymbol{w}_{m,n+1} \cdot \boldsymbol{w}_m^* = \begin{cases} \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* - \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} & \text{(Case 2)} \\ \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* + \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} & \text{(Case 3)} \end{cases} \tag{52}$$

So, we can obtain the following two claims:

Claim 1. $\boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* = \boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + \sum_{k=0}^{n-1} \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_k} \mid$

Claim 2. $\mid \boldsymbol{w}_{m,n} \mid = \sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + \sum_{k=0}^{n-1} \left(-2 \mid \boldsymbol{w}_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_k} \mid^2\right)}$

Using Claim 2, Cauchy-schwarz inequality and Claim1, we have

$$\begin{aligned} &\sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + \sum\nolimits_{k=0}^{n-1} \left(-2 \mid \boldsymbol{w}_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_n} \mid^2\right)} \mid \boldsymbol{w}_m^* \mid \\ &\quad = \mid \boldsymbol{w}_{m,n} \mid \mid \boldsymbol{w}_m^* \mid \\ &\quad \geq \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* \\ &\quad = \boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + \sum\nolimits_{k=0}^{n-1} \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_k} \mid . \end{aligned}$$

Then

$$\mid \boldsymbol{w}_m^* \mid \geq \frac{\boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + \sum\limits_{k=0}^{n-1} \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_k} \mid}{\sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + \sum\limits_{k=0}^{n-1} \left\{-2 \mid w_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_n} \mid^2\right\}}}.$$

Let $\delta = \min\limits_{i \in I} \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^i \mid$ and $\chi = \max\limits_{i \in I} \mid \boldsymbol{d}^i \mid^2$. Then we have

$$
\begin{aligned}
\mid \boldsymbol{w}_m^* \mid &\geq \frac{\boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + \sum_{k=0}^{n-1} \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_k} \mid}{\sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + \sum_{k=0}^{n-1} \left\{ -2 \mid w_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_n} \mid^2 \right\}}} \\
&\geq \frac{\boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + \sum_{k=0}^{n-1} \delta}{\sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + \sum_{k=0}^{n-1} \chi}} \\
&= \frac{\boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + n\delta}{\sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + n\chi}}.
\end{aligned}
$$

Now take the limit as n goes to infinity

$$
\mid \boldsymbol{w}_m^* \mid = \lim_{n\to\infty} \mid \boldsymbol{w}_m^* \mid \geq \lim_{n\to\infty} \frac{\boldsymbol{w}_{m,0} \cdot \boldsymbol{w}_m^* + n\delta}{\sqrt{\mid \boldsymbol{w}_{m,0} \mid^2 + n\chi}} = \infty,
$$

which is a contradiction. Therefore $\{\boldsymbol{w}_{m,n}\}$ is finite sequence.

Proof of Claim 1.

Case 2: $T(\boldsymbol{d}^{\xi_n}) = 0$. Then $\boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} < 0$, which gives

$$
\begin{aligned}
\boldsymbol{w}_{m,n+1} \cdot \boldsymbol{w}_m^* &= \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* - \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} \\
&= \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* + \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} \mid.
\end{aligned} \tag{53}
$$

Case 3: $T(\boldsymbol{d}^{\xi_n}) = 1$. Then $\boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} > 0$, which gives

$$
\begin{aligned}
\boldsymbol{w}_{m,n+1} \cdot \boldsymbol{w}_m^* &= \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* + \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} \\
&= \boldsymbol{w}_{m,n} \cdot \boldsymbol{w}_m^* + \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_n} \mid.
\end{aligned} \tag{54}
$$

Using (52), (53) and (54), we obtain

$$
\boldsymbol{w}_{m,k+1} \cdot \boldsymbol{w}_m^* = \boldsymbol{w}_{m,k} \cdot \boldsymbol{w}_m^* + \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_k} \mid.
$$

Taking a summation, we have that

$$
\sum_{k=0}^{n-1} (\boldsymbol{w}_{m,k+1} \cdot \boldsymbol{w}_m^*) = \sum_{k=0}^{n-1} \left\{ \boldsymbol{w}_{m,k} \cdot \boldsymbol{w}_m^* + \mid \boldsymbol{w}_m^* \cdot \boldsymbol{d}^{\xi_k} \mid \right\}
$$

and then

$$\sum_{k=0}^{n-1}(\boldsymbol{w}_{m,k+1}\cdot\boldsymbol{w}_m^*)=\sum_{k=0}^{n-1}(\boldsymbol{w}_{m,k}\cdot\boldsymbol{w}_m^*)+\sum_{k=0}^{n-1}\mid\boldsymbol{w}_m^*\cdot\boldsymbol{d}^{\xi_k}\mid.$$

This gives

$$\sum_{k=1}^{n-1}(\boldsymbol{w}_{m,k}\cdot\boldsymbol{w}_m^*)+(\boldsymbol{w}_{m,n}\cdot\boldsymbol{w}_m^*)=(\boldsymbol{w}_{m,0}\cdot\boldsymbol{w}_m^*)+\sum_{k=1}^{n-1}(\boldsymbol{w}_{m,k}\cdot\boldsymbol{w}_m^*)+\sum_{k=0}^{n-1}\mid\boldsymbol{w}_m^*\cdot\boldsymbol{d}^{\xi_k}\mid$$

and hence

$$\boldsymbol{w}_{m,n}\cdot\boldsymbol{w}_m^*\geq\boldsymbol{w}_{m,0}\cdot\boldsymbol{w}_m^*+\sum_{k=0}^{n-1}\mid\boldsymbol{w}_m^*\cdot\boldsymbol{d}^{\xi_k}\mid.$$

Proof of Claim 2.
Case 2: $S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^{\xi_n})=1$. Then $\boldsymbol{w}_{m,n}\cdot\boldsymbol{d}^{\xi_n}>0$, which gives

$$\begin{aligned}\mid\boldsymbol{w}_{m,n+1}\mid^2&=\mid\boldsymbol{w}_{m,n}-\boldsymbol{d}^{\xi_n}\mid^2=(\boldsymbol{w}_{m,n}-\boldsymbol{d}^{\xi_n})\cdot(\boldsymbol{w}_{m,n}-\boldsymbol{d}^{\xi_n})\\&=\mid\boldsymbol{w}_{m,n}\mid^2-2\boldsymbol{w}_{m,n}\cdot\boldsymbol{d}^{\xi_n}+\mid\boldsymbol{d}^{\xi_n}\mid^2\\&=\mid\boldsymbol{w}_{m,n}\mid^2-2\mid\boldsymbol{w}_{m,n}\cdot\boldsymbol{d}^{\xi_n}\mid+\mid\boldsymbol{d}^{\xi_n}\mid^2.\end{aligned}$$

Taking a summation, we have

$$\sum_{k=0}^{n-1}\mid\boldsymbol{w}_{m,k+1}\mid^2=\sum_{k=0}^{n-1}\Big\{\mid\boldsymbol{w}_{m,k}\mid^2-2\mid\boldsymbol{w}_{m,k}\cdot\boldsymbol{d}^{\xi_k}\mid+\mid\boldsymbol{d}^{\xi_k}\mid^2\Big\}$$

and then

$$\sum_{k=0}^{n}\mid\boldsymbol{w}_{m,k+1}\mid^2+\mid\boldsymbol{w}_{m,n}\mid^2=\sum_{k=0}^{n-1}\mid\boldsymbol{w}_{m,k}\mid^2+\sum_{k=0}^{n-1}\Big\{-2\mid\boldsymbol{w}_{m,k}\cdot\boldsymbol{d}^{\xi_k}\mid+\mid\boldsymbol{d}^{\xi_k}\mid^2\Big\}.$$

