New Criteria for the Linear Binary Separability in the Euclidean Normed Space

Yeong-Jeu Sun*

Department of Electrical Engineering, I-Shou University, Kaohsiung 840, Taiwan, Republic of China

Abstract: In this paper, the classical binary classification problem is investigated. Necessary and sufficient criterion is presented to guarantee the linear binary separability of the training data in the Euclidean normed space. A suitable hyperplane that correctly classifies the training data is also constructed provided that the necessary and sufficient is satisfied. Based on the main result, we offer an easy-to-check criterion for the linear binary separability of the training set. Finally, a numerical example is provided to illustrate the feasibility and effectiveness of the obtained result.

Keywords: Binary classification, necessary and sufficient criterion, pattern recognition, theory of learning.

INTRODUCTION

For convenience, we define $\underline{p} := \{1, 2, \cdots, p\}$ and $A - B := \{x \mid x \in A \text{ and } x \notin B\}$ that will be used throughout this paper. Let $X \subseteq \Re^n$ and $Y := \{1, -1\}$. Suppose we are given the training set

$$S := \{(\mathbf{z}_i, y_i)\}_{i=1}^{l} \subseteq X \times Y .$$

If the primal form (or dual form) of Rosenblatt's perceptron algorithm converges in a finite number of iterations, then the training set is linearly separable [1]. Furthermore, in this case, the modified Rosenblatt's perceptron algorithm can be used to find a separating hyperplane [2]. However, if we find that the primal form (or dual form) of Rosenblatt's perceptron algorithm does not converge in five days, we know nothing about the linear separability or linear inseparability of the training set. It is clear that the given training set may not be linearly separable. In this case, it is common to transform the training data *via* a nonlinear map, called the feature map, into another space, called the feature space, so that the transformed training data may possibly be separable. If the training set is linearly separable and the modified Rosenblatt's perceptron algorithm does not converge in five days (or even in one month), we ought to find other criteria to guarantee the linear separability of the training set. It is the purpose of this paper to investigate other criteria for the linear separability of the training data in the Euclidean normed space.

We wish to point out that we will use in this paper only the Euclidean space $\Re^n$, equipped with the usual inner product $\langle \mathbf{x}, \mathbf{z} \rangle := \mathbf{x}^T \mathbf{z}$, $\mathbf{x}, \mathbf{z} \in \Re^n$, and normed product $\|\mathbf{z}\| := \sqrt{\mathbf{z}^T \mathbf{z}}$, $\mathbf{z} \in \Re^n$, with the metric space $d(\mathbf{x}, \mathbf{z}) := \|\mathbf{x} - \mathbf{z}\|$, $\mathbf{x}, \mathbf{z} \in \Re^n$, since it is the simplest to work with and if the original input space X is finite-dimensional, it is isomorphic to some Euclidean space $\Re^n$.

*Address correspondence to this author at the Department of Electrical Engineering, I-Shou University, Kaohsiung, Taiwan 840, Republic of China; Tel: 886-7-6577711, Ext. 6626; Fax: 886-7-6577205; E-mail: yjsun@isu.edu.tw

MAIN RESULTS

For the simplicity of notation, suppose the training set is given by $S = S_1 \cup S_2$, where

$$S_1 := \{(\mathbf{z}_i, y_i)\}_{i=1}^{l_1} \neq \varnothing,\ y_i = 1$$

for all $i = 1, 2, ..., l_1$, (1a)

$$S_2 = \{(\mathbf{z}_i, y_i)\}_{i=l_1+1}^{l} \neq \varnothing,\ y_i = -1$$

for all $i = l_1 + 1, l_1 + 2, ..., l$, (1b)

$\mathbf{z}_i \neq \mathbf{z}_j$ for all $i \neq j$,

$$\{\mathbf{z}_{1,j}\}_{j=1}^{l_1} := \{\mathbf{z}_i\}_{i=1}^{l_1}, \tag{1c}$$

$$\{\mathbf{z}_{2,j}\}_{j=1}^{l-l_1} := \{\mathbf{z}_i\}_{i=l_1+1}^{l}. \tag{1d}$$

