RANKED MODELING ON TIME-DATA WITH INEXACT TIMING[1]

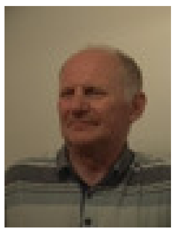
Leon Bobrowski[1,2], **Tomasz Łukaszuk**[1]

[1]Faculty of Computer Science, Bialystok Technical University
15-351 Bialystok, ul. Wiejska 45A
[2]Institute of Biocybernetics and Biomedical Engineering, PAS
02-109 Warsaw, ul. Ks. Trojdena 4, Poland
leon@ibib.waw.pl

Abstract

Ranked linear models can be designed on the basis of time-dependent data with inexact timing. Such data has the form of a sequence of multivariate feature vectors representing subsequent states of dynamical objects or development of events in time. The ordering relation between selected pairs of feature vectors (e.g. "a given feature vector appeared later than another one") is defined on the basis of an observed sequence. The same relation can be designated for a variety of observed sequences, even when the moment of appearance of a given vector is not defined precisely. A set of ordering relations between selected feature vectors allows for designing a ranked linear model. The ranked model has the form of linear transformation of multidimensional feature vectors into points on a line which preserves a set of ordering relations in the best possible manner. The ranked regression models can be designed by means of minimization of the convex and piecewise linear (*CPL*) criterion functions defined on differences of such feature vectors that are related by an ordering relation. The linear ranked model reflects all ranking relations between feature vectors if and only if two sets of the positive and negative differences between these vectors are linearly separable. This way the problem of ranked modeling can be transferred into the problem of linear separability of two sets of feature vectors. Ranked regression models can have many applications. Among others, the problem of different dynamical systems comparison can be addressed in this way.

Keywords: temporal data, inexact timing, ranked relations, convex and piecewise linear (CPL) criterion functions, linear separability.



Leon Bobrowski: Research interests include data mining, pattern recognition, neural networks, and medical diagnosis support. The main results concern basis exchange algorithms, designing of hierarchical neural networks, multivariate decision trees and visualizing transformations based on convex and piecewise linear functions.
Prof. Bobrowski is currently the Dean of the Faculty of Computer Science, Bialystok University of Technology and, additionally, works in the Laboratory of Biomedical Data Analysis at the Institute of Biocybernetics and Biomedical Engineering PAS in Warsaw.

[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.*

1. Introduction

Properties of dynamical objects or events can be explored, among others, using methods of multivariate statistical analysis [1] or pattern recognition [2]. Such properties can be examined on the basis of a sequence of multidimensional feature vectors, representing results of measurements of a given object or event. A feature vector can be treated as a "picture" of the explored object or event in a given time. But in some cases the exact timing of such a picture is unknown. Instead it is only possible to know that a given feature vector had appeared earlier than another one. The ordering relations inside pairs of the feature vectors can be defined on the basis of that.

A set of ordering relations between selected feature vectors allows for designing a ranked linear model. A ranked model has the form of a linear transformation of multidimensional feature vectors on points on a line which preserves a given set of ordering relations in the best possible manner.

Designing ranked linear transformations can be carried out through minimizing the convex and piecewise linear (*CPL*) criterion functions [3]. These *CPL* criterion functions are defined by differences of such feature vectors which are related by an ordering relation.

The linear ranked model reflects all ranking relations between feature vectors if and only if two sets of the positive and negative differences between these vectors are linearly separable. This way the problem of ranked modelling has been transformed into the problem of linear separability of two sets of feature vectors.

Exploration of the linear separability also allows to address the feature selection problem and the problem of enlarging linear model with nonlinear components. In this paper particular attention is paid to the assessment of systems' linearity on the basis of time dependent data with inexact timing.

