Diagnosis supporting rules of the Hepar system

Article *in* Studies in health technology and informatics · February 2001

DOI: 10.3233/978-1-60750-928-8-1309 · Source: PubMed

Biocybernetics and Biomedical Engineering
2007, Volume 27, Number 3, pp. 3–14

Induction of Diagnostic Support Rules through Data Mapping — on the Example of the *Hepar* System

LEON BOBROWSKI [1,2,*], **HANNA WASYLUK** [2,3]

[1]*Faculty of Computer Science, Białystok Technical University*
[2]*Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland*
[3]*Medical Center of Postgraduate Education, Warsaw*

Induction of similarity measures through data transformations is considered in the paper. Particular attention is paid to the separable linear transformation of multidimensional data on a plane and the data visualization on diagnostic maps. Diagnostic support in the *Hepar* computer system is based on diagnostic maps. Separable linear transformations in the *Hepar* system are based on minimization of the convex and piecewise linear (CPL) criterion functions. The CPL criterion functions give a possibility for a flexible and efficient design of the visualizing transformations which well separates the disease on the diagnostic maps.

Keywords: decision support, similarity measures, linear transformations, data visualization

1. Introduction

Medical diagnostic support systems are often based on the case based reasoning (*CBR*) scheme [1]. In such a scheme, the record of a new patient is used to identify the most similar records of the patients with the confirmed disease in the system's database. The new patient is linked to a disease that has been confirmed for the most similar records. A similar paradigm is also used in the nearest neighbors (*K-NN*) technique originated from the pattern recognition [2, 3]. One of the central problems in implementation of the *CBR* or the *K-NN* scheme is the choice of the similarity measure or the distance function between the database records. The quality of the decision support rules can be improved by adjusting the similarity measures or adequately tailoring the distance functions [4].

* Correspondence to: Leon Bobrowski, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, 4 Trojdena St., 02-109 Warsaw, Poland, e-mail: leon@ibib.waw.pl

Here we are analyzing possibilities of applying separable linear transformations of learning data sets to induction of the similarity measures and diagnostic support rules. Designing a linear transformation scheme based on separability postulates is considered here [4, 5]. The separability postulates are reinforced through minimization of the convex and piecewise linear (*CPL*) criterion functions. The basis exchange algorithms, similar to linear programming, allow to efficiently find the minimum of the *CPL* criterion functions, even in the case of large, multidimensional data sets [6]. Among others, the visualizing transformations can be designed this way.

Separable linear transformations have been used for induction of similarity measures in the *Hepar* system for diagnostic support [7, 8]. This system combines the hepathological database with the data analysis and diagnostic support tools. These tools include diagnostic maps, which result from linear data transformations aimed at a possibly good separation of selected diseases on the visualizing plane.

2. Separable Learning Sets

Let us assume that patient descriptions O_j stored in a clinical database can be represented as the so called feature vectors $\mathbf{x}_j = [x_{j1}, ..., x_{jn}]^T$, points in the n-dimensional feature space $F[n]$. The components x_{ji} of the vectors $\mathbf{x}_j$ are numerical results of a variety of medical examinations of a given patient O_j. The feature vectors $\mathbf{x}$ can be of the mixed, qualitative-quantitative type, because they contain both symptoms and signs ($x_i \in \{0,1\}$) as well as numerical results of laboratory tests ($x_i \in R$).

We assume that a clinical database allows to identify patients $O_j(k)$ related to the k-th disease (*class, category*) ω_k ($k = 1 ,..., K'$) and to represent such patients by *labelled* feature vectors $\mathbf{x}_j(k)$. The learning set C_k contains m_k labelled feature vectors $\mathbf{x}_j(k)$ linked to the k-th disease ω_k:

$$C_k = \{\mathbf{x}_j(k)\} \quad (j \in I_k) \tag{1}$$

where I_k is the set of indices j of m_k feature vectors $\mathbf{x}_j(k)$ belonging to the class $\omega_k (k = 1, ..., K')$.

DEFINITION 1: The learning sets C_k (1) are *separable* in the feature space $F[n]$ if they are disjoined in this space. It means that each of the feature vectors $\mathbf{x}_j$ belongs to only one set C_k:

$$(\forall \mathbf{x}_j(k) \in C_k) \; and \; (\forall \mathbf{x}_{j'}(k') \in C_{k'}, k' \neq k) \; \mathbf{x}_{j'}(k') \neq \mathbf{x}_j(k). \tag{2}$$

We are also considering separation of the sets C_k (1) by the hyperplanes $H(\mathbf{w}_k, \theta_k)$ in the feature space $F[n]$

$$H(\mathbf{w}_k, \theta_k) = \{\mathbf{x}: (\mathbf{w}_k)^T\mathbf{x} = \theta_k\} \tag{3}$$

where $\mathbf{w}_k = [w_{k1},...,w_{kn}]^T \in R^n$ is the weight vector, $\theta_k \in R^1$ is the threshold, and $(\mathbf{w}_k)^T\mathbf{x}$ is the inner product. The feature vector $\mathbf{x}$ is situated on the *positive side* of the hyperplane $H(\mathbf{w}_l,\theta_l)$ if and only if $(\mathbf{w}_k)^T\mathbf{x} > \theta_l$. Similarly, vector $\mathbf{x}$ is situated on the *negative side* of $H(\mathbf{w}_l,\theta_l)$ if and only if $(\mathbf{w}_k)^T\mathbf{x} < \theta_l$.