Define

$$l_2 := l - l_1,\ \overline{\mathbf{z}}_1 := \frac{1}{l_1}\sum_{i=1}^{l_1} \mathbf{z}_i,\ \overline{\mathbf{z}}_2 := \frac{1}{l_2}\sum_{i=l_1+1}^{l} \mathbf{z}_i, \tag{2a}$$

$$r_1 := \max_{i \in \underline{l_1}} \|\overline{\mathbf{z}}_1 - \mathbf{z}_{i,1}\|,\ r_2 := \max_{i \in \underline{l_2}} \|\overline{\mathbf{z}}_2 - \mathbf{z}_{i,2}\|. \tag{2b}$$

Obviously, $\overline{\mathbf{z}}_1$ is the mean of all elements of S_1 and $\overline{\mathbf{z}}_2$ is the mean of all elements of S_2. In addition, r_1 is the farthest distance between the point $\overline{\mathbf{z}}_1$ and the elements of S_1 and r_2 is the farthest distance between the point $\overline{\mathbf{z}}_2$ and the elements of S_2.

Definition 2.1 [3]

The training set S is said to be linearly separable if there is a hyperplane that correctly classifies the training data, i.e., there exist $\mathbf{w} \in \Re^n$ and $b \in \Re$ such that

$y_i \cdot \left[\langle \mathbf{w}, \mathbf{z}_i \rangle + b \right] > 0$ for all $i \in \underline{l}$.

Now we present the modified Rosenblatt's algorithm as follows.

Lemma 2.1 [4] Modified Rosenblatt's Algorithm

Data:the training set $S := \left\{ \left(\mathbf{z}_i, y_i \right) \right\}_{i=1}^{l} \subseteq X \times Y$ and a learning rate $\eta > 0$.

Goal: a hyperplane $(\mathbf{w}, b)$ that correctly classifies the training set.

Step 1: $\mathbf{w}_0 \leftarrow 0$; $b_0 \leftarrow 0$; $k \leftarrow 0$;

$I_S^+ := \left\{ i \in \underline{l} : y_i = 1 \right\}$;

$I_S^- := \left\{ j \in \underline{l} : y_j = -1 \right\}$;

Step 2: Choose $Q \ge V := \max\left(\min_{i \in I_S^+} \|x_i\|, \min_{j \in I_S^-} \|x_j\| \right)$;

Step 3: repeat

for $i = 1$ to l

if $y_i \cdot \left[\langle \mathbf{w}_k, \mathbf{z}_i \rangle + b_k \right] \le 0$, then

$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + \eta y_i \mathbf{z}_i$;

$b_{k+1} \leftarrow b_k + \eta y_i Q^2$;

$k \leftarrow k + 1$;

end if

end for

until no misclassification within the *for* loop

return *k*, $(\mathbf{w}_k, b_k)$ where *k* is the number of mistakes

Before presenting the main result, we introduce the necessary and sufficient criterion for the linear binary separability of the training set *S*.

(A1) There exist $\mathbf{O}_1 \in \Re^n$, $\mathbf{O}_2 \in \Re^n$, and two non-negative numbers r_1 and r_2 such that the following conditions are satisfied.

(i) $\left\| \mathbf{O}_1 - \mathbf{z}_{i,1} \right\| \le r_1, \quad \forall\, i \in \underline{l_1}$;

(ii) $\left\| \mathbf{O}_2 - \mathbf{z}_{i,2} \right\| \le r_2, \quad \forall\, i \in \underline{l - l_1}$;

(iii) $r_1 + r_2 < \left\| \mathbf{O}_1 - \mathbf{O}_2 \right\|$.

Lemma 2.2 If the training set of *S* is linearly separable, then (A1) is satisfied.

