ALMOST SEPARABLE DATA AGGREGATION BY LAYERS OF FORMAL NEURONS[1]

Leon Bobrowski

Abstract: Information extraction or knowledge discovery from large data sets should be linked to data aggregation process. Data aggregation process can result in a new data representation with decreased number of objects of a given set. A deterministic approach to separable data aggregation means a lesser number of objects without mixing of objects from different categories. A statistical approach is less restrictive and allows for almost separable data aggregation with a low level of mixing of objects from different categories. Layers of formal neurons can be designed for the purpose of data aggregation both in the case of deterministic and statistical approach. The proposed designing method is based on minimization of the of the convex and piecewise linear (CPL) criterion functions.

Key words: data aggregation, layers of formal neurons, separability principles

1. Introduction

Data exploration or data mining tools should allow for efficient discovering of regularities (*patterns*) in large data sets. Data models can be designed on the basis of such patterns. Data exploration tools can be based on variety of methods of multivariate data analysis or pattern recognition [1], [2], [3]. In these approaches, each object or event is typically represented as a feature vector or as a point in a multidimensional feature space. Feature vectors are often divided by experts into categories in such a way that each vector belongs to no more than one category (*class*). In this way the reference (*learning*) set can be generated for each category. For example, teams of medical experts want to obtain such representative learning set for each important disease. Such a set should contain a large number of multidimensional vectors representing particular patients linked to this disease.

In the presented paper, data aggregation term means a reduction of numbers of different feature vectors in learning sets resulting from nonlinear transformation of these vectors. Such transformations may cause merging of a large number of different feature vectors into the same transformed vector. In the case of separable data aggregation, only some feature vectors belonging to the same category are merged. In the case of almost separable data aggregation, a low fraction of feature vectors from different categories can be merged.

Data aggregation can be performed by a layer of formal neurons [4]. The feature vectors are transformed by a layer of formal neurons into vectors with binary components. The dipolar and the ranked strategies of designing separable layers of formal neurons were proposed earlier [5]. The possibility of applying the dipolar and ranked strategies to designing almost separable layers is analyzed in the presented paper.

[1] This work was partially supported by the grants KBN 3T11F011 30, by the grant W/II/1/2007 from the Białystok University of Technology and by the grant 16/St/2007 from the Institute of Biocybernetics and Biomedical Engineering PAS.

2. Separable learning sets

Let us assume that m objects O_j (j = 1,….,m) are represented as the so-called feature vectors $x_j = [x_{j1},......,x_{jn}]^T$, or as points in the n-dimensional feature space $F[n]$ ($x_j \in F[n]$). Components (*features*) x_i of the feature vector x represent numerical results of different measurements on a given object O ($x_i \in \{0,1\}$ or $x_i \in R$).

We assume that the feature vector $x_j(k)$ (j = 1,......, m) has been labelled in accordance with the object $O_j(k)$ *category* (*class*) ω_k (k = 1,….,K). The learning set C_k contains m_k feature vectors $x_j(k)$ assigned to the k-th category ω_k

$$C_k = \{x_j(k)\} \quad (j \in I_k) \tag{1}$$

where I_k is the set of indices j of the feature vectors $x_j(k)$ assigned to the class ω_k.

Definition 1: The learning sets C_k (1) are *separable* in the feature space $F[n]$, if they are disjoined in this space ($C_k \cap C_{k'} = \emptyset$, if $k \neq k'$). It means that the feature vectors $x_j(k)$ and $x_{j'}(k')$ belonging to different learning sets C_k and $C_{k'}$ cannot be equal:

$$(k \neq k') \Rightarrow (\forall j \in I_k) \textit{ and } (\forall j' \in I_{k'}) \;\; x_j(k) \neq x_{j'}(k') \tag{2}$$

We are also considering the separation of the sets C_k (1) by the hyperplanes $H(w_k,\theta_k)$ in the feature space $F[n]$:

$$H(w_k,\theta_k) = \{x: w_k^Tx = \theta_k\}. \tag{3}$$

where $w_k = [w_{k1},....,w_{kn}]^T \in R^n$ is the weight vector, $\theta_k \in R^1$ is the threshold, and $(w_k)^Tx$ is the inner product.