DEFINITION 2: The learning sets (1) are *linearly separable* if each of the sets C_k can be fully separated from the sum of the remaining sets $C_{k'}$ by some hyperplane $H(\mathbf{w}_k,\theta_k)$ (4):

$$(\forall k \in \{1,.....,K'\})\ (\exists\ \mathbf{w}_k,\theta_k)\ (\forall \mathbf{x}_j(k) \in C_k)\ (\mathbf{w}_k)^T\mathbf{x}_j(k) > \theta_k \\ \textit{and}\ (\forall \mathbf{x}_{j'}(k') \in C_{k'},\ k' \neq k)\ (\mathbf{w}_k)^T\mathbf{x}_{j'}(k') < \theta_k \tag{4}$$

In accordance with the relation (4), all the vectors $\mathbf{x}_j(k)$ belonging to the learning set C_k are situated on the positive side of the hyperplane $H(\mathbf{w}_k,\theta_k)$ (3) and all the feature vectors $\mathbf{x}_{j'}(k')$ from the remaining sets $C_{k'}$ are situated on the negative side of this hyperplane.

3. Transformations of the Feature Vectors on the Visualizing Plane

Let us consider transformations $\mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x})]^T$ of n-dimensional feature vectors $\mathbf{x}_j(k)$ (1) on points $\mathbf{y}_j(k) = [y_1(k),y_2(k)]^T$ of the plane R^2:

$$\mathbf{y} = [y_1,y_2]^T = [F_1(\mathbf{x}), F_2(\mathbf{x})]^T. \tag{5}$$

If the transformations $F_i(\mathbf{x})$ are linear, then they can be represented in the manner below

$$\mathbf{y} = [y_1,y_2]^T = [\mathbf{w}_1{}^T\mathbf{x},\ \mathbf{w}_2{}^T\mathbf{x}\ ]^T. \tag{6}$$

where $\mathbf{w}_i = [w_{i1},...,w_{in}]^T$ $(i = 1, 2)$ are linearly independent weight vectors.

The linear transformation (5) can be represented in the matrix form

$$\mathbf{y}_j = \boldsymbol{W}^T \mathbf{x}_j\ (j = 1,...,m), \tag{7}$$

where $\boldsymbol{W} = [\mathbf{w}_1,\mathbf{w}_2]$ is the matrix of dimension $(n \times 2)$.

The relation (7) allows to generate the transformed learning sets C'_k, where

$$C'_k = \{\mathbf{y}_j(k)\}\ (j \in I_k). \tag{8}$$

4. The Distance Functions Induced by Linear Data Transformations

The nearest neighbours diagnosis support rules are based on the distances $\delta(\mathbf{x}_0,\mathbf{x}_j(k))$ between the feature vector $\mathbf{x}_0$ of a newly diagnosed patient and the labelled vectors $\mathbf{x}_j(k)$ from the clinical database. Let us assume, for a moment, that the labelled feature vectors $\mathbf{x}_j(k)$ are *ranked* $\{\mathbf{x}_{j(1)}, \mathbf{x}_{j(2)},...,\mathbf{x}_{j(m)}\}$ in respect to the distances $\delta(\mathbf{x}_0,\mathbf{x}_j(k))$ between the vectors $\mathbf{x}_0$ and $\mathbf{x}_j(k)$:

$$(\forall i \in \{1,...,m-1\})\ \delta(\mathbf{x}_0,\mathbf{x}_{j(i)}) \leq \delta(\mathbf{x}_0,\mathbf{x}_{j(i+1)}). \tag{9}$$

Let use define the ball $B_x(\mathbf{x}_0,K)$ which is centered in a point $\mathbf{x}_0$ and contains only K vectors $\mathbf{x}_{j(i)}(k)$ that are nearest to $\mathbf{x}_0$:

$$B_x(\mathbf{x}_0,K) = \{\mathbf{x}_j(k):\quad \delta(\mathbf{x}_0,\mathbf{x}_{j(i)}) \leq \delta(\mathbf{x}_0,\mathbf{x}_{j(K)})\}. \tag{10}$$

We can assume that the ball $B_x(\mathbf{x}_0,K)$ contains K vectors $\mathbf{x}_{j(i)}(k)$ that are the most *similar* to the vector $\mathbf{x}_0$.

In accordance with the K-nearest neighbours (K-NN) classification rule, the object $\mathbf{x}_0$ is allocated into this class ω_k ($k = 1,......,K'$) where the most of the labelled vectors $\mathbf{x}_j(k)$ from the ball $B_x(\mathbf{x}_0,K)$ (10) belong [2]:

$$\textbf{\textit{if}}\ (\forall l \in \{1,...,K'\})\ n_k \geq n_l\ \textbf{\textit{then}}\ \mathbf{x}_0 \in \omega_k \tag{11}$$

where n_k is the number of the vectors $\mathbf{x}_j(k)$ from the set C_k (1) contained in the ball $B_\mathrm{x}(\mathbf{x}_0,K)$.