This gives

$$\sum_{k=1}^{n-1}\mid\boldsymbol{w}_{m,k}\mid^2+\mid\boldsymbol{w}_{m,n}\mid^2=\sum_{k=1}^{n-1}\mid\boldsymbol{w}_{m,k}\mid^2+\mid\boldsymbol{w}_{m,0}\mid^2+\sum_{k=0}^{n-1}\Big\{-2\mid\boldsymbol{w}_{m,k}\cdot\boldsymbol{d}^{\xi_k}\mid+\mid\boldsymbol{d}^{\xi_k}\mid^2\Big\}$$

and hence

$$| \boldsymbol{w}_{m,n} |^2 = | \boldsymbol{w}_{m,0} |^2 + \sum_{k=0}^{n-1} \Big\{ -2 \mid \boldsymbol{w}_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_k} \mid^2 \Big\}. \tag{55}$$

Case 3: $S_m^{\boldsymbol{w}_n}(\boldsymbol{d}^{\xi_n}) = 0$. Then $\boldsymbol{w}_{m,n} \cdot \boldsymbol{d}^{\xi_n} < 0$, which gives

$$\begin{aligned} | \boldsymbol{w}_{m,n+1} |^2 &= | \boldsymbol{w}_{m,n} + \boldsymbol{d}^{\xi_n} |^2 = (\boldsymbol{w}_{m,n} + \boldsymbol{d}^{\xi_n}) \cdot (\boldsymbol{w}_{m,n} + \boldsymbol{d}^{\xi_n}) \\ &= | \boldsymbol{w}_{m,n} |^2 + 2\boldsymbol{w}_{m,n} \cdot \boldsymbol{d}^{\xi_n} + | \boldsymbol{d}^{\xi_n} |^2 \\ &= | \boldsymbol{w}_{m,n} |^2 - 2 \mid \boldsymbol{w}_{m,n} \cdot \boldsymbol{d}^{\xi_n} \mid + \mid \boldsymbol{d}^{\xi_n} \mid^2. \end{aligned}$$

Then we have

$$| \boldsymbol{w}_{m,n} |^2 = | \boldsymbol{w}_{m,0} |^2 + \sum_{k=0}^{n-1} \Big\{ -2 \mid \boldsymbol{w}_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_k} \mid^2 \Big\}. \tag{56}$$

Using (55) and (56), we obtain the desired result

$$| \boldsymbol{w}_{m,n} | = \sqrt{| \boldsymbol{w}_{m,0} |^2 + \sum_{k=0}^{n-1} \Big\{ -2 \mid w_{m,k} \cdot \boldsymbol{d}^{\xi_k} \mid + \mid \boldsymbol{d}^{\xi_k} \mid^2 \Big\}}.$$

□

5 Numerical examples

Example 5.1. In order to find $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ defined in Theorem 3.1, let $\Omega = \{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\} = \{(1,1),(2,4),(3,1)\}$, where $\boldsymbol{x^1} = (1,1)$, $\boldsymbol{x^2} = (2,4)$ and $\boldsymbol{x^3} = (3,2)$. Let $\boldsymbol{w}^* = (-1,1) \in R^2$ and $\theta^* = 1 \in R$. Then $\min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* \mid = 1 > 0$. Let T be a function from Ω to $\{0,1\}$ satisfying

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* < 0, \end{cases} \qquad \text{for all } 1 \leq i \leq 3.$$

Let $\boldsymbol{w}_0 = (1,1) \in R^2$ and $\theta_0 = 1 \in R$. Since $\boldsymbol{w}^* \cdot \boldsymbol{x}^1 - \theta^* = (-1,1) \cdot (1,1) - 1 = -1$, we have $T(\boldsymbol{x^1}) = 0$. Similarly we have $(T(\boldsymbol{x^1}), T(\boldsymbol{x^2}), T(\boldsymbol{x^3})) = (0,1,0)$. Since $\boldsymbol{w}_0 \cdot \boldsymbol{x^1} - \theta_0 = (1,1) \cdot (1,1) - 1 = 1$, we have $S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^1) = 1 > T(\boldsymbol{x^1})$, which is Case 2 and then $(\boldsymbol{w_1}, \theta_1) = (\boldsymbol{w_0} - \boldsymbol{x^1}, \theta_0 + 1) = ((0,0),2)$. Repeating the same process we get $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ in Table 5.1, where $\boldsymbol{w} = \boldsymbol{w}_4 = (0,2)$ and $\theta = \theta_4 = 3$.

n	w_n	θ_n	$\left(T(x^1), T(x^2), T(x^3)\right)$	$\left(S_{w_n,\theta_n}(x^1), S_{w_n,\theta_n}(x^2), S_{w_n,\theta_n}(x^3)\right)$	Case
0	$(1,1)$	1	$(0,1,0)$	$(1,1,1)$	Case 2
1	$(0,0)$	2	$(0,1,0)$	$(0,0,0)$	Case 3
2	$(2,4)$	1	$(0,1,0)$	$(1,1,1)$	Case 2
3	$(1,3)$	2	$(0,1,0)$	$(1,1,1)$	Case 2
4	$(0,2)$	3	$(0,1,0)$	$(0,1,0)$	Case 1

Table 5.1: Construction of $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$

Example 5.2. In order to find $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ defined in Theorem 3.2, let $\Omega = \{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\} = \{(1,1),(2,4),(3,1)\}$, where $\boldsymbol{x^1} = (1,1)$, $\boldsymbol{x^2} = (2,4)$ and $\boldsymbol{x^3} = (3,1)$. Let $\boldsymbol{w}^* = (-1,1) \in R^2$, $\theta^* = 1 \in R$ and $\phi(\boldsymbol{x},\boldsymbol{y}) = (\phi_1(\boldsymbol{x},\boldsymbol{y}), \phi_2(\boldsymbol{x},\boldsymbol{y})) = ((tanh(\boldsymbol{x}), tanh(\boldsymbol{y}))$. Then we have $\min_\Omega \mid \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* \mid = 1.23346 > 0$. Let T

be a function from Ω to $\{0,1\}$ satisfying

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^i) - \theta^* < 0 \end{cases} \qquad \text{for all } 1 \leq i \leq 3.$$

Let $\boldsymbol{w}_0 = (1,1) \in R^2$ and $\theta_0 = 1 \in R$. Since $\boldsymbol{w}^* \cdot \phi(\boldsymbol{x}^1) - \theta^* = (-1,1) \cdot (tanh(1), tanh(1)) - 1 = -1$, we have $T(\boldsymbol{x}^1) = 0$. Similarly we have $(T(\boldsymbol{x^1}), T(\boldsymbol{x^2}), T(\boldsymbol{x^3})) = (0,0,0)$. Since $\boldsymbol{w}_0 \cdot \phi(\boldsymbol{x^1}) - \theta_0 = (1,1) \cdot \phi(1,1) - 1 = 0.5232$, we have $S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^1) = 1 > T(\boldsymbol{x}^1)$, which is Case 2 and then $(\boldsymbol{w_1}, \theta_1) = (\boldsymbol{w_0} - \phi(\boldsymbol{x}^1), \theta_0 + 1) = ((0.238406, 0.238406), 2)$. Since $\boldsymbol{w}_1 \cdot \phi(\boldsymbol{x^1}) - \theta_1 = (0.238406, 0.238406) \cdot \phi(1,1) - 2 = -1.6369$, we have $S_{\boldsymbol{w}_1,\theta_1}(\boldsymbol{x}^1) = 0 = T(\boldsymbol{x}^1)$, which is Case 1. Repeating the same process, we get $(S_{\boldsymbol{w}_1,\theta_1}(\boldsymbol{x}^1), S_{\boldsymbol{w}_1,\theta_1}(\boldsymbol{x}^2), S_{\boldsymbol{w}_1,\theta_1}(\boldsymbol{x}^3)) = (0,0,0)$, which leads to Case 1, where $\boldsymbol{w} = (0.238406, 0.238406)$ and $\theta = 2$.

