KYBERNETIKA — VOLUME 16 (1980), NUMBER 2

Convergence Properties of Adaptive Threshold Elements in Respect to Application and Implementation

WERNER SCHOENBORN, GERD STANKE

Referring to the convergence proof of the error correction algorithm estimations of the upper bound t_g of correction steps and of the sufficient weight interval $(-w_g, +w_g)$ in digital implementation are derived and illustrated. Both t_g and w_g depend on the gap $2\hat{\Delta}$ between the convex hulls of classes in pattern space. The estimations give $t_g \sim |x|^2_{\max}/\hat{\Delta}^2$ and $w_g \sim |x|^2_{\max}/\hat{\Delta}$, where $|x|_{\max}$ is the largest pattern vector of both classes.

1. INTRODUCTION

Adaptive threshold elements (ATE) are used to realize or to modify linear discriminant functions in a training (learning) process. ATE are applied as basic building blocks for learning systems (hardware) and for iterative determination of discriminant functions with aid of computers (software). They are simple realizable. There exist numerous detailed descriptions of the convergence properties of training algorithms and of the classification properties, e.g. [1, 2, 5, 6 and 9].

The convergence proofs show that for the case of linear separability a solution is found after a finite number of correction steps. For practical purposes it is unsatisfactory proving only the finiteness of the number of correction steps without giving a bound. In the third part of this paper an estimation of the upper bound in relation to the gap between the convex hulls of pattern classes is derived.

The fast progress of digital circuits has an important influence on the recent implementations of ATE, e.g. [3, 13 and 11]. These constructions are characterized by a limited number of bits for the parameter values of the discriminant function. Therefore the possibilities for the implementation are restricted. In the fourth part a relation is derived between the gap width and the weight interval ensuring convergence. The results of the 3rd and 4th part are connected in the 5th part.

For illustration the estimations of t_g and w_g are applied on the special case of threshold functions in BOOLEAN space.

In the following part the properties of ATE are illustrated and basic definitions are introduced.

2. PATTERN SPACE AND *δ-SEPARABILITY*

Patterns are characterized by n-component vectors $\mathbf{x}$ in $\mathscr{X}$-space. Let $\mathscr{K}1$ and $\mathscr{K}2$ be two pattern classes. In the linearly-separable case a pair of hyperplanes exists separating these classes. The hyperplanes are determined by the normal vector $\mathbf{w}^+/|\mathbf{w}^+|$, the signed distance a from origin and the distance $2\Delta > 0$ between the hyperplanes (Fig. 1).*) $2\hat{\Delta}$ denotes the distance between the convex hulls of $\mathscr{K}1$ and $\mathscr{K}2$.

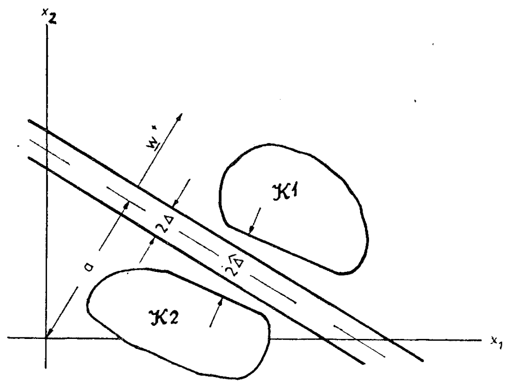

Fig. 1. Two-class problem with hyperplane distance 2Δ and gap width $2\hat{\Delta}$ in $\mathscr{X}$-space.

The decision rule is

$$(1) \qquad \frac{\mathbf{w}^+}{|\mathbf{w}^+|}\mathbf{x} - a \begin{cases} \geqq \Delta & \Rightarrow \mathbf{x} \in \mathscr{K}1 \\ \leqq -\Delta & \Rightarrow \mathbf{x} \in \mathscr{K}2 \\ \text{otherwise} & \Rightarrow \text{rejection.} \end{cases}$$

The inner product of vectors is not especially indicated. For the following considerations it is useful to introduce $\mathscr{Y}$-space (defined by the augmented pattern vector $\mathbf{y}$), the weight vector $\mathbf{w}$ and a parameter δ:

*) Notice that in all figures underlined letters denote vectors (printed by bold letters in the text).

