Ranked modelling with feature selection based on the *CPL* criterion functions[1]

Leon Bobrowski[a,b]

[a]Faculty of Computer Science, Bialystok Technical University

[b]Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

Abstract: Ranked transformations should preserve a priori given ranked relations (order) between some feature vectors. Designing ranked models includes feature selection tasks. Components of feature vectors which are not important for preserving the vectors order should be neglected. This way unimportant dimensions are greatly reduced in the feature space. It is particularly important in the case of "long" feature vectors, when a relatively small number of objects is represented in a high dimensional feature space. In the paper, we describe designing ranked models with the feature selection which is based on the minimisation of convex and piecewise linear (*CPL*) functions.

Key words: ranked linear models, feature selection, convex and piecewise linear (CPL) criterion functions, linear separability of data sets

1 Introduction

Special tools for data exploration are based on a variety of methods including: mulitivariate data analysis [1], data mining [2], pattern recognition [3], fuzzy sets [5], rough sets [6], or machine learning [7].

Data exploration goals may include trends for extraction on the basis of a known order between selected objects represented as feature vectors in a data set. For example, we could know that some objects are older (more developed, more efficient, more expensive, ...) than any object from the first set and they are younger (less developed, less efficient, less expensive, ...) than any object from the second set. This kind of a priori information about the order relation between selected pairs of objects can be the basis for ranked model designing. We assume here the ranked model is such a linear transformation, which preserves in an satisfactory manner the a priori knowledge on a line in the form of the order relations between selected pairs of feature vectors. The process of ranked model designing can be seen as trend induction from data sets which is based on a priori information about the data ordering.

The procedure of the ranked models design which is based on the minimisation of the convex and piecewise linear (*CPL*) criterion functions is described in

[1] This work was partially supported by the W/II/1/2005from the Białystok University of Technology and by the 16/St/2005 grant from the Institute of Biocybernetics and Biomedical Engineering PAS.

the paper. These criterion functions are the sums of the positive and the negative *CPL* penalty functions which are defined through differences between the feature vectors constituting referencing dipoles [8]. This way, the task of the ranked model design can be linked to the problem of the linear separability of two sets in a given feature space. The enlargement of the criterion function by the feature cost functions allows one to include the feature selection into the procedure of designing ranked models [9].

2 Feature vectors and ranked relations

We are taking into consideration the data set *C* built from *m* feature vectors $\mathbf{x}_j$ with the fixed indexing *j*

$$C = \{\mathbf{x}_j\} \ \ (j = 1,.......,m) \tag{1}$$

The components (*features*) x_{ji} of the vector $\mathbf{x}_j = [x_{j1},......,x_{jn}]^T$ are numerical results of the *j*-th object O_j examinations ($i = 1,...,n$). The feature vectors $\mathbf{x}_j$ are often of a mixed type, because they represent different types of measurements (e.g. $x_i \in \{0,1\}$) or ($x_i \in R$)).

Let the symbol "$\prec$" mean the ranked relation "*follows*" which may be fulfilled between selected feature vectors $\mathbf{x}_j$ and $\mathbf{x}_k$:

$$\mathbf{x}_j \prec \mathbf{x}_k \Leftrightarrow \mathbf{x}_k \ \textit{follows} \ \mathbf{x}_j \tag{2}$$

The relation "$\prec$" between the feature vectors x_j and x_k means that the pair $\{x_j,x_k\}$ is *ranked*. The ranked relations between particular feature vectors x_j and x_k could result from additional information about the objects O_j and O_k.

Our aim is to design such a transformation of feature vectors x_j on the *ranked line* $y = w^T x$, which preserves the relation "$\prec$" (2) as precisely as possible

$$y_j = y_j(\mathbf{w}) = \mathbf{w}^T\mathbf{x}_j \tag{3}$$

where $\mathbf{w} = [w_1,......,w_n]^T$ is the vector of parameters.

The relation "$\prec$" (2) is preserved on the line (3) if and only if the following implication holds:

$$(\forall (j,k)) \quad \mathbf{x}_j \prec \mathbf{x}_k \Rightarrow y_j(\mathbf{w}) < y_k(\mathbf{w}) \tag{4}$$

The procedure of the ranked line design can be based on the concept of positively and negatively oriented dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$[8].