Definition 2: The feature vector x_j is situated on the *positive side* of the hyperplane $H(w_k,\theta_k)$ (3) if and only if $(w_k)^Tx_j > \theta_k$. Similarly, vector x_j is situated on the *negative side* of $H(w_k,\theta_k)$ if and only if $(w_k)^Tx_j < \theta_k$.

Definition 3: The learning sets (1) are *linearly separable* in the n-dimensional feature space $F[n]$ if each of the sets C_k can be fully separated from the sum of the remaining sets C_i by some hyperplane $H(w_k,\theta_k)$ (3):

$$\begin{aligned}&(\exists k \in \{1,...,K\}) \;\; (\exists\, w_k,\theta_k) \;\; (\forall x_j(k) \in C_k) \;\;\; w_k^Tx_j(k) > \theta_k. \\ &\qquad\qquad\qquad \text{and} \;\; (\forall x_j(i) \in C_i,\, i \neq k) \;\; w_k^Tx_j(i) < \theta_k\end{aligned} \tag{4}$$

In accordance with the relation (4), all the vectors $x_j(k)$ from the set C_k are situated on the positive side of the hyperplane $H(w_k,\theta_k)$ (3) and all vectors $x_j(i)$ from the remaining sets C_i are situated on the negative side of this hyperplane.

Linear independence of the feature vectors $x_j(k)$ is a sufficient condition for linear separability of the learning sets C_k (1) [5]:

Remark 1: If the feature vectors $x_j(k)$ constituting the learning sets C_k (1) are linearly independent in given feature space $F[n]$, then the sets C_k (1) are linearly separable (4) in this space.

3. Separable layers of formal neurons

The formal neuron $NF(w,\theta)$ can be defined by the threshold activation function $r_t(w,\theta; x)$

$$r = r(w_k,\theta_k; x) = \begin{cases} 1 & \text{if} \quad w_k^T x \geq \theta_k \\ 0 & \text{if} \quad w_k^T x < \theta_k \end{cases} \tag{5}$$

where $w = [w_1,....,w_n]^T \in R^n$ is the weight vector, $\theta \in R^1$ is the threshold, and *r* is the output.

The layer of *L* formal neurons *NF*(w_k,θ_k) transforms feature vectors x into output vectors $r = [r_1,.....,r_L]^T$ with *L* binary components $r_i \in \{0,1\}$:

$$r = r(W; x) = [r_t(w_1,\theta_1;x),......., r_t(w_L,\theta_L;x)]^T \tag{6}$$

where $W = [w_1^T, \theta_1,......., w_L^T, \theta_L]^T$ is the vector of the layer parameters.

The relation (6) determines the transformed vectors $r_j(k)$ with binary components $r_t(w_i,\theta_i;x)$.

$$(\forall k \in \{1,...,K\}) \quad (\forall x_j(k) \in C_k) \quad r_j(k) = r(W;x_j(k)) \tag{7}$$

The transformed learning sets C'_k (1) constitute of the vectors $r_j(k)$:

$$C'_k = \{r_j(k)\} \quad (j \in I_k) \tag{8}$$

We are examining the properties of the transformation (7) which assure the separability $((k \neq k') \Rightarrow x_{j,}(k) \neq x_{j,}(k'))$ (2) of the transformed sets C'_k (8). Such a property can be based on the concept of the mixed dipoles separation [5].

Definition 3: A pair of feature vectors $(x_j(k),x_{j'}(k'))$ creates a *mixed dipole* if and only if these vectors belong to different classes ω_k $(k \neq k')$. Similarly, a pair of vectors from the same class ω_k create the *clear dipole* $(x_j(k),x_{j'}(k))$.

Definition 4: The formal neuron *NF*(w_k,θ_k) (5) separates the dipole $(x_j(k),x_{j'}(k'))$ if <u>only one</u> vector $x_j(k)$ or $x_{j'}(k')$ is situated on the positive side of the hyperplane *H*(w_k,θ_k) (3).