$$
(2) \qquad \begin{aligned} &w = (w^+, w_{n+1}) \\ &y = (x, x_{n+1}) \quad \text{for} \quad x \in \mathcal{K}1 \\ &y = (-x, -x_{n+1}) \quad \text{for} \quad x \in \mathcal{K}2 \\ &\delta > 0 . \end{aligned}
$$

Using Eq. (2) we can write for Eq. (1)

$$
(3) \qquad wy \geqq \delta ,
$$

$$
(4) \qquad \text{with} \quad a = -x_{n+1}w_{n+1}/|w^+| \quad \text{and} \quad \Delta = \delta/|w^+| .
$$

To separate $\mathcal{K}1$ and $\mathcal{K}2$ we have to find a vector w satisfying $wy > 0$ for all y. One of the algorithms doing this is the following well known error correction algorithm

$$
(5) \qquad \begin{aligned} &w_1 \text{ arbitrary} \\ &w_{t+1} = w_t + \gamma y_t \quad \text{if} \quad w_t y_t \leqq 0 \\ &\quad \text{with } 0 < \varepsilon \leqq \gamma \leqq c < \infty , \end{aligned}
$$

where γ is the correction factor and t denotes the correction step. Only patterns y misclassified during the adaption process cause corrections and are accounted to the y_t sequence.

Bounds for the number of correction steps and the sufficient weight interval can be obtained only when some assumptions are made concerning the type of separability. Starting point is a definition of separability with gap. A definition given in [10] will be expanded.

Definition 1. Two classes $\mathcal{K}1$ and $\mathcal{K}2$ are said to be δ_1-*separable* if there exists a weight vector w with

$$
wy \geqq \delta_1 > 0 \quad \text{and} \quad |w_i| \leqq 1 , \quad i = 1, 2, \ldots, n + 1 .
$$

Definition 2. Two classes $\mathcal{K}1$ and $\mathcal{K}2$ are said to be *max-*δ_1*-separable* if they are δ_1-separable but not $(\delta + \varepsilon)_1$-separable with $\varepsilon > 0$.
Figure 2 shows examples for the Definitions 1 and 2.

3. UPPER BOUND OF CORRECTION STEPS

Linear separability is assumed. Then the proof shows that algorithm (5) yields a solution vector after a limited number of correction steps, and the correction bound is especial a function of the gap $2\hat{\Delta}$.

Let w_δ be a solution vector of the region $\mathcal{W}_\delta$ containing all w with $wy \geqq \delta$ (Fig. 3). Then each αw_δ for $\alpha > 0$ is a solution vector of the solution region $\mathcal{W}_0$ containing all w with $wy > 0$ (Fig. 3). We only consider such y of the (cyclical) pattern sequence

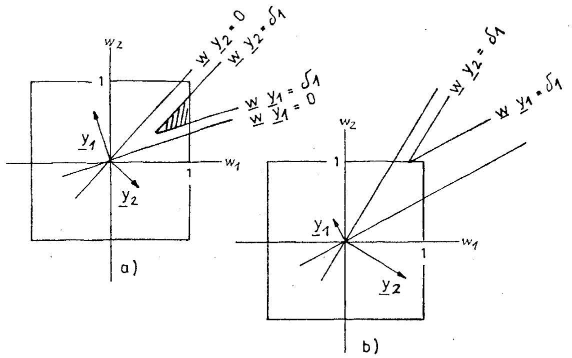

Fig. 2. Illustration of the Definitions 1 and 2 in the weight space. Each class contains only one pattern. The hatched region for δ_1-*separability* in a) degenerates to one point for *max*-δ_1-*separability* in b).