Definition 1: The ranked pair $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j<j'$) of the feature vectors $\mathbf{x}_j$ and $\mathbf{x}_{j'}$ constitutes the *positively oriented dipole* $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($\forall (j, j') \in I^+$) if and only if $\mathbf{x}_j \prec \mathbf{x}_{j'}$

$$(\forall \ (j,j') \in I^+) \quad \mathbf{x}_j \prec \mathbf{x}_{j'} \tag{5}$$

Definition 2: The ranked pair $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j<j'$) of the feature vectors $\mathbf{x}_j$ and $\mathbf{x}_{j'}$ constitutes the *negatively oriented dipole* $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ $(\forall(j, j') \in I^-)$, if and only if $\mathbf{x}_{j'} \prec \mathbf{x}_j$.

$$(\forall\ (j,j') \in I^-) \quad \mathbf{x}_{j'} \prec \mathbf{x}_j \tag{6}$$

Definition 3: The line $y(\mathbf{w}) = \mathbf{w}^T\mathbf{x}$ (3) is fully consistent (*ranked*) with the dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ orientations if and only if

$$\begin{aligned}&(\forall\ (j,j') \in I^+) \quad y_j(\mathbf{w}) < y_{j'}(\mathbf{w}) \qquad \textit{and}\\&(\forall\ (j,j') \in I^-) \quad y_j(\mathbf{w}) > y_{j'}(\mathbf{w})\end{aligned} \tag{7}$$

where I^+ and I^- are the sets of the positively and negatively oriented dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j<j'$).

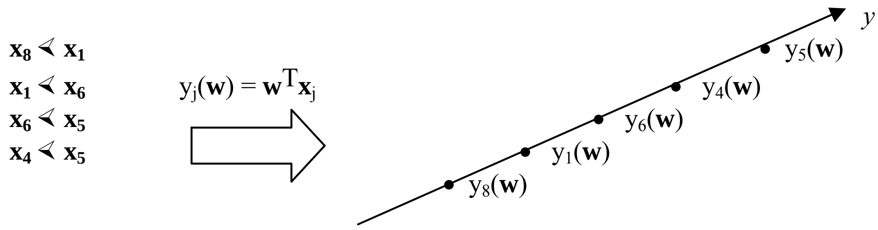

Fig. 1. An example of the order relations (2) and the ranked line (7), where $I^+ = \{(1,6), ((4,5)\}$ and $I^- = \{(1,8), (5,6)\}$

Let us introduce two sets C^+ and C^- of the differential vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ which are given by

$$\begin{aligned}C^+ &= \{\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)\text{: } (j,j') \in I^+\}\\C^- &= \{\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)\text{: } (j,j') \in I^-\}\end{aligned} \tag{8}$$

We will examine the possibility of the sets separation C^+ and C^- by the hyperplane $H(\mathbf{w})$, which passes through the origin $\mathbf{0}$ of the feature space:

$$H(\mathbf{w}) = \{\mathbf{x}\text{: } \mathbf{w}^T\mathbf{x} = 0\} \tag{9}$$

where $\mathbf{w} = [w_1,......,w_n]^T$ is the vector of parameters.

Definition 4: The sets C^+ and C^- (8) are linearly separable with the threshold equal to zero if and only if there exists such a parameter vector $\mathbf{w}^*$ that:

$$\begin{aligned}&(\forall\ (j,j') \in I^+) \quad (\mathbf{w}^*)^T\ \mathbf{r}_{jj'} > 0\\&(\forall\ (j,j') \in I^-) \quad (\mathbf{w}^*)^T\ \mathbf{r}_{jj'} < 0\end{aligned} \tag{10}$$

The above inequalities can be represented in the following manner:

$$\begin{aligned}(\exists \mathbf{w}^*)\ &(\forall\ (j,j') \in I^+) \quad (\mathbf{w}^*)^T\ \mathbf{r}_{jj'} \geq\ 1\\&(\forall\ (j,j') \in I^-) \quad (\mathbf{w}^*)^T\mathbf{r}_{jj'} \leq -1\end{aligned} \tag{11}$$

Remark 1: If the parameter vector $\mathbf{w}^*$ linearly separates (11) the sets C^+ and C^- (8), then the line $y_j(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}_j$ is fully consistent (7) with the dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ orientation.

3 *CPL* criterion functions

`Designing the separating hyperplane H(**w**) could be carried out through the minimisation of the convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w})$ similar to the perceptron criterion function [2]. Let us introduce for this purpose the positive $\varphi_{jj'}^{+}(\mathbf{w})$ and negative $\varphi_{jj'}^{-}(\mathbf{w})$ penalty functions (Fig.2)

$$(\forall\ (j,j') \in I^+)\qquad \varphi_{jj'}^{+}(\mathbf{w}) = \begin{cases} 1 - \mathbf{w}^T\mathbf{r}_{jj'} & \textbf{\textit{if}}\ \mathbf{w}^T\mathbf{r}_{jj'} < 1 \\ 0 & \textbf{\textit{if}}\ \mathbf{w}^T\mathbf{r}_{jj'} \geq 1 \end{cases} \qquad (12)$$

$$\text{and}\ (\forall\ (j,j') \in I^-)\qquad \varphi_{jj'}^{-}(\mathbf{w}) = \begin{cases} 1 + \mathbf{w}^T\mathbf{r}_{jj'} & \textbf{\textit{if}}\ \mathbf{w}^T\mathbf{r}_{jj'} > -1 \\ 0 & \textbf{\textit{if}}\ \mathbf{w}^T\mathbf{r}_{jj'} \leq -1 \end{cases} \qquad (13)$$

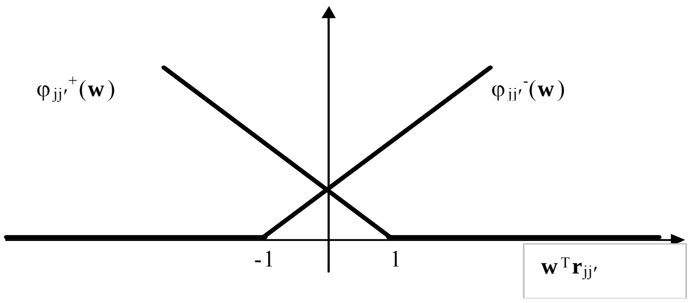

Fig. 2. The penalty functions $\varphi_{jj'}^{+}(\mathbf{w})$ (12) and $\varphi_{jj'}^{-}(\mathbf{w})$ (13).

The criterion function $\Phi(\mathbf{w})$ is the weighted sum of the above penalty functions

$$\Phi(\mathbf{w}) = \sum_{(j,j')\in I^+} \gamma_{jj'}\, 1_{jj'}^{+}(\mathbf{w}) + \sum_{(j,j')\in I^-} \gamma_{jj'}\, 1_{jj'}^{-}(\mathbf{w}) \qquad (14)$$

where $\gamma_{jj'}$ ($\gamma_{jj'} \geq 0$) is a nonnegative parameter (*price*) related to the dipole $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j<j'$).

The criterion function $\Phi(w)$ (14) is the convex and piecewise linear (*CPL*) function as the sum of such type of the penalty functions $\phi_{jj'}^{+}(w)$ (12) and $\phi_{jj'}^{-}(w)$ (13). The basis exchange algorithms, similar to linear programming, allow one to find a minimum of such functions efficiently, even in the case of large, multidimensional data sets C^{+} and C^{-} [10]:

$$\Phi^{*} = \Phi(\mathbf{w}^{*}) \; = min\; \Phi(\mathbf{w}) \geq 0 \qquad \mathbf{(15)}$$

The optimal parameter vector $\mathbf{w}^{*}$ and the minimal value Φ^{*} of the criterion function $\Phi(\mathbf{w})$ (11) can be applied to a variety of data ranking problems. In particular, the vector $\mathbf{w}^{*}$ defining the best ranked line $y = (\mathbf{w}^{*})^{T}\mathbf{x}$ (3) can be found this way.

Lemma 1: The minimal value Φ^{*} (15) of the criterion function $\Phi(\mathbf{w})$ (14) is nonnegative and equal to zero if and only if there exists such a vector $\mathbf{w}$ that the ranking of the points $y_j(\mathbf{w})$ on the line (3) are fully consistent (*Def.* 3) with the relations "$\prec$" (4).

Prove: The function $\Phi(\mathbf{w})$ (14) is nonnegative as the sum of the nonnegative components $\varphi_{jj'}^{+}(\mathbf{w})$ (12) and $\varphi_{jj'}^{-}(\mathbf{w})$ (13). If there exists such a vector $\mathbf{w}^{*}$ that the ranking of the points $y_j(\mathbf{w}^{*})$ on the line (3) is fully consistent (*Def.* 3) with the relations "$\prec$" (4), then the sets C^{+} and C^{-} (8) can be separated (10) by the hyperplane $H(\mathbf{w}^{*})$ (9). In this case, the minimal value of the perceptron criterion function $\Phi(\mathbf{w})$ (14) is equal to zero as it results from pattern recognition theory [2]. On the other hand, if the minimal value of the criterion function $\Phi(\mathbf{w})$ (14) is equal to zero in the point $\mathbf{w}^{*}$, then the values $\phi_{jj'}^{+}(\mathbf{w}^{*})$ and $\phi_{jj'}^{-}(\mathbf{w}^{*})$ of all the penalty functions $1_{jj'}^{+}(\mathbf{w})$ (12) and $1_{jj'}^{-}(\mathbf{w})$ (13) have to be equal to zero. It means that the sets C^{+} and C^{-} (8) can be separated (6) by the hyperplane $H(\mathbf{w}^{*})$ (9). As the result, the ranking of the points $y_j(\mathbf{w}^{*})$ on the line (3) is fully consistent (*Def.* 3) with the relations "$\prec$" (4). □

Let us introduce the below hyperplanes $h^{+}_{jj'}$ and $h^{-}_{jj'}$ defined in the parameter space by the difference vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ $(j<j')$

$$(\forall\, (j,j') \in I^{+}) \quad h^{+}_{jj'} = \{\mathbf{w}: (\mathbf{r}_{jj'})^{T}\, \mathbf{w} = 1\}$$
$$(\forall\, (j,j') \in I^{-}) \quad h^{-}_{jj'} = \{\mathbf{w}: (\mathbf{r}_{jj'})^{T}\, \mathbf{w} = -1\} \qquad \mathbf{(16)}$$

Definition 5: The parameter vector $\mathbf{w}$ is situated on the *positive side* of the hyperplane $h^{+}_{jj'}$ if the inequality $(\mathbf{w})^{T}\mathbf{r}_{jj'} \geq 1$ is fulfilled. Similarly, the parameter vector $\mathbf{w}$ is situated on the positive side of the hyperplane $h^{-}_{jj'}$ if the inequality $(\mathbf{w})^{T}\mathbf{r}_{jj'} \leq -1$ holds.

The penalty functions $1_{jj'}^{+}(\mathbf{w})$ (12) and $1_{jj'}^{-}(\mathbf{w})$ (13) are equal to zero if and only if the parameter vector $\mathbf{w}$ is situated on the positive side of the hyperplanes $h^{+}_{jj'}$ and $h^{-}_{jj'}$ (16). The minimal value $\Phi^{*} = \Phi(\mathbf{w}^{*})$ (15) of the criterion function $\Phi(\mathbf{w})$ (14) is equal to zero if the optimal parameter vector $\mathbf{w}^{*}$ is situated on the positive side of all hyperplanes $h^{+}_{jj'}$ and $h^{-}_{jj'}$. Such a solution $\mathbf{w}^{*}$ $(\Phi(\mathbf{w}^{*}) = 0)$ exists if the sets C^{+} and C^{-} (8) are linearly separable (10).

Remark 2: Linear independence of the vectors $\mathbf{r}_{jj'}$ constituting the sets C^{+} and C^{-} (8) is the sufficient condition for the linear separability (10) of these sets [9].

4 Modified crterion function with feature costs

The criterion function $\Phi(\mathbf{w})$ (14) can be modified by introducing the cost function $\phi_i(\mathbf{w})$ (Fig. 3) for each feature x_i in order to search for the best feature subspace $F_I^*[m]$ [9].

$$\phi_i(\mathbf{w}) = \begin{cases} -(\mathbf{e}_i)^T\mathbf{w} & \textit{if} \quad (\mathbf{e}_i)^T\,\mathbf{w} < 0 \\ (\mathbf{e}_i)^T\mathbf{w} & \textit{if} \quad (\mathbf{e}_i)^T\,\mathbf{w} \geq 0 \end{cases} \qquad \textbf{(17)}$$

where $\mathbf{e}_i = [0,...,0,1,0,...,0]^T$ are the unit vectors (i=1,.....,n)

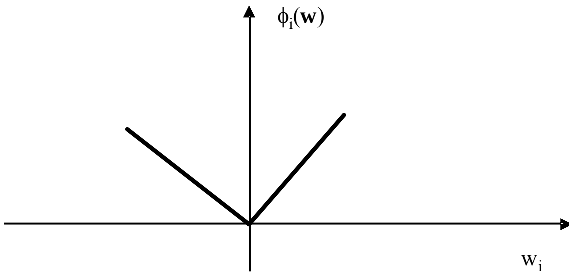

Fig. 3. The cost function $\phi_i(\mathbf{w})$ (17)

The modified criterion function $\Psi_\lambda(\mathbf{w})$ can be given in the following form [9]:

$$\Psi_\lambda(\mathbf{w}) = \Phi(\mathbf{w}) + \lambda \sum_{i \in I} \gamma_i\, \phi_i(\mathbf{w}) \qquad \textbf{(18)}$$

where $\Phi(\mathbf{w})$ is given by (14), $\lambda \geq 0$, $\gamma_i > 0$, and $I = \{1,....,n\}$.

The function $\Psi_\lambda(\mathbf{w})$ is the sum of the perceptron criterion function $\Phi(\mathbf{w})$ (14) and the cost functions $\phi_i(\mathbf{w})$ (17) multiplied by the positive parameters γ_i. The parameters γ_i. represent the *costs* of particular features x_i. These costs γ_i can be chosen a priori, according to our preferences.

The criterion function $\Psi_\lambda(\mathbf{w})$ (18) is the convex and piecewise linear (*CPL*) function as the sum of the *CPL* functions $\Phi(\mathbf{w})$ (14) and $\lambda\ \gamma_i\ \phi_i(\mathbf{w})$ (18). Like previously in (15), we are taking into account the point $\mathbf{w}_\lambda^*$ constituting the minimal value of the criterion function $\Psi_\lambda(\mathbf{w})$:

$$\Psi_\lambda^* = \Psi_\lambda(\mathbf{w}_\lambda^*) = \min_{\mathbf{w}} \Psi_\lambda(\mathbf{w}) \qquad \textbf{(19)}$$

The basis exchange algorithms allow one to solve efficiently also this minimisation problem [10].

The below hyperplanes h_i in the feature space can be linked to the cost functions $\phi_i(\mathbf{w})$ (17)

$$(\forall i \in I = \{1,....,n\}) \quad h_i = \{\mathbf{w}: (\mathbf{e}_i)^T \mathbf{w} = 0\} \tag{20}$$

The cost function $\phi_i(\mathbf{w})$ (18) is equal to zero if the point $\mathbf{w}$ is situated on the hyperplane h_i (20).

The k-th *basis* $\mathbf{B}_k[n]$ of the n-dimensional feature space $F[n]$ can be constituted by any set $S_k[n]$ of n linearly independent vectors $\mathbf{r}_{jj'}$ ($(j,j') \in I^+ \cup I^-$) (16)) and $\mathbf{e}_i$ ($i \in I = \{1,....,n\}$). The basis $\mathbf{B}_k[n]$ is the nonsingular matrix with the n rows $\mathbf{b}_l$ constituted by the vectors $\mathbf{r}_{jj'}$ or $\mathbf{e}_i$.

$$\mathbf{B}_k^T[n] = [\mathbf{b}_1,......, \mathbf{b}_n] \tag{21}$$

where

$$\begin{aligned} &\mathbf{b}_l = \mathbf{r}_{jj'} \quad \textit{if} \quad \text{the vector } \mathbf{r}_{jj'} \text{ constitutes the } l\text{- th row of the matrix } \mathbf{B}_k[n] \\ &\mathbf{b}_l = \mathbf{e}_i \quad \textit{if} \quad \text{the vector } \mathbf{e}_i \text{ constitutes the } l\text{- th row of the matrix } \mathbf{B}_k[n] \end{aligned} \tag{22}$$

The basis $\mathbf{B}_k[n]$ defines the point (the *vertex*) $\mathbf{w}_k[n]$ in the feature space in accordance with the below equation

$$\mathbf{w}_k[n] = \mathbf{B}_k^{-1}[n] \, \mathbf{c}_k[n] \tag{23}$$

where $\mathbf{c}_k[n] = [c_1,......,c_n]^T$ is the *margin* vector with the components c_l defined by the following conditions

$$\begin{aligned} c_l &= 1 \quad \textit{if} \quad \mathbf{r}_{jj'} \ ((j,j') \in I^+) \text{ (8) constitutes the } l\text{-th row of the matrix } \mathbf{B}_k[n] \\ c_l &= -1 \quad \textit{if} \quad \mathbf{r}_{jj'} \ ((j,j') \in I^-) \text{ (8) constitutes the } l\text{-th row of the matrix } \mathbf{B}_k[n] \\ c_l &= 0 \quad \textit{if} \quad \text{the unit vector } \mathbf{e}_i \text{ constitutes the } l\text{-th row of the matrix } \mathbf{B}_k[n] \end{aligned} \tag{24}$$

It could be seen that the vertex $\mathbf{w}_k[n]$ is the point of intersection of n hyperplanes $h^+_{jj'}$ and $h^-_{jj'}$ (16) or h_i (20) in accordance with the conditions of (25).

It can be proved by applying results of linear programming theory [2], that the global minimum (19) of the criterion function $\Psi_\lambda(\mathbf{w})$ (18) can be found in one of the vertices $\mathbf{w}_k[n]$.

$$(\exists \mathbf{w}_k^*[n]) \ (\forall \mathbf{w}) \quad \Psi_\lambda(\mathbf{w}) \geq \Psi_\lambda(\mathbf{w}_k^*[n]) \tag{25}$$

The optimal vertex $\mathbf{w}_k^*[n]$ and the related basis, the basis $\mathbf{B}_k^*[n]$, can be used in the feature selection problem.

5 Feature selection for the ranked models

The optimal vertex $\mathbf{w}_k^*[n] = [w_1^*,.....,w_n^*]^T$ (25) related to the basis $\mathbf{B}_k^*[n]$ (23) defines the ranked model (3) in the n-dimensional feature space $F[n]$

$$y_j = (\mathbf{w}_k^*[n])^T \mathbf{x}_j[n] \quad \textbf{(26)}$$

Remark 3: If the unit vector $\mathbf{e}_i$ constitutes the *l*-th row (22) of the optimal basis $\mathbf{B}_k^*[n]$, then the *i*-th feature x_i can be omitted from the feature vectors $\mathbf{x}_j$ without the changing of the order of the points y_j on the line (26).

In order to justify the above statement let us remark, that the unit vector $\mathbf{e}_i$ in the basis $\mathbf{B}_k^*[n]$, means that the *i*-th component w_i^* of the weight vector $\mathbf{w}_k^*[n]$ is equal to zero. The feature x_i related to the weight w_i^* equal to zero can be omitted without the changing of the inner products (26) value.

Remark 4: If the number *m* of the linearly independent vectors $\mathbf{r}_{jj'}[n] = (\mathbf{x}_{j'}[n] - \mathbf{x}_j[n])$ $((j, j') \in I^+ \cup I^-)$ (8) is less than the dimension *n* of the feature space F[*n*] ($m < n$), then at least $n - m$ features x_i can be omitted from the vectors $\mathbf{x}_j[n]$ without changing the points y_j order on the line (26).

Neglecting the features x_i related to the unit vectors $\mathbf{e}_i$ in the basis $\mathbf{B}^*_k[n]$ of the optimal vertex $\mathbf{w}_k^*[n]$ (26) is linked to the reduction of the feature space F[*n*] dimension *n*. The reduced basis $\mathbf{B}_k^*[n']$ contains only differential vectors $\mathbf{r}_{jj'}[n'] = (\mathbf{x}_{j'}[n'] - \mathbf{x}_j[n'])$ from the feature subspace $F_l[n']$ of dimension n'.