Example 5.3. In order to find $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ defined in Theorem 3.3, let $\Omega = \{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\} = \{(1,1), (2,4), (3,1)\}$, where $\boldsymbol{x^1} = (1,1)$, $\boldsymbol{x^2} = (2,4)$ and $\boldsymbol{x^3} = (3,1)$. Let $\boldsymbol{w}^* = (-1,1) \in R^2$, $\theta^* = 1 \in R$ and $\psi(\boldsymbol{x}) = tanh(\boldsymbol{x})$. Then we have $\mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \psi^{-1}(\theta^*) \mid = 1.964 > 0$. Let T be a function from Ω to $\{0,1\}$ satisfying

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i) - \theta^* > 0, \\ 0 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i) - \theta^* < 0 \end{cases} \qquad \text{for all } 1 \leq i \leq 3.$$

Let $\boldsymbol{w}_0 = (1,1) \in R^2$ and $\theta_0 = 1 \in R$. Since $\psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^1) - \theta^* = \psi((-1,1) \cdot (1,1)) - 1 = -1$, we have $T(\boldsymbol{x^1}) = 0$ and same with $i = 2,3$ and then $(T(\boldsymbol{x^1}), T(\boldsymbol{x^2}), T(\boldsymbol{x^3})) = (0,0,0)$. Since $\psi(\boldsymbol{w}_0 \cdot \boldsymbol{x^1}) - \theta_0 = \psi((1,1) \cdot (1,1)) - 1 = -0.03597$, we have $S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^1) = 0 = T(\boldsymbol{x^1})$, which is Case 1. Repeating the same process we get $(S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^1), S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^2), S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^3)) = (0,0,0)$, which leads to Case 1 and then $\boldsymbol{w} = (1,1)$ and $\theta = 1$.

Example 5.4. In order to find $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ defined in Theorem 3.4, let $\Omega = \{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \leq i \leq r\} = \{(1,1), (2,4), (3,1)\}$, where $\boldsymbol{x^1} = (1,1)$, $\boldsymbol{x^2} = (2,4)$ and $\boldsymbol{x^3} = (3,1)$. Let $\boldsymbol{w}^* = (-1,1) \in R^2$, $\theta^* = 1 \in R$ and $\psi(\boldsymbol{x}) = tanh(\boldsymbol{x})$.Then

$\min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* \mid = 1 > 0$. Let T be a function from Ω to $\{0,1\}$ satisfying

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) > 0, \\ & \\ 0 & \text{if} \quad \psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^*) < 0 \end{cases} \qquad \text{for all } 1 \le i \le 3.$$

Let $\boldsymbol{w}_0 = (1,1) \in R^2$ and $\theta_0 = 1 \in R$. Since $\psi(\boldsymbol{w}^* \cdot \boldsymbol{x}^1 - \theta^*) = \psi((-1,1) \cdot (1,1) - 1) = tanh(-1) = -0.7616$, we have $T(\boldsymbol{x^1}) = 0$ and same with $i = 2,3$ and then $(T(\boldsymbol{x^1}), T(\boldsymbol{x^2}), T(\boldsymbol{x^3})) = (0,1,0)$. Since $\psi(\boldsymbol{w}_0 \cdot \boldsymbol{x^1} - \theta_0) = \psi((1,1) \cdot (1,1) - 1) = tanh(1) = 0.7616$, we have $S_{\boldsymbol{w}_0,\theta_0}(\boldsymbol{x}^1) = 1$, which is Case 2. Then $(\boldsymbol{w_1}, \theta_1) = (\boldsymbol{w_0} - \boldsymbol{x^1}, \theta_0 + 1) = ((0,0),2)$. Repeating the same process we get $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$ in Table 5.2, where $\boldsymbol{w} = \boldsymbol{w}_4 = (0,2)$ and $\theta = \theta_4 = 3$.

n	w_n	θ_n	$\left(T(x^1), T(x^2), T(x^3)\right)$	$\left(S_{w_n,\theta_n}(x^1), S_{w_n,\theta_n}(x^2), S_{w_n,\theta_n}(x^3)\right)$	Case
0	$(1,1)$	1	$(0,1,0)$	$(1,1,1)$	Case 2
1	$(0,0)$	2	$(0,1,0)$	$(0,0,0)$	Case 3
2	$(2,4)$	1	$(0,1,0)$	$(1,1,1)$	Case 2
3	$(1,3)$	2	$(0,1,0)$	$(1,1,1)$	Case 2
4	$(0,2)$	3	$(0,1,0)$	$(0,1,0)$	Case 1

Table 5.2: Construction of $\{\boldsymbol{w}_n\}$ and $\{\theta_n\}$

From Examples 5.5 to 5.7, let $M = 1$ and then we do not use the subscript m, $a = 1$ and $b = 0$.

Example 5.5. In order to find $\boldsymbol{w}$, θ that leads to Case 1 and the number of iterations defined in Theorem 3.5, let $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \le i \le r\right\}$ such that $N = 8$ and $r = 40$, where $\boldsymbol{x}^2, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^r$ are in Table 5.3.

x^{1}	0.44272	-0.69902	0.909558	0.999933	-0.90059	-0.6037	0.948375	-0.82722
x^{2}	-0.00677	-0.29784	-0.63347	-0.39371	-0.67471	-0.0709	0.592109	0.033357
x^{3}	0.431855	0.716738	0.458921	0.797586	0.785902	0.739946	-0.19779	-0.64469
x^{4}	0.058985	-0.42012	-0.98384	-0.95925	-0.30428	-0.41613	-0.81827	0.963871
x^{5}	0.634855	0.668843	-0.81776	-0.05379	0.864826	0.318731	-0.88349	0.812109
x^{6}	0.419503	-0.29161	-0.36516	0.994753	0.316163	-0.16754	-0.53998	0.267101
x^{7}	0.31048	-0.88587	-0.73828	-0.9999	-0.98824	-0.82213	0.864173	0.910545
x^{8}	0.455246	-0.48992	0.392558	0.963671	-0.53593	-0.12044	0.705057	-0.54854
x^{9}	0.531183	-0.17446	-0.49782	0.983906	0.296514	-0.2296	-0.48844	0.428006
x^{10}	-0.02897	-0.17137	-0.9097	0.710488	-0.00362	-0.13496	-0.49437	0.638931
x^{11}	0.561937	-0.37556	0.629851	0.997869	-0.55949	-0.03214	0.899936	-0.71128
x^{12}	0.635684	0.910769	-0.89435	-0.94524	0.980058	0.803153	-0.97635	0.847138
x^{13}	0.581569	0.51616	0.546931	0.998924	0.69866	0.268211	-0.41202	-0.1819
x^{14}	0.480978	-0.62797	-0.30449	-0.93841	-0.92211	-0.51968	0.778296	0.716912
x^{15}	0.679279	-0.03162	-0.53604	0.997346	0.31815	-0.32859	-0.41015	0.598961
x^{16}	0.521988	-0.07088	-0.74965	0.554874	0.31291	-0.03698	-0.50226	0.655467
x^{17}	0.53143	-0.87798	-0.59905	-0.99906	-0.9856	-0.75454	0.929371	0.867073
x^{18}	0.339508	-0.28165	0.175025	0.155895	-0.68307	-0.25625	0.788899	-0.51905
x^{19}	0.354661	0.307192	-0.84761	0.997344	0.746837	0.039741	-0.81946	0.656389
x^{20}	0.467657	-0.47586	0.945785	0.998901	-0.87908	-0.10909	0.983356	-0.95427
x^{21}	0.096565	0.60532	-0.09686	0.899977	0.655663	0.619522	-0.28118	-0.67485
x^{22}	0.427007	-0.68598	-0.173	0.14405	-0.86564	-0.6212	0.622508	0.637621
x^{23}	0.330051	-0.29608	0.685491	0.999544	-0.5849	-0.2587	0.813085	-0.77179
x^{24}	0.176966	-0.42137	-0.1656	-0.99993	-0.96272	-0.19926	0.921883	0.313346
x^{25}	0.281953	0.548729	0.364887	-0.9122	0.006079	0.517735	0.504913	-0.59972
x^{26}	0.43732	-0.30187	-0.00423	-0.95839	-0.70415	0.051552	0.664572	0.293177
x^{27}	0.574156	-0.54568	0.516704	0.999543	-0.54386	-0.50762	0.528823	-0.21029
x^{28}	0.596989	-0.31944	0.291238	0.990407	0.017806	0.125437	0.308341	-0.27502
x^{29}	0.232824	-0.44555	0.621784	0.999842	-0.60958	-0.31764	0.779583	-0.7473
x^{30}	0.602654	-0.4502	-0.71431	-0.87881	-0.64501	-0.32602	0.410362	0.724883
x^{31}	0.353448	0.53548	-0.80249	-0.9991	0.44722	0.399881	-0.73014	0.801815
x^{32}	0.580819	0.806498	-0.75229	-0.29988	0.965447	0.701746	-0.95769	0.715671