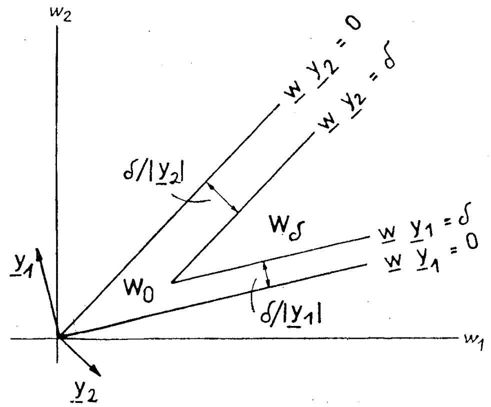

Fig. 3. Solution region $\mathscr{W}_0$ for $wy > 0$ and $\mathscr{W}_\delta$ for $wy \geqq \delta$.

which lead to a correction. Then the following relations are valid (compare also [2]):

$$
(6)\qquad \begin{aligned}
w_{t+1} &= w_t + \gamma y_t\,,\\
w_{t+1} - \alpha w_\delta &= w_t - \alpha w_\delta + \gamma y_t\,,\\
|w_{t+1} - \alpha w_\delta|^2 &= |w_t - \alpha w_\delta|^2 + 2\gamma w_t y_t - 2\gamma\alpha w_\delta y_t + \gamma^2 |y_t|^2\,,
\end{aligned}
$$

with $w_t y_t \leqq 0$, $w_\delta y_t \geqq \delta$ and $|y_t^2| \leqq |y|^2_{max}$ Eq. (6) yields

(7) $$|w_{t+1} - \alpha w_\delta|^2 \leqq |w_t - \alpha w_\delta|^2 - 2\gamma\alpha\delta + \gamma^2 |y|^2_{max} .$$

Starting with the weight vector w_1 we obtain after t corrections

(8) $$|w_{t+1} - \alpha w_\delta|^2 \leqq |w_1 - \alpha w_\delta|^2 - t(2\gamma\alpha\delta - \gamma^2 |y|^2_{max}) .$$

If $w_1 = 0$ is chosen and $0 \leqq |w_{t+1} - \alpha w_\delta|$ is taken into consideration, then from Eq. (8) follows

$$0 \leqq \alpha^2 |w_\delta|^2 - t(2\gamma\alpha\delta - \gamma^2 |y|^2_{max}) .$$

After rearrangement we obtain

(9) $$t \leqq \frac{\alpha^2 |w_\delta|^2}{2\gamma\alpha\delta - \gamma^2 |y|^2_{max}} .$$

Since t must be positive the same follows for the denominator of Eq. (9) and finally

(10) $$\alpha > \frac{\gamma |y|^2_{max}}{2\delta} .$$

Relation (10) ensures the step by step approach of w_t to αw_δ. With t_g as upper bound for the number of correction steps in Eq. (9) the minimization in respect to α yields for

(11) $$\alpha = \frac{\gamma |y|^2_{max}}{\delta}$$

the correction bound

(12) $$t_g = \frac{|y|^2_{max} |w_\delta|^2}{\delta^2}$$

in $\mathscr{Y}$-space. The estimation of correction bound t_g in Eq. (12) gives a minimum if, for a given δ, the shortes solution vector is taken as w_δ or, for *max-δ_1-separability*, δ_1 and $\hat{w}(\delta_1)$ are put in.

However, a relation between the correction bound t_g and the gap $2\hat{\Delta}$ in $\mathscr{X}$-space would be quite clearer. Eq. (12) can be written as

(13) $$t_g = (|x|^2_{max} + x^2_{n+1}) \frac{|w^+|^2 + w^2_{n+1}}{\delta^2} .$$

With $\Delta = \delta / |w^+|$ and $a = -w_{n+1} x_{n+1} / |w^+|$ follows

$$t_g = (|x|^2_{max} + x^2_{n+1}) \frac{x^2_{n+1} + a^2}{x^2_{n+1} \Delta^2} .$$

 We only know $|a| < |x|_{max}$ about the value of a and with $\hat{\Delta}$ for Δ we obtain

(14) $$t_g = \frac{(|x|^2_{max} + x^2_{n+1})^2}{x^2_{n+1}\hat{\Delta}^2}.$$

Figure 4 illustrates Eq. (14).