Proof: Without loss of generality, we assume that $\langle \mathbf{w}, \mathbf{z}_{i,1} \rangle + b > 0$, $\forall\, i \in \underline{l_1}$, and $\langle \mathbf{w}, \mathbf{z}_{i,2} \rangle + b < 0$, $\forall\, i \in \underline{l - l_1}$. Define $P := \left\{ x \mid \langle \mathbf{w}, \mathbf{x} \rangle + b = 0 \right\}$. Apparently, there exists a $\mathbf{z}_0 \in P$ such that $\mathbf{z}_i \notin L := \left\{ t\mathbf{w} + \mathbf{z}_0 \mid t \in \Re \right\}$, $\forall\, i \in \underline{l}$. Let us define

$$a_1 := \max_{i \in \underline{l_1}} d\left(\mathbf{z}_{i,1}, L \right), \; a_2 := \max_{i \in \underline{l - l_1}} d\left(\mathbf{z}_{i,2}, L \right),$$

$$c_1 := \frac{1}{2} \min_{i \in \underline{l_1}} d\left(\mathbf{z}_{i,1}, P \right), \; c_2 := \frac{1}{2} \min_{i \in \underline{l - l_1}} d\left(\mathbf{z}_{i,2}, P \right),$$

$$b_1 := \frac{a_1}{2c_1}, \qquad b_2 := \frac{a_2}{2c_2},$$

$$\mathbf{O}_1 := \frac{b_1 + 2c_1}{\|\mathbf{w}\|} \mathbf{w} + \mathbf{z}_0, \qquad \mathbf{O}_2 := -\frac{b_2 + 2c_2}{\|\mathbf{w}\|} \mathbf{w} + \mathbf{z}_0,$$

$$r_1 := b_1 + c_1, \; r_2 := b_2 + c_2,$$

$$\mathbf{q}_{i,1} := \left(\frac{b + \langle \mathbf{z}_{i,1}, \mathbf{w} \rangle}{\|\mathbf{w}\|^2} \right) \mathbf{w} + \mathbf{z}_0, \; \mathbf{q}_{i,2} := \left(\frac{b + \langle \mathbf{z}_{i,2}, \mathbf{w} \rangle}{\|\mathbf{w}\|^2} \right) \mathbf{w} + \mathbf{z}_0.$$

Thus, one has $\mathbf{q}_{i,1} \in L$, $\mathbf{q}_{i,2} \in L$,

$$d\left(\mathbf{z}_{i,1}, P \right) = d\left(\mathbf{q}_{i,1}, L \right) = \frac{\langle \mathbf{w}, \mathbf{z}_{i,1} \rangle + b}{\|\mathbf{w}\|},$$

$$d\left(\mathbf{z}_{i,2}, P \right) = d\left(\mathbf{q}_{i,2}, L \right) = \frac{\langle \mathbf{w}, \mathbf{z}_{i,2} \rangle + b}{\|\mathbf{w}\|},$$

$\langle \mathbf{z}_{i,1} - \mathbf{q}_{i,1}, r\mathbf{w} \rangle = 0, \forall\, r \in \Re$, and $\langle \mathbf{z}_{i,2} - \mathbf{q}_{i,2}, r\mathbf{w} \rangle = 0, \forall\, r \in \Re$. It can be readily obtained that

$$\begin{aligned}
&\left\| \mathbf{O}_1 - \mathbf{z}_{i,1} \right\|^2 \\
&= \left\| \left(\mathbf{O}_1 - \mathbf{q}_{i,1} \right) + \left(\mathbf{q}_{i,1} - \mathbf{z}_{i,1} \right) \right\|^2 \\
&\le \left\| \mathbf{O}_1 - \mathbf{q}_{i,1} \right\|^2 + \left\| \mathbf{q}_{i,1} - \mathbf{z}_{i,1} \right\|^2 \\
&= \left[\left\| \mathbf{O}_1 - \mathbf{z}_0 \right\| - \left\| \mathbf{z}_0 - \mathbf{q}_{i,1} \right\| \right]^2 + \left\| \mathbf{q}_{i,1} - \mathbf{z}_{i,1} \right\|^2 \\
&\le \left[\left(b_1 + 2c_1 \right) - 2c_1 \right]^2 + a_1^2 \\
&= b_1^2 + a_1^2 \\
&= b_1^2 + 2b_1 c_1 \\
&\le \left(b_1 + c_1 \right)^2 \\
&= r_1^2, \quad \forall\, i \in \underline{l_1}, \qquad (3)
\end{aligned}$$