x^{33}	0.045781	0.188572	-0.41702	0.996414	0.504292	0.222294	-0.27646	-0.38928
x^{34}	0.471454	0.088356	-0.26911	0.999918	0.518818	-0.17207	-0.54161	0.300503
x^{35}	0.411829	0.367499	-0.96168	0.957686	0.90004	0.085952	-0.97106	0.905397
x^{36}	0.72806	0.044155	-0.58074	0.939977	0.587735	0.166467	-0.46193	0.586048
x^{37}	0.473735	-0.0302	-0.3807	-0.99919	-0.73134	0.254257	0.628837	0.535523
x^{38}	0.237162	0.668166	-0.53674	-0.10169	0.812937	0.454635	-0.86589	0.364313
x^{39}	0.432439	0.062215	-0.21667	0.882506	-0.1216	0.235702	0.597194	-0.34654
x^{40}	0.349254	-0.03274	-0.83492	-0.98987	-0.18896	-0.21626	-0.5644	0.849977

Table 5.3: Data set Ω with $N = 8$ and $r = 40$

Let $\boldsymbol{w}^*$ =(0.06552036, -1.6806747, 1.712667, -2.7097487, -3.393183, -1.0610393, 3.2309122, -1.1821818)$\in R^8$ and $\theta^* = -0.8439462$. Then $\min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* \mid = 0.159001671262379 > 0$. Let T be a function from Ω to $\{0, 1\}$ satisfying

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i - \theta^* < 0, \end{cases} \qquad \text{for all } 1 \le i \le 40.$$

Let $\boldsymbol{w}_0 = (0, 0, 0, 0, 0, 0, 0, 0) \in R^8$ and $\theta_0 = 0$. By using a MATLAB code we get the number of iterations 8, where $\boldsymbol{w}$ = (-0.316182588, -1.94431553, 0.78118317, -1.863868405, -2.988500945, -1.29417711, 1.58361996,-0.52080689) and $\theta = -1$.

Example 5.6. In order to find $\boldsymbol{w}$ that leads to Case 1 and the number of iterations defined in Theorem 4.2, let $\Omega = \left\{\boldsymbol{x}^i | \boldsymbol{x}^i \in R^N, 1 \le i \le r\right\}$ such that $N = 8$ and $r = 40$, where $\boldsymbol{x}^2, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^r$ are the same in Example 5.5. Let $\boldsymbol{w}^*$ =(0.06552036, -1.6806747, 1.712667, -2.7097487, -3.393183, -1.0610393, 3.2309122, -1.1821818)$\in R^8$. Then $\min_\Omega \mid \boldsymbol{w}^* \cdot \boldsymbol{x}^i \mid = 0.337028247412895 > 0$. Let T be a function from Ω to $\{0, 1\}$ satisfying

$$T(\boldsymbol{x}^i) = \begin{cases} 1 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i > 0, \\ 0 & \text{if} \quad \boldsymbol{w}^* \cdot \boldsymbol{x}^i < 0, \end{cases} \qquad \text{for all } 1 \le i \le 40.$$

Let $\boldsymbol{w}_0 = (0,0,0,0,0,0,0,0) \in R^8$. By using a MATLAB code we get 7 iterations without using the sorting in Theorem 4.2, where $\boldsymbol{w}$ = (-0.960143586, 0.19196406, 1.21689047, -1.73142355, -1.4117441496, -0.66065294, 1.85266342, -1.19607602). Applying the sorting in Theorem 4.2 with

$$\xi_n = \underset{i \in I, S^{\boldsymbol{w}_n}(\boldsymbol{x}^i) \neq T(\boldsymbol{x}^i)}{\operatorname{argmax}} \left| \frac{\boldsymbol{w}^*}{|\boldsymbol{w}^*|} \cdot \boldsymbol{x}^i \right|,$$

the update stops after 2 iterations with $\boldsymbol{w}$ = (-0.41182914, -0.3674993, 0.9616763, -0.95768565, -0.90004015, -0.085952446, 0.971062, -0.90539664). Here $\{\xi_n\}$ is the sequence of 35, 19, 32, 5, 17, 24, 7, 12, 14, 38, 36, 1, 26, 20, 34, 21, 13, 15, 18, 33, 16, 9, 3, 6, 37, 10, 22, 30, 29, 31, 23, 11, 2, 25, 8, 27, 4, 40, 28, 39, which is obtained by using a MATLAB code.

Example 5.7. In this example we show the application of our theorems to linearly non-separable data. At first we generate 50 input and target data sets of Boolean states and next add 10% same input data with different target data for each data set to make linearly non-separable data as Table 5.4, where the dimension and the number of input data are denoted by 'Dim. of data' and 'No. of data', respectively. To select linearly separable data Ω from the non-separable data as well as T definded on Ω, we use a deep neural network, which is net=nn.Sequential(nn.Linear (dim,10), nn.Tanh(), nn.Linear(10,8), nn.LeakyReLU(), nn.Linear(8,1)).

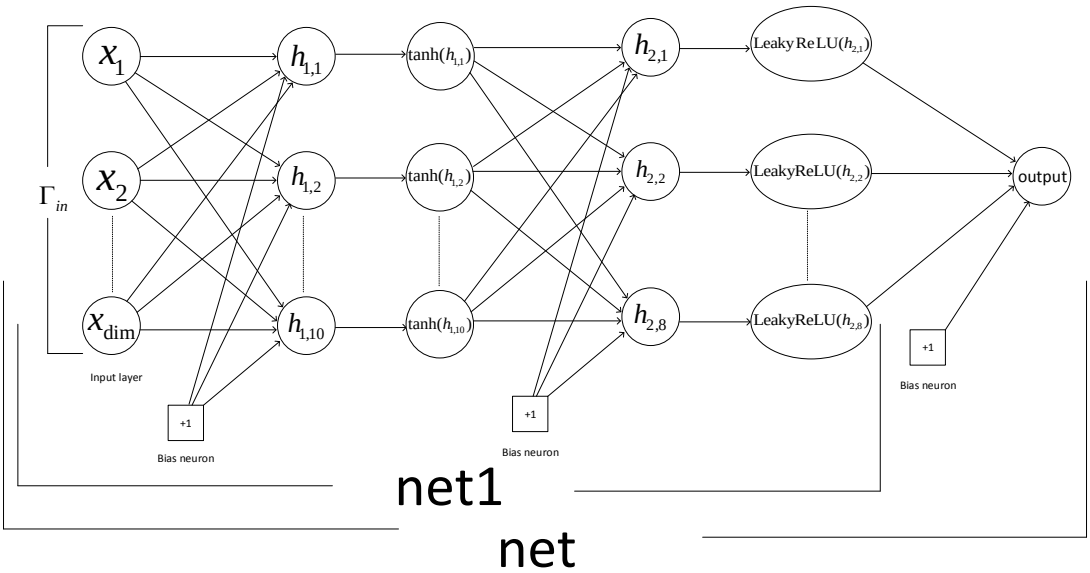


Figure 5.1: Neural Network architectures of net and net1

In case of the first input and target data set $(\Gamma_{in,1}, \Gamma_{tar,1})$ in Table 5.4, we define the linearly separable data set Ω_1 as $net1(\Gamma_{in,1})$. T_1 is defined by using parameters in nn.Linear(8,1) and $\Gamma_{tar,1}$, where net1= nn.Sequential(nn.Linear(dim,10), nn.Tanh(), nn.Linear(10,8), nn.LeakyReLU()) as in Figure 5.1 . The number of data of Ω_1 and the update number are 224 and 446 which are denoted by 'No. of selected data' and 'Update No.' in Table 5.4, respectively. Sorting the data of Ω_1 with ξ_n in Theorem 4.2, we can reduce the update number 446 to 104 ('Update No. with sorting' in Table 5.4). Applying Remark 4.1 on Theorem 4.2 then the update stops if

$$n > \mid \boldsymbol{w}^* \mid^2 \frac{\max\limits_{i \in I} \mid \boldsymbol{x}^i \mid^2}{\min\limits_{i \in I} \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* \mid^2}.$$

In order to satisfy the inequality faster we introduce threshold $\zeta = 0.01$ to make $\min\limits_{i \in I} \mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* \mid$ bigger by removing the data $\boldsymbol{x}^i$ such that have $\mid \boldsymbol{x}^i \cdot \boldsymbol{w}^* \mid < \zeta$. Applying the threshold to Ω_1 gives the set $\Omega_{1,\zeta}$ of 242 data as in 'No. of threshold data'. In the case of $\Omega_{1,\zeta}$, the update numbers without and with sorting are 487 and 30, respectively.