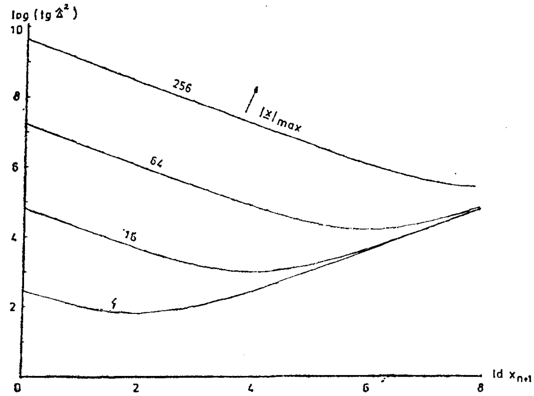

Fig. 4. log $(t_g\Delta^2)$ as function of ld (x_{n+1}) with $|x|_{max}$ as parameter.

Figure 4 shows that an inadequate choice of the $(n + 1)$ th component can increase t_g at some orders. In the literature x_{n+1} is in general chosen independently on $|x|_{max}$. The minimum of t_g with respect to x_{n+1} is

(15) $$t_g = 4\frac{|x|^2_{max}}{\hat{\Delta}^2} \quad \text{for} \quad x_{n+1} = |x|_{max}.$$

Eq. (15) gives a direct dependence of the correction bound t_g on the gap $2\hat{\Delta}$ and the largest pattern vector $|x|_{max}$ in $\mathscr{X}$-space.

The rearrangement of Eq. (15) results in

(16) $$\hat{\Delta} \leqq 2|x|_{max}/\sqrt{t}.$$

After t corrections the following statement can be made, the gap is less than or equal to the right side of Eq. (16).

Theorem 1 gives the basis for the convergence of algorithm (5) in case of a limited number of bits for each weight.

Theorem 1. If $\mathscr{K}1$ and $\mathscr{K}2$ are δ_1-separable and if γ is chosen

$$0 < \gamma < \frac{2\delta_1}{|y|^2_{\max}} \tag{17}$$

then algorithm (5) with the additional condition $|w_i| \leqq 1$, $i = 1, 2, \ldots, n + 1$, leads to a solution after a finite number of corrections.

For proving this we go back to part 3. From Eqs. (7) resp. (10) follows that for

$$0 < \gamma < \frac{2\delta}{|y|^2_{\max}}$$

w approaches any $w \in \widetilde{\mathscr{W}}_\delta$ at each correction. For proving this put $\alpha = 1$ and determine γ for

$$\gamma \quad \text{to} \quad |w_{t+1} - w_\delta|^2 - |w_t - w_\delta|^2 < 0\,.$$

How is the influence of weight limitation? Let w_a be a vector outside the bounded hypercube and w_b the nearest vector to w_a inside the bounded hypercube, then we can show for all vectors w inside

$$|w_b - w| < |w_a - w|\,,$$

particularly for $w \in \widetilde{\mathscr{W}}_\delta$, too, where $\widetilde{\mathscr{W}}_\delta$ is the bounded region of $\mathscr{W}_\delta$ (hatched in Fig. 2). The vector w enters the region $\mathscr{W}_0$ during the step by step approach of w to $\widetilde{\mathscr{W}}_\delta$. This completes the proof.

For

$$\gamma > \frac{2\delta_1}{|y|^2_{\max}\,(1 - \delta_1^2)}$$

examples can be given, where w oscillates during training (Figure 5). Therefore estimation (17) can not be essentially improved.

For digital implementations (pattern components and weighting coefficients are digital) algorithm (5) has to be chosen with integer γ. An appropriate choice is $\gamma = \tilde{\gamma} = 1$. From Theorem 1 follows:

Corollary 1. Supposing δ_1-*separability* and putting $\gamma = \tilde{\gamma} = 1$ in algorithm (5) the following interval $(-w_g, +w_g)$ of weights is sufficient for convergence:

$$w_g > \frac{|y|^2_{\max}}{2\delta_1}\,. \tag{18}$$

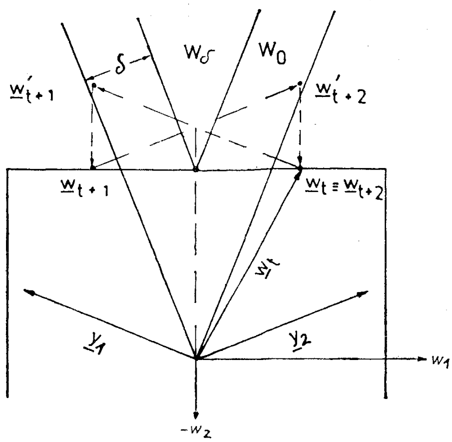

Fig. 5. Example for a not convergent training with $w_{t+2} \equiv w_t \notin \mathscr{W}_0(|y_1| = |y_2| = 1)$.