which implies that $\left\| \mathbf{O}_1 - \mathbf{z}_{i,1} \right\| \le r_1, \quad \forall\, i \in \underline{l_1}$. Similarly, for every $i = l_1 + 1, l_1 + 2, \cdots, l$, it is easy to see that

$$\begin{aligned}
&\left\| \mathbf{O}_2 - \mathbf{z}_{i,2} \right\|^2 \\
&= \left\| \left(\mathbf{O}_2 - \mathbf{q}_{i,2} \right) + \left(\mathbf{q}_{i,2} - \mathbf{z}_{i,2} \right) \right\|^2 \\
&\le \left\| \mathbf{O}_2 - \mathbf{q}_{i,2} \right\|^2 + \left\| \mathbf{q}_{i,2} - \mathbf{z}_{i,2} \right\|^2 \\
&= \left[\left\| \mathbf{O}_2 - \mathbf{z}_0 \right\| - \left\| \mathbf{z}_0 - \mathbf{q}_{i,2} \right\| \right]^2 + \left\| \mathbf{q}_{i,2} - \mathbf{z}_{i,2} \right\|^2 \\
&\le \left[\left(b_2 + 2c_2 \right) - 2c_2 \right]^2 + a_2^2
\end{aligned}$$

$$= b_2^2 + a_2^2$$
$$= b_2^2 + 2b_2c_2$$
$$\le (b_2 + c_2)^2$$
$$= r_2^2, \quad (4)$$

which implies that

$$\|\mathbf{O}_2 - \mathbf{z}_{i,2}\| \le r_2, \quad \forall\, i \in l_1 + 1, l_1 + 2, \cdots, l\,.$$

In addition, one has

$$\|\mathbf{O}_1 - \mathbf{O}_2\| = b_1 + 2c_1 + b_2 + 2c_2$$
$$> b_1 + c_1 + b_2 + c_2 = r_1 + r_2. \quad (5)$$

This completes the proof in view of (2)-(5). Δ

Lemma 2.3 The training set of S is linearly separable provided that (A1) is satisfied. In this case, a suitable hyperplane that correctly classifies the training set of (1) is given by $[\langle \mathbf{w}, \mathbf{x}\rangle + b] = 0$, where $\mathbf{w} := \mathbf{O}_1 - \mathbf{O}_2$ and $b := \langle \mathbf{w}, -\alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$, with

$$\frac{r_1}{\|\mathbf{O}_1 - \mathbf{O}_2\|} < \alpha < \frac{\|\mathbf{O}_1 - \mathbf{O}_2\| - r_2}{\|\mathbf{O}_1 - \mathbf{O}_2\|}.$$