Definition 5: The layer of formal neurons *NF*(w_k,θ_k) (5) is *separable* in respect to the learning sets C_k (1) if and only if the transformed sets C'_k (8) are separable (2) and each feature vector $x_j(k)$ (1) is situated on the positive side of at least one of the hyperplane *H*(w_k,θ_k) (3).

Lemma 1: The necessary condition for separability (*Def*. 5) of the layer of formal neurons *NF*(w_k,θ_k) (5) is the separation (*Def*. 4) of each mixed dipole $(x_j(k),x_{j'}(k'))$ by at least one neuron *NF*(w_k,θ_k) (9) of this layer. [5].

The formal neuron *NF*(w_k,θ_k) (5) is *activated* by the vector $x_j(k)$ ($r_t(w_k,\theta_k; x_j(k)) = 1$) if and only if this vector is situated on the positive side of the hyperplane *H*(w_k,θ_k) (3). In accordance with *Definition* 5, the separable layer should assure that each feature vector $x_j(k)$ (1) activates at least one neuron *NF*(w_k,θ_k) (5). The separable layer can be designed in a multistage procedure, when at one stage a successive neuron *NF*(w_k,θ_k) (5) is added to the layer. In order to increase the generality of the designed neural layers, the following postulate has been introduced [6]:

Postulate of dipolar designing: The separating hyperplane $H(w_k,\theta_k)$ (3) should divide the highest possible number of mixed and undivided yet dipoles $(x_j(k),x_{j'}(k'))$ and at the same time the lowest possible number of the clear dipoles $(x_j(k),x_{j'}(k))$ should be divided.

The linear separability (4) of the sets C'_l (11) could also be achieved in the layer of formal neurons $NF(w_k,\theta_k)$ (5):

Definition 6: The layer of formal neurons $NF(w_k,\theta_k)$ (5) is *linearly separable* in respect to the learning sets C_k (1) if and only if the transformed sets C'_k (8) are separated (4) by the hyperplanes $H(w_k,\theta_k)$ (3) and each feature vector $x_j(k)$ activates at least one of these neurons.

The linear separability (4) of the sets C'_l (11) can be achieved by applying a multistage designing process consistent with the following postulate []:

Postulate of ranked designing: The hyperplane $H(w_k,\theta_k)$ (3) designed during the *l*-th stage should separate $((w_k)^T x_j(k) > \theta_k)$ as many as possible feature vectors $x_j(k)$ from one set $C_k[l]$ under the condition that no vector $x_{j'}(k')$ from the remaining sets $C_{k'}[l]$ $(k' \neq k)$ is separated.

The symbol $C_k[l]$ in the above postulate means that the learning set C_k (1) has been reduced as a result of neglecting feature vectors $x_j(k)$ which have been separated by the hyperplanes $H(w_k,\theta_k)$ (3) during previous l - 1 steps. In the deterministic approach, the designing procedure is stopped during the *L*-th step if all the sets $C_k[L]$ become empty.

Lemma 2: The layer of *L* formal neurons $NF(w_k,\theta_k)$ (5) designed in accordance with the above ranked postulate results in linearly separable (4) transformed sets C'_k (8) [5].

4. Convex and piecewise linear criterion functions (CPL)

The procedure of designing a separable layer can be based on a sequence of minimisation of the convex and piecewise linear (*CPL*) criterion functions $\Psi_k(w,\theta)$ [3], [4]. The perceptron criterion function belongs to the *CPL* family. Let us define the function $\Psi_k(w,\theta)$ by using the positive G_k^+ and the negative G_k^- sets of the feature vectors $x_j = [x_{j1},.....,x_{jn}]^T$ (1):

$$G_k^+ = \{x_j\} \quad (j \in J_k^+) \quad \textit{and} \quad G_k^- = \{x_j\} \quad (j \in J_k^-) \tag{9}$$

