УДК 681.14

В.М. Коцовський, Ф.Е. Гече, О.В. Міца, А.Є. Батюк

ПРО ПОВЕДІНКУ АЛГОРИТМУ НАВЧАННЯ ПЕРЦЕПТРОНА У НЕСЕПАРАБЕЛЬНОМУ ВИПАДКУ

Анотація. У роботі доводиться узагальнення теореми про зациклювання перцептрона для навчання поліноміальних нейронних елементів. Отриманий результат може бути використаний для побудови алгоритмів навчання несепарабельних множин.

Ключові слова: перецептрон, алгоритм навчання, теорема про зациклювання.

Вступ

Поліноміальний нейронний елемент (ПНЕ) був уведений у розгляд у 60-их роках XX століття. На ідейному рівні ПНЕ із n входами відрізняється від НЕ тим, що у ньому замість звичайної лінійної зваженої суми входів $\sum_{i=1}^{n} w_i x_i$ використовується поліноміальна зважена сума вигляду

$$\sum_{k=1}^{m} w_k x_1^{j_{k1}} \dots x_n^{j_{kn}}, \quad j_{ki} \in \mathrm{N},\ i = 1, \dots, n. \tag{1}$$

Дослідження властивостей ПНЕ було зумовлено прагненням отримати клас порогових функцій більшої потужності ніж клас звичайних порогових функцій. У більш загальному випадку вивчалася можливість відокремлювати множини точок у R^n за допомогою поліноміальних гіперповерхонь, можливості яких по здійсненню дихотомій є більш універсальними, ніж відповідні можливості гіперплощин, які використовувалися у класичних перцептронах. Теоретичним підґрунтям доцільності використання ПНЕ є відома теорема Ковера [1], у якій стверджується, що з переходом до простору більшої розмірності ймовірність лінійної сепарабельності множин може тільки збільшуватися. Класичний алгоритм навчання перцептрона з незначними змінами може бути використаний для навчання ПНЕ. Відомо [1], що із зростанням розмірності частка порогових (поліноміально порогових) дихотомій швидко зменшується. Цей факт зумовлює інтерес до вивчення поведінки алгоритму навчання перцептрона у несепарабельному випадку. У роботі буде показано, що за певних припущень вагові вектори, які отримуються згідно алгоритму навчання, є обмеженими. Цей результат є узагальненням відомої теореми Ефрона про «зациклювання» перцептрона і може бути використаний для обґрунтування pocket-алгоритмів навчання ПНЕ.

Ітераційний пакетний алгоритм навчання поліноміальних нейронних елементів.

Розглянемо алгоритм навчання ПНЕ, який багато в чому схожий до алгоритму навчання перцептрона, запропонованому в [2]. Будемо розглядати множину одночленів $\Pi = \left\{P_1(\mathbf{x}),\dots,P_m(\mathbf{x})\right\}$, елементи якої є мономами вигляду

$$P_i(\mathbf{x}) = x_1^{i_1}\cdot\ldots\cdot x_n^{i_n},\ \ \mathbf{x} = (x_1,\dots,x_n)\in \mathrm{R}^n,\ \ i_j \ge 0,\ j = 1,\dots,n,\ i = 1,\dots,m.$$

Ваговий вектор w назвемо П-допустимим вектором, якщо $(\mathbf{w}, P(\mathbf{a})) \neq 0$, де $(\mathbf{w}, P(\mathbf{a}))$ — скалярний добуток векторів w та $P(\mathbf{a})$, а $P(\mathbf{a}) = (P_1(\mathbf{a}),\dots,P_m(\mathbf{a}))$.

Нехай підмножини A^+, A^- — множини n-вимірних векторів простору R^n, які задовольняють умову $A^+ \cap A^- = \varnothing$. Якщо знайдеться такий П-допустимий ваговий вектор $\mathbf{w}\in \mathrm{R}^m$, що для всіх $\mathbf{a}\in A^+$ виконується умова $(\mathbf{w}, P(\mathbf{a})) > 0$, а для всіх $\mathbf{a}\in A^-$ $(\mathbf{w}, P(\mathbf{a})) < 0$, то множини A^+ і A^- назвемо П-сепарабельними і будемо казати, що ПНЕ з ваговим вектором w відокремлює ці множини відносно системи поліномів П. Якщо крім того знайдеться таке $\delta > 0$, що для всіх $\mathbf{a}\in A^+ \cup A^-$ виконується умова $\left|(\mathbf{w}, P(\mathbf{a}))\right| > \delta$, то множини A^+ і A^- назвемо сильно П-сепарабельними, а величину δ — допуском . Слід зауважити, що сильна сепарабельність є бажаною при програмній чи технічній реалізації ПНЕ, оскільки вона дозволяє уникнути небажаного впливу похибок заокруглень або завад. Легко переконатися, що довільні П-сепарабельні скінченні множини задовольняють умову сильної П-сепарабельності. Обернене твердження не завжди є вірним.

Нехай множини A^+ і A^- є сильно П-сепарабельними і нехай $A = A^+ \cup A^-$. Опишемо алгоритм навчання ПНЕ, який дозволяє отримати ваговий вектор ПНЕ, який сильно відокремлює множини A^+ і A^- із заданим допуском $\varepsilon > 0$. Під навчаючою послідовністю будемо розуміти нескінченну послідовність векторів $\{\mathbf{a}^k\}$, яка задовольняє наступні дві умови:

$$\mathbf{a}^k \in A;$$

для всіх $r \in \mathrm{N}$ множина $\{\mathbf{a}^r, \mathbf{a}^{r+1}, \mathbf{a}^{r+2}, \ldots\}$ всюди щільна у множині A.

Для скінченної множини умова 2 рівносильна тому, що кожний елемент множини A повторюється у навчаючій послідовності безліч разів.

Нехай функція Rsign_b $(b > 0)$ обчислюється за наступним правилом:

$$\mathrm{Rsign}_b\, x = \begin{cases} \mathrm{Rsign}\, x, & |x| > b, \\ 0, & |x| \le b. \end{cases}$$

Виходячи з довільного (не обов'язкового П-допустимого) початкового наближення $\mathbf{w}^0$ будемо будувати послідовність вагових за наступним алгоритмом:

ПЕРЦЕПТРОН$\left(\{\mathbf{a}^k\}, \{\beta^k\}, l, \varepsilon\right)$

ПОЧАТОК: $k \leftarrow 0$. Вибрати в якості початкового наближення заданий вектор $\mathbf{w}^0$.

ПЕРЕВІРКА: $k \leftarrow k+1$, $j \leftarrow 0$, $\Delta\mathbf{w}^k = \mathbf{0}$.

КОЕФІЦІЄНТ: $j \leftarrow j+1$, $\gamma_{kj} \leftarrow 0$.

якщо $\mathrm{Rsign}_{2\varepsilon}\left(\mathbf{w}^{k-1}, P\left(\mathbf{a}^{(k-1)l+j}\right)\right) < 1$ і $\mathbf{a}^{(k-1)l+j} \in A^+$, то $\gamma_{kj} \leftarrow 1$ і перейти до ПРИРІСТ.