Proof: For every $i \in \underline{l_1}$, one has

$$y_i\left[\langle \mathbf{w}, \mathbf{z}_i\rangle + b\right]$$
$$= \langle \mathbf{w}, \mathbf{z}_i\rangle + b$$
$$= \langle \mathbf{w}, \mathbf{z}_i\rangle + \langle \mathbf{w}, -\alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$$
$$= \langle \mathbf{w}, \mathbf{z}_i - \alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$$
$$= \langle \mathbf{w}, \mathbf{O}_1 + (\mathbf{z}_i - \mathbf{O}_1) - \alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$$
$$= \langle \mathbf{w}, \alpha(\mathbf{O}_1 - \mathbf{O}_2) + (\mathbf{z}_i - \mathbf{O}_1)\rangle$$
$$= \langle \mathbf{w}, \alpha(\mathbf{O}_1 - \mathbf{O}_2) + (\mathbf{z}_i - \mathbf{O}_1)\rangle$$
$$= \langle \mathbf{w}, \alpha\mathbf{w} + (\mathbf{z}_i - \mathbf{O}_1)\rangle$$
$$= \alpha\langle \mathbf{w}, \mathbf{w}\rangle + \langle \mathbf{w}, (\mathbf{z}_i - \mathbf{O}_1)\rangle$$
$$\ge \alpha\langle \mathbf{w}, \mathbf{w}\rangle - \|\mathbf{w}\| \cdot \|\mathbf{z}_i - \mathbf{O}_1\|$$
$$\ge \alpha\|\mathbf{O}_1 - \mathbf{O}_2\|^2 - \|\mathbf{O}_1 - \mathbf{O}_2\| \cdot r_1$$
$$= \|\mathbf{O}_1 - \mathbf{O}_2\|^2\left[\alpha - \frac{r_1}{\|\mathbf{O}_1 - \mathbf{O}_2\|}\right] > 0. \quad (6)$$

For every $i \in \underline{l} - \underline{l_1}$, one has

$$y_i\left[\langle \mathbf{w}, \mathbf{z}_i\rangle + b\right]$$
$$= -\langle \mathbf{w}, \mathbf{z}_i\rangle - b$$
$$= -\langle \mathbf{w}, \mathbf{z}_i\rangle - \langle \mathbf{w}, -\alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$$
$$= -\langle \mathbf{w}, \mathbf{z}_i - \alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$$
$$= -\langle \mathbf{w}, \mathbf{O}_2 + (\mathbf{z}_i - \mathbf{O}_2) - \alpha\mathbf{O}_2 + (\alpha-1)\mathbf{O}_1\rangle$$
$$= -\langle \mathbf{w}, (\alpha-1)(\mathbf{O}_1 - \mathbf{O}_2) + (\mathbf{z}_i - \mathbf{O}_2)\rangle$$
$$= -\langle \mathbf{w}, (\alpha-1)\mathbf{w} + (\mathbf{z}_i - \mathbf{O}_2)\rangle$$
$$= (1-\alpha)\langle \mathbf{w}, \mathbf{w}\rangle - \langle \mathbf{w}, (\mathbf{z}_i - \mathbf{O}_2)\rangle$$
$$\ge (1-\alpha)\langle \mathbf{w}, \mathbf{w}\rangle - \|\mathbf{w}\| \cdot \|\mathbf{z}_i - \mathbf{O}_2\|$$
$$\ge (1-\alpha)\|\mathbf{O}_1 - \mathbf{O}_2\|^2 - \|\mathbf{O}_1 - \mathbf{O}_2\| \cdot r_2$$
$$= \|\mathbf{O}_1 - \mathbf{O}_2\|^2\left[1 - \alpha - \frac{r_2}{\|\mathbf{O}_1 - \mathbf{O}_2\|}\right] > 0. \quad (7)$$

This completes the proof in view of (6) and (7). Δ

Based on Lemma 2.2 and Lemma 2.3, the necessary and sufficient criterion for the linear binary separability of the training set S is stated as follows.

Theorem 2.1 The training set of (1) is linearly separable if and only if (A1) is satisfied.

Thus, based on Theorem 2.1 with Lemma 2.3, we present an easy-to-check criterion for the linear binary separability of the training set of (1).

Corollary 2.1 The training set of (1) is linearly separable provided that $r_1 + r_2 < \|\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2\|$. In this case, a suitable hyperplane that correctly classifies the training set of (1) is given by $[\langle \mathbf{w}, \mathbf{x}\rangle + b] = 0$, where $\mathbf{w} := \bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2$ and $b := \langle \mathbf{w}, -\alpha\bar{\mathbf{z}}_2 + (\alpha-1)\bar{\mathbf{z}}_1\rangle$, with

$$\frac{r_1}{\|\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2\|} < \alpha < \frac{\|\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2\| - r_2}{\|\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2\|}.$$