Each element x_j of the set G_k^+ defines the positive penalty function $\varphi_j^+(w,\theta)$

$$(\forall x_j \in G_k^+) \qquad \varphi_j^+(w,\theta) = \begin{cases} 1 - w^T x_j + \theta & \textit{if} \quad w^T x_j - \theta \leq 1 \\ 0 & \textit{if} \quad w^T x_j - \theta > 1 \end{cases} \tag{10}$$

Similarly, each element x_j of the set G_l^- defines the negative penalty function $\varphi_j^-(w,\theta)$

$$(\forall x_j \in G_k^-) \qquad \varphi_j^-(w,\theta) = \begin{cases} 1 + w^T x_j - \theta & \textit{if} \quad w^T x_j - \theta \geq -1 \\ 0 & \textit{if} \quad w^T x_j - \theta < -1 \end{cases} \tag{11}$$

The penalty function $\varphi_j^+(w,\theta)$ is aimed at positioning the vector x_j from the set G_k^+ ($x_j \in G_k^+$) on the positive side of the hyperplane $H(w_k,\theta_k)$ (3). Similarly, the function $\varphi_j^-(w,\theta)$ should set the vector x_j from the set G_k^- ($x_j \in G_k^-$) on the negative side of this hyperplane.

The criterion function $\Psi_k(w,\theta)$ is the positively weighted sum of the penalty functions $\varphi_j^+(w,\theta)$ and $\varphi_j^+(w,\theta)$

$$\Psi_k(w,\theta) = \sum_{j \in J_k^+} \alpha_j \varphi_j^+(w,\theta) + \sum_{j \in J_k^-} \alpha_j \varphi_j^-(w,\theta) \tag{12}$$

where α_j ($\alpha_j > 0$) are the positive parameters (*prices*).

The criterion function $\Psi_k(w,\theta)$ belongs to the family of the convex and piecewise linear (*CPL*) criterion functions. Minimization of the function $\Psi_k(w,\theta)$ allows to find optimal parameters (w_k^*,θ_k^*):

$$\Psi_k^* = \Psi_k(w_k^*,\theta_k^*) = min\, \Psi_k(w,\theta) \geq 0 \tag{13}$$

The basis exchange algorithms which are similar to the linear programming allow to find the minimum of the criterion function $\Psi_k(w,\theta)$ efficiently, even in the case of large, multidimensional data sets G_k^+ and G_k^- (29) [5].

The parameters (w_k^*, θ_k^*) constituting the minimum of the function $\Psi_k(w,\theta)$ (12) define the *k*-th neuron $NF(w_k,\theta_k)$ (5) of the layer and the separating hyperplane $H(w_k,\theta_k)$ (3). The criterion functions $\Psi_k(w,\theta)$ can be specified both for the dipolar and for the ranked designing postulates. The specification of the criterion function $\Psi_k(w,\theta)$ (12) is performed through the choice of adequate sets G_k^+ and G_k^- (9) and the prices α_j related to particular vectors x_j from these sets.

It has been proved that the minimal value Ψ_k^* (13) of the criterion function $\Psi_k(w,\theta)$ (12) is equal to zero ($\Psi_k^* = 0$) if and only if the positive G_k^+ and the negative G_k^- sets (9) are linearly separable (4). In this case, all elements x_j of the set G_k^+ (9) are located on the positive side of the hyperplane $H(w_k^*,\theta_k^*)$ (3) and all elements x_j of the set G_k^- are located on the negative side:

$$(\forall x_j \in G_k^+) \quad (w_k^*)^T x_j > \theta_k^*$$
$$and \quad (\forall x_{j'} \in G_k^-) \quad (w_k^*)^T x_{j'} < \theta_k^* \tag{14}$$

If the sets G_k^+ and G_k^- (9) are not linearly separable (4), then $\Psi_k^* > 0$ and the inequalities (14) are fulfilled only partly, not by all, but by a majority of elements x_j of the sets (9).