якщо $\mathrm{Rsign}_{2\varepsilon}\left(\mathbf{w}^{k-1}, P\left(\mathbf{a}^{(k-1)l+j}\right)\right) > -1$ і $\mathbf{a}^{(k-1)l+j} \in A^-$, то $\gamma_{kj} \leftarrow -1$.

ПРИРІСТ: $\Delta\mathbf{w}^k \leftarrow \Delta\mathbf{w}^k + \gamma_{kj} P\left(\mathbf{a}^{(k-1)l+j}\right)$.

Якщо $j < l$, то перейти до КОЕФІЦІЄНТ.

Якщо $\Delta\mathbf{w}^k \ne \mathbf{0}$, то перейти до КОРЕКЦІЯ, інакше перейти до ПЕРЕВІРКА

КОРЕКЦІЯ: $\mathbf{w}^k \leftarrow \mathbf{w}^{k-1} + \beta_k \Delta\mathbf{w}^k$.

Перейти до ПЕРЕВІРКА.

Рисунок 1 – Пакетний алгоритм навчання ПНЕ з допуском та змінними коефіцієнтами .

Як видно з рис. 1, у алгоритмі ПЕРЦЕПТРОН вагові вектори обчислюються за формулою:

$$\mathbf{w}^k = \mathbf{w}^{k-1} + \beta_k \Delta\mathbf{w}^k,\ \ k = 1, 2, \ldots, \qquad (2)$$

де додатний коефіцієнт β_k відповідає за швидкість навчання, а вектор корекції $\Delta\mathbf{w}^k$ обчислюється так:

$$\Delta\mathbf{w}^k = \gamma_{k1} P\left(\mathbf{a}^{(k-1)l+1}\right) + \ldots + \gamma_{kl} P\left(\mathbf{a}^{kl}\right).$$

З правила вибору коефіцієнтів γ_{kj} випливає, що у випадку $\mathbf{a}^{(k-1)l+j} \in A^+$ і $\mathrm{Rsign}_{2\varepsilon}\left(\mathbf{w}^{k-1}, P\left(\mathbf{a}^{(k-1)l+j}\right)\right) = 1$ або $\mathbf{a}^{(k-1)l+j} \in A^-$ і $\mathrm{Rsign}_{2\varepsilon}\left(\mathbf{w}^{k-1}, P\left(\mathbf{a}^{(k-1)l+j}\right)\right) = -1$. корекція по вектору $\mathbf{a}^{(k-1)l+j}$ відсутня, оскільки ваговий вектор $\mathbf{w}^{k-1}$ правильно класифікує вхідний вектор $\mathbf{a}^{(k-1)l+j}$. У інших випадках відбувається корекція і для вектора $\mathbf{w}^k$ скалярний добуток $\left(\mathbf{w}^k, \Delta\mathbf{w}^k\right)$ або вже має потрібне для класифікації значення, або принаймні ближче до потрібного значення, ніж скалярний добуток $\left(\mathbf{w}^{k-1}, \Delta\mathbf{w}^k\right)$. Алгоритм навчання ПНЕ ПЕРЦЕПТРОН $\left(\left\{\mathbf{a}^k\right\}, \left\{\beta^k\right\}, l, \varepsilon\right)$, ітерація якого описується співвідношенням (2) належить до пакетних алгоритмів навчання, оскільки при кожній ітерації (по k) у алгоритмі аналізується реакція ПНЕ на l векторів $\mathbf{a}^{(k-1)l+1}, \ldots, \mathbf{a}^{kl}$. Індекс k природно називати номером кроку алгоритму. Параметр алгоритму l будемо назвати довжиною навчального пакета даних. У випадку l = 1 ми отримаємо он-лайн версію алгоритму. Якщо при цьому $\varepsilon = 0$, $\Pi = \left\{1, x_1, \ldots, x_n\right\}$, $\beta_k = 1,\ k = 1, 2, \ldots$ то (2) перетворюється на звичайний алгоритм навчання перцептрона. Наступна теорема є узагальненням добре відомої теореми про збіжність навчання перцептрона [2] і доводиться схожим чином (див. [3]).

Теорема 1. Якими б не були початкове наближення $\mathbf{w}^0$, довжина пакета l і навчаюча послідовність $\left\{\mathbf{a}^k\right\}$, через скінчену кількість кроків алгоритму ПЕРЦЕПТРОН$\left(\left\{\mathbf{a}^k\right\}, \left\{\beta^k\right\}, l, \varepsilon\right)$ ми отримаємо ваговий вектор $\mathbf{w}^k$, який з допуском ε відокремлює обмежені, сильно П-сепарабельні множини A^{+} і A^{-} за умови, що для коефіцієнтів β_k справджується рівність

$$\lim_{k\to\infty}\frac{\sum_{j=1}^{k}\beta_j^2}{\left(\sum_{j=1}^{k}\left|\beta_j\right|\right)^2}=0$$

Зауваження. Для класичного ПЕ допуск і різні способи вибору коефіцієнтів швидкості навчання β_k вивчалися у [4]. Можливі різноманітні варіанти побудови навчаючої послідовності $\left\{\mathbf{a}^k\right\}$. Якщо $l=\mathrm{Card}\,A$ і $\beta_k=1/l$, то теорема забезпечує скінченність алгоритму навчання, для якого на кожному кроці проводиться лише одна усереднююча корекція по всім елементам скінченної множини $A=A^{+}\cup A^{-}$. Слід зазначити, що якщо початкове наближення є цілочисловим вектором і коефіцієнти β_k — цілі, то й усі вектори $\mathbf{w}^k$, які отримуються згідно (2) також є цілочисловими, що є важливим для більшості застосувань.

Обмеженість вагових векторів у випадку навчання несепарабельних множин.

Цікавим є питання поведінки величини $\left\|\mathbf{w}^k\right\|$ для несепарабельного випадку. У випадку он-лайн алгоритму навчання звичайних НЕ (l = 1, $\beta_j=\mathrm{const}$) по розпізнаванню скінченних множин спочатку експериментально, а потім і строго теоретично [5] було встановлено, що норма $\left\|\mathbf{w}^k\right\|$ є обмеженою. Це факт відомий у літературі під назвою «теорема про зациклювання перцептрона». Ми покажемо, що теорему про зациклювання можна узагальнити на випадок пакетного алгоритму навчання ПНЕ з допуском за умови, що вагові вектори $\left\{\mathbf{w}^k\right\}$ отримуються за формулою (2), а невід'ємні коефіцієнти β_k обмежені зверху. На Рис. 2 наведені факти, які наштовхують на ідею обмеженості вагових векторів у випадку m = 2, l = 1.