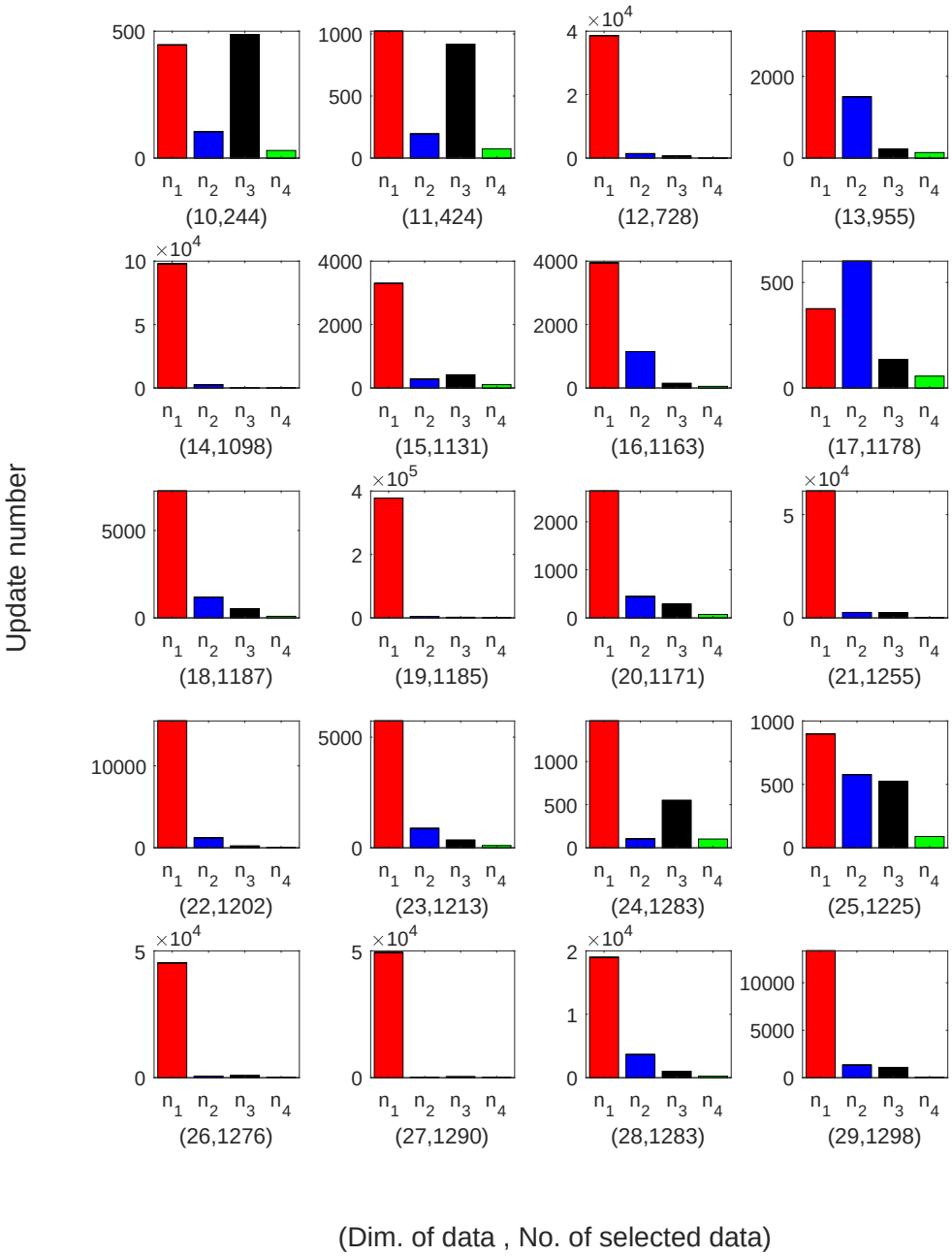
Similar process is applied to other data sets, which gives in Table 5.4. This show that update can stop faster if our sorting and threshold approaches are applied.

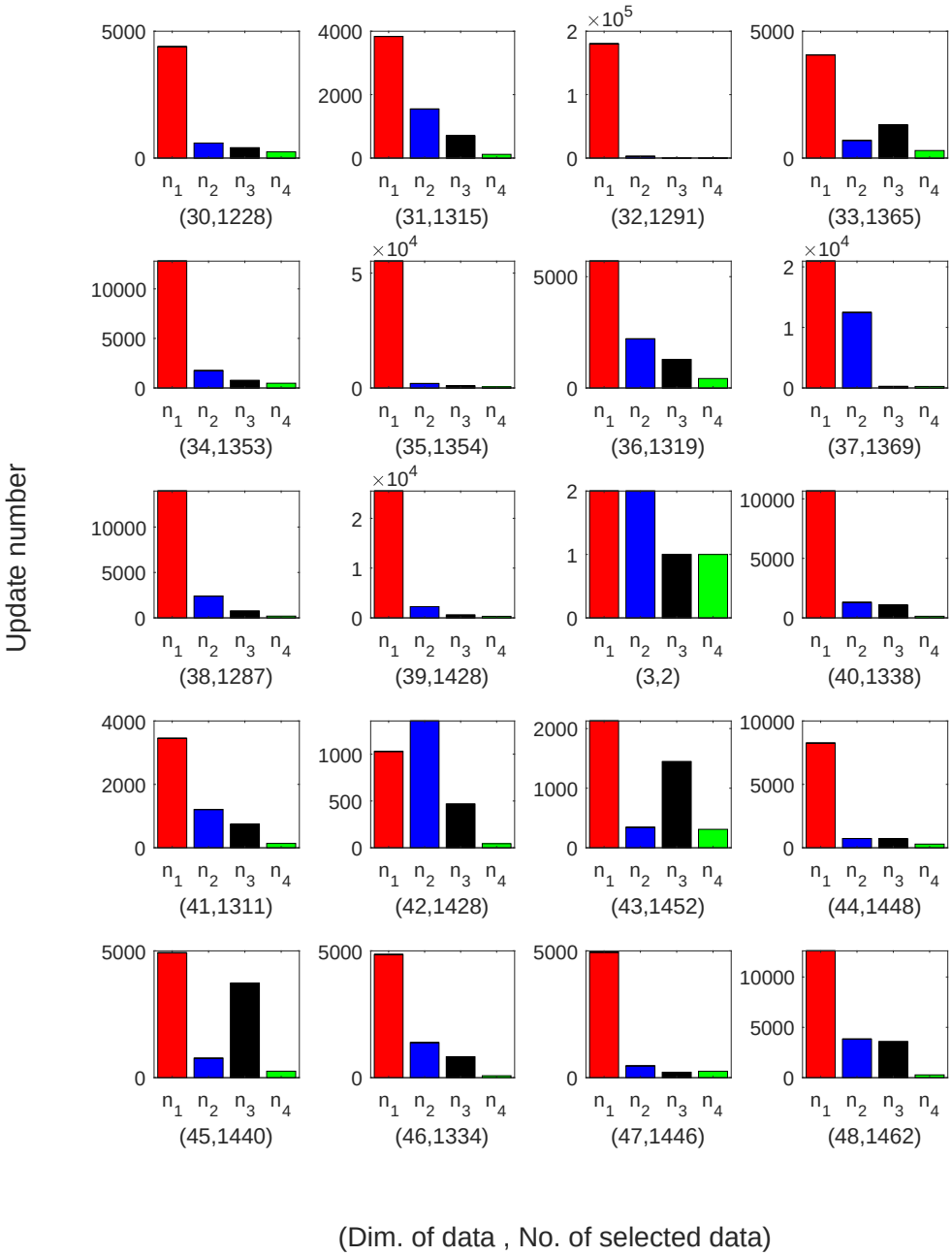
Data set ID	Dim. of data	No. of data	No. of selected data	Update No.	Update No. with sorting	No. of threshold data	Update No. with threshold	Update No. with threshold and sorting
			N1	n1	n2 (n2/n1*100)	N2(N2/N1*100)	n3(n3/n1*100)	n4 (n4/n1*100)
1	10	344	244	446	104(23.3%)	242(99.18%)	487(109.19%)	30(6.73%)
2	11	683	424	1022	195(19%)	419(98.8%)	918(89.82%)	75(7.34%)
3	12	1356	728	38619	1395(3.6%)	701(96.29%)	613(1.59%)	42(0.108%)
4	13	1752	955	3103	1499(48.3%)	917(96.02%)	222(7.154%)	135(4.35%)
5	14	1977	1098	97889	2571(2.6%)	1050(95.62%)	371(0.38%)	156(0.159%)
6	15	2092	1131	3298	278(8.43%)	1109(98.05%)	416(12.61%)	105(3.184%)
7	16	2155	1163	3946	1147(29.1%)	1113(95.7%)	145(3.674%)	51(1.3%)
8	17	2190	1178	375	601(160.3%)	1167(99.07%)	135(36%)	56(14.93%)
9	18	2204	1187	7259	1184(16.3%)	1137(95.7%)	521(7.18%)	91(1.254%)
10	19	2212	1185	377753	4794(1.2%)	1128(95.2%)	1957(0.52%)	585(0.155%)
11	20	2217	1171	2638	446(16.9%)	1146(97.9%)	287(10.88%)	72(2.79%)
12	21	2222	1255	61435	2627(4.27%)	1225(97.6%)	2572(4.187%)	208(0.339%)
13	22	2222	1202	15484	1215(7.8%)	1136(94.5%)	219(1.414%)	20(0.13%)
14	23	2221	1213	5733	878(15.3%)	1187(97.9%)	348(6.07%)	100(1.744%)
15	24	2220	1283	1470	104(7.07%)	1275(99.4%)	550(37.415%)	102(6.9%)
16	25	2222	1225	898	577(64.3%)	1195(97.6%)	525(58.46%)	89(9.9%)