For $|w_i| \leqq w_g$ the correction factor γw_g with $\gamma < 2\delta_1/|y|^2_{\max}$ according to Eq. (17) ensures that w enters the bounded solution region, and fixing w_g by $\gamma w_g = 1$, for digital implementation of algorithm (5), results in

$$w_g > \frac{|y|^2_{\max}}{2\delta_1}.$$

Demanding a separation gap $\delta < \delta_1$ the same arguments lead to

$$(19) \qquad w_g > \frac{|y|^2_{\max}}{2(\delta_1 - \delta)}.$$

The Ineqs. (18) and (19) are related to the $\mathscr{Y}$-space. In the following the dependence of w_g from $|x|_{\max}$ and $2\hat{\Delta}$ in $\mathscr{X}$-space will be determined. $\Theta = \delta/|w|$ is introduced as auxilliary variable. With $\Delta = \delta/|w^+|$ and $a = -w_{n+1}x_{n+1}/|w^+|$ we find

$$(20) \qquad \frac{\Delta}{\Theta} = \frac{|w|}{|w^+|} = \sqrt{\left(1 + \frac{w^2_{n+1}}{|w^+|^2}\right)} = \sqrt{\left(1 + \frac{a^2}{x^2_{n+1}}\right)}.$$

As in Eq. (14) also in Eq. (20) only $|a| < |x|_{\max}$ can be assumed:

$$\frac{\Delta}{\Theta} < \sqrt{\left(1 + \frac{|x|^2_{\max}}{x^2_{n+1}}\right)}$$

resp.

$$\Theta > \frac{x_{n+1}\Delta}{|y|_{\max}}.$$

In case of *max-δ_1-separability* there is at least one $|w_i| = 1$ and hence $|w| \geqq 1$ as well as

$$\delta_1 \geqq \Theta > x_{n+1}\Delta/|y|_{\max}.$$

Inserted in Eq. (18) and with $\hat{\Delta}$ for Δ we obtain

$$w_g = \frac{|y|^3_{\max}}{2x_{n+1}\hat{\Delta}}$$

resp.

$$w_g = \frac{(x^2_{\max} + x^2_{n+1})^{3/2}}{2x_{n+1}\hat{\Delta}}. \tag{21}$$

Figure 6 illustrates the influence of x_{n+1} on the sufficient interval of weights. The

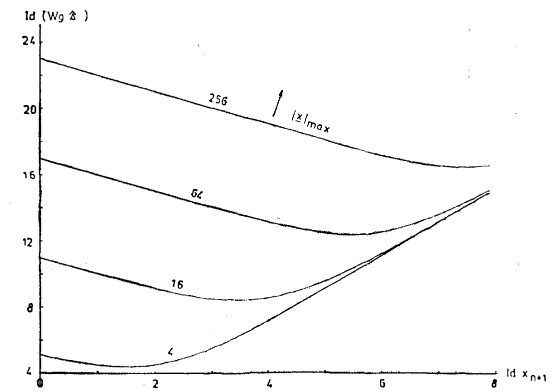

Fig. 6. ld $(w_g\hat{\Delta})$ as function of ld (x_{n+1}) with $|x|_{\max}$ as parameter.

minimum of Eq. (21) with respect to x_{n+1} is

$$w_g = \frac{\sqrt{27}}{4}\frac{|x|^2_{\max}}{\hat{\Delta}} \quad \text{for} \quad x_{n+1} = |x|_{\max}/\sqrt{2}. \tag{22}$$

If w_g is fixed by the implementation then the convergence of algorithm (5) is ensured if

$$\hat{\Delta} \geqq \frac{\sqrt{27}}{4} \frac{|x|^2_{\max}}{w_g} \quad \text{and} \quad x_{n+1} = |x|_{\max}/\sqrt{2}\,.$$

Demanding the separation of classes with a zone width $2\tilde{\Delta}$ near $2\hat{\Delta}$ Eq. (19) yields for the sufficient weight interval

$$w_g = \frac{(x^2_{\max} + x^2_{n+1})^{3/2}}{2x_{n+1}(\hat{\Delta} - \tilde{\Delta})} \quad \text{for} \quad 0 \leqq \tilde{\Delta} < \hat{\Delta} \tag{23}$$

resp.

$$w_g = \frac{\sqrt{27}}{4} \frac{|x|^2_{\max}}{\hat{\Delta} - \tilde{\Delta}} \quad \text{for} \quad 0 \leqq \tilde{\Delta} < \hat{\Delta} \quad \text{and} \quad x_{n+1} = |x|_{\max}/\sqrt{2}\,. \tag{24}$$

Figure 7 shows the hyperbolic growth of w_g for $\tilde{\Delta} \to \hat{\Delta}$.

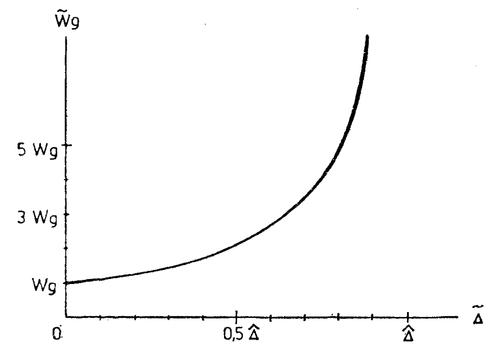

Fig. 7. Sufficient interval of weights $(-w_g, +w_g)$ depending on $\hat{\Delta}$ and $\tilde{\Delta}$.

5. INTERVAL OF WEIGHTS AND CORRECTION BOUND

In what respect does the interval of weights influence the statements for the correction bound? The basis for estimating the correction bound t_g is the step by step approach of w to a vector αw_δ with $w_\delta \in \mathcal{W}_\delta$. Is w_δ specified by $\hat{w}$, where $\hat{w}$ is the solution vector for *max-δ_1-separability*, Eq. (11) yields for $\gamma = 1$ an $\alpha = |y|^2_{\max}/\delta_1$. As in Figure 8 is shown, $\alpha\hat{w}$ lies in a hypercube with the half edge

$$w_\alpha = \frac{|y|^2_{\max}}{\delta_1}\,. \tag{25}$$

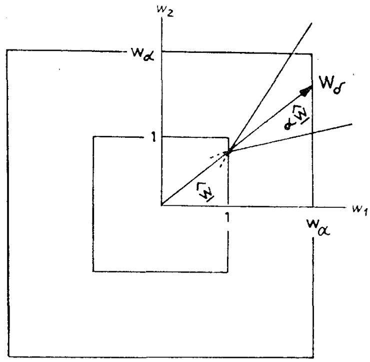

Fig. 8. Hypercube with $\hat{w}$ for *max-δ_1-separability* and $\alpha\hat{w}$ of the convergence proof to algorithm (5).

By a limitation of weight to $|w_i| \leqq |y|^2_{\max}/\delta_1$ the correction bound t_g according to Eq. (12) is not increased since the impact of limitation at a correction step only leads to a stronger approach of w to $\alpha\hat{w}$ in this step (compare the proof of Theorem 1).

The comparison of w_α from Eq. (25) and w_g from Eq. (18) leads to

$$w_\alpha = 2w_g \,. \tag{26}$$

The statements for the correction bound in part 4 allow a limited interval for the weights of $\pm w_\alpha = \pm 2w_g$.

6. EXAMPLE: BOOLEAN THRESHOLD FUNCTIONS

The separation problem in BOOLEAN space is a well known special case. For lower dimensionality the interval $(-w_{\max}, +w_{\max})$ of weights is known being sufficient for the realization of all threshold functions according to

$$wy \geqq 1 \quad \text{with} \quad y_i = \pm 1$$

under the condition

$$\sum_{i=1}^{n+1} |w_i| = \text{minimum}$$

[4, 7 and 12]. Therefore we have all information to determine t_g and w_g.