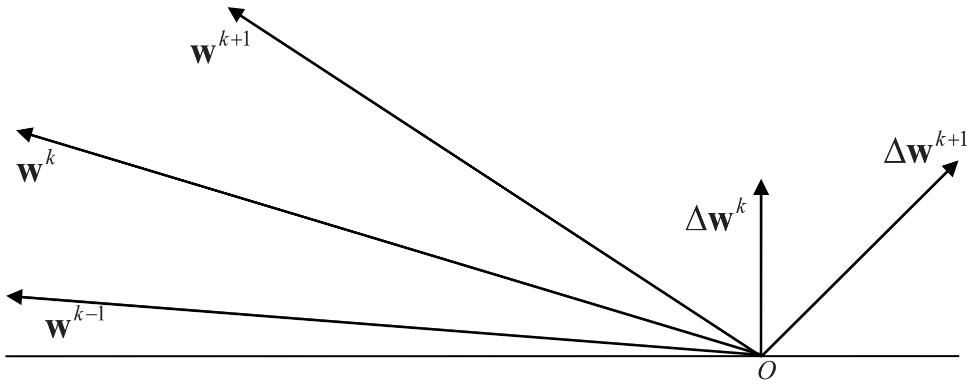

Рисунок 2

Якщо довжина вектора $\mathbf{w}^{k-1}$ набагато більша за довжину вектора $\beta_k \Delta\mathbf{w}^k$, $\mathbf{a}^k \in A^+$, то $\left\|\mathbf{w}^k\right\| > \left\|\mathbf{w}^{k-1}\right\|$ лише у тому випадку, коли кут між $\mathbf{w}^{k-1}$ і $\Delta\mathbf{w}^k$ «близький» до прямого кута. Причому при виконанні умови $\left\|\mathbf{w}^{k-1}\right\| \gg \beta_k \left\|\Delta\mathbf{w}^k\right\|$ значення різниці $\left\|\mathbf{w}^k\right\| - \left\|\mathbf{w}^{k-1}\right\|$ близьке до нуля і кут між векторами $\mathbf{w}^k$ і $\mathbf{w}^{k-1}$ також мало відрізняється від нуля. Тоді за умов корекції у алгоритмі навчання ПНЕ кут між векторами $\mathbf{w}^k$ і $\Delta\mathbf{w}^k$ не менший за деякий фіксований тупий кут (вважаємо, що $\mathbf{a}^{k+1} \in A^+$). Тоді за умови обмеженості відношень довжин сусідніх приростів $\beta_k P\left(\mathbf{a}^k\right)$ і $\beta_{k+1} P\left(\mathbf{a}^{k+1}\right)$ у алгоритмі навчання ПНЕ $\left\|\mathbf{w}^{k+1}\right\| < \left\|\mathbf{w}^{k-1}\right\|$ (для простоти міркувань на Рис. 2 зображено випадок $\beta_k = \beta_{k+1} = 1$). Однак у випадку $\beta_k \gg \beta_{k+1}$ може виявитися, що $\left\|\mathbf{w}^{k+1}\right\| > \left\|\mathbf{w}^{k-1}\right\|$. Далі ми покажемо, що і в цьому випадку згідно алгоритму (2) ми отримуємо таку послідовність векторів $\mathbf{w}^k, \mathbf{w}^{k+1}, \dots, \mathbf{w}^{k+r}$, $r \geq 1$, що $\left\|\mathbf{w}^k\right\| > \left\|\mathbf{w}^{k+1}\right\| > \dots > \left\|\mathbf{w}^{k+r}\right\|$ і $\left\|\mathbf{w}^{k+r}\right\| \leq \left\|\mathbf{w}^{k-1}\right\|$. Для строгого доведення попереднього твердження нам знадобиться ряд допоміжних означень і лем.

Нехай B — скінченна множина у R^m. Назвемо (B, β, ε)-ланцюгом послідовність векторів $\mathbf{w}^0, \mathbf{w}^1, \dots, \mathbf{w}^k$, яка задовольняє умови:

$$\left(\mathbf{w}^{j-1}, \mathbf{b}^j\right) \leq 2\varepsilon, \quad \mathbf{w}^j = \mathbf{w}^{j-1} + \beta_j \mathbf{b}^j, \quad \mathbf{b}^j \in B, \ j = 1, \dots, k.$$

(B, β, ε)-ланцюг називається правильним, якщо для всіх j $\left\|\mathbf{w}^j\right\| \geq \left\|\mathbf{w}^0\right\|$.

Лема 1. Нехай L — підпростір евклідового простору R^m, у якому лежить скінченна множина B. Проекція $\tilde{\mathbf{w}}^0, \tilde{\mathbf{w}}^1, \ldots, \tilde{\mathbf{w}}^k$ правильного (B, β, ε)-ланцюга $\mathbf{w}^0, \mathbf{w}^1, \ldots, \mathbf{w}^k$ на L також є правильним (B, β, ε)-ланцюгом. Крім того $\|\mathbf{w}^k\| - \|\mathbf{w}^0\| \le \|\tilde{\mathbf{w}}^k\| - \|\tilde{\mathbf{w}}^0\|$.

Доведення аналогічне до доведення леми 4 [5, с. 184].

Лема 2. Для довільних векторів $\mathbf{a}, \mathbf{b}$ ($\mathbf{a} \neq \mathbf{0}$) дійсного евклідового простору мають місце нерівності

$$\frac{(\mathbf{a}, \mathbf{b})}{\|\mathbf{a}\|} \le \|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| \le \frac{(\mathbf{a}, \mathbf{b}) + \|\mathbf{b}\|^2}{\|\mathbf{a}\|}.$$

Доведення. Доведення лівої частини нерівності:

$$(\mathbf{a}, \mathbf{b}) = (\mathbf{a}, \mathbf{a} + \mathbf{b}) - \|\mathbf{a}\|^2 \le \|\mathbf{a}\| \|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\|^2 = \|\mathbf{a}\| \left(\|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| \right).$$

Доведемо праву частину нерівності. Маємо

$$\|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| = \frac{\|\mathbf{a} + \mathbf{b}\|^2 - \|\mathbf{a}\|^2}{\|\mathbf{a} + \mathbf{b}\| + \|\mathbf{a}\|} = \frac{(\mathbf{a}, \mathbf{b})}{\|\mathbf{a}\|} + \frac{\|\mathbf{b}\|^2}{\|\mathbf{a} + \mathbf{b}\| + \|\mathbf{a}\|} + \frac{\left(\|\mathbf{a}\| - \|\mathbf{a} + \mathbf{b}\| \right)(\mathbf{a}, \mathbf{b})}{\left(\|\mathbf{a} + \mathbf{b}\| + \|\mathbf{a}\| \right) \|\mathbf{a}\|}.$$

Якщо $(\mathbf{a}, \mathbf{b}) \ge 0$, то $\|\mathbf{a} + \mathbf{b}\| \ge \|\mathbf{a}\|$. Тоді

$$\|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| \le \frac{(\mathbf{a}, \mathbf{b})}{\|\mathbf{a}\|} + \frac{\|\mathbf{b}\|^2}{2\|\mathbf{a}\|} \le \frac{(\mathbf{a}, \mathbf{b}) + \|\mathbf{b}\|^2}{\|\mathbf{a}\|}.$$

