COMMENT ON "CONVERGENCE PROPERTIES OF ADAPTIVE THRESHOLD ELEMENTS..."

SVATOPLUK BLÁHA

Results published by authors W. Schoenborn and G. Stanke in [1] can be simplified if normalized pattern weigh vectors are used.

First of all the following convention for denoting relations will be introduced: the relations taken over from [1] have their origin numbers or numbers with apostrophe, if relations have been modified. The new relations are signed by numbers with leading zero.

Let us consider the two-class pattern recognition problem, which is in [1] described in Fig. 1. In that case there exists the solution vector $\mathbf{w}_{2\Delta}$ which determine the hyperplane separating two max -2Δ-separable classes $\mathscr{K}1$ and $\mathscr{K}2$ according to definition (2) in [1]. The maximum distance between two parallel hyperplanes separating both classes is 2Δ. The decision rule is

$$\frac{\mathbf{w}_{2\Delta}^+}{|\mathbf{w}_{2\Delta}^+|}\,\mathbf{x} - a \quad \begin{cases} \geqq \Delta \Rightarrow \mathbf{x} \in \mathscr{K}1 \\ \leqq \Delta \Rightarrow \mathbf{x} \in \mathscr{K}2 \end{cases} \tag{1'}$$

Definition (2) can be rearranged at first by changing the scales of individual components x_i of pattern vectors, so that their values will belong to the interval $\langle 0, 1 \rangle$:

$$\bar{x}_i = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}}\,, \quad i = 1, \ldots, n \tag{01}$$

$$0 \leqq \bar{x}_i \leqq 1$$

Augmented pattern vector $\mathbf{y}$ and weight vector $\mathbf{w}$ will be

$$\begin{aligned} \mathbf{w} &= (\mathbf{w}^+, w_{n+1})\frac{1}{|\mathbf{w}|} \\ \mathbf{y} &= (\ \bar{\mathbf{x}},\ \ x_{n+1}) \quad \text{for} \quad \mathbf{x} \in \mathscr{K}1 \\ \mathbf{y} &= (-\bar{\mathbf{x}}, -x_{n+1}) \quad \text{for} \quad \mathbf{x} \in \mathscr{K}2 \end{aligned} \tag{2'}$$

The decision rule can be then written in the very simple form

(3) $$\mathbf{w}\mathbf{y} \geqq \delta\,.$$

Let the value of additional component be $x_{n+1} = 1$ and as $|\mathbf{w}| = 1$, then

(4′) $$w_{n+1} = -a|\mathbf{w}^+|\,, \quad \delta = \Delta/|\mathbf{w}^+|\,.$$

The training algorithm, concerning only patterns $\mathbf{y}$ misclassified by vector $\mathbf{w}_t$, leads to the following correction of vector $\mathbf{w}_t$:

(5′) $$\begin{aligned} &\mathbf{w}_1 \quad \text{arbitrary, but} \quad |\mathbf{w}_1| = 1 \\ &\mathbf{w}_{t+1} = \overline{\mathbf{w}}_t + \gamma \mathbf{y}_t \\ &\overline{\mathbf{w}}_{t+1} = \mathbf{w}_{t+1}/|\mathbf{w}_{t+1}|\,,\ \bar{\gamma} = \gamma/|\mathbf{w}_{t+1}| \end{aligned}$$

In (6) the solution vector $\mathbf{w}_{2\Delta}$ instead of $\alpha\mathbf{w}_\delta$ will be used:

(6′) $$|\overline{\mathbf{w}}_{t+1} - \overline{\mathbf{w}}_{2\Delta}|^2 = |\overline{\mathbf{w}}_t/|\mathbf{w}_{t+1}| - \overline{\mathbf{w}}_{2\Delta}|^2 + 2\bar{\gamma}\overline{\mathbf{w}}_t\mathbf{y}_t/|\mathbf{w}_{t+1}| - 2\bar{\gamma}\overline{\mathbf{w}}_{2\Delta}\mathbf{y}_t + \bar{\gamma}^2\,|\,\mathbf{y}_t|^2$$

and after t corrections starting with $\mathbf{w}_1$ and considering that maximum length of pattern vector is $|\mathbf{y}_{\max}|^2 = n$, as it is hypercube diagonal, we obtain using $\bar{\gamma} \doteq \gamma$

(8′) $$|\overline{\mathbf{w}}_{t+1} - \overline{\mathbf{w}}_{2\Delta}|^2 \leqq |\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_{2\Delta}|^2 - t(2\gamma\delta - \gamma^2 n)\,.$$

The most unconvenient starting vector $\overline{\mathbf{w}}_1$ is perpendicular to the solution vector $\overline{\mathbf{w}}_{2\Delta}$, and as $|\overline{\mathbf{w}}| = 1$ the difference

(02) $$|\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_{2\Delta}|^2 \leqq 2\,,$$

where equality is valid for the most unconvenient case. When the solution is attained, the difference $|\overline{\mathbf{w}}_{t+1} - \overline{\mathbf{w}}_{2\Delta}| = 0$ and using (02) we obtain from (8′)

(03) $$\begin{aligned} &0 \leqq 2 - t(2\gamma\delta - \gamma^2 n) \\ &t \geqq \frac{2}{2\gamma\delta - \gamma^2 n} \end{aligned}$$

As $t > 0$ then necessarily $2\gamma\delta - \gamma^2 n > 0$ and, consequently, γ must be chosen so that the relation

(04) $$\gamma < \frac{2\delta}{n}$$

holds. Let us minimize t considering equality in (03)

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}(2\gamma\delta - \gamma^2 n) = 2\delta - 2\gamma n = 0$$

We obtain

(05) $$\gamma = \frac{\delta}{n}$$

and the minimum of t, as upper bound for the number of correction steps, is evidently given by

$$t_g = \frac{2n}{\delta^2} = \frac{2}{\gamma^2 n} \tag{06}$$

The minimum value of δ can depend on a technical equipment and in that case it is limited by accuracy of measurement. If the accuracy is for example 1% of the range of possible values, then $\delta = 0{\cdot}005 \sqrt{n}$ and the upper limit of the correction steps is $t_g = 80000$ and the optimum value of γ is $\gamma = 0{\cdot}005/\sqrt{n}$.

(Received July 1, 1980.)

REFERENCES

[1] W. Schoenborn, G. Stanke: Convergence of adaptive threshold elements in respect to application and implementation. Kybernetika *16* (1980), 2, 159—171.

Ing. Svatopluk Bláha, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.