Minimization of the function $\Psi_k(w,\theta)$ (12) allows one to find the optimal parameters (w_k^*,θ_k^*) defining such hyperplane $H(w_k^*,\theta_k^*)$ (3) which relatively well separates two sets G_l^+ and G_l^- (9). The parameters (w_k^*,θ_k^*) can be also used in the definition of the *l*-th element $NF(w_k^*,\theta_k^*)$ (5) of a neural layer.

5. Almost separable layers

A layer of *L* formal neurons $NF(w_k,\theta_k)$ (6) can transform a large number of feature vectors $x_j(k)$ ($k = 1,\ldots,K$) (1) into the same output vector $r_l = [r_l1,\ldots,r_lL]^T$ (6) with *L* binary components $r_l \in \{0,1\}$, where different indexes l and l′ mean different vectors **rl** and **rl′**:

$$(l \neq l') \Rightarrow (\mathbf{r}l \neq \mathbf{r}l') \qquad (15)$$

The set of such feature vectors xj(k) (1) which are transformed into the same output vector $\mathbf{r}_l$ is called the l-th *activation field* Sl of the layer of formal neurons:.

$$Sl = \{ \mathrm{x}j(k): [r_t(\mathrm{w}_1,\theta_1;\mathrm{x}j(k)),\ldots\ldots, r_t(\mathrm{w}L,\theta L;\mathrm{x}j(k))]T = \mathbf{r}_l\} \qquad (16)$$

Definition 7: The set Sl (16) is the *clear activation field* if all feature vectors xj(k) (1) from this set (xj(k)$\in$ Sl) belong to the same learning set Ck (1). Similarly, the set Sl is the *mixed activation field* if it contains feature vectors xj(k) from different sets Ck.

Lemma 3: The layer of L formal neurons $NF(\mathrm{w}_k,\theta_k)$ (5) is separable in respect to the learning sets C_k (*Def*. 5) if and only if all the activation fields Sl (16) of this layer are clear.

It results from the above *Lemma* that the layer with all the clear activation fields Sl (16) is separable and preserves the learning sets Ck (1) separabilty during data aggregation. Even one mixed field Sl (16) results in a nonseparbilty of the layer. The concept of mixed activation fields Sl (16) is useful in the analysis of nearly separable layers.

Definition 8: The class ω_k is *dominant* in the active the set Sl (16) if and only if the <u>most</u> of the feature vectors xj(k) (1) from this set are assigned to the class ω_k.

The activation field Sl(k) (16) and the output vector **r**l(k) is assigned to the dominant class ωk.

Definition 9: The set Sl (16) is the ε-*clear* (*almost clear*) activation field if the fraction fl of the feature vectors xj(k') (1) from nondominant classes ωk′ in this set (xj(k')$\in$ Sl(k)) is less than ε.

$$fl(k) = ml' / (ml(k) + ml') < \varepsilon \qquad (17)$$

where ml(k) is the number of elements xj(k) of the set Sl(k) (16) belonging to the dominant class ω_k and ml′ is the number of elements xj(k') of the set Sl(k) belonging to non-dominant classes $\omega_{k'}$ ($k' \neq k$).

All feature vectors xj(k) from the l-th activation field Sl are *aggregated* by the layer of formal neurons into one output vector **r**l. In other words, the vector **r**l. *generalizes* all feature vectors xj(k) from the field Sl (26). It can be expected that the layer of formal neurons with a large and ε-clear activation fields Sl (26) could have a great *generalization power*. Such layer could be also used as a classifier with the following decision rule:

$$\textit{if } (\mathrm{x}0 \in Sl(k)) \textit{ then } \mathrm{x}0 \in \omega_k \qquad (18)$$

The quality of a layer of formal neurons $NF(\mathrm{w}_k,\theta_k)$ (5) can be evaluated by the *error rate* of the decision rule (18). The error rate is often estimated through an *apparent error er* [1]:

$$\mathrm{er} = \mathrm{me} / \mathrm{m} \qquad (19)$$

where me is the number of such feature vector xj(k) from the sets Ck (1) which are wrongly allocated by the decision rule (18).