Data set ID	Dim. of data	No. of data	No. of selected data	Update No.	Update No. with sorting	No. of threshold data	Update No. with threshold	Update No. with threshold and sorting
			N1	n1	n2 (n2/n1*100)	N2(N2/N1*100)	n3(n3/n1*100)	n4 (n4/n1*100)
17	26	2222	1276	45260	635(1.4%)	1252(98.12%)	877(1.938%)	114(0.252%)
18	27	2222	1290	49352	249(0.5%)	1277(99%)	412(0.835%)	78(0.158%)
19	28	2222	1283	18977	3662(19.3%)	1266(98.7%)	1009(5.32%)	236(1.244%)
20	29	2222	1298	13361	1339(10%)	1272(98%)	1086(8.12%)	54(0.4%)
21	30	2222	1228	4389	591(13.5%)	1201(97.8%)	406(9.25%)	244(5.56%)
22	31	2222	1315	3836	1547(40.3%)	1302(99.01%)	704(18.35%)	117(3.05%)
23	32	2222	1291	180148	2758(1.5%)	1259(97.52%)	443(0.246%)	207(0.115%)
24	33	2222	1365	4067	696(17.11%)	1346(98.6%)	1306(32.11%)	297(7.3%)
25	34	2222	1353	12797	1752(13.7%)	1343(99.26%)	760(5.939%)	480(3.75%)
26	35	2222	1354	55223	1939(3.5%)	1333(98.45%)	920(1.67%)	516(0.93%)
27	36	2222	1319	5697	2206(38.7%)	1302(98.711%)	1279(22.45%)	425(7.46%)
28	37	2222	1369	20944	12471(59.5%)	1344(98.17%)	265(1.26%)	257(1.23%)
29	38	2222	1287	13962	2400(17.19%)	1252(97.3%)	756(5.414%)	196(1.4%)
30	39	2222	1428	25478	2257(8.86%)	1415(99.1%)	583(2.29%)	317(1.24%)
31	3	2	2	2	2(100%)	2(100%)	1(50%)	1(50%)
32	40	2222	1338	10655	1303(12%)	1308(97.8%)	1075(10.09%)	136(1.28%)
33	41	2222	1311	3451	1207(35%)	1299(99.08%)	742(21.5%)	138(4%)
34	42	2222	1428	1030	1355(131.5%)	1417(99.23%)	468(45.43%)	44(4.272%)
35	43	2222	1452	2128	343(16.11%)	1446(99.6%)	1446(67.9%)	310(14.57%)
36	44	2222	1448	8252	733(8.9%)	1439(99.4%)	711(8.61%)	292(3.54%)
37	45	2222	1440	4925	768(15.6%)	1432(99.44%)	3724(75.6%)	257(5.22%)
38	46	2222	1334	4855	1383(28.5%)	1320(98.95%)	824(16.97%)	78(1.6%)
39	47	2222	1446	4947	466(9.4%)	1437(99.37%)	212(4.29%)	254(5.13%)
40	48	2222	1462	12592	3837(30.5%)	1453(99.38%)	3584(28.46%)	279(2.22%)
41	49	2222	1408	4063	560(13.8%)	1403(99.645%)	1384(34.06%)	355(8.74%)
42	4	5	5	2	2(100%)	5(100%)	2(100%)	2(100%)
43	50	2222	1465	4049	1213(30%)	1457(99.45%)	1213(29.96%)	300(7.41%)
44	51	2222	1337	2241	1125(50.2%)	1323(98.952%)	771(34.4%)	162(7.23%)
45	52	2222	1471	86607	2464(2.85%)	1453(98.77%)	3989(4.6%)	527(0.61%)
46	5	10	9	2	2(100%)	9(100%)	2(100%)	2(100%)
47	6	17	16	2	2(100%)	16(100%)	2(100%)	2(100%)
48	7	44	40	10	5(50%)	40(100%)	7(70%)	2(20%)
49	8	87	79	4	9(225%)	79(100%)	4(100%)	9(225%)
50	9	171	152	163	34(20.8%)	152(100%)	152(93.25%)	45(27.61%)

Table 5.4: Update numbers with sorting and threshold. Dim. and No. denote dimension and number, respectively.

The update numbers n_1, n_2, n_3, n_4 in Table 5.4 can be represented as Figure 5.2.


Update number
(10,244)
(11,424)
(12,728)
(13,955)
(14,1098)
(15,1131)
(16,1163)
(17,1178)
(18,1187)
(19,1185)
(20,1171)
(21,1255)
(22,1202)
(23,1213)
(24,1283)
(25,1225)
(26,1276)
(27,1290)
(28,1283)
(29,1298)
(Dim. of data , No. of selected data)


Update number
(30,1228)
(31,1315)
(32,1291)
(33,1365)
(34,1353)
(35,1354)
(36,1319)
(37,1369)
(38,1287)
(39,1428)
(3,2)
(40,1338)
(41,1311)
(42,1428)
(43,1452)
(44,1448)
(45,1440)
(46,1334)
(47,1446)
(48,1462)
(Dim. of data , No. of selected data)