Нехай тепер $(\mathbf{a}, \mathbf{b}) < 0$. Якщо $\|\mathbf{a} + \mathbf{b}\| \le \|\mathbf{a}\|$, то

$$\|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| \le \frac{(\mathbf{a}, \mathbf{b})}{\|\mathbf{a}\|} + \frac{\|\mathbf{b}\|}{\|\mathbf{a} + \mathbf{b}\| + \|\mathbf{a}\|} \le \frac{(\mathbf{a}, \mathbf{b}) + \|\mathbf{b}\|^2}{\|\mathbf{a}\|}.$$

Якщо $\|\mathbf{a} + \mathbf{b}\| > \|\mathbf{a}\|$, то

$$\begin{aligned} \|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| &= \frac{(\mathbf{a}, \mathbf{b})}{\|\mathbf{a}\|} + \frac{\|\mathbf{b}\|^2}{\|\mathbf{a} + \mathbf{b}\| + \|\mathbf{a}\|} + \frac{\left(\|\mathbf{a}\| - \|\mathbf{a} + \mathbf{b}\| \right)(\mathbf{a}, \mathbf{b})}{\left(\|\mathbf{a} + \mathbf{b}\| + \|\mathbf{a}\| \right) \|\mathbf{a}\|} \le \\ &\le \frac{(\mathbf{a}, \mathbf{b})}{\|\mathbf{a}\|} + \frac{\|\mathbf{b}\|^2}{2\|\mathbf{a}\|} + \frac{\|\mathbf{b}\|(\mathbf{a}, \mathbf{b})}{2\|\mathbf{a}\|^2} \le \frac{(\mathbf{a}, \mathbf{b}) + \|\mathbf{b}\|^2}{\|\mathbf{a}\|} \end{aligned}.$$

Лема доведена.

Наслідок. Якщо $(\mathbf{a}, \mathbf{b}) = 0$, то $\|\mathbf{a} + \mathbf{b}\| - \|\mathbf{a}\| \le \|\mathbf{a}\|^{-1} \|\mathbf{b}\|^2$.

Лема 3. Якщо у дійсному евклідовому просторі кут φ між вектором $\mathbf{b}$ і одиничним вектором e є тупим, то для довільного $\delta \le -\frac{1}{2}\|\mathbf{b}\|\cos\varphi$ і довільного $\lambda \ge \frac{\|\mathbf{b}\|}{2|\cos\varphi|}$ виконується нерівність $\|\mathbf{b}+\lambda\mathbf{e}\| \le \lambda - \delta$.

Доведення. Скористаємося попередньою лемою, поклавши $\mathbf{a} = \lambda\mathbf{e}$. Тоді

$$\|\mathbf{b}+\lambda\mathbf{e}\| - \lambda \le \|\mathbf{b}\|\cos\varphi - \frac{1}{2}\|\mathbf{b}\|\cos\varphi \le -\delta .$$

Лема 4. Для довільного $\mathbf{w}^0 \in \mathrm{R}^m$ і довільної послідовності $\{\beta^k\}$, елементи якої задовольняють умову

$0 < \beta_{\min} \le \beta_k,\ \ k = 1,2,\dots,$ (3)

$0 < \beta_k \le \beta_{\max} < \infty,\ \ k = 1,2,\dots,$ (4)

знайдеться таке число $N = N(m, B, \beta_{\min}, \beta_{\max}, \varepsilon)$, що якщо $\mathbf{w}^0, \mathbf{w}^1 \dots, \mathbf{w}^s$ — правильний (B,β,ε)-ланцюг, то $\|\beta_1\mathbf{b}^1 + \dots + \beta_s\mathbf{b}^s\| \le N(m, B, \beta_{\min}, \beta_{\max}, \varepsilon)$.

Доведення. Скористаємося індукцією по m — розмірності евклідового простору, у якому міститься множина B. Нехай $d_1 = \min_{\mathbf{b}\in B}\|\mathbf{b}\|,\ d_2 = \max_{\mathbf{b}\in B}\|\mathbf{b}\|$.

При $m = 1$ з того, що $\Delta\mathbf{w}^k \ne \mathbf{0}$ випливає, що $\mathbf{w}^{k-1}\cdot\mathbf{b}^k \le 2\varepsilon$, а отже $\|\mathbf{w}^{k-1}\| > \max\{2\varepsilon / d_1, \beta_{\max}d_2\} \Rightarrow \|\mathbf{w}^k\| < \|\mathbf{w}^{k-1}\|$. Тому для довільного правильного (B,β,ε)-ланцюга $\|\beta_1\mathbf{b}^1 + \dots + \beta_s\mathbf{b}^s\| \le \max\{2\varepsilon / d_1, \beta_{\max}d_2\} + 2\beta_{\max}d_2$.

Припустимо, що теорема справджується для простору R^m і доведемо її у випадку $\dim B = m+1$ методом від супротивного.

Припустимо, що для кожного додатного L знайдеться правильний (B,β,ε)-ланцюг $\mathbf{w}^0, \mathbf{w}^1 \dots, \mathbf{w}^s$, такий, що $\|\beta_1\mathbf{b}^1 + \dots + \beta_s\mathbf{b}^s\| > L$. Неважко, показати, що тоді для довільного натурального N величина $\|\beta_1\mathbf{b}^1 + \dots + \beta_s\mathbf{b}^s\|$ є необмеженою зверху для класу (B,β,ε)-ланцюгів, які задовольняють умову $\|\mathbf{w}^0\| \ge N$. Доведемо це твердження. Нехай M — довільне додатне число, $\|\mathbf{w}^0\| < N$, $\|\beta_1\mathbf{b}^1 + \dots + \beta_s\mathbf{b}^s\| > M + 3N$ і

 ISSN 1562-9945

$N \geq \beta_{\max} d_2$.. Серед векторів $\mathbf{w}^k$, $k = 0,\ldots,s$, які задовольняють нерівність $\left\|\mathbf{w}^k\right\| \leq 2N$ виберемо вектор з найбільшим індексом. Нехай цей індекс рівний k. Тоді $\left\|\mathbf{w}^k\right\| \geq N$, $\left\|\beta_1 \mathbf{b}^1 + \ldots + \beta_k \mathbf{b}^k\right\| = \left\|\mathbf{w}^k - \mathbf{w}^0\right\| < 3N$. Тому $\left\|\mathbf{b}^{k+1} + \ldots + \mathbf{b}^s\right\| > M$. Отже, $\mathbf{w}^k,\ldots,\mathbf{w}^s$ — шуканий правильний ланцюг.

З використанням теореми Больцано-Веєрштраса можна довести існування такого одиничного вектора e, що для довільних $N > 0, M > 0$ будь-який окіл e містить такий одиничний вектор $\dfrac{\mathbf{w}^0}{\left\|\mathbf{w}^0\right\|}$, що для правильного (B,β,ε)-ланцюга $\mathbf{w}^0, \mathbf{w}^1 \ldots, \mathbf{w}^s$ виконуються нерівності $\left\|\mathbf{w}^0\right\| \geq N$ і $\left\|\beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s\right\| > M$. Позначимо через H_{e} ортогональне доповнення до підпростору, породженого вектором e у R^{m+1}. Із скінченності множини B випливає, що знайдеться таке додатне $\delta_{\mathrm{e}} > 0$, що для всіх $\mathbf{b} \in B \cap \left(R^{m+1} \setminus H_{\mathrm{e}}\right)$ виконується нерівність $\left|(\mathbf{b},\mathbf{e})\right| > 2\delta_{\mathrm{e}}$. Тоді знайдеться така відкрита куля $U_{\mathrm{e}} = B(\mathbf{e}, 2r_{\mathrm{e}})$ з центром у точці e, що для довільних $\mathbf{x} \in U_{\mathrm{e}}$ і $\mathbf{b} \in U_{\mathrm{e}} \cap \left(R^{m+1} \setminus H_{\mathrm{e}}\right)$ $\left|(\mathbf{b},\mathbf{x})\right| > \delta_{\mathrm{e}}$. Покажемо, що знайдуться такі числа $\delta > 0$ і $L_1(\mathbf{e})$, що якщо

$$\lambda > L_1(\mathbf{e}),\ \beta_{\min} < \beta < \beta_{\max}\ \ \mathbf{b} \in B \cap \left(R^{m+1} \setminus H_{\mathrm{e}}\right),\ \mathbf{x} \in U_{\mathrm{e}} \text{ і } \lambda(\mathbf{b},\mathbf{x}) \leq 2\varepsilon, \tag{5}$$

то

$$\left\|\lambda \mathbf{x} + \beta \mathbf{b}\right\| < \lambda - 2\delta. \tag{6}$$

Справді, для достатньо великих λ з того, що $\left|(\mathbf{b},\mathbf{x})\right| > \delta_{\mathrm{e}}$ і (5) випливає, що $(\mathbf{b},\mathbf{x}) < -\delta_{\mathrm{e}}$. Застосувавши лему 3, отримаємо (6).

Розглянемо довільний правильний (B,β,ε)-ланцюг $\mathbf{w}^0, \mathbf{w}^1, \ldots, \mathbf{w}^k$ і припустимо, що вектор $\left\|\mathbf{w}^0\right\|^{-1} \mathbf{w}^0$ «достатньо близький» до вектора e. Ми будемо вимагати, щоб $\left\|\mathbf{w}^0\right\|^{-1} \mathbf{w}^0 \in V_{\mathrm{e}} = B(\mathbf{x}, r_{\mathrm{e}})$. Позначимо $L(\mathbf{e}) = \max\left\{L_1(\mathbf{e}), tL_2(\mathbf{e})\right\}$, де

$$L_2(\mathbf{e}) = \frac{2}{\delta} N^2 \left(m, H_{\mathbf{e}} \cap B, \beta_{\min}, \beta_{\max}, \varepsilon \right),$$

а множник $t \geq 1$ вибирається таким чином, щоб для довільного вектора $\mathbf{b} \in H_{\mathbf{e}}$, довжина якого не більша за $L\left(m, H_{\mathbf{e}} \cap B, \beta_{\min}, \beta_{\max}, \varepsilon\right)$ кінець орт-вектора, відповідного вектору $\mathbf{w}^0 + \mathbf{b}$ потрапив у $U_{\mathbf{e}}$ (існування такого множника випливає з наслідку до леми 2). Припустимо також, що $\left\|\mathbf{w}^0\right\| > L(\mathbf{e})$.

Згідно до (5)-(6), (B, β, ε)-ланцюг $\mathbf{w}^0, \mathbf{w}^1, \ldots, \mathbf{w}^k$ не може бути правильним, якщо $\mathbf{b}^1 \notin H_{\mathbf{e}}$. Покажемо, що не тільки $\mathbf{b}^1$, а й усі вектори $\mathbf{b}^1, \ldots, \mathbf{b}^k$ належать $H_{\mathbf{e}}$. Припустимо, що $\mathbf{b}^1, \ldots, \mathbf{b}^s \in H_{\mathbf{e}}$ і вектор $\mathbf{b}^{s+1}$ не належить підпростору $H_{\mathbf{e}}$. Тоді за лемою 1 проекції $\tilde{\mathbf{w}}^0, \ldots, \tilde{\mathbf{w}}^s$ векторів ланцюга утворюють правильний ланцюг у m-вимірному просторі, а тому за припущенням індукції

$$\left\|\beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s\right\| \leq N\left(m, B \cap H_{\mathbf{e}}, \beta_{\min}, \beta_{\max}, \varepsilon\right). \tag{7}$$

Поклавши у лемі 2 $\mathbf{a} = \mathbf{w}^0$, $\mathbf{b} = \beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s$, отримаємо, що

$$\left\|\mathbf{w}^s\right\| \leq \left\|\mathbf{w}^0\right\| + \delta / 2 < \left\|\mathbf{w}^0\right\| + \delta$$

$$\text{і } \mathbf{w}^s \in U_{\mathbf{e}}.$$

Оскільки $\mathbf{b}^{s+1} \notin H_{\mathbf{e}}$, то з урахуванням (5)-(6) і попередньої нерівності отримуємо

$$\left\|\mathbf{w}^{s+1}\right\| < \left\|\mathbf{w}^s\right\| - 2\delta < \left\|\mathbf{w}^0\right\| - \delta .$$

Таким чином (B, β, ε)-ланцюг не може бути правильним, якщо усі вектори $\mathbf{b}^1, \ldots, \mathbf{b}^k$ не належать гіперплощині $H_{\mathbf{e}}$. Отже, для довільного правильного (B, β, ε)-ланцюга, початок якого $\mathbf{w}^0$ «достатньо близький» до e і $\left\|\mathbf{w}^0\right\| > L(\mathbf{e})$, величина $\left\|\beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s\right\|$ задовольняє (7). Це суперечить вибору e. Лема доведена.

Слід зауважити, що частинний випадок попередньої леми ($\varepsilon = 0$, $\beta_k = 1$) неявно і без доведення використовувався у [5] при доведенні теореми про зациклювання перцептрона. Тому доведення леми

4 заповнює прогалини у доведенні згаданої теореми і робить його коректним. Лема 4 перестає бути вірною у випадку, коли у (4) $\beta_{\max} = +\infty \left(\overline{\lim_{k\to\infty}} \beta_k = +\infty \right)$. Умова (3) використовувалася тільки для доведення того, що з (5) випливає (6). Шляхом ускладнення міркувань можна обійтися без (3).

Лема 5. Для довільного довільної невід'ємної послідовності $\left\{\beta^k\right\}$, елементи якої задовольняють (4) знайдеться таке число $N = N\left(m, B, \beta_{\max}, \varepsilon\right)$, що якщо $\mathbf{w}^0, \mathbf{w}^1 \ldots, \mathbf{w}^s$ — правильний $\left(B, \beta, \varepsilon\right)$-ланцюг, то $\left\| \beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s \right\| \le N\left(m, B, \beta_{\max}, \varepsilon\right)$.

Доведення. Так само, як і у доведенні леми 4 скористаємося індукцією по m. Випадок $m = 1$ доводиться аналогічно. Припустимо, що теорема є вірною для простору R^m і доведемо, що тоді вона справджується і для R^{m+1}. Нехай e — довільний одиничний вектор простору R^{m+1}. Використаємо ті самі позначення, що й при доведенні леми 4 і визначимо значення величин $L_1\left(\mathbf{e}\right)$ і $\delta_{\mathbf{e}}$ та знайдемо такий відкритий окіл $U_{\mathbf{e}}$, що

$$\text{для } \forall \lambda \ge L_1\left(\mathbf{e}\right) \text{ і } \forall \mathbf{x} \in U_{\mathbf{e}}, \forall \mathbf{b} \in B \setminus H_{\mathbf{e}} \; \left|\left(\lambda \mathbf{x}, \mathbf{b}\right)\right| \le 2\varepsilon \Rightarrow \left(\lambda \mathbf{x}, \mathbf{b}\right) < 0 . \quad (8)$$

Також будемо вимагати, щоб для всіх $\mathbf{x} \in U_{\mathbf{e}} \cap B\left[\mathbf{0}, 1\right]$ $\left(\mathbf{x}, \mathbf{e}\right) > 1/2$. Крім того, знайдеться таке $\tau > 0$, що для всіх $\mathbf{b} \in B \setminus H_{\mathbf{e}}$ $\tau_{\mathbf{b}} \le -\tau d_2$, де $\mathbf{b} = \tau_{\mathbf{b}} \mathbf{e} + \tilde{\mathbf{b}}, \tilde{\mathbf{b}} \in H_{\mathbf{e}}$.

Розглянемо довільний правильний $\left(B, \beta, \varepsilon\right)$-ланцюг $\mathbf{w}^0, \mathbf{w}^1, \ldots, \mathbf{w}^k$ і припустимо, що вектор $\left\|\mathbf{w}^0\right\|^{-1} \mathbf{w}^0$ не тільки «достатньо близький» до вектора e, як це було при доведенні леми 4, але й «достатньо віддалений» від нього. Ми будемо вимагати, щоб $\left\|\mathbf{w}^0\right\|^{-1} \mathbf{w}^0 \in V_{\mathbf{e}} = B\left(\mathbf{x}, r_{\mathbf{e}}\right) \setminus B\left(\mathbf{x}, \frac{1}{2} r_{\mathbf{e}}\right)$. Виберемо число δ таким чином, щоб $\delta = \dfrac{\tau\varepsilon}{16 d_2}$ у випадку $\varepsilon > 0$ і $\delta = \dfrac{1}{d_2^2}$ у іншому випадку. Нехай $\Lambda = \left\{0, \varepsilon, \dfrac{\varepsilon}{2}, \dfrac{1}{d_2^2}\right\}$. Позначимо $L\left(\mathbf{e}\right) = \max\left\{L_1\left(\mathbf{e}\right), t L_2\left(\mathbf{e}\right)\right\} + \dfrac{8\delta}{\tau} + \beta_{\max} d_2$, де

$$L_2(\mathbf{e}) = \frac{1}{\delta}\max_{\eta\in\Lambda}\left\{\exp\left\{\left(\frac{8\delta}{\tau}+\beta_{\max}\right)\right\}N\left(m, H_{\mathbf{e}}\cap B, \beta_{\max}, \eta\right)+\frac{8\delta d_2}{\tau}+\beta_{\max}\right\}^2,$$

а множник $t \geq 1$ вибирається таким чином, щоб для довільного вектора $\mathbf{b}\in H_{\mathbf{e}}$, довжина якого не більша за $\exp\left\{\left(\frac{8\delta}{\tau}+\beta_{\max}\right)\right\}N\left(m, H_{\mathbf{e}}\cap B, \beta_{\max}, \eta\right)+\frac{8\delta d_2}{\tau}+\beta_{\max}$ кінець орт-вектора, відповідного вектору $\mathbf{w}^0 - \left(\frac{8\delta}{\tau}+\beta_{\max}d_2\right)\mathbf{e}+\mathbf{b}$ потрапив у $U_{\mathbf{e}}$ (існування такого множника випливає з наслідку до леми 2). Припустимо, що $\left\|\mathbf{w}^0\right\| > L(\mathbf{e})$ і $\left\|\tilde{\mathbf{w}}^0\right\| > d_2$, де $\tilde{\mathbf{w}}^0$ — проекція вектора $\mathbf{w}^0$ на підпростір $H_{\mathbf{e}}$.

Згідно до (8) і леми 3 (B,β,ε)-ланцюг $\mathbf{w}^0,\mathbf{w}^1,\ldots,\mathbf{w}^k$ не може бути правильним, якщо $\mathbf{b}^1 \notin H_{\mathbf{e}}$.

Припустимо, що $\mathbf{b}^1,\ldots,\mathbf{b}^{s_1}\in H_{\mathbf{e}}$ і вектор $\mathbf{b}^{s_1+1}$ не належить підпростору $H_{\mathbf{e}}$. Тоді $\mathbf{b}^{s_1+1}=\tau_1\mathbf{e}+\tilde{\mathbf{b}}^{s_1}$, де $\tilde{\mathbf{b}}^{s_1}$ — проекція $\mathbf{b}^{s_1+1}$ на $H_{\mathbf{e}}$ і $\tau_1 \leq -\tau d_2$. Нехай $\tilde{\mathbf{c}}^{s_1}=\tilde{\mathbf{w}}^0+\beta_1\mathbf{b}^1+\ldots+\beta_{s_1}\mathbf{b}^{s_1}$, $\tilde{\mathbf{b}}^{s_1}=\mu_1\tilde{\mathbf{c}}^{s_1}+\mathbf{c}^{s_1+1}, |\mu_1|<1$, де $\mathbf{c}^{s_1+1}\perp\tilde{\mathbf{c}}^{s_1}$. Тоді вектори $\left(1+\mu_1\beta_{s_1+1}\right)\tilde{\mathbf{w}}^0$, $\left(1+\mu_1\beta_{s_1+1}\right)\tilde{\mathbf{w}}^1,\ldots,\left(1+\mu_1\beta_{s_1+1}\right)\tilde{\mathbf{w}}^{s_1}$ за умови $\beta'_j=\beta_j\times\left(1+\mu_1\beta_{s_1+1}\right)$, $j=0,1,\ldots s_1$ утворюють правильний (B,β,ε)-ланцюг.

Розглянемо вектор $\mathbf{b}^{s_1+2}$. Якщо $\mathbf{b}^{s_1+2}\in H_{\mathbf{e}}$, то нерівність $\left(\mathbf{w}^{s_1+1},\mathbf{b}^{s_1+2}\right)\leq 2\varepsilon$ в силу (8) імплікує $\left(1+\mu_1\beta_{s_1+1}\right)\left(\tilde{\mathbf{w}}^{s_1},\mathbf{b}^{s_1+2}\right)+\beta_{s_1+1}\left(\mathbf{c}^{s_1+1},\mathbf{b}^{s_1+2}\right)<0$. Звідси у свою чергу випливає, що $\left(1+\mu_1\beta_{s_1+1}\right)\left(\tilde{\mathbf{w}}^{s_1},\mathbf{b}^{s_1+2}\right)\leq\eta$ при умові $\beta_{s_1+1}d_2^2\leq\eta$, де $\eta=\varepsilon/2$ якщо $\varepsilon>0$ і $\eta=1/d_2^2$ у протилежному випадку. Тому у цьому випадку вектори $\left(1+\mu_1\beta_{s_1+1}\right)\tilde{\mathbf{w}}^0$, $\left(1+\mu_1\beta_{s_1+1}\right)\tilde{\mathbf{w}}^1,\ldots,\left(1+\mu_1\beta_{s_1+1}\right)\tilde{\mathbf{w}}^{s_1}$ і $\tilde{\mathbf{w}}^{s_1+1}+\beta_{s_1+2}\mathbf{b}^{s_1+2}$ утворюють правильний (B,β',η)-ланцюг і

$$\beta_1\mathbf{b}^1+\ldots\beta_{s_1}\mathbf{b}^{s_1}+$$
$$+\beta_{s_1+1}\mathbf{b}^{s_1+1}+\beta_{s_1+2}\mathbf{b}^{s_1+2}=\beta'_1\mathbf{b}^1+\ldots+\beta'_{s_1}\mathbf{b}^{s_1}+\beta_{s_1+1}\mathbf{c}^{s_1+1}+\beta_{s_1+2}\mathbf{b}^{s_1+2}+\beta_{s_1+1}\tau_1\mathbf{e}.$$

 ISSN 1562-9945

Якщо $\mathbf{b}^{s_1+2} \notin H_{\mathbf{e}}$, то застосувавши міркування аналогічні до наведених у попередньому абзаці, отримаємо правильний (B, β', ε)-ланцюг $(1+\mu_1\beta_{s_1+1})\times \times(1+\mu_2\beta_{s_2+1})\tilde{\mathbf{w}}^1,\ldots,(1+\mu_1\beta_{s_1+1})(1+\mu_2\beta_{s_2+1})\tilde{\mathbf{w}}^{s_1}$, де $|\mu_2|<1$, $s_2 = s_1+1$, $\beta'_j = \beta_j(1+\mu_1\beta_{s_1+1})(1+\mu_2\beta_{s_2+1})$ і

$\beta_1\mathbf{b}^1 + \ldots + \beta_{s_1+1}\mathbf{b}^{s_1+1} + \beta_{s_1+2}\mathbf{b}^{s_1+2} = \beta'_1\mathbf{b}^1 + \ldots +$

$+\beta'_{s_1}\mathbf{b}^{s_1} + \beta_{s_1+1}\mathbf{c}^{s_1+1} + \beta_{s_2+1}\mathbf{c}^{s_2+1} + (\beta_{s_1+1}\tau_1 + \beta_{s_2+1}\tau_2)\mathbf{e}$.

Припустимо, що

$$\beta_{s_1+1} + \ldots + \beta_{s_k+1} < \frac{8\delta}{\tau}. \tag{9}$$

Повторне застосування вищезгаданих міркувань дозволяє отримати (B, β', η)-ланцюг

$$\zeta_k\tilde{\mathbf{w}}^0, \zeta_k\tilde{\mathbf{w}}^1,\ldots,\zeta_k\tilde{\mathbf{w}}^j,\ldots,\zeta_k\tilde{\mathbf{w}}^{s_k},\; j\in J_k, \tag{10}$$

$\zeta_k = (1+\mu_1\beta_{s_1+1})\ldots(1+\mu_k\beta_{s_k+1})$, $\eta=\varepsilon$ або $\eta=\varepsilon/2$ або $\eta = 1/d_2^2$ (у випадку нульового допуску), $\beta'_j \in \{\beta_j\zeta_j, \beta_j\}$, $j\in J_k$, $J_k = \{1,\ldots,s_k\} \setminus \{s_1+1,\ldots,s_{k-1}+1\}$.

Для того, щоб переконатися, що послідовність (10) є правильним (B,β',η)-ланцюгом досить перевірити, що з (9) випливає, що $\sum_{j=1}^{k-1}\beta_{s_j+1}(\mathbf{c}^{s_j+1}, \mathbf{b}^{s_j+2}) \le \eta$, а тому $\zeta_j(\tilde{\mathbf{w}}^{s_j}, \mathbf{b}^{s_1+2}) \le \eta$, $j=1,\ldots,k-1$.

Крім того

$$\beta_1\mathbf{b}^1 + \ldots + \beta_{s_1}\mathbf{b}^{s_1} + \ldots + \beta_{s_k+1}\mathbf{b}^{s_k+1} = \Sigma_1 + \Sigma_2 + \Sigma_3,$$

де

$$\Sigma_1 = \sum_{j\in J_k}\beta'_j\mathbf{b}^j, \quad \beta'_j \in \{\beta_j\zeta_j, \beta_j\},\; \Sigma_2 = \sum_{j=1}^{k}\beta_{s_j+1}\mathbf{c}^{s_j+1},\; \Sigma_3 = \sum_{j=1}^{k}\tau_j\beta_{s_j+1}\mathbf{e}.$$

Тоді згідно до леми 2

$$\|\Sigma_1 + \Sigma_2 + \Sigma_3\| \le \|\Sigma_1 + \Sigma_2\| \le \zeta_k \max_{\eta\in\Lambda} N(m, H_{\mathbf{e}} \cap B, \beta_{\max}, \eta) + \frac{8\delta}{\tau}.$$

Для оцінки величини ζ_k скористаємося відомою нерівністю $1 + x < e^x \ (x > 0)$. Тоді $\zeta_k < \left(1 + \beta_{s_1+1}\right) \ldots \left(1 + \beta_{s_k+1}\right) < \exp\left\{\beta_{s_1+1} + \ldots + \beta_{s_k+1}\right\}$. Остаточно

$$\left\|\Sigma_1 + \Sigma_2 + \Sigma_3\right\| \le \exp\left\{\frac{8\delta}{\tau}\right\} \max_{\eta\in\Lambda} N\left(m, H_{\mathbf{e}} \cap B, \beta_{\max}, \eta\right) + \frac{8\delta}{\tau}. \qquad \textbf{(11)}$$

Припустимо тепер, що $\beta_{s_1+1} + \ldots + \beta_{s_{k-1}+1} < \dfrac{8\delta}{\tau}$ і $\beta_{s_1+1} + \ldots + \beta_{s_k+1} \ge \dfrac{8\delta}{\tau}$.