2. Time dependent data with inexact timing

Let us take into consideration a dynamical object or event $O(t)$, where symbol $O(t)$ means that properties of this object can change during the time t. The state of the object $O(t)$ at the time t_j (*picture time*) is represented as the n-dimensional feature vector $\mathbf{x}(t_j)$:

$$(\forall j \in \{1,\ldots, m\}) \quad \mathbf{x}(t_j) = \mathbf{x}_j = [x_{j1},\ldots\ldots,x_{jn}]^T \tag{1}$$

The i-th component x_{ji} of the vector $\mathbf{x}_j$ is the numerical result of the i-th measurement or observation of the explored object $O(t)$ at the time t_j. We assume here that the indexing j of the feature vectors $\mathbf{x}_j$ is consistent with the observation time t_j:

$$(\forall j \in \{1,\ldots,m\text{-}1\}) \quad j < k \Rightarrow t_j \leq t_k \tag{2}$$

The feature vectors $\mathbf{x}_j$ are ordered in accordance with the below ranked relation "$\prec$" on the basis of picture times t_j and the *margins* δ_j ($\delta_j > 0$):

$$(\forall j \in \{1,\ldots, m\text{-}1\}) \quad t_j + \delta_j \leq t_k \Rightarrow \mathbf{x}_j \prec \mathbf{x}_k \tag{3}$$

We have assumed here that the relation "$\prec$" was defined at the beginning despite the fact that the times t_j are not exactly known. For example inexact timing t_j occurs when particular measurements x_{ji} which are a part of the j-th picture of the explored object $O(t)$ were made at slightly different times t_j. The margin δ has been introduced in order to exclude times t_j and t_k which are too close to each other and could result in a false relation "$\prec$".

The ranked relation "*is older than*" is fulfilled inside pairs of the feature vectors $\mathbf{x}_j$ and $\mathbf{x}_k$ with the indices (j, k) from the set J_p:

$$(\forall (j, k) \in J_p) \quad j < k \quad \textit{and} \quad (\mathbf{x}_j \prec \mathbf{x}_k) \Leftrightarrow (\mathbf{x}_k \ \textit{is older than} \ \mathbf{x}_j) \tag{4}$$

Let us remark that the set J_p can contain only a part of the indices (j, k) fulfilling the relation $t_j + \delta \leq t_k$ (3). The problem is how to design a linear transformation $y = \mathbf{w}^T\mathbf{x}$ which preserves the relation "$\prec$" (3) for all or almost all pairs of indices (j,k) from the set J_p (4). Similar problem has been analysed in the paper [3].

The set C of differential vectors $\mathbf{r}_{jk} = (\mathbf{x}_k - \mathbf{x}_j)$ is built from the feature vectors $\mathbf{x}_j$ and $\mathbf{x}_k$ with the indices (j, k) from the set J_p (4):

$$C = \{\mathbf{r}_{jk} = (\mathbf{x}_k - \mathbf{x}_j): (j,k) \in J_p\} \tag{5}$$

Definition 1: The line $y(\mathbf{w}) = \mathbf{w}^T\mathbf{x}$, where $\mathbf{w} = [w_1,\ldots\ldots,w_n]^T$, is *ranked* in respect to the set J_p if and only if

$$(\forall (j, k) \in J_p) \quad \mathbf{w}^T\mathbf{x}_j < \mathbf{w}^T\mathbf{x}_k \tag{6}$$

Let us consider a hyperplane $H(\mathbf{w})$ in the feature space that passes through the point **0**:

$$H(\mathbf{w}) = \{\mathbf{x}: \mathbf{w}^T\mathbf{x} = 0\} \tag{7}$$

Definition 2: The differential set C (5) can be situated on the positive side of the hyperplane $H(\mathbf{w})$ (7) if and only if the below inequalities hold

$$(\exists \mathbf{w}) \quad (\forall (j,k) \in J_p) \quad \mathbf{w}^T \mathbf{r}_{jk} > 0 \tag{8}$$

Remark 1: The line $y(\mathbf{w}) = \mathbf{w}^T\mathbf{x}$ is *ranked* (6) in accordance with all the ordering relations "$\prec$" from the set J_p (4) if and only if the differential set C (5) is situated on the positive side of the hyperplane $H(\mathbf{w})$ (7)

The above *Remark* results directly from the *Definition* 1 and the *Definition* 2.

3. Convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w})$

The ranked line $y(\mathbf{w}) = \mathbf{w}^T\mathbf{x}$ can be designed on the basis of differential vectors $\mathbf{r}_{jk} = (\mathbf{x}_k - \mathbf{x}_j)$ from the set C (5) through minimisation of the convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w})$ [3]:

$$\Phi(\mathbf{w}) = \sum_{(j,k)\in J_p} \gamma_{jk}\, \varphi_{jk}(\mathbf{w}) \tag{9}$$

where γ_{jk} ($\gamma_{jk} > 0$) is a positive parameter (*price*) related to the pair $\{\mathbf{x}_j, \mathbf{x}_k\}$ ($j < k$) and $\varphi_{jk}(\mathbf{w})$ is the below penalty function:

$(\forall (j, k) \in J_p)$

$$\varphi_{jk}(\mathbf{w}) = \begin{cases} 1 - (\mathbf{r}_{jk})^T \mathbf{w} & \text{if } (\mathbf{r}_{jk})^T \mathbf{w} < 1 \\ 0 & \text{if } (\mathbf{r}_{jk})^T \mathbf{w} \geq 1 \end{cases} \tag{10}$$

The basis exchange algorithms, similar to the linear programming, allow to find the minimum of the function $\Phi(\mathbf{w})$ (9) efficiently, even in the case of large data set C (5) [4], [5]:

$$\Phi^* = \Phi(\mathbf{w}^*) = \min_{\mathbf{w}} \Phi(\mathbf{w}) \geq 0 \tag{11}$$

The optimal parameter vector $\mathbf{w}^*$ and the minimal value Φ^* of the criterion function $\Phi(\mathbf{w})$ (9) can be applied to a variety of ranking modelling problems. In particular, the best ranked line $y = (\mathbf{w}^*)^T\mathbf{x}$ can be found in this way. The below Lemma can be proved:

Lemma 1: The minimal value $\Phi(\mathbf{w}^*)$ (11) is equal to zero if and only all the inequalities (8) are fulfilled on the line $y(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}$.

As a consequence (*Remark* 1), the line $y(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}$. is *fully ranked* (6) in accordance with all the ordering relations "$\prec$" from the set J_p (4) if and only if $\Phi(\mathbf{w}^*) = 0$.

Let us consider an affine transformation of the feature vectors $\mathbf{x}_j$:

$$(\forall \mathbf{x}_j) \quad \mathbf{y}_j = A\, \mathbf{x}_j + \mathbf{b} \tag{12}$$

where A is a non-singular matrix of dimension n x n (A^{-1} exists) an $\mathbf{b}$ is a vector.

Lemma 2 (*affine invariance*): The minimal value Φ^* (11) of the criterion function $\Phi(\mathbf{w})$ (9) does not depend on non-singular, affine transformations (12):

$$\Phi_Y^* = \Phi_X^* \tag{13}$$

where Φ_Y^* is the minimal value of the criterion function $\Phi(\mathbf{w})$ (9) defined on the differential vectors $\mathbf{r}_{jk}' = (\mathbf{y}_k - \mathbf{y}_j)$ (12), where $((j,k) \in J_p)$.

Lemma 3 (*monotonocity property*): The minimal value Φ_{F2}^* (11) of the criterion function $\Phi_{F2}(\mathbf{w})$ (9) defined on features x_i from a subset $F2$ cannot decrease as a result of adding of some features x_i.

$$\textbf{\textit{if}}\ F_1 \supset F_2,\ \textbf{\textit{then}}\ \ \Phi_{F1}^* \leq \Phi_{F2}^* \tag{14}$$

where Φ_{F1}^* is the minimal value of the criterion function $\Phi_{F2}(\mathbf{w})$ (14) defined on the features x_i from the futures' subset F_1.

The above *Lemmas* describe the most important properties of the minimal value Φ^* (11). These properties justify using Φ^* (11) as the measure of linearity of given set J_p of ordering relations (4).

An arbitrary set C_B ($C_B \subset C$ (5)) of n linearly independent vectors $\mathbf{r}_{ii'} = (\mathbf{x}_{i'} - \mathbf{r}_i)$ ($(i,i')\in J_B$) can constitute the *basis* of the n-dimensional feature space $X[n]$ ($\mathbf{x}_j[n] \in X[n]$), and:

$$(\forall \mathbf{r}_{jk} \in C) \quad \mathbf{r}_{jk} = \sum_{(i,i')\in J_B} \alpha_{ii'}\, \mathbf{r}_{ii'} \tag{15}$$

where $\alpha_{ii'}$ are the combination′ parameters ($\alpha_{ii'} \in R^1$)