It could be seen, that the vectors $\mathbf{r}_{jj'}[n']$ constituting the basis $\mathbf{B}_k^*[n']$ are linearly separable (11). In result, if the all vectors $\mathbf{r}_{jj'}[n']$ from the sets C^+ and C^- (8) are used in the optimal basis $\mathbf{B}_k^*[n']$, then these sets are linearly separable (10).

In the case of *m* linearly independent, "long" vectors $\mathbf{r}_{jj'}[n]$ ($n >> m$) there can exist many feature subspaces $F_k[m]$ of dimension *m*, which assure the linear separability (11) of the sets C^+ and C^- (8) formed by the vectors $\mathbf{r}_{jj'}[n]$. The minimisation (25) of the criterion function $\Psi_\lambda(\mathbf{w})$ (18) with a small, positive values of the parameter λ $(\forall\lambda\in(0,\lambda^+))$ allows one to find the optimal feature subspace $F_l^*[m]$. It can be proved, that the minimal value $\Psi_\lambda(\mathbf{w}_k^*[m])$ (26) of the criterion function $\Psi_\lambda(\mathbf{w})$ (18) could be expressed in the below manner [11]:

$$(\forall\lambda\in [0, \lambda^+]) \quad \Psi_\lambda(\mathbf{w}_k^*[m]) = \lambda \sum_{i\in I^*_l[m]} \gamma_i \mid w_i^*| \quad \textbf{(27)}$$

where w_i* are the components of the optimal, *m*-dimensional vertex $\mathbf{w}_k^*[m]$) (26) and $I_l^*[m]$ is the set of the indices *i* of such features x_i which are included in this vertex. All included features x_i have the weights w_i^* greater than zero
$((\forall i\in I_l^*[m])\ w_i^* > 0)$.

If the costs γ_i are equal to one, then the minimal value $\Psi_\lambda(\mathbf{w}_k^*[m])$ (27) of the function $\Psi_\lambda(\mathbf{w})$ (18) can be expressed as:

$$\Psi_\lambda^* = \Psi_\lambda(\mathbf{w}_k^*[m]) = \lambda \sum_{i\in I_l} \mid w_i^*| = \lambda \, ||\mathbf{w}_k^*[m]||_{L1} \quad \textbf{(28)}$$

where $\| \mathbf{w}_k^*[m] \|_{L1}$ is the L_1 norm of the vector $\mathbf{w}_k^*[m]$.

In the case of such sets C^+ and C^- (8) which are linearly separable (10), the minimisation problem (19) with the function $\Psi_\lambda(\mathbf{w})$ (18) could by solved by using the following formulation [9]

$$\min_{\mathbf{w}} \{\|\mathbf{w}\|_{L1}: \ \mathbf{w} \text{ separates linearly (11) the sets } C^+ \text{ and } C^- \text{ (8)}\} \qquad \textbf{(29)}$$

The above formulation is similar to those used in the Support Vector Machines (*SVM*) method [12]. One of the important differences is such that the *SVM* method is based on the Euclidean norm $\| \mathbf{w} \|_{L2}$, where

$$\| \mathbf{w} \|_{L2} = (\mathbf{w}^T \mathbf{w})^{1/2} \qquad \textbf{(30)}$$

The similarity of the expression (29) to the *SVM* approach allows one to explain in a better manner properties of the optimal vector $\mathbf{w}_k^*[m]$ which constitutes solution of the problem (29).

An efficient algorithm of the feature subspaces $F_l[m]$ exchange has been developed in order to find the optimal subspace $F_k^*[m]$ or solve the problem (29) through computations in the *m*-dimensional parameter spaces $F_k[m]$ instead of the initial, high dimensional feature space $F[n]$ [11].

The optimal vertex $\mathbf{w}_k^*[m]$ (25) related to the basis $\mathbf{B}_k^*[m]$ defines the ranked model $y = (\mathbf{w}_k^*[m])^T\mathbf{x}[m]$ (26) in the *m*-dimensional feature subspace $F_l^*[m]$. Such ranked model allows one to put new objects $\mathbf{x}[m]$ on the ranked (trend) line (3) and provides additional information concerning features x_i ($i \in I_l^*[m]$) which are the most important for preserving the discovered trend.