Поклавши у лемі 2 $\mathbf{a} = \mathbf{w}^0 + \Sigma_3$, $\mathbf{b} = \Sigma_1 + \Sigma_2$, з урахуванням правила вибору $tL_2(\mathbf{e})$, отримаємо, що

$$\left\|\mathbf{w}^{s_k+1}\right\| \le \left\|\mathbf{w}^0 + \Sigma_3\right\| + \delta \text{ і } \mathbf{w}^0 + \Sigma_3 \in U_{\mathbf{e}}.$$

Оскільки $\left\|\mathbf{w}^0\right\| \ge \dfrac{8\delta}{\tau}$ і косинус кута між $\dfrac{\mathbf{w}^0}{\left\|\mathbf{w}^0\right\|}$ і $\dfrac{\Sigma_3}{\left\|\Sigma_3\right\|}$ менший за $-\frac{1}{2}$, то з урахуванням леми 3 і попередньої нерівності отримуємо

$$\left\|\mathbf{w}^{s_k+1}\right\| \le \left\|\mathbf{w}^0 + \Sigma_3\right\| + \delta \le \left\|\mathbf{w}^0\right\| + \delta - 2\delta = \left\|\mathbf{w}^0\right\| - \delta.$$

Таким чином (B, β, ε)-ланцюг не може бути правильним, якщо не виконується нерівність (9). Отже, для довільного правильного (B, β, ε)-ланцюга $\mathbf{w}^0, \mathbf{w}^1 \ldots, \mathbf{w}^s$, початок якого задовольняє умову $\left\|\mathbf{w}^0\right\|^{-1} \mathbf{w}^0 \in V_{\mathbf{e}}$ і $\left\|\mathbf{w}^0\right\| > L(\mathbf{e})$, величина $\left\|\beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s\right\|$ задовольняє **(11)**. Множини $V_{\mathbf{e}}$ утворюють відкрите покриття одиничної сфери. Це покриття за лемою Бореля містить скінченне підпокриття $V_{\mathbf{e}_1}, \ldots, V_{\mathbf{e}_p}$. Тому можна для довільного правильного (B, β, ε)-ланцюга $\mathbf{w}^0, \mathbf{w}^1 \ldots, \mathbf{w}^s$ $\left\|\beta_1 \mathbf{b}^1 + \ldots + \beta_s \mathbf{b}^s\right\| < N(m+1, B, \beta_{\max}, \varepsilon)$, де

$$N(m+1, B, \beta_{\max}, \varepsilon) = \max\left\{L_1(\mathbf{e}), \ldots, L_p(\mathbf{e})\right\} + \\ + \exp\left\{\frac{8\delta}{\tau}\right\} \cdot \max_{\eta\in\Lambda} N\left(m, H_{\mathbf{e}} \cap B, \beta_{\max}, \eta\right) + \frac{8\delta}{\tau}.$$

Лема доведена.

Тепер ми вже можемо довести теорему про обмеженість вагових векторів ПНЕ для пакетного алгоритму навчання з допуском.

 ISSN 1562-9945

Теорема 2. Якщо множини A^+ і A^- є скінченними, послідовність вагових векторів $\left\{\mathbf{w}^k\right\}$ будується згідно до (2), то для довільних допуску $\varepsilon \geq 0$, початкового наближення $\mathbf{w}^0$, навчаючої послідовності $\left\{\mathbf{a}^k\right\}$ і послідовності $\left\{\beta_k\right\}$, послідовність $\left\|\mathbf{w}^k\right\|$ є обмеженою за умови, що коефіцієнти $\left\{\beta_k\right\}$ задовольняють умову (4).

Доведення. Не втрачаючи загальності, будемо вважати, що у процесі навчання $\Delta\mathbf{w}^k \neq \mathbf{0}$. Крім того, можна вважати, що $A^- = \varnothing$, оскільки у протилежному випадку усі вектори $P\left(\mathbf{a}^k\right)$, де $\mathbf{a}^k \in A^-$ можна замінити на вектори $-P\left(\mathbf{a}^k\right)$. Ця замінна є допустимою, оскільки навчання ПНЕ по розпізнаванню підмножин множини $A \subset \mathrm{R}^n$ можна розглядати, як навчання НЕ по розпізнаванню підмножин множини $P(A) \subset \mathrm{R}^n$, де $P(A) = \left\{P(\mathbf{a}) \mid \mathbf{a} \in A\right\}$. Тому для всіх $\mathrm{k} > 0$ $\gamma_{kj} \in \left\{0,1\right\}$, $j = \overline{1,l}$. Покажемо тепер, що доведення теореми досить провести для он-лайн алгоритму навчання звичайного НЕ. Для цього розглянемо множину

$$B = \left\{\alpha_1 P\left(\mathbf{a}^{j_1}\right) + \ldots + \alpha_l P\left(\mathbf{a}^{j_l}\right) \middle| \mathbf{a}^{j_i} \in A,\ \alpha_i \in \mathrm{Z}_2,\ i = 1,\ldots,l\right\}.$$

Легко переконатися, що множина B скінченна, оскільки $\mathrm{Card}\, B \leq \sum_{i=0}^{l} \overline{C}_t^i = \sum_{i=0}^{l} C_{t+i-1}^i = C_{t+l}^t$, де $t = \mathrm{Card}\, A$. Для кожної навчаючої послідовності $\left\{\mathbf{a}^k\right\}$ ПНЕ можна побудувати навчаючу послідовність $\left\{\mathbf{b}^k\right\}$ НЕ, вектори приростів яких співпадають. Тим самим навчання ПНЕ по розпізнаванню підмножин множини A зводиться до навчання звичайного НЕ по розпізнаванню підмножин множини B, для якого $\Delta\mathbf{w}^k = \mathbf{b}^k \neq \mathbf{0}$.

Висновки

У роботі було показано, виконання умови (4) забезпечує обмеженість вагових векторів, які отримуються згідно до алгоритму навчання перцептрона. Цей результат може бути використаний при побудові pocket-алгоритмів навчання ПНЕ.

ЛІТЕРАТУРА

1. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – 2-е изд. – М. : Вильямс-Телеком, 2006. – 1104 с.
2. Розенблатт, Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. – М. : Мир, 1965. – 480 с.
3. Гече, Ф. Е. Алгоритми навчання узагальнених нейронних елементів відносно системи характерів / Ф. Е. Гече, В. М. Коцовський, А. Є. Батюк // Збірник наукових праць інституту проблем моделювання в енергетиці НАН України. -- К., 2007.~-- Вип. 41. -- С. 124-136.
4. Дуда, Р. Распознавание образов и анализ сцен /Р. Дуда, П. Харт. – М. : Мир, 1976. – 507 с.
5. Минский, М. Р. Персептроны /М. Минский, С. Пайперт. – М. : Мир, 1971. – 261 с.

 ISSN 1562-9945