Lemma 4: The error rate er (19) of the decision rule (18) is equal to zero if and only if the layer of formal neurons is separable (*Def*. 5).

The error rate evaluation (19) is positively biased (*optimistic bias*) [1]. The unbiased error rate *er* evaluations can be based on the technique of cross-validation [3].

6. Designing almost separable layers of formal neurons

The separable layer of formal neurons with the decision rule (18) assures the correct classification of all the feature vectors xj(*k*) from the learning sets *C*k (1) and the apparent error rate *er* (19) is equal to zero (*er* = 0). As it results from previous considerations, if the learning sets *C*k (1) are separable (2), the separable layers (*Def*. 5) can be designed in accordance with the dipolar strategy, and in accordance with the ranked strategy. Unfortunately, a separable layer can cause the *overfitting problem* [3]. This problem is manifested in such a way that despite the fact that the apparent error rate *er* (19) of the rule (18) evaluated on the vectors xj(*k*) from learning sets *C*k (1) is equal to zero, the error rate of this rule on new feature vectors x that does not belong to the sets *C*k (1) is too large.

A layer of formal neurons with the apparent error rate *er* (19) greater than zero (*er* $>$ 0) could have a larger *discriminative power* than a separable layer (*er* = 0). We shall take into considerations almost separable (ε-separable) layers of formal neurons.

Definition 10: The layer of formal neurons is ε-separable (*almost separable*) if and only if the apparent error rate *er* (19) is less than ε ($er < \varepsilon$).

Lemma 5: If all the activation fields *S*l(*k*) (16) of the layer of formal neurons are of the ε-*clear* type (*Def*. 9), then this layer is ε-separable

Proof: The apparent error rate *er* (19) can be represented in the below manner:

$$er = me \,/\, m = (m1' + \ldots\ldots + mM') \,/\, m \tag{20}$$

where *mi*′ (17) is the number of elements xj(*k*′) of the activation field *S*i(*k*) (16) belonging to non-dominant classes $\omega_{k'}$ ($k' \neq k$), and *M* is the number of activation fields. Thus

$$\begin{aligned} er &= m1'/(ml(k(1)) + ml')\ (ml(k(1)) + ml') \,/\, m + \ldots\ \ldots \\ &\ldots\ \ldots.\ m1'/(ml(k(M)) + mM')\ (mM(k(M)) + mM') \,/\, m < \varepsilon \end{aligned} \tag{21}$$

and the thesis is proved. □

The *Lemma* 5 gives indications for designing almost separable layers of formal neurons $NF(w_k,\theta_k)$ (6). Designing process can be based on a generation of such activation fields *S*l(*k*) (16) which are of the ε-*clear* type (*Def*. 9). As a consequence, the *Ranked designing postulate* can be modified in the below manner:

Postulate of almost ranked designing: The hyperplane $H(w_k,\theta_k)$ (3) designed during the *l*-th stage should separate ($(w_k)^T x_j(k) > \theta_k$) as many as possible feature vectors $x_j(k)$ from one set $C_k[l]$ under the condition that the fraction *f*l (17) of the separated vectors xj(*k*′) (1) from other sets Ck′[*l*] ($k' \neq k$) is less than ε.

One can see that all the activation fields *S*l(*k*) (16) of the layer of formal neurons designed in accordance with the above postulate are of the ε-*clear* type (*Def*. 9).

The *Postulate of dipolar designing* can be also modified in a similar manner, for example by neglecting such mixed dipoles ($x_j(k), x_{j'}(k')$) that the feature vectors $x_j(k)$ and $x_{j'}(k')$ belong to the activation field *S*l(*k*) (16) of the ε-*clear* type (*Def*. 9).

The data aggregation process can be based on a layer of formal neurons *NF*(w_k, θ_k) (6). Let us define the *aggregation coefficient* ηa of such layer in the following manner

$$\eta a = (m - m(\mathbf{rl}) / (m - K) \tag{22}$$

where, *m* is the number of the feature vectors xj(*k*) in the sets *C*k (1), *m*(**rl**) is he number of different output vectors **rl** (15) from a separable layer, and *K* is the number of the classes ω_k or the learning sets Ck(1).