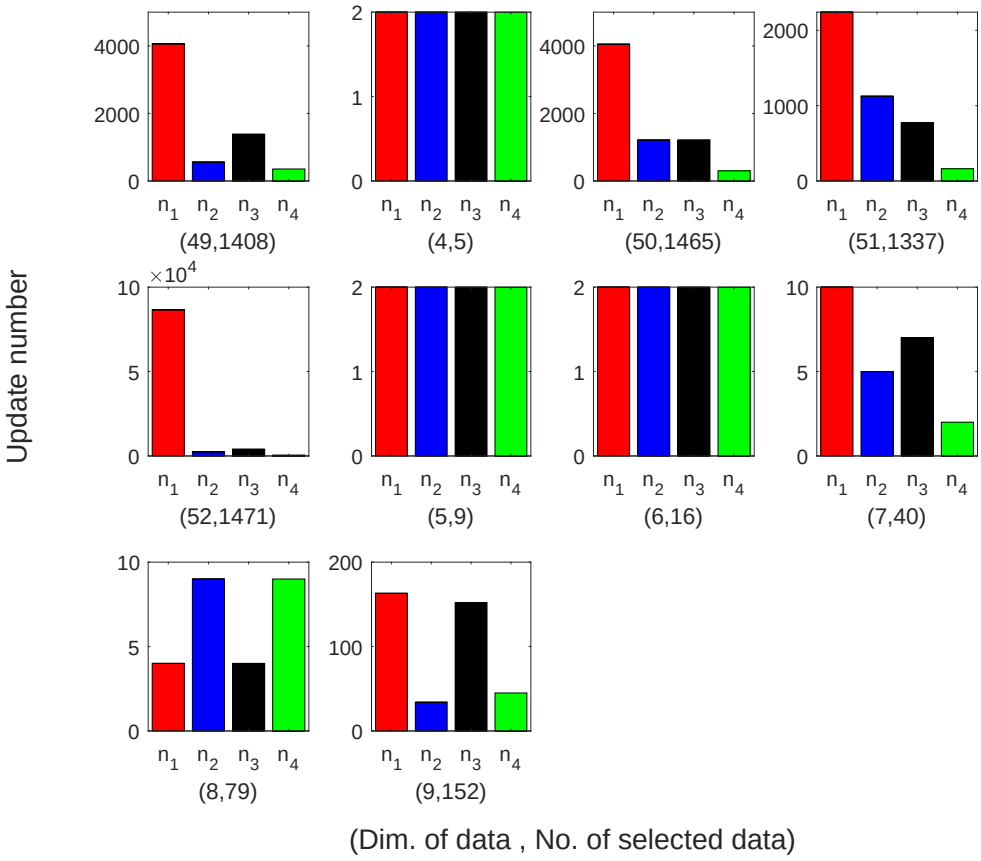


Figure 5.2: Update numbers with sorting and threshold in Table 5.4

Figure 5.3 below is a subfigure of Figure 5.2 that shows the update numbers decrease when sorting or threshold for data set is applied, where Dim. of data and and No. of selected data are 25 and 1125, respectively.

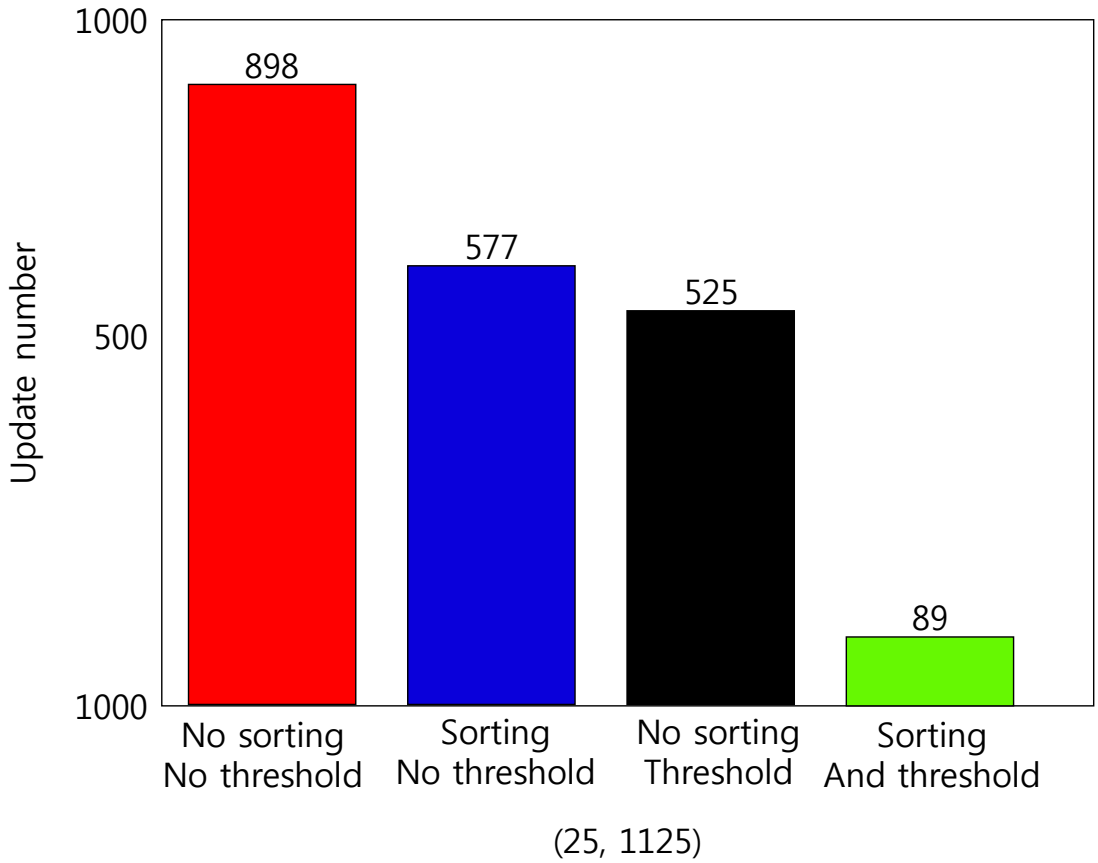


Figure 5.3: Update numbers with sorting and threshold for data set with Dim. of data = 25 and No. of selected data = 1125

Matlab code for generationg input and target

```
%Generating input and target
clear all, close all, clc
n=[];
R=[];
maxN=50;
K=randperm(maxN,maxN)+1;
for i= 1:maxN
    %Step1: Generate (N,r)
    N=K(i);
    r=min(2000,0.9*2^N);
    %Step2: Generate numbers 0 or 1 in each cell
    input=randi([0,1],fix(r),N);
    output=randi([0,1],fix(r),1);
    %make all different input_target
    [diff_input,ia,ib]=unique(input,'rows','stable');%check for uniqueness between rows
    diff_out=output(ia);
    %make 10% repeated input with different output (nois data)
    q=size(diff_out);
    new_r=q(1);
    numberOfRepeated=fix(new_r*0.1)
    for k=1:numberOfRepeated
        a=diff_input(k,:);
        diff_input(new_r-k,:)=a;
        diff_out(k)=1;
        diff_out(new_r-k)=0;
    end
    last_input_output=[diff_input,diff_out];
    [r,N]=size(diff_input);
    %Step3: save the generated data with csv file
    csvwrite(".\input_N_r_target\input_"+N+"_"+r+"_target.csv",last_input_output)
    n=[n,N];
    R=[R,r];
end
```

```
Nr=[n ; R]';
csvwrite(".\input_N_r_target\Nr.csv",Nr)
```

PyTorch code for finding parameters

```
import os
import glob
path = 'c:\\'
extension = 'csv'
csvfiles = glob.glob('*.{}'.format(extension))
import numpy as np;import pandas as pd
import torch;import torch.nn as nn;torch.manual_seed(1)
from sklearn.metrics import accuracy_score
AccOfT=[]
num_inputfiles=0

for x in csvfiles:
    feature_response=pd.read_csv(x,header=None)
    feature=torch.tensor(feature_response.values[:, :-1],
        dtype=torch.float32)
    #unique_data, index_data = np.unique(feature_response.
        values, axis=0, return_index=True)
    response=torch.tensor(feature_response.values[:,-1],
        dtype=torch.float32)
    r=feature.shape[0]
    response1=response.view(r,1)
    dim=feature.shape[1]
    dim=feature.shape[1]
    net=nn.Sequential(nn.Linear(dim,10),nn.Tanh(),nn.Linear
        (10,8),nn.LeakyReLU(),nn.Linear(8,1))
    params=list(net.parameters())
    optimizer = torch.optim.SGD(net.parameters(), lr=0.1)
    loss_ftn=nn.BCEWithLogitsLoss()
    print("loss")
    losses = []
    for iter in range(1,50000):
        batch_loss = 0.0
        optimizer.zero_grad()
        loss=loss_ftn(net(feature),response1)
```

```
31            loss.backward()
32            optimizer.step()
33            batch_loss += loss.item()
34            losses.append(batch_loss)
35        output=(net(feature)>0.0).type(torch.uint8)
36        acc=accuracy_score(response1, output)
37        AccOfT+= [acc,]
38        print(acc)
39        print(AccOfT)
40 #finding WTheta_star
41        learned_params=net.state_dict()
42        W_star=params[-1].data.numpy()
43        #Theta_star=params[-1].data.numpy()
44 #sellect xi
45        output=(net(feature)>0.0).type(torch.uint8)
46        response1=response.view(r,1)
47        L=output==response1
48        xi_index=[i for i, x in enumerate(L) if x]
49        xi=feature[xi_index]
50 #Extract X^i
51        net1=nn.Sequential(nn.Linear(dim,10),nn.Tanh(),nn.
             Linear(10,8),nn.LeakyReLU())
52        net1.load_state_dict(learned_params, strict=False)
53        Xi=net1(xi).data.numpy()#num Xi=r
54        paramss=list(net1.parameters())
55 #save W*  and X^i
56        import pandas as pd
57        pd.DataFrame(Xi).to_csv(r'C:\Users\Uou\Desktop\thesis\
             train_data\input_N_r_target_ params\input.csv'+x,
             index=False, header=False)
58        pd.DataFrame(W_star).to_csv(r"C:\Users\Uou\Desktop\
             thesis\train_data\input_N_r_target_ params\params.
             csv"+x,index=False, header=False)
59        print(num_inputfiles)
60 pd.DataFrame(csvfiles).to_csv(r'C:\Users\Uou\Desktop\thesis
      \train_data\input_N_r_target_ params\csvfiles.csv',index
      =False, header=False)
```

Matlab function code for teacher

```
function q = Teacher_Thm1(W,x)
    if dot(W,x) > 0
        q=1;
    else
        q=0;
    end
end
```

Matlab function code for student

```
function p = Student_Thm1(W,x)
    if dot(W,x)> 0
        p=1;
    elseif dot(W,x)< 0
        p=0;
    else
        p= randi([0,1]);
    end
end
```

Matlab sorting code

```
clear all; close all; clc;
%csvfiles=readtable('csvfiles.csv')
[totalData,str,raw] =xlsread('.\input_N_r_target_ params\
    csvfiles.csv');
allr=[];
Alln=[];
for q=1:50
    tic
    params=xlsread(".\input_N_r_target_ params\params.csv"+
        str(q));%read (w*; theta*) of Teacher
    W_star=params(1,:);%read w^* of Teacher
    X=xlsread(".\input_N_r_target_ params\input.csv"+str(q)
        );%read inputs Xi=(T2oT1)(xi) in R^N(1<=i<=r, r=984,
         N=8)
    r=size(X,1);
    N=size(X,2); %dimension of Xi
    %%%%% threshold
```

```
    for i=1:r
        temp1(i)=dot(W_star,X(i,:));
    end
    find_idx=find(abs(temp1)>=0.01);
    X_del=X(find_idx,:);
    X=[];X=X_del;
    temp1=[];
    r=size(X,1); %number of Xi_del
    N=size(X,2); %dimension of Xi

    %%%%%% Initial W,Theta   %%%%%%%%%%%%%%
    W = zeros(1,N); %Theta = 0;
    W0=W; %Theta0=Theta;
    %%%%%%%%%%%%%%%% sorting X^i
    XW=[];
    for z=1:r
        XW=[XW,abs( dot(W_star,X(z,:))/norm(W_star))];
    end
    [XW_sort,I] = sort(XW,'descend');%sort as descending order
    X=X(I,:);
    %%%%%%%%%% read all inputs with targets
    allinput=xlsread(".\input_N_r_target_ params\allinput.csv"+str(q));
    num_inputs=size(allinput,1);
    %%%%%%%%%%%%%%%
    allr=[allr,num_inputs];
    %%%%%%% Teacher
    for k=1:r
        Teacher(k)=Teacher_Thm1(W_star,X(k,:));
    end
    %%%%%%%%%%%%%%%
    W_k=W0;
    %%%%%%
    lr=1.0; n=0;accS=[];all_W=[];
    while true
        n=n+1 ;
```

```
        for idx=1:r
            ST=Student_Thm1(W,X(idx,:))-Teacher(idx);
            switch ST
                case 1 % Student(i)=1 , Teacher(i)=0
                    W = W - X(idx,:);
                    break
                case -1 % Student(i)=0 , Teacher(i)=1
                    W = W + X(idx,:);
                    break
            end
        end
        if idx==r
            break
        end
    end
    disp(q)
    Alln=[Alln,n];
    toc
    TOC(q)=toc;
end
allN =readtable('.\input_N_r_target_ params\csvfiles.csv');
allN=table2array(allN(:,2));
allr=allr';
Alln=Alln';
TOC=TOC';
tabelwewant=[ allN,allr,Alln,TOC];
header = {'N','r','n','toc'};
output = [header; num2cell(tabelwewant)];
%xlswrite('with sorting and threshold.xls',output);
xlswrite('with threshold and no sorting.xls',output)
%xlswrite('sorting with no threshold.xls',output)
%xlswrite('No sorting with no threshold.xls',output)
```

References

[1] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[2] F. Rosenblatt, *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

[3] A. C. Coolen, "A beginner's guide to the mathematics of neural networks," in *Concepts for Neural Networks*, pp. 13–70, Springer, 1998.

[4] J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: an overview," vol. 1142, no. 1, p. 012012, 2018.

[5] C. C. Aggarwal *et al.*, "Neural networks and deep learning," *Springer*, vol. 10, pp. 978–3, 2018.

[6] A. L. Caterini and D. E. Chang, *Deep Neural Networks in a Mathematical Framework*. Springer, 2018.

[7] A. G. Ivakhnenko and V. G. Lapa, "Cybernetic predicting devices," tech. rep., DTIC Docmument, 1966.

[8] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and cooperation in neural nets*, pp. 267–285, Springer, 1982.

[9] P. J. Werbos, "Applications of advances in nonlinear sensitivity analysis," in *System modeling and optimization*, pp. 762–770, Springer, 1982.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural computation*, vol. 1, no. 2, pp. 270–280, 1989.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[14] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, JMLR Workshop and Conference Proceedings, 2010.

[15] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[16] S.-M. Choo, L. M. Almomani, and K.-H. Cho, "Boolean feedforward neural network modeling of molecular regulatory networks for cellular state conversion," *Frontiers in physiology*, vol. 11, p. 1544, 2020.

[17] P. Jain, J. M. Garibaldi, and J. D. Hirst, "Supervised machine learning algorithms for protein structure classification," *Computational biology and chemistry*, vol. 33, no. 3, pp. 216–223, 2009.

[18] T. O. Ayodele, "Types of machine learning algorithms," *New advances in machine learning*, vol. 3, pp. 19–48, 2010.

[19] P. Timothy Jason Shepard, "Decision fusion using a multi-linear classifier," in *Proceedings of the International Conference on Multisource-Multisensor Information Fusion*, 1998.

[20] G. Dreyfus, *Neural networks: methodology and applications*. Springer Science & Business Media, 2005.

[21] S. Haykin, *Neural networks and learning machines, 3/E*. Pearson Education India, 2010.

Abstract

Artificial neural networks have been used in diverse area and played an important role. However there are a few results on mathematical analysis of neural networks. Especially, as far as we know, there is no theoretical approach to construct the order of training data for accelerating the convergence speed of neural network algorithms.

For the construction, we consider the single layer perceptron convergence algorithm and make new convergence algorithms for different structures of the perceptron as well as their convergence proofs.

We present the order of training data for the acceleration of convergence speed based on the convergence proof. Finally, we provide numerical examples of our extended convergence theorems and the order of training data.