The number of all existing threshold functions and the maximum weight $w_{\max}$ for all dimensions $\leqq 8$ are given in the 2nd and 3rd row of Table 1. For the dimension n we obtain $|y|^2_{max} = n + 1$. With $\delta_1 = 1/|w|_{max}$ Eqs. (12) and (18) lead to

(27) $$t_g = (n + 1)^2 \, w^2_{max}$$

and

(28) $$w_g = \frac{n + 1}{2} \, w_{max} \, .$$

Table 1

dimension n	1	2	3	4	5	6	7	8
threshold function	4	14	104	1882	94572	15028134	$8{,}38 \, . \, 10^9$	$1{,}76 \, . \, 10^{13}$
maximum weight w_{max}	1	1	2	3	5	9	18	42
correction bound t_g	4	9	32	225	900	3969	20736	142884
sufficient weight w_g	1	1,5	4	7,5	15	31,5	72	189
bits for $(-w_g, +w_g)$	2	3	4	5	5	7	8	9

The 4th and 5th row of Table 1 contain t_g and w_g. The 6th row contains the needed bits for the interval $(-w_g, +w_g)$.

It should be remarked, that the estimation (27) and (28) are not related to optimum conditions for t_g or w_g and accordingly they give too large values.

For higher dimensions n an upper bound of w_{max} can be used [7].

7. SUMMARY

Basing on the convergence proof of the error correction algorithm for ATE (5) and referring to the gap $2\hat{\Delta}$ in $\mathscr{X}$-space estimations of the correction bound t_g (15) and of the sufficient interval $(-w_g, +w_g)$ of weights (22) for digital implementation were derived. The estimation of t_g yields an upper bound for $2\hat{\Delta}$ in dependence on the actual number of correction steps (16). In case of bounded interval of weights the estimation of t_g is preserved if the interval ensuring separation is doubled. The interval of weights ensuring separation is also given for adapting a zone $2\tilde{\Delta} < 2\hat{\Delta}$ (24). As an example the well known BOOLEAN threshold functions are used for a concrete determination of values for t_g and w_g.

(Received May 13, 1977.)

[1] S. Bláha: The convergence of one group of correction training procedures. Kybernetika *5* (1969), 2, 136—153.

[2] R. O. Duda, P. E. Hart: Pattern classification and scene analysis. J. Wiley, New York 1973.

[3] W. H. Dunn, R. Bank, C. F. Fey, C. Chase, L. N. Kanal, W. L. McDermid, D. F. Smith: Pattern recognition systems using digital logic. US PS 3.275. 985 of 27. 9. 1966.

[4] B. Fimmel: Bestimmung der Anzahl der Schwellwertfunktionen von Schwellwertelementen mit beschränkten Gewichten. Kybernetikforschung *6*, 31—40, Deutscher Verlag d. Wiss., Berlin 1976.

[5] G. Meyer-Brötz, J. Schürmann: Methoden der automatischen Zeichenerkennung. R. Oldenbourg Verlag, München 1970.

[6] M. Minsky, S. Papert: Perceptrons. MIT Press, Massachusetts 1969.

[7] S. Muroga, I. Toda, S. Takasu: Theory of majority decision elements. J. Franklin Inst. *271* (1961), 376—418.

[8] S. Muroga, T. Tsuboi, C. R. Baugh: Enumeration of seven-argument threshold functions. IEEE Trans. *C-19* (1970), 818—825.

[9] N. J. Nilsson: Learning machines. McGraw-Hill, New York 1965.

[10] A. Pospisil: AMOS: The "learning" multiclass pattern classifier. Pattern Recognition *3* (1971), 253—268.

[11] W. Schoenborn, B. Fimmel: Digital arbeitender belehrbarer matrixförmiger Zuordner. DDR WP 108.393 of 4. 4. 1974.

[12] R. O. Winder: Enumeration of seven-argument threshold functions. IEEE Trans. *EC-14* (1965), 315—325.

[13] S. A. White: Digital adaptive-element-building blocks for MOS large-scale integration. IEEE Trans. *C-18* (1969), 699—706.

Dr. Werner Schoenborn, Dr. Gerd Stanke, Akademie der Wissenschaften der DDR, Zentralinstitut für Kybernetik und Informationsprozesse (Academy of Sciences of the GDR, Central Institute for Cybernetics and Information Processing), 1199 Berlin—Adlerhof, Rudower Chaussee 5. G.D.R.