The Euclidean distance $\delta_\mathrm{E}(\mathbf{x}_0,\mathbf{x}_{j(i)})$ between the feature vectors $\mathbf{x}_0$ and $\mathbf{x}_{j(i)}$ is commonly used in the nearest neighbours classification rule (11):

$$\delta_\mathrm{E}^2(\mathbf{x}_0,\mathbf{x}_{j(i)}) = (\mathbf{x}_0 - \mathbf{x}_{j(i)})^T(\mathbf{x}_0 - \mathbf{x}_{j(i)}). \tag{12}$$

The quality of the rule (11) can be improved in some cases by modification of the distance function $\delta_\mathrm{E}(\mathbf{x}_0,\mathbf{x}_{j(i)})$. Let us use for this purpose the *induced* distance function $\delta_\mathrm{I}(\mathbf{x}_0,\mathbf{x}_{j(i)})$ between the feature vectors $\mathbf{x}_0$ and $\mathbf{x}_{j(i)}$. The *induced distance* $\delta_\mathrm{I}(\mathbf{x}_0,\mathbf{x}_{j(i)})$ between feature vectors $\mathbf{x}_0$ and $\mathbf{x}_{j(i)}$ is defined as equal to the Euclidean distance $\delta_\mathrm{E}(\mathbf{y}_0,\mathbf{y}_{j(i)})$ between the transformed points $\mathbf{y}_0$ and $\mathbf{y}_{j(i)}$ (7):

$$\delta_\mathrm{I}^2(\mathbf{x}_0,\mathbf{x}_{j(i)}) = \delta_\mathrm{E}^2(\mathbf{y}_0,\mathbf{y}_{j(i)}) = (\mathbf{y}_0 - \mathbf{y}_{j(i)})^T(\mathbf{y}_0 - \mathbf{y}_{j(i)}) = (\mathbf{x}_0 - \mathbf{x}_j(k))^T \boldsymbol{W}\boldsymbol{W}^T(\mathbf{x}_0 - \mathbf{x}_j(k)). \tag{13}$$

The *induced ball* $B_\mathrm{I}(\mathbf{x}_0,K)$ can be defined by using the distance function $\delta_\mathrm{I}(\mathbf{x}_0,\mathbf{x}_{j(i)})$ (13).

$$B_I(\mathbf{x}_0,K) = \{\mathbf{x}_j(k): \delta_I^2(\mathbf{x}_0,\mathbf{x}_{j(i)}) \le \delta_I^2(\mathbf{x}_0,\mathbf{x}_{j(K)})\} = \{\mathbf{x}_j(k): \delta_E^2(\mathbf{y}_0,\mathbf{y}_{j(i)}) \le \delta_E^2(\mathbf{y}_0,\mathbf{y}_{j(K)})\} \quad (14)$$

where points $\mathbf{y}_{j(i)}$ on the visualizing plane are *ranked* $\{\mathbf{y}_{j(1)}, \mathbf{y}_{j(2)},..., \mathbf{y}_{j(n)}\}$ (9) in accordance with the Euclidean distance function $\delta_E(\mathbf{y}_0,\mathbf{y}_{j(i)})$ (12). The induced ball $B_I(\mathbf{x}_0,K)$ contains K feature vectors $\mathbf{x}_j(k)$ with the smallest induced distance $\delta_I^2(\mathbf{x}_0,\mathbf{x}_{j(i)})$ (13) (Fig. 1).

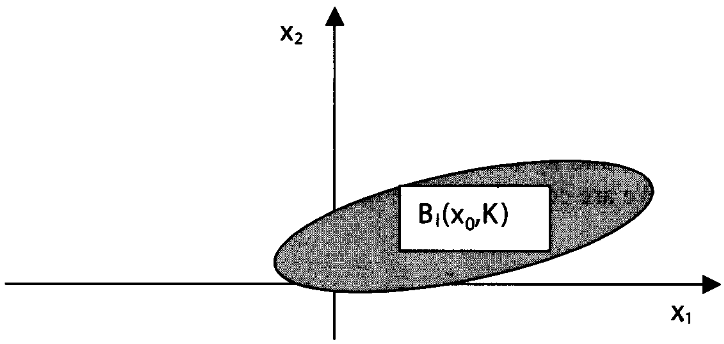

Fig. 1. An example of a ball $B_I(\mathbf{x}_0,K)$ (14) induced in a two-dimensional feature space by the transformation (19)

The K nearest neighbours (*K-NN*) type of diagnosis support rule (11) can be based on the induced ball $B_I(\mathbf{x}_0,K)$ (14):

If the most of the labelled vectors $\mathbf{x}_j(k)$ *from the induced ball* $B_I(\mathbf{x}_0,K)$ (14) (15) *belong to the disease* ω_k, ***then the** patien*t represented by $\mathbf{x}_0$ *should be related to this disease.*

We examine here procedures aimed at optimization of the rule (15) through a special choice of the parameter vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ in the linear transformation (6) of the feature vectors $\mathbf{x}_j(k)$ (1) on the visualizing plane.

5. Decorrelation of Transformed Data Sets

An important role in the classification is played by such linear transformations (7) which reduce correlation or whitening of the learning sets C_k (1) [2]. Such transformations can be built on the basis of the eigenvectors $\mathbf{k}_i$ and the eigenvalues λ_i of the covariance matrix Σ. Let us take into consideration the covariance matrix Σ_k estimated on the set C_k (1)

$$\Sigma_k = \sum_{j \in I_k} (\mathbf{x}_j - \boldsymbol{\mu}_k)(\mathbf{x}_j - \boldsymbol{\mu}_k)^T / (m_k - 1) \quad (16)$$

where $\boldsymbol{\mu}_k$ is the mean vector in the set C_k

$$\boldsymbol{\mu}_k = \left(\sum_{j \in I_k} \mathbf{x}_j\right) / m_k \tag{17}$$

The eigenvalue problem with the covariance matrix Σ_k is formulated as the search for the eigenvectors $\mathbf{k}_i$ and the eigenvalues λ_i which fulfil the below equation

$$\Sigma_k \, \mathbf{k}_i = \lambda_i \, \mathbf{k}_i . \tag{18}$$

The eigenvectors $\mathbf{k}_i$ can be chosen as orthogonal vectors ($\mathbf{k}_k^T\mathbf{k}_l = 0$ if $k \neq l$) of the unit length ($\mathbf{k}_k^T\mathbf{k}_k = 1$).

Let us assume that the linear transformation (7) is defined by two orthogonal eigenvectors $\mathbf{k}_l$ and $\mathbf{k}_{l'}$ ($l \neq l'$) with eigenvalues λ_l and $\lambda_{l'}$ ($\lambda_l > \lambda_{l'} > 0$). Typically, the eigenvectors $\mathbf{k}_l$ and $\mathbf{k}_{l'}$ with the greatest eigenvalues λ_l and $\lambda_{l'}$ are taken into considerations. We are considering the linear transformation (7) of the following form:

$$\mathbf{y}_j(k) = \boldsymbol{B}^T (\mathbf{x}_j(k) - \boldsymbol{\mu}_k) \quad (j = 1,\ldots,m) \tag{19}$$

where matrix $\boldsymbol{B}$ of the dimension ($n \times 2$) has the columns formed by the vectors $\mathbf{k}_l/(\lambda_l)^{1/2}$ and $\mathbf{k}_{l'}/(\lambda_{l'})^{1/2}$.

$$\boldsymbol{B} = [\mathbf{k}_l/(\lambda_l)^{1/2}, \mathbf{k}_{l'}/(\lambda_{l'})^{1/2}]. \tag{20}$$

The transformed vectors $\mathbf{y}_j(k)$ (19) form the sets C_k' (8) with the mean vectors $\boldsymbol{\mu}_k'$ (17). The correlation matrix Σ_k' (13) is defined on the transformed vectors $\mathbf{y}_j$ (19) from one set C_k'

$$\begin{aligned} \Sigma_k' &= \sum_{j \in I_k} (\mathbf{y}_j - \mu_k')(\mathbf{y}_j - \mu_k')^T / (m_k - 1) = \\ &\boldsymbol{B}^T \sum_{j \in I_k} (\mathbf{x}_j - \mu_k)(\mathbf{x}_j - \mu_k)^T \boldsymbol{B} / (m_k - 1) = \boldsymbol{B}^T \Sigma_k \boldsymbol{B} = \boldsymbol{I}_{2\times 2} \end{aligned} \tag{21}$$

where $\boldsymbol{I}_{2\times 2}$ is the unit matrix of dimension (2 x 2).

In accordance with the equation (21), the transformation (19) decorrelates data set C_k' (8) on the visualizing plane. In this case, the transformed vectors $\mathbf{y}_j(k)$ (19) from the set C_k' have the unit correlation matrix Σ_k' (21). Generally, the classification rule (11) with the Euclidean distance $\delta_E^2(\mathbf{y}_0, \mathbf{y}_{j(i)})$ (13) can be matched in the best manner to data set C_k' (8) with the unit correlation matrix Σ_k' (21).

6. Perceptron Criterion Functions (*CPL*)

Designing data transformation (7) can be also based on minimization of the convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w},\theta)$ [5, 6]. It is convenient to define the functions $\Phi(\mathbf{w},\theta)$ by using the positive G^+ and the negative G^- sets of the feature vectors $\mathbf{x}_j$ (1).

$$G^+ = \{\mathbf{x}_j\} \; (j \in J^+) \; and \; G_1^- = \{\mathbf{x}_j\} \; (j \in J^-). \tag{22}$$

Each element $\mathbf{x}_j$ of the set G^+ defines the positive penalty function $\varphi_j^+(\mathbf{w}, \theta)$

$$\varphi_j^+(\mathbf{w}, \theta) = \begin{cases} 1 - \mathbf{w}^T\mathbf{x}_j + \theta & if \quad \mathbf{w}^T\mathbf{x}_j - \theta \le 1 \\ 0 & if \quad \mathbf{w}^T\mathbf{x}_j - \theta > 1 \end{cases} \tag{23}$$

Similarly, each element $\mathbf{x}_j$ of the set G^- defines the negative penalty function $\varphi_j^-(\mathbf{w}, \theta)$ (Fig. 2)

$$\varphi_j^-(\mathbf{w}, \theta) = \begin{cases} 1 + \mathbf{w}^T\mathbf{x}_j - \theta & if \quad \mathbf{w}^T\mathbf{x}_j - \theta \ge -1 \\ 0 & if \quad \mathbf{w}^T\mathbf{x}_j - \theta < -1 \end{cases} \tag{24}$$

The penalty function $\varphi_j^+(\mathbf{w},\theta)$ is aimed at situating vector $\mathbf{x}_j$ ($\mathbf{x}_j \in G_1^+$) on the positive side of the hyperplane $H(\mathbf{w},\theta)$ (3). Similarly, the function $\varphi_j^-(\mathbf{w},\theta)$ should situate the vector $\mathbf{x}_j$($\mathbf{x}_j \in G_1^-$) on the negative side of this hyperplane.

The *perceptron criterion function* $\Phi(\mathbf{w},\theta)$ is defined as the weighted sum of the penalty functions $\varphi_j^+(\mathbf{w},\theta)$ (23) and $\varphi_j^-(\mathbf{w},\theta)$ (24) [9]:

$$\Phi(\mathbf{w},\theta) = \sum_{j \in J^+} \alpha_j^+ \varphi^+(\mathbf{w},\theta) + \sum_{j \in J^-} \alpha_j^- - \varphi^-(\mathbf{w},\theta) \tag{25}$$

where α_j^+ ($\alpha_j^+ > 0$) and α_j^- ($\alpha_j^- > 0$) are positive parameters (*prices*), J_1^+ and J_1^- are the sets of indices j of feature vectors $\mathbf{x}_j$ (22).

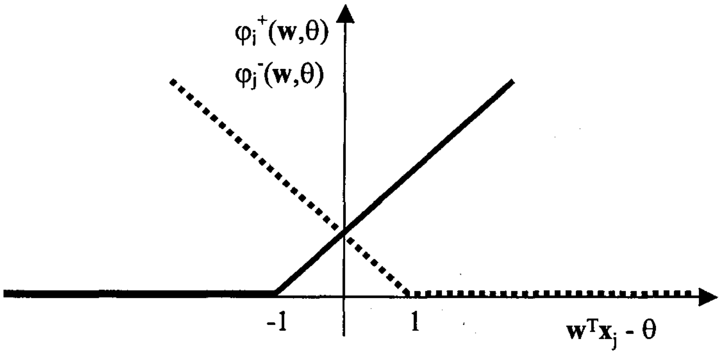

Fig. 2. The positive $\varphi_j^+(\mathbf{w}, \theta)$ (23) and the negative $\varphi_j^-(\mathbf{w}, \theta)$ (24) penalty functions. (••••• — $\varphi_j^+(\mathbf{w}, \theta)$ and —— — $\varphi_j^-(\mathbf{w}, \theta)$)

The perceptron criterion function $\Phi(\mathbf{w},\theta)$ belongs to the family of the convex and piecewise linear (*CPL*) functions. Minimization of the function $\Phi(\mathbf{w},\theta)$ allows to find parameters $(\mathbf{w}^*,\theta^*)$ which define such hyperplane $H(\mathbf{w}^*,\theta^*)$ (3) that optimally separates two sets G_l^+ and G_l^- (22).

$$\Phi^* = \Phi(\mathbf{w}^*,\theta^*) = \min_{\mathbf{w},\theta} \Phi(\mathbf{w},\theta) \geq 0. \tag{26}$$

The basis exchange algorithms which are similar to linear programming allow to find the minimum of the criterion function $\Phi(\mathbf{w},\theta)$ efficiently even in cases of large, multidimensional data sets G^+ and G^- (22) [7].

It has been proved that the minimal value Φ^* $(0 \leq \Phi^* \leq 1)$ of the criterion function $\Phi(\mathbf{w},\theta)$ (25) is equal to zero $(\Phi^* = 0)$ if and only if the positive G_l^+ and the negative G_l^- sets (22) are linearly separable (4) [3]. The value Φ^* is near to one if the sets G_l^+ and G_l^- are completly mixed (overlapping) in a given fetaure space $F[n]$. The quantity Φ^* can be used as a measure of the linear nonseparabilty of the sets G_l^+ and G_l^- in the fetaure space $F[n]$ [9]. One of the most important properties of the measure Φ^* is their invariancy in respect to linear, nonsingular data transformations. It means, that the minimal value Φ^* of the function $\Phi(\mathbf{w},\theta)$ remains constant if the feature vectors $\mathbf{x}_j$ are transformed in an affine, nonsingular manner

$$(\forall \mathbf{x}_j \in \{G^+ \cup G_l^-\} \;\; \mathbf{y}_j = A\,\mathbf{x}_j + \mathbf{b}) \Rightarrow \Phi_x^* = \Phi_y^* \quad (\textit{invariancy property}) \tag{27}$$

where A is a nonsingular matrix (A^{-1} exists), Φ_y^* is the minimal value of the criterion function $\Phi_y(\mathbf{w},\theta)$ (25) defined on the vectors $\mathbf{y}_j$.

Another property of the measures Φ^* is their monotonocity. Let symbol S_k mean some subset $(S_k \subset S)$ of the feature set $S = \{x_1,\ldots,x_n\}$. The feature vectors $\mathbf{x}'_j$ in the feature subspace $F_k[n']$ $(F_k[n'] \subset F[n])$ are obtained from the vectors $\mathbf{x}_j$ $(\mathbf{x}_j \in F[n])$ by neglecting such features x_i which does not belong to the subset $S_k (x_i \notin S_k)$.

$$(S_k \subset S_{k'}) \Rightarrow \Phi^*(S_k) \geq \Phi^*(S_{k'}) \qquad (\textit{monotonocity property}) \tag{28}$$

The monotonocity property gives a jusification for using the *branch and bound* strategy in the feature selection [2].

7. Quadrifield Diagnostic Maps of the *Hepar* System

The *Hepar* computer system aggregates the clinical database with tools for data exploration and diagnosis support [8]. The database of the *Hepar* system contains hepathological data. An essential part of the system is the data visualization module. For the purpose of the data visualization linear transformations are used (6) from multidimensional feature space $F[n]$ on a plane are used. Such transformations allow for inducing the distance function $\delta_{\mathrm{I}}^2(\mathbf{x}_0,\mathbf{x}_{j(i)})$ (13) based both on the Euclidean distance $\delta_{\mathrm{E}}^2(\mathbf{y}_0,\mathbf{y}_{j(i)})$ (12) between transformed vectors $\mathbf{y}_0$ and $\mathbf{y}_{j(i)}$ as well as on a subjective measures of similarity of these vectors on the visualizing plane.

The parameters $\mathbf{w}^*$ and θ^* determining minimum (26) of the criterion function $\Phi(\mathbf{w},\theta)$ (25) are used in the *Hepar* system for definition of the affine transformation of the feature vectors $\mathbf{x}$ on a co-ordinate line y:

$$(\mathbf{x}_j \in \{G^+ \cup G^-\}) \quad y_j = (\mathbf{w}^*)^T \mathbf{x}_j - \theta^* \tag{29}$$

where $\mathbf{w}^* = [w^*_1,...,w^*_n]^T$ is the parameter vector which determines the direction of the line.

The transformation (29) has been applied in the *Hepar* system for definition of the visualizing planes [8]. This system allows for designing pairs of special visualizing transformations (29) which result in the so called *diagnostic maps*. Two transformations (29) defined by linearly independent vectors $\mathbf{w}_1^*$ and $\mathbf{w}_1^*$ give a possibility to produce such a visualizing plane (*diagnostic map*) which relatively well separates four groups of patients. The diagnostic maps are used for inducing the similarity measure between the feature vector of a new patient $\mathbf{x}_0$ and the vectors $\mathbf{x}_j(k)$ from the reference sets C_k (1).

The affine transformation of the feature vectors $\mathbf{x}_j$ (1) on a plane can be represented in the below manner

$$\mathbf{y}_j = [y_{j1},y_{j2}]^T = [(\mathbf{w}_1^*)^T\mathbf{x}_j - \theta_1^*, (\mathbf{w}_2^*)^T\mathbf{x}_j - \theta_2^*]^T \tag{30}$$

where $\mathbf{w}_i^* = [w_{i1},.......,w_{in}]^T$ $(i = 1,2)$ are the parameter vectors that span a plane.

The *maps* or, in other words, the *scatterplots* of data, can be generated as a result of visualization of the transformed points $\mathbf{y}_j(k)$ (30). If the vectors $\mathbf{w}_i^*$ are orthogonal $((\mathbf{w}_1^*)^T\mathbf{w}_2^* = 0)$ and have the unit length $((\mathbf{w}_1^*)^T \mathbf{w}_1^* = (\mathbf{w}_2^*)^T \mathbf{w}_2^* = 1)$ then the transformations (30) describes the *projection* of the feature vectors $\mathbf{x}_j(k)$ on the visualizing plane $P(\mathbf{w}_1^*,\mathbf{w}_2^*;\boldsymbol{\theta}^*)$, where

$$P(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta}) = \{\mathbf{x}\colon \mathbf{x} = \alpha_1\,\mathbf{w}_1 + \alpha_2\,\mathbf{w}_2 + \boldsymbol{\theta}, \quad where \quad \alpha_i \in R^1 \; and \; \boldsymbol{\theta} = [\theta_1,\theta_2]^T\} \tag{31}$$

EXAMPLE 1: Let us explain on an example the basic principles of the diagnostic map designing in the framework of the system *Hepar*: We have taken into consideration seven learning sets C_k (1) describing 814 patients from the *Hepar* database [8].