In accordance with the equality (15) each vector $\mathbf{r}_{jk}$ from the set C (5) can be expressed as a *linear combination* of the vectors $\mathbf{r}_{ii'}$ from the base set C_B ($(i,i')\in J_B$). The equality (15) describes *non-negative linear combination* if and only if all the parameters $\alpha_{ii'}$ are non-negative:

$$(\forall \mathbf{r}_{jk}\in C\ (15))\ \ (\forall (i,i')\in J_B)\ \ \alpha_{ii'} \geq 0 \qquad (16)$$

Lemma 4 (*non-negative linear combination*): If each vector r_{jk} from the set C (5) can be expressed as a *non-negative linear combination* (15), (16) of the vectors $r_{ii'}$ from some base set C_B ($(i,i')\in J_B$), then all the inequalities (8) are fulfilled on the line $y(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}$ defined by the optimal parameter vector $\mathbf{w}^*$ (11).

In particular, the thesis of the *Lemma* 4 is fulfilled if the set C (5) is built from n linearly independent vectors $\mathbf{r}_{jk} = (\mathbf{x}_k - \mathbf{x}_j)$ $(\forall (j,k)\in J_p)$.

The *Lemma* 4 specifies a sufficient condition that allows to design the fully ranked line $y(\mathbf{w}) = \mathbf{w}^T\mathbf{x}$ (6).

4. Cost sensitive criterion function $\Psi_\lambda(\mathbf{w})$

The criterion function $\Phi(\mathbf{w})$ (9) can be modified by introducing the cost function $\phi_i(\mathbf{w})$ for each feature x_i ($i = 1,\ldots, n$) in order to search for the best feature subspace F_l^* [3].

$$(\forall (i \in \{1,\ldots,n\})$$

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

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

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

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

where $\Phi(\mathbf{w})$ is given by (9), $\lambda \geq 0$, $\gamma_i > 0$, and $I = \{1,\ldots,n\}$.

The criterion function $\Psi_\lambda(\mathbf{w})$ (18) is the convex and piecewise linear (*CPL*) as the sum of the *CPL* functions $\Phi(\mathbf{w})$ (9) and $\lambda\, \gamma_i\, \phi_i(\mathbf{w})$ (18). Like previously (11), we are taking into account the point $\mathbf{w}_\lambda^*$ constituting the minimal value of the criterion function $\Psi_\lambda(\mathbf{w})$:

$$(\exists \mathbf{w}_\lambda^*)\ \ (\forall \mathbf{w})\ \ \Psi_\lambda(\mathbf{w}) \geq \Psi_\lambda(\mathbf{w}_\lambda^*) \qquad (19)$$

Let us introduce the below hyperplanes h_{jk} defined in the parameter space by difference vectors $\mathbf{r}_{jk} = (\mathbf{x}_k - \mathbf{x}_j)$ ($(j,k)\in J_p$ (4)).

$$(\forall (j,k)\in J_p)\ \ h_{jk} = \{\mathbf{w}: (\mathbf{r}_{jk})^T\, \mathbf{w} = 1\} \qquad (20)$$

Similarly, the unit vectors $\mathbf{e}_i$ define the below hyperplanes h^0_i ($\forall i\in I = \{1,\ldots,n\}$):

$$(\forall i\in I)\qquad h^0_i = \{\mathbf{w}: (\mathbf{e}_i)^T\mathbf{w} = 0\} \qquad (21)$$

The *vertex* $\mathbf{w}_m$ is defined as the point of intersection of n hyperplanes h_{jk} (20) or h^0_i (21) in the n-dimensional parameter space:

$$(\forall (j,k)\in J(\mathbf{w}_m))\qquad (\mathbf{r}_{jk})^T\, \mathbf{w}_m = 1 \qquad (22)$$

and

$$(\forall i\in I(\mathbf{w}_m))\qquad (\mathbf{e}_i)^T\, \mathbf{w}_m = 0 \qquad (23)$$

where $J(\mathbf{w}_m)$ is a subset of indices (j,k) of such hyperplanes h_{jk} (20) that pass through the point $\mathbf{w}_m$ ($J(\mathbf{w}_m) \subset J_p$ (4)). Similarly, $I(\mathbf{w}_m)$ is a subset of indices i of such hyperplanes h^0_i (21) which pass through the point $\mathbf{w}_m$ ($I(\mathbf{w}_m) \subset I$).

It can be proved, that the criterion function $\Psi_\lambda(\mathbf{w})$ (13) has the minimal value in one of the vertices $\mathbf{w}_m$:

$$(\exists \mathbf{w}_m^*)\ \ (\forall \mathbf{w})\ \ \Psi_\lambda(\mathbf{w}) \geq \Psi_\lambda(\mathbf{w}_m^*) \qquad (24)$$

Definition 3: The optimal vertex $\mathbf{w}_m^*$ has the *rank* equal to r if and only if the set $J(\mathbf{w}_m^*)$ (22) contains r pairs of the indices (j,k).

The indices (j,k) from the set $J(\mathbf{w}_m^*)$ (22) define those hyperplanes h_{jk} (20) which pass through the vertex $\mathbf{w}_m^*$. The set $I(\mathbf{w}_m^*)$ (23) contains $n - r$ indices i which define hyperplanes h^0_i (21) which pass through the optimal vertex $\mathbf{w}_m^*$ of the rank r.

The equations (22) and (23) can be presented in the below matrix for the optimal vertex $\mathbf{w}_m^*$ of the rank r:

$$\boldsymbol{B}_m^T\mathbf{w}_m^* = \mathbf{1}_m \qquad (25)$$

where

$$\boldsymbol{B}_m = [\mathbf{r}_{j(1),k(1)},\ldots, \mathbf{r}_{j(r),k(r)}, \mathbf{e}_{i(f+1)},\ldots,\mathbf{e}_{i(n)}] \qquad (26)$$

and

ISBN 978-3-901608-32-2

$$\mathbf{1}_m^T = [1,\ldots, 1, 0,\ldots\ldots,0] \qquad (27)$$

$\boldsymbol{B}_m$ is a nonsingular matrix (the optimal *basis*) with r first columns built from the vectors $\mathbf{r}_{jk}$ ($(j,k)\in J(\mathbf{w}_m^*)$) and the last n - r columns built from the unit vectors $\mathbf{e}_i$ ($i\in I(\mathbf{w}_m^*)$ (23)).

Remark 2: If the optimal vertex $\mathbf{w}_m^*$ (24) has the rank equal to r, then the $n - r$ features x_i ($i\in I(\mathbf{w}_m^*)$ (23)) can be neglected in the feature vectors $\mathbf{x}_j$ without changing the ordering relation (6) on the optimal line $y(\mathbf{w}_m^*) = (\mathbf{w}_m^*)^T\mathbf{x}$.

We can also remark that the rank r of the optimal vertex $\mathbf{w}_m^*$ (24) can be decreased by increasing of the value of the parameter λ (18). An increase of the value of the parameter λ (18) results in an increase of the costs of features x_i.

Example: Let us consider K learning sets C_k composed of the feature vectors $\mathbf{x}_j$:

$$(\forall k\in \{1,\ldots, K\})\ (\forall j \in I_k) \quad C_k = \{\mathbf{x}_j\} \qquad (28)$$

where I_k is the set of indices j of the feature vectors $\mathbf{x}_j$ belonging to the set C_k.

We assume the following relation (4) between the elements $\mathbf{x}_j$ of the learning sets C_k:

$$(\forall \mathbf{x}_j\in C_k)\ (\forall \mathbf{x}_{j'} \in C_{k'}) \qquad (29)$$
$$k < k' \Rightarrow \mathbf{x}_j \prec \mathbf{x}_{j'} \ \ (\mathbf{x}_{j'} \textit{ is older than } \mathbf{x}_j)$$

In accordance with the above relation, each element $\mathbf{x}_{j'}$ of the set $C_{k'}$ is older than an arbitrary element $\mathbf{x}_{j'}$ of the set C_k, if $k < k'$.

For some learning sets C_k (28) there exists such an unknown parameter vector $\mathbf{a} = [a_1,\ldots\ldots,a_n]^T$ and such unknown scalars τ_k ($k = 1,\ldots..,K$) that each feature vector $\mathbf{x}_j$ fulfils the below conditions:

$$(\exists \mathbf{a}\in R^n)\ (\forall k\in \{1,\ldots..,K\})\ (\exists \tau_k\in R^1) \qquad (30)$$
$$(\exists \mathbf{x}_j\in C_k)\ \ \tau_k < \mathbf{a}^T\mathbf{x}_j < \tau_{k+1}$$

In accordance with the above conditions, all the feature vectors $\mathbf{x}_j$ from one learning set C_k (28) are contained in the layer (*slice*) $L(\mathbf{a},\tau_k,\tau_{k+1})$ in the feature space $\boldsymbol{X}[n]$, where:

$$L(\mathbf{a},\tau_k,\tau_{k+1}) = \{\mathbf{x}\colon \tau_k < \mathbf{a}^T\mathbf{x} < \tau_{k+1}\} \qquad (31)$$

Learning sets C_k (28) consistent with the conditions (30) can represent a dynamical event, when successive layers describe a temporal development of this event. An unknown vector $\mathbf{a}$ can be interpreted as a *trend* of this event (*Fig.* 1).

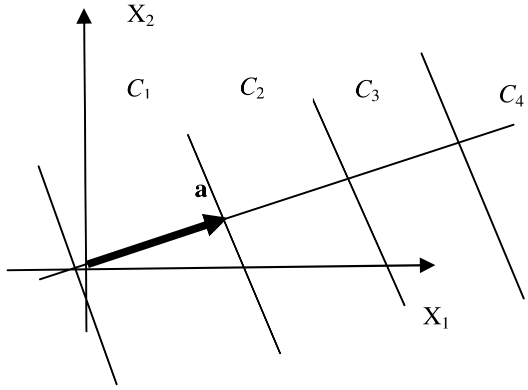


Fig 1: The layers $L(\mathbf{a},\tau_k,\tau_{k+1})$ (31) of the four learning sets C_k (28).

The conditions (30) can describe a generation of the feature vectors $\mathbf{x}_j$ from successive "slices" in the feature space $\boldsymbol{X}[n]$. For example, the development of the cancer can be characterized by a temporal sequence of images with the relations (30). The Cox model used in the survival analysis also leads to similar relations [6].

It can be proved that in the case of the feature vectors $\mathbf{x}_j$ consistent with the conditions (30) and the ranked relation (29), the minimal value $\Phi(\mathbf{w}^*)$ (11) of the criterion function $\Phi(\mathbf{w})$ (9) defined on the vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ (where $\mathbf{x}_j \prec \mathbf{x}_{j'}$) is equal to zero. As a result, all the ordering relations (29) are preserved on the line $y(\mathbf{w}^*) = (\mathbf{w}^*)^T\ \mathbf{x}$, defined by the optimal parameters vector $\mathbf{w}^*$ (11).

The vector $\mathbf{w}^*$ (11) can be also used in estimating unknown trend vector $\mathbf{a}$ (30) which determines an evolution of the feature vectors $\mathbf{x}_j$.

5. Concluding remarks

The ranked regression models $y(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}$ (4) can be designed through minimisation (11) of the convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w})$ (9) or $\Psi_\lambda(\mathbf{w})$ (18). Such criterion functions can be defined on the basis of time dependent data with inexact timing.

Ranked regression models have a variety of interesting applications, for example in survival analysis [7]. In general, ranked regression models can be used for the purpose of prognosis when forecasting future development of dynamical events or systems. The minimization of the cost sensitive criterion function $\Psi_\lambda(\mathbf{w})$ (18) also allows to identify such features x_i, that will be most influential in the future behaviour of a given system.

An evaluation of linearity of a given dynamical system can be also based on the minimization of the *CPL* criterion functions.

5. References

[1] R. A. Johnson, D. W. Wichern: *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

[3] L. Bobrowski: "Ranked modeling with feature selection based on the CPL criterion functions", in *Machine Learning and Data Mining in Pattern Recognition*, Eds. P. Permer et al., *Lecture Notes in Computer Science* vol. 3587, Springer Verlag, Berlin 2005

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

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

[6] J. P. Klein, M. L. Moeschberger: *Survival Analysis, Techniques for Censored and Truncated Data*, Springer, NY 1997

[7] L. Bobrowski, T. Łukaszuk: "Ranked linear modeling in survival analysis", pp. 61 - 67 in: Lecture Notes of the VI-th ICB Seminar: *Statistics and Clinical Practice*, ed. by L. Bobrowski, J. Doroszewski, C. Kulikowski, N. Victor, June 2005, Warsaw