Proof: This proof can be immediately obtained by Lemma 2.3 and Theorem 2.1 with the choice of $\mathbf{O}_1 := \bar{\mathbf{z}}_1$, $\mathbf{O}_2 := \bar{\mathbf{z}}_2$, $r_1 := d_1$, and $r_1 := d_2$. Δ

Remark 2.2 Support vector learning is one of the most exciting tools for machine learning, data mining, hand-written character recognition, image classification, biosequence analysis, etc; see, for example [1, 5-16], and the references therein. We wish to point out that Lemma 2.3 (or Corollary 2.1) provides only a suitable hyperplane that correctly classifies the training set of (1). Such a hyperplane may not be “optimal” from the computational point of view. In principle, the powerful support vector learning algorithm can be employed to find the maximal margin hyperplane (or called the optimal hyperplane) in the Euclidean normed space of $\Re^n$ for the sake of generalization performance [3]. Suppose the training set of S is linearly separable (or equivalently the condition of (A1) is satisfied). Using the support vector learning method of [3], the maximal margin hyperplane is given by $[\langle \mathbf{w}^*, \mathbf{x}\rangle + b^*] = 0$, where

$$\mathbf{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i z_i = \sum_{i \in I_{sv}} \alpha_i^* y_i \mathbf{z}_i ,$$

$$b^* = y_k - \langle \mathbf{w}^*, \mathbf{z}_k \rangle = y_k - \left\langle \sum_{i \in I_{sv}} \alpha_i^* y_i \mathbf{z}_i, \mathbf{z}_k \right\rangle = y_k - \sum_{i \in I_{sv}} \alpha_i^* y_i \langle \mathbf{z}_i, \mathbf{z}_k \rangle,$$

$$I_{sv} := \left\{ i \in \underline{l} : \alpha_i^* > 0 \right\},$$

and α_i^*, $i = 1, 2, \cdots, l$, is the solution of the following optimization problem:

$$\text{maximize } \sum_{i=1}^{l} \alpha_i - 2^{-1} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$\text{subject to } \sum_{i=1}^{l} \alpha_i y_i = 0 \text{ and } \alpha_i \ge 0 \text{ for all } i \in \underline{l}.$$

It is noted that the sequential minimal optimization algorithm [3] can be used to solve the foregoing optimization problem.

ILLUSTRATIVE EXAMPLE

In this section, we provide a simple example just to illustrate the main result. The training data is given as follows.

	$\mathbf{z}_i$	y_i
i=1	(2.7,-1,7,0.2,-0.7)	1
i=2	(-0.1,-1.5,-4,-5)	1
i=3	(-0.7,1.7,-4.2,-5.3)	1
i=4	(2.1,1.5,0,-1)	1
i=5	(3.8,-3,1.3,0.2)	-1
i=6	(5.2,-4.2,3.7,2.7)	-1
i=7	(3,-1.8,1,0.1)	-1

Comparison of the training data with (1), it can be readily obtained that $n = 4$, $l = 7$, $l_1 = 4$, $l_2 = 3$,

$$S_1 = \{(2.7,-1,7,0.2,-0.7), (-0.1,-1.5,-4,-5), (-0.7,1.7,-4.2,-5.3), (2.1,1.5,0,-1)\},$$

$$S_2 = \{(3.8,-3,1.3,0.2), (5.2,-4.2,3.7,2.7), (3,-1.8,1,0.1)\}.$$

In addition, from (2), one has

$$\bar{\mathbf{z}}_1 = (\mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_3 + \mathbf{z}_4)/4 = \begin{bmatrix} 1 & 0 & -2 & -3 \end{bmatrix}^T,$$

$$r_1 = 3.99,$$

$$\bar{\mathbf{z}}_2 = (\mathbf{z}_5 + \mathbf{z}_6 + \mathbf{z}_7)/3 = \begin{bmatrix} 4 & -3 & 2 & 1 \end{bmatrix}^T,$$

$$r_2 = 2.94.$$

This implies that $\|\bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_2\| = \sqrt{50} > 6.93 = r_1 + r_2$. Consequently, by Corollary 2.1 with the choice $\alpha = 0.58$, we conclude that the hyperplane of $-3x_1 + 3x_2 - 4x_3 - 4x_4 + 12 = 0$ correctly classifies this training set in $\Re^4$. In this case, the support vector learning algorithm can be used to find the maximal margin hyperplane in $\Re^4$ for high generalization ability. To save the space, the details refer to the Remark 2.2 and omitted here.