C_1 – *Cirrhosis hepatis* – 382 patients (32)
C_2 – *Hepatitis chronica* – 373 patients
C_3 – *Carcinoma* – 20 patients
C_4 – *H-biopsy negative* – 16 patients
C_5 – *Hepatitis acuta* – 9 patients

C_6 – *Hepatitis subacuta* – 9 patients

C_7 – *HBV-positive* – 5 patients

Patients from the sets C_k have been described by the feature vectors $\mathbf{x}_j(k)$ of dimensionality n equal to 40. The components x_i of the vectors $\mathbf{x}_j(k)$ were numerical results of various diagnostic examinations of a given patient. Numerical results of both laboratory tests ($x_i \in$ R) as well as patients symptoms ($x_i \in \{0,1\}$) have been taken as features x_i. The above sets C_k have been used in designing the diagnostic map (Fig. 3). This map resulted from the affine transformation (30) of the 40 — dimensional feature vectors $\mathbf{x}_j(k)$ on a visualizing plane.

The affine transformation (30) of the feature vectors $\mathbf{x}_j(k)$ on a visualizing plane is determined by two pairs of parameters $(\mathbf{w}_1^*, \theta_1^*)$ and $(\mathbf{w}_2^*, \theta_2^*)$. These parameters have been induced from the sets (32) through minimization of two perceptron criterion functions $\Phi_1(\mathbf{w},\theta)$ and $\Phi_2(\mathbf{w},\theta)$ (25). Each function $\Phi_k(\mathbf{w},\theta)$ ($k = 1, 2$) was defined by their own pair of the sets G_k^+ and G_k^- (22), where

$$G_1^+ = C_1 \cup C_3 \cup C_5 \cup C_6 \cup C_7 \ \textit{and}\ G_1^- = C_2 \cup C_4 \quad (33)$$

$$G_2^+ = C_2 \cup C_3 \cup C_5 \cup C_6 \cup C_7 \ \textit{and}\ G_2^- = C_1 \cup C_4 \quad (34)$$

If each pair of the sets G_k^+ and G_k^- ($k = 1, 2$) is linearly separable (4), then the transformation (30) defined by parameters $(\mathbf{w}_1^*, \theta_1^*)$ and $(\mathbf{w}_2^*, \theta_2^*)$ assures the exact

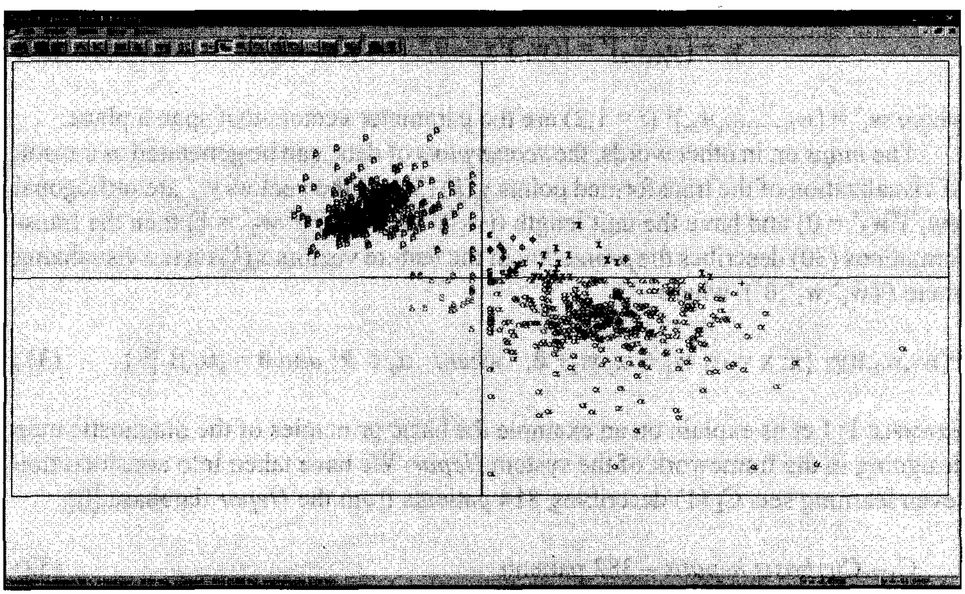

Fig. 3. The quadrifield diagnostic map with a below structure: *the upper-left quarter* – C_2, *the upper-right quarter* – $C_3 \cup C_5 \cup C_6 \cup C_7$, *the lower-right quarter* – C_1, *the lower-left quarter* – C_4
The horizontal axis y_1 of the map is defined by the transformation $y_1 = (\mathbf{w}_1^*)^T\mathbf{x} - \theta_1^*$ (30).
The vertical axis y_2 of the map is defined by the transformation $y_2 = (\mathbf{w}_2^*)^T\mathbf{x} - \theta_2^*$ (30)

placements of the learning sets C_k (32) in an adequate quarter of the diagnostic map as it can be seen on the Fig. 3.

The transformation (30) defines the coordinates $\mathbf{y}_j(k) = [y_{j1}(k), y_{j2}(k)]$ on the map of particular feature vectors $\mathbf{x}_j(k)$ from the reference sets C_k (32). The vector of $\mathbf{x}_0$ of a new patient can be located on the map as $\mathbf{y}_0$ by using the transformation (30). As a result, the map can be also used in the diagnosis support in accordance with the *CNR* or or the *K-NN* scheme (15), which is based on the Euclidean distances $\delta_E(\mathbf{y}_0,\mathbf{y}_j(k))$ (12) between the transformed (30) vectors $\mathbf{y}_0$ and $\mathbf{y}_j(k)$.

A quality of the diagnostic map from the Fig. 3 has been evaluated by using twice the same $K-NN$ rule (15) with $K = 10$ both on the original feature vectors $\mathbf{x}_j(k)$ as well as on the transformed vectors $\mathbf{y}_j(k)$ (30) from the visualizing plane. The $K-NN$ rule has been used for allocation of the feature vectors $\mathbf{x}_j(k)$ and the transformed vectors $\mathbf{y}_j(k)$ (30) to one of the four classes *A*, *B*, *C* or *D*: class *A* has been represented by the set C_2, class *B* by the set $C_3 \cup C_5 \cup C_6 \cup C_7$, class *C* by C_1, and class *D* by C_4 (Fig. 3). The crossvalidation method *leave-one-out* has been applied in evaluation of the allocation rule [3]. The results af this evaluation are summarised in the two contingency tables given below:

Table 1. Allocation of the feature vectors $x_j(k)$ by the $K-NN$ rule (15) with K = 10

	Allocation *A*	Allocation *B*	Allocation *C*	Allocation *D*	Success rate (%)
Class *A*	**2**	8	0	33	4.7
Class *B*	0	**353**	9	20	94.6
Class *C*	0	2	**7**	7	43.6
Class *D*	4	18	9	**357**	93.4
TOTAL					**88.3**

Table 2. Allocation of the transformed vectors $y_j(k)$ (30) on the map (Fig. 1) by the $K-NN$ rule (15) with K = 10

	Allocation *A*	Allocation *B*	Allocation *C*	Allocation *D*	Success rate (%)
Class *A*	**31**	2	1	9	72.1
Class *B*	0	**369**	1	3	98.9
Class *C*	0	3	**11**	2	68.8
Class *D*	6	2	3	**371**	97.1
TOTAL					**96.0**

The above results show that introducing of the diagnostic maps allowed for a few percent reduction of the error rate of the the $K-NN$ rule (15). The error reduction has been achieved despite of a significant reduction of dimensionality (from $n = 40$ to $n' = 2$).

8. Concluding Remarks

It has been shown on data sets that the classification accuracy of the *K-NN* rules can be improved through linear transformations of the learning sets (Table 1). Such transformations allow for modification (*induction*) of the similarity measure or distance function used in the *K-NN* rule or in the *Case Based Reasoning* scheme. Useful similarity measures can be induced from the learning data sets through separable linear transformations. Separable transformations of multivariate feature vectors on a visualizing plane have been used in the Hepar system for designing diagnostic maps (Fig. 3).

Separable linear transformations can be designed not only by the solutions of eignevalue problems used in the *Principal Components Analysis* or in the *Discriminant Analysis* [4], but also through minimization of the convex and piecewise linear (*CPL*) criterion functions. The *Perceptron* and the *Differential* criterion functions from the *CPL* family belong, among others, to the *CPL* family [9].

Functions from the *CPL* family give possibility for flexible modelling and solving many problems of the exploratory data analysis [9]. In particular, the feature selection problem can be solved through minimization of the *CPL* criterion functions. The basis exchange algorithms, which are similar to the linear programming, allow to find the minimum of the *CPL* criterion functions efficiently even in the case of large, multidimensional data sets [7].

Acknowledgments

This work was partially supported by grant W/II/1/2007 from the Białystok University of Technology, the KBN grant 3T11F01130 and by grant 16/St/2007 from the Institute of Biocybernetics and Biomedical Engineering Polish Academy of Sciences.

References

1. Perner P.: Data Mining on Multimedia Data, Springer, Berlin 2002.
2. Fukunaga K.: Statistical Pattern Recognition, Academic Press, Inc., San Diego 1990.
3. Duda O. R., Hart P. E.: Pattern Classification, Sec. Edition. J. Wiley, New York 2001.
4. Johnson R. A., Wichern D. W.: Applied Multivariate Statistical Analysis, Prentice-Hall, Inc., Englewood Cliffs, New York 1991.
5. Bobrowski L., Topczewska M.: Improving the K-NN classification with the Euclidean distance through linear data transformations, in: ICDM 2004, Eds. P. Perner et al., Lipsk Germany, Springer Verlag 2004, Springer Lecture Notes in Artificial Intelligence 3275, 22–32.
6. Bobrowski L., Topczewska M.: Linear visualising transformations and convex, piecewise linear criterion functions, Bioc. and Biom. Eng., 2002, 1, 22, 69–78.
7. Bobrowski L.: Design of piecewise linear classifiers from formal neurons by some basis exchange, Pattern Recognition, 1991, 24(9), 863–870.
8. Bobrowski L., Wasyluk H.: Diagnosis supporting rules of the Hepar system, in: MEDINFO 2001, Eds: V. L. Petel, R. Rogers, R. Haux, IOS Press, Amsterdam 2001, 1309–1313.
9. Bobrowski L.: Data mining based on convex and piecewise linear (CPL) criterion functions (in Polish), Technical University Białystok 2005.