6 Concluding remarks

The concept of ranked linear transformations (2) of the feature space *X* on the line is examined in the paper. Such lines reflect (3), to a possible extent, the relations "$\prec$" (4) between the feature vectors $\mathbf{x}_j$ in the selected pairs $\{\mathbf{x}_j, \mathbf{x}_{j'}\}$ ($(j,j') \in I^+$) or $(j,j') \in I^-$). It has been shown that the ranked linear transformations (2) are linked to the concept of linear separability of some data sets.

Designing ranked linear transformations (2) can be based on minimisation of the convex and piecewise linear (*CPL*) criterion function $\Psi_\lambda(\mathbf{w})$ (18). The basis exchange algorithms, similar to linear programming, allow one to find the minimum of this function [10].

Designing ranked linear transformations allows for sequencing the feature vectors $\mathbf{x}_j$ in a variety of manners, depending on the choice of sets I^+ and I^- (8) of oriented dipoles $\{\mathbf{x}_j, \mathbf{x}_{j'}\}$.Such approach allows for the experimental verification of different sequencing models. The models could be defined on the basis of the selected dipoles sets I^+ and I^- (8). Next, such a model could be verified on the basis of the dipoles from the testing sets and used as a tool for sequencing new feature vectors $\mathbf{x}_j$.

The feature selection approach could indicate which features x_i are the most important in the ranked model.

The ranked linear transformations may have many applications. One of the most interesting applications could be the sequencing of genomic data or phylogenetic classification [13]. We are using a similar approach in designing tools for medical diagnosis support in the system *Hepar* [14].

Bibliography

1. Johnson R. A., Wichern D. W. *Applied Multivariate Statistical Analysis*, Prentice- Hall, Inc., Englewood Cliffs, New York, 1991
2. O. R. Duda and P. E. Hart, D. G. Stork: *Pattern Classification,* J. Wiley, New York, 2001
3. K. Fukunaga: *Statistical Pattern Recognition*, Academic Press, Inc., San Diego, 1990.
4. Fayyad U. M., Piatetsky-Shapiro G., Smyth P., Uthursamy R. (Eds.): *Advances in Knowledge Discovery and Data Mining*, AAAI Press The MIT Press, London (1996)
5. Zadeh L.A., Kacprzyk J. (Eds.), *Fuzzy Logic for the Management of Uncertainty*, John Wiley & Sons, New York, 1992
6. Pawlak Z. *Rough Sets - Theoretical Aspects of Reasoning about Data* Kluwer Academic Publishers, Dordrecht 1991
7. Michalski R. S., Bratko I., Kubat M., *Machine Learning and Data Mining*, J. Wiley, New York 1998
8. 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, Wienna, Austria
9. Bobrowski L., Łukaszuk T., "Selection of the linearly separable feature subsets", pp. 544-549 in: *Artificial Intelligence and Soft Computing - ICAISC 2004*, Eds. L. Rutkowski et al., Springer Verlag 2004, *Springer Lecture Notes in Artificial Intelligence 3070*
10. L. Bobrowski and W.Niemiro, "A method of synthesis of linear discriminant function in the case of nonseparabilty". *Pattern Recognition* **17,** pp.205-210,1984
11. Bobrowski L.: *Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych* (*Data mining based on convex and piecewise linear (CPL) criterion functions*) (in Polish), Technical University Białystok, 2005
12. Vapnik V. N. *Statistical Learning Theory*, J. Wiley, New York 1998
13. Bichindaritz I., Potter S., Knowledge Based Phylogenetic Classification Mining, pp. 163-172 in: *ICDM* 2004, Eds. P. Permer et al., Lipsk, Germany, Springer Verlag 2004, *Springer Lecture Notes in Artificial Intelligence 3275*
14. Bobrowski L., Wasyluk H. "Diagnosis supporting rules of the *Hepar* system", pp. 1309 – 1313 in: *MEDINFO 2001*, Proceedings of the 10-th World Congress on Medical Informatics, Ed. by V. L. Patel, R. Rogers, R Haux, IMIA, IOS Press Amsterdam 2001