CONCLUSIONS

In this paper, the classical binary classification problem has been investigated. Necessary and sufficient criterion has been presented to guarantee the linear binary separability of the training data in the Euclidean normed space. A suitable hyperplane that correctly classifies the training data has also been constructed provided that the necessary and sufficient criterion is satisfied. Based on the main result, an easy-to-check criterion has been offered to guarantee the linear binary separability of the training set. Finally, a numerical example has been given to illustrate the use of the main result. The necessary and sufficient criterion for the linear multi-class classification in the Euclidean normed space is still remains unanswered. This constitutes an interesting future research problem.

ACKNOWLEDGEMENTS

The author would like to thank the National Science Council of Republic of China for supporting this work under grants NSC-96-2221-E-214-035.

REFERENCES

[1] S. Haykin, *Neural Networks: A Comprehensive Foundation*, New Jersey: Prentice-Hall, 1999.
[2] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychol. Rev.*, vol. 65, pp. 386-408, November 1958.
[3] N. Cristianini, and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge: Cambridge University Press, 2000.
[4] H. C. Chen, Y. L. Lin, Y. J. Sun, and J. G.. Hsieh, (2002). "Modified Rosenblatt's perceptron algorithm and Novikoff's Theorem", *Proceeding of 2002 IEEE International Conference on Industrial Technology*, Bangkok, Thailand, pp. 1282-1284, December 2002.
[5] O. Barzilay, and V. J. Brailovsky, "On domain knowledge and feature selection using a support vector machine", *Pattern Recogn. Lett.*, vol. 20, pp. 475-484, May 1999.
[6] V. J. Brailovsky, O. Barzilay, and R. Shahave, "On global, local, mixed and neighborhood kernels for support vector machines", *Pattern Recogn. Lett.*, vol. 20, pp. 1183-1190, November 1999.
[7] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", *Data Min. Knowl. Disc.*, vol. 2, pp. 121-167, June 1998.
[8] C. C. Chang, C. W. Hsu, and C. J. Lin, "The analysis of decomposition methods for support vector machines", *IEEE Trans. Neural Networ.*, vol. 11, pp. 1003-1008, July 2000.
[9] C. Cortes, and V. Vapnik, "Support vector networks", *Mach. Learn.*, vol. 20, pp. 273-297, September 1995.
[10] F. Girosi, "An equivalence between sparse approximation and support vector machines", *Neural Comput.*, vol. 10, pp. 1455-1480, August 1998.
[11] M. A. Hearst, B. Schölkopf, S. Dumais, E. Osuna, and J. Platt, "Trends and controversies-support vector machines", *IEEE Intell. Syst. App.*, vol. 13, pp. 18-28, July 1998.

[12] R. Herbrich, *Learning Kernel Classifiers: Theory and Algorithms*, Cambridge: MIT Press, 2002.
[13] V. Kecman, *Learning and Soft Computing*, Cambridge: MIT Press, 2001.
[14] B. Schölkopf, and A. J. Smola, *Learning with Kernels: Sgupport Vector Machines, Regularization, Optimization, and Beyond*, Cambridge: MIT Press, 2002.
[15] V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer, 1995.
[16] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, and K. R. Müller, "Engineering support vector machine kernels that recognize translation initiation sites", *Bioinform.*, vol. 16, pp. 799-807, September 2000.

Received: March 19, 2008 Revised: April 3, 2008 Accepted: April 5, 2008