The minimal number *m*(**rl**) of the output vectors **rl** (15) from a separable layer is equal to *K* (*m*(**rl**) = *K*). The aggregation coefficient ηa (22) takes the maximal value equal to one (ηa = 1) in this ideal situation. The aggregation coefficient ηa (22) of a layer of formal neurons *NF*(wi,θi) (5) can take the maximal value ηa = 1 if and only if the learning sets *C*k (1) are linearly separable. The maximal value of the number *m*(**rl**) is equal to *m*. There is no aggregation in this case and the aggregation coefficient ηa (29) takes the minimal value equal to 0 (ηa = 0). As a result:

$$0 \le \eta a \le 1 \tag{23}$$

It can be noted that a solution of the *Optimization problem* leads to the maximisation of the aggregation coefficient ηa (22).

Optimization problem: To design such ε-separable (*Def*. 10) layer of formal neurons *NF*(w_k, θ_k) (6) which has a minimal number *M* of activation fields *S*l(*k*) (16) or different output vectors **rl** (15).

The minimal number *M* of the activation fields *S*l (16) can not be less than the number *K* of the classes ω_k (*M* ≥*K*).

7. Concluding remarks

Separable layer of formal neurons *NF*(w_k, θ_k) (5) can be induced from the learning sets C_k (1) only when these sets are separable (2) in a given feature space. The dipolar strategy allows for preserving the separability of the sets C_k (1) during their transformation by the induced layer of formal neurons. The ranked strategy also allows to achieve the linear separability (4) of the transformed sets C'_k (8).

Separable layers of formal neurons with the decision rule (18) secure correct classification of all the feature vectors xj(*k*) from the learning sets *C*k (1) and the apparent error rate *er* (19) is equal to zero (*er* = 0) in this case. Unfortunately, such property of the designed layers is often linked with the occurrence of overfitting [3]. Despite the fact that apparent error rate *er* (19) evaluated on the basis of elements $x_j(k)$ of the learning sets *C*k (1) is equal to zero, the decision rule (18) can be burdened with a too large error rate on new feature vectors x (x ∉ *C*k). In this case, the generalization power of the designed network is too low.

Almost separable layer of formal neurons *NF*(w_k, θ_k) (5) allow apparent error rate *er* (19) greater than zero. Designing almost separable layers should allow for achieving a larger generalization power and level of data aggregation in comparison to strictly separable layers. There are still many problems to resolve in the search for efficient strategy for designing almost separable layer.

Bibliography

[1]. Johnson R. A., Wichern D. W.: Applied Multivariate Statistical Analysis, Prentice-Hall, Inc., Englewood Cliffs, New York, 1991

[2]. Duda O. R. and Hart P. E., Stork D. G.: Pattern Classification, J. Wiley, New York, 2001,

[3]. Fukunaga K., Introduction to Statistical Pattern Recognition, Academic Press 1990,

[4]. Rosenblatt F.: Principles of neurodynamics, Spartan Books, Washington 1962

[5]. Bobrowski L.: Data mining based on convex and piecewise linear (CPL) criterion functions, (in Polish), Technical University Białystok, 2005,

[6]. Bobrowski L.: "Piecewise-Linear Classifiers, Formal Neurons and separability of the Learning Sets", Proceedings of ICPR'96, pp. 224-228, (13th International Conference on Pattern Recognition", August 25-29, 1996, Vienna)

[7]. Bobrowski L,: "Design of piecewise linear classifiers from formal neurons by some basis exchange technique"' Pattern Recognition, 24(9), pp. 863-870, 1991

[8]. Bobrowski L., 'The ranked neuronal networks", Biocybernetics and Biomedical Engineering, Vol. 12, No. 1-4, pp. 61-75, 1992

Author's Information

Leon Bobrowski – leon@ii.pb.bialystok.pl

Faculty of Computer Science, Bialystok Technical University

Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland