Prognostic Models Based on Linear Separability

Leon Bobrowski

Faculty of Computer Science, Bialystok Technical University,
ul. Wiejska 45A, Bialystok
and
Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland
leon@ibib.waw.pl

Abstract. Prognostic models are often designed on the basis of learning sets in accordance with multivariate regression methods. Recently, the interval regression and the ranked regression methods have been developed. Both these methods are useful in modeling censored data used in survival analysis. Designing the interval regression models as well as the ranked regression models can be treated similarly as the problem of linear classifier designing and linked to the concept of *linear separability* used in pattern recognition. The term linear separability refers to the examination of separation of two sets by a hyperplane in a given feature space.

Keywords: linear prognostic models, interval regression, ranked regression, *CPL* criterion functions.

1 Introduction

We are considering prognostic models based on linear multivariate regression models [1], [2]. In this case, the value of dependent (target) variable is predicted on the basis of linear combination of some independent variables. Problems of regression models designing on the basis of data sets are considered in the paper. The term *designing* means here a computation of parameters of the considered linear combination from available data (learning) set.

The classical and commonly used the last-square linear regression models are estimated on the basis of learning sequence in the form of feature vectors combined with exact values of the dependent (target) variable [1]. The exact value of target variable can be treated as an additional knowledge about a particular object represented by a given feature vector. The logistic regression is typically used when the target variable is a categorical one. If the target variable is a binary one, the logistic regression model is linked to a linear division of a given set of feature vectors [2].

The ranked regression models are designed on the basis of a set of feature vectors with an additional knowledge (information) in the form of an ordering relation in selected pairs of these vectors [3]. The ranked model is the linear transformation (projection) of multidimensional feature vectors on such a line which preserves the ordering relations in selected pairs as precisely as possible.

P. Perner (Ed.): ICDM 2011, LNAI 6870, pp. 11–24, 2011.

The interval regression models are designed on the basis of a set of feature vectors with an additional knowledge about predicted (dependent) variable in the form of intervals [4]. Each interval determines the minimal and the maximal value of the dependent variable which is linked to the given feature vector. The exact values of the predicted variable is a missing information in this case.

Prognostic models developed in the framework of the survival analysis are important in many biomedical applications. Such models are designed on the basis of the so called *censored* data sets. The Cox model plays a basic role in the survival analysis [5]. In the case of censored data sets, an additional information can be represented by intervals with only one constraint (border). It can mean the infinite minimal value (*left censoring*) or the infinite maximal value (*right censoring*) of the target variable interval linked to selected feature vectors.

Censored data set could be treated as a special case of interval data set. In consequence, the interval regression models can be designed also on the basis of censored data sets by using the convex and piecewise linear (*CPL*) criterion functions [6]. An ordering relation in selected pairs of feature vectors can be determined also on the basis of the censored data sets [7]. So, also the ranked regression models can be designed on the basis of censored data sets.

The concept of *linear separability* is used in theory of neural networks or in pattern recognition methods [2], [8]. The term linear separability is referring to exploration of two sets separation by a hyperplane in a given feature space. It has been shown that the problem of designing of both ranked models as well as interval regression models can be represented and solved as the problem of examination of linear separability. Consequences of this property are analyzed in the presented paper.

2 Linear Regression Models and Learning Sets with Different Structure

We take into considerations a set of m feature vectors $\mathbf{x}_j[n] = [x_{j1},\ldots,x_{jn}]^T$ belonging to a given n-dimensional feature space $F[n]$ ($\mathbf{x}_j[n] \in F[n]$). Feature vectors $\mathbf{x}_j[n]$ represent a family of m objects (events, patients) O_j ($j = 1,\ldots,m$). Components x_{ji} of the vector $\mathbf{x}_j[n]$ could be treated as the numerical results of n standardized examinations of the given object O_j ($x_{ji} \in \{0,1\}$ or $x_{ji} \in R^1$). Each vector $\mathbf{x}_j[n]$ can be treated also as a point in the n-dimensional feature space $F[n]$.

We are considering regression models based on linear (affine) transformations of n-dimensional feature vectors $\mathbf{x}[n]$ ($\mathbf{x}[n] \in F[n]$) on the points y of the line ($y \in R^1$):

$$y(\mathbf{x}) = \mathbf{w}[n]^T\mathbf{x}[n] + \theta \tag{1}$$

where $\mathbf{w}[n] = [w_1,\ldots, w_n]^T \in R^n$ is the parameters (*weight*) vector and θ is the *threshold* ($\theta \in R^1$).

Properties of the model (1) depend on the choice of the parameters $\mathbf{w}[n]$ and θ. The weights w_i and the threshold θ are usually computed on the basis of the available data (learning) sets. In the classical regression analysis the learning sets have the below structure [1]:

$$C_1 = \{\mathbf{x}_j[n];\, y_j\} = \{x_{j1}, \ldots., x_{jn}, ;\, y_j\}, \quad where\ j = 1, \ldots.., m \tag{2}$$

Each of m objects O_j is characterized in the set C_1 by values x_{ji} of n *independent* variables (*features*) X_i, and by the observed value y_j ($y_j \in R^1$) of the *dependent* (*target*) variable Y.

In the case of *classical regression*, the parameters $\mathbf{w}[n]$ and θ are chosen in such a manner that the sum of the squared differences $(y_j - \hat{y}_j)^2$ between the observed target variable y_j and the modeled variable $\hat{y}_j = \mathbf{w}[n]^T\mathbf{x}_j[n] + \theta$ (1) is minimal [1].

In the case of *interval regression*, an additional knowledge about particular objects O_j is represented by the intervals $[y_j^-, y_j^+]$ ($y_j^- < y_j^+$) instead of the exact values y_j (2) [4], [5]:

$$C_2 = \{\mathbf{x}_j[n], [y_j^-, y_j^+]\}, \quad where\ j = 1, \ldots.., m \tag{3}$$

where y_j^- is the lower bound ($y_j^- \in R^1$) and y_i^+ is the upper bound ($y_j^+ \in R^1$) of unknown value y of the target variable Y ($y_j^- < y < y_j^+$).

Let us remark, the classical learning set C_1 (2) can be transformed into the interval learning set C_2 (3) by introducing the boundary values $y_j^- = y_j - \varepsilon$ and $y_j^+ = y_j + \varepsilon$, where ε is a small positive parameter (ε > 0). Imprecise measurements of dependent variable y can be represented in such a manner.

Definition 1. The transformation (1) constitutes the *interval regression model* if the below linear inequalities are fulfilled in the best way possible for feature vectors $\mathbf{x}_j[n]$ from the set C_2 (3):

$$y_j^- < \mathbf{w}[n]^T\mathbf{x}_j[n] + \theta < y_j^+ \tag{4}$$

In the case of *ranked regression***,** additional knowledge about particular objects O_j and O_k ($j \neq k$) represented by feature vectors $x_j[n]$ and $x_k[n]$ is given in the form of ordering relation $''\mathbf{x}_j[n] \prec \mathbf{x}_k[n]''$, which could be read as $''x_j[n]$ is *before* $x_k[n]''$. For example, the relation $''\mathbf{x}_j[n] \prec \mathbf{x}_k[n]''$ could mean that the event O_j represented by the feature vector $\mathbf{x}_j[n]$ has occurred earlier, before the event O_k represented by the feature vector $\mathbf{x}_k[n]$. The relation $''\mathbf{x}_j[n] \prec \mathbf{x}_k[n]''$ between the feature vectors $x_j[n]$ and $\mathbf{x}_k[n]$ mans that the pair $\{\mathbf{x}_j[n], \mathbf{x}_k[n]\}$ has been *ranked*. It is natural to assume that the ordering relation $''\mathbf{x}_j[n] \prec \mathbf{x}_k[n]''$ should be *transitive:*

$$(\mathbf{x}_j[n] \prec \mathbf{x}_k[n])\ and\ (\mathbf{x}_k[n] \prec \mathbf{x}_l[n]) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_l[n]) \tag{5}$$

Example 1. Let us consider the relation $''O_j$ is *less risky* than O_k'' between selected patients O_j and O_k represented by the feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_k[n]$. Such relation between patients O_j and O_k, may reflect, for example, knowledge of medical experts. This relation between patients O_j and O_j can implicate the *ranked relation* $''\mathbf{x}_j[n] \prec \mathbf{x}_k[n]''$ between adequate feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_k[n]$.

$$(O_j \text{ is } \textit{less risky} \text{ than } O_k) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \tag{6}$$

The ranked learning set C_3 is constituted from the set $\{\mathbf{x}_j[n]\}$ of feature vectors $\mathbf{x}_j[n]$ (j = 1,….., m), and the set $\{\mathbf{x}_j[n] \prec \mathbf{x}_k[n]\}$ of ranked pairs $\{\mathbf{x}_j[n], \mathbf{x}_k[n]\}$:

$$C_3 = \{\{\mathbf{x}_j[n]\}, \{\mathbf{x}_j[n] \prec \mathbf{x}_k[n]\}\}, \ (j, k) \in I_r \tag{7}$$

where I_r is the set of indices (j, k) of the ranked pairs $\{x_j[n], x_k[n]\}$.

We can remark that usually not all the pairs $\{\mathbf{x}_j[n], \mathbf{x}_k[n]\}$ are ranked and can be used in regression model designing.

Definition 2. The transformation $y(\mathbf{x}) = \mathbf{w}[n]^T\mathbf{x}[n]$ (1) constitutes the *ranked regression model* if exists such weight vector $\mathbf{w}'[n]$, that the below implication is fulfilled in the best way possible for ranked pairs $\{\mathbf{x}_j[n] \prec \mathbf{x}_k[n]\}$ from the set C_3 (7):

$$(\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \Rightarrow (\mathbf{w}'[n]^T\mathbf{x}_j[n] < \mathbf{w}'[n]^T\mathbf{x}_k[n]) \tag{8}$$

The above implication means that the feature vectors $\mathbf{x}_j[n]$ preserve the ranked relations ″$\mathbf{x}_j[n] \prec \mathbf{x}_k[n]$″ also after their projection $\mathbf{w}'[n]^T\mathbf{x}_j[n]$ on the line $y(\mathbf{x}) = \mathbf{w}'[n]^T\mathbf{x}[n]$ (1), where $||\mathbf{w}'[n]|| = 1$.

3 Learning Sets in Survival Analysis

Traditionally, the *survival analysis* data sets C_s have the below structure [5]:

$$C_s = \{\mathbf{x}_j[n], t_j, \delta_j\} \ \ (j = 1,\ldots.., m) \tag{9}$$

where t_j is the *observed survival time* between the entry of the j-th patient O_j into the study and the end of the observation, δ_j is an indicator of failure of this patient ($\delta_j \in \{0,1\}$): δ_j= 1 - means the end of observation in the event of interest (*failure*), δ_j = 0 - means that the follow-up on the j-th patient ended before the event (*the right censored observation*). In this case (δ_j = 0) information about survival time t_j is *not complete*. The *real survival time* T_j can be defined in the below manner on the basis of the set C_s (9):

$$\begin{aligned} (\forall j = 1,\ldots\ldots.,m) \quad & \textbf{\textit{if}}\ \delta_j = 1,\ \textbf{\textit{then}}\ T_j = t_j,\ \textbf{\textit{and}} \\ & \textbf{\textit{if}}\ \delta_j = 0,\ \textbf{\textit{then}}\ T_j > t_j \end{aligned} \tag{10}$$

Assumption: If the survival time T_j of the j-th patients O_j is longer then the time T_k of the k-th patients O_k, then the patients O_j was *less risky* then the patients O_k [7]:

$$(T_j > T_k) \Rightarrow (O_j \text{ is } \textit{less risky} \text{ than } O_k) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \tag{11}$$

This implication can be expressed also by using the observed survival times t_j and t_k :

$$(t_j > t_k \text{ and } \delta_k = 1) \Rightarrow (O_j \text{ is } \textit{less risky} \text{ than } O_k) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \tag{12}$$

The right censoring means that an unknown survival time T_j of the of the j-th patient O_j is **longer** than the observed time t_j. The left censoring means that an unknown survival time T_j of the j-th patient O_j is **shorter** than the observed time t_j.

The censored survival times T_j can be represented also by intervals (3) through introducing two numbers (parameters) – the *lower bound* t_j^- ($t_j^- \in R^1$) and the *upper bound* t_j^+ ($t_j^- \in R^1$), where $t_j^- < t_j^+$. These parameters define the time interval $[t_j^-, t_j^+]$ for an unknown survival time T_j ($T_j \in [t_j^-, t_j^+]$ (3)). In the case of the right censoring, an unknown survival time T_j is greater than the given (known) lower bound t_j^- ($T_j > t_j^-$). It could mean, that $T_j \in [t_j^-, +\infty)$. In the case of the left censoring, an unknown survival time T_j is less than the given (known) upper bound t_j^+ ($T_j < t_j^+$). It could mean, that $T_j \in (-\infty, t_j^+]$. In accordance with such data representation, the right censoring means the replacement of the upper bound t_j^+ by the positive infinity $+\infty$. Similarly, the left censoring means the replacement of the lower bound t_j^- by the negative infinity $-\infty$. In the context of the survival time t_j^+ meaning, the more reasonable representation of the left censoring could be $[0, t_j^+]$ ($T_j \in [0, t_j^+]$).

Both the right censored and the left censored times T_j can be represented by using the interval data set C_2 (3) with the below structure:

$$C_4 = \{\mathbf{x}_j[n], [t_j^-, t_j^+], \delta_j'\} \quad (j = 1,\ldots.., m) \tag{13}$$

where δ_j' is the *indicator of censoring* of the survival time T_j of the patient O_j ($\delta_j' \in \{-1,0,1\}$): $\delta_j = -1$ means the left censoring ($T_j \in [0, t_j^+]$), $\delta_j = 1$ means the right censoring ($T_j \in [t_j^-,+\infty)$) , and $\delta_j = 0$ means that $T_j \in [t_j^-, t_j^+]$, where $0 < t_j^- < t_j^+ < \infty$.

Let us assume, that the prognostic model T(**x**) of an unknown survival time T is linear (1):

$$T(\mathbf{x}) = \mathbf{w}[n]^T\mathbf{x}[n] + \theta \tag{14}$$

In this case we can use the below linear inequalities for the purpose of the model (14) designing from the censored data C_4 (13):

$$\textbf{\textit{if}}\ \ \delta_j = -1, \textbf{\textit{then}}\ \ \mathbf{w}[n]^T\mathbf{x}_j[n] + \theta < t_j^+ \tag{15}$$

$$\textbf{\textit{if}}\ \ \delta_j = 1, \textbf{\textit{then}}\ \ \mathbf{w}[n]^T\mathbf{x}_j[n] + \theta > t_j^- \tag{16}$$

$$\textbf{\textit{if}}\ \ \delta_j = 0, \textbf{\textit{then}}\ \ t_j^- < \mathbf{w}[n]^T\mathbf{x}_j[n] + \theta < t_j^+ \tag{17}$$

The term model (14) designing means finding such parameters w[n] and θ that the above linear inequalities are fulfilled in the best way possible for feature vectors $x_j[n]$ from the set C_4 (13).

The parameters **w**[n] and θ of the interval regression model (14) are typically estimated from the data set C_s (9) by using the *Expectation Maximization* (*EM*) algorithms [4]. There are rather troublesome procedures with serious drawbacks concerning among others a low efficiency, particularly in the case of high dimensional feature space $F[n]$.

In the next section we examine the problem of prognostic models designing on the basis of the data set C_4 (13) by using the concept of the linear separability [2]. The linear separability of two data sets is evaluated through the minimisation of the convex and piecewise linear (*CPL*) criterion functions defined on these sets [8].

4 Linear Separability of Two Data Sets

Let us take into considerations two data sets: the *positive set* G^+ and the *negative set* G^- which are composed of n-dimensional feature vectors $\mathbf{x}_j[n]$ ($\mathbf{x}_j[n] \in F[n]$):

$$G^+ = \{\mathbf{x}_j[n]: j \in J^+\} \textit{ and } G^- = \{\mathbf{x}_j[n]: j \in J^-\} \tag{18}$$

where J^+ and J^- are disjoined sets ($J^+ \cap J^- = \emptyset$) of indices j.

Definition 3. The data sets G^+ and G^- (19) are *linearly separable*, if and only if there exists such a weight vector $\mathbf{w}[n]$ ($\mathbf{w}[n] \in R^n$) and a threshold θ ($\theta \in R$), that all the below inequalities with the inner products $\mathbf{w}[n]^T\mathbf{x}_j[n]$ are fulfilled:

$$\begin{aligned} (\exists\, \mathbf{w}[n], \theta)\ \ (\forall \mathbf{x}_j[n] \in G^+) \quad & \mathbf{w}[n]^T\mathbf{x}_j[n] > \theta \\ \textbf{\textit{and}}\ \ (\forall \mathbf{x}_j[n] \in G^-) \quad & \mathbf{w}[n]^T\mathbf{x}_j[n] < \theta \end{aligned} \tag{19}$$

The parameters $\mathbf{w}[n]$ and θ define the below hyperplane $H(\mathbf{w}[n],\theta)$ in the feature space $F[n]$ ($\mathbf{x}[n] \in F[n]$):

$$H(\mathbf{w}[n],\theta) = \{\mathbf{x}[n]: \mathbf{w}[n]^T\mathbf{x}[n] = \theta\} \tag{20}$$

If all the inequalities (19) are fulfilled, then each feature vector $\mathbf{x}_j[n]$ from the set G^+ is situated on the *positive side* ($\mathbf{w}[n]^T\mathbf{x}_j[n] > \theta$) of the hyperplane $H(\mathbf{w}[n],\theta)$ (20) and each vector from the set G^- is situated on the *negative side* ($\mathbf{w}[n]^T\mathbf{x}_j[n] < \theta$) of this hyperplane.

The concept of *linear separability* is used from many years in the theory of neural networks and in pattern recognition methods [2]. Among others, the linear separability has been used in the proof of the convergence of the error-correction algorithm – classic learning algorithm of neural networks. The linear classifiers can be designed through exploration of the linear separability of the data sets G^+ and G^- (19) [8].

The augmented vectors $\mathbf{z}_j^+[n+2]$ and $\mathbf{z}_j^-[n+2]$ have been introduced for the purpose of interval regression [6]:

$$(\forall j \in \{1,\ldots., m\})$$

$$\begin{aligned} &\textbf{\textit{if}}\ \ (y_j^- > -\infty), \textbf{\textit{ then }}\ \mathbf{z}_j^+[n+2] = [\mathbf{x}_j[n]^T, 1, -y_j^-]^T \ \textbf{\textit{ else }}\ \mathbf{z}_j^+[n+2] = \mathbf{0}, \\ &\textbf{\textit{and}} \\ &\textbf{\textit{if}}\ \ (y_j^+ < +\infty), \textbf{\textit{ then }}\ \mathbf{z}_j^-[n+2] = [\mathbf{x}_j[n]^T, 1, -y_j^+]^T \ \textbf{\textit{ else }}\ \mathbf{z}_j^-[n+2] = \mathbf{0} \end{aligned} \tag{21}$$

and

$$\mathbf{v}[n+2] = [\mathrm{v}_1,\ldots,\mathrm{v}_{n+2}]^{\mathrm{T}} = [\mathbf{w}[n]^{\mathrm{T}}, \theta, \beta]^{\mathrm{T}} \tag{22}$$

where β is the *interval weight* ($\beta \in R^1$).

The linear separability of the augmented vectors $\mathbf{z}_{\mathrm{j}}^{+}[n+2]$ and $\mathbf{z}_{\mathrm{j}}^{-}[n+2]$ means, that

$$\begin{gathered}(\exists \mathbf{v}[n+2])\ \ (\forall j \in \{1,\ldots., m\}) \\ (\forall \mathbf{z}_{\mathrm{j}}^{+}[n+2] \neq \mathbf{0})\ \ \mathbf{v}[n+2]^{\mathrm{T}} \mathbf{z}_{\mathrm{j}}^{+}[n+2] > 0, \ \ \textit{and} \\ (\forall \mathbf{z}_{\mathrm{j}}^{-}[n+2] \neq \mathbf{0})\ \ \mathbf{v}[n+2]^{\mathrm{T}}\ \mathbf{z}_{\mathrm{j}}^{-}[n+2] < \ 0\end{gathered} \tag{23}$$

or (23)

$$\begin{gathered}(\exists \mathbf{v}'[n+2])\ \ (\forall j \in \{1,\ldots., m\}) \\ (\forall \mathbf{z}_{\mathrm{j}}^{+}[n+2] \neq \mathbf{0})\ \ \mathbf{v}'[n+2]^{\mathrm{T}} \mathbf{z}_{\mathrm{j}}^{+}[n+2] \geq\ 1, \ \ \textit{and} \\ (\forall \mathbf{z}_{\mathrm{j}}^{-}[n+2] \neq \mathbf{0})\ \ \mathbf{v}'[n+2]^{\mathrm{T}}\ \mathbf{z}_{\mathrm{j}}^{-}[n+2] \leq -1\end{gathered} \tag{24}$$

Let us introduce the *positive set* H^{+} and the *negative set* H^{-} composed of such vectors $\mathbf{z}_{\mathrm{j}}^{+}[n+2]$ and $\mathbf{z}_{\mathrm{j}}^{-}[n+2]$ (21) which are different from zero:

$$H^{+} = \{\mathbf{z}_{\mathrm{j}}^{+}[n+2]\} \ \text{ and } \ H^{-} = \{\mathbf{z}_{\mathrm{j}}^{-}[n+2]\} \tag{25}$$

The *positive set* H^{+} is composed of m^{+} augmented vectors $\mathbf{z}_{\mathrm{i}}^{+}[n+2]$ ($\mathbf{z}_{\mathrm{i}}^{+}[n+2] \neq \mathbf{0}$) and the *negative set* H^{-} is composed of m^{-} augmented vectors $\mathbf{z}_{\mathrm{j}}^{-}[n+2]$ ($\mathbf{z}_{\mathrm{j}}^{-}[n+2] \neq \mathbf{0}$).

Definition 4. The sets H^{+} and H^{-} (25) of the augmented feature vectors $\mathbf{z}_{\mathrm{j}}^{+}[n+2]$ and $\mathbf{z}_{\mathrm{j}}^{-}[n+2]$ are *linearly separable*, if and only if there exists such augmented vector of parameters $\mathbf{v}'[n+2]$, that all the inequalities (24) are fulfilled.

Lemma 1. All the interval inequalities $y_{\mathrm{i}}^{-} < \mathrm{w}'[n]^{\mathrm{T}}\mathrm{x}_{\mathrm{i}}[n] + \theta' < y_{\mathrm{j}}^{+}$ (4) can be fulfilled by some parameters vector $\mathrm{v}'[n+2] = [\mathrm{w}'[n]^{\mathrm{T}}, \theta', 1]$ (25) if and only if the sets H^{+} and H^{-} (25) are linearly separable (24).

The ranked regression models (*Definition* 2) can be designed by using the ranked learning set C_3 (7). The expected implications (8) allows to transform the set $\{(\mathrm{x}_{\mathrm{j}}[n] \prec \mathrm{x}_{\mathrm{k}}[n])\}$ of ranked pairs $\{\mathrm{x}_{\mathrm{j}}[n], \mathrm{x}_{\mathrm{k}}[n]\}$ into the below set of desired linear inequalities:

$$(\exists \mathbf{w}'[n])\ \ (\forall (j, k) \in I_{\mathrm{r}})\ \ (\mathbf{x}_{\mathrm{j}}[n] \prec \mathbf{x}_{\mathrm{k}}[n]) \Rightarrow \mathbf{w}'[n]^{\mathrm{T}}\mathbf{x}_{\mathrm{j}}[n] < \ \mathbf{w}'[n]^{\mathrm{T}}\mathbf{x}_{\mathrm{k}}[n] \tag{26}$$

or

$$(\exists \mathbf{w}'[n])\ \ (\forall (j, k) \in I_{\mathrm{r}})\ \ (\mathbf{x}_{\mathrm{j}}[n] \prec \mathbf{x}_{\mathrm{k}}[n]) \Rightarrow \mathbf{w}'[n]^{\mathrm{T}}(\mathbf{x}_{\mathrm{j}}[n] - \mathbf{x}_{\mathrm{k}}[n]) < 0 \tag{27}$$

Let us introduce the *differential vectors* $\mathrm{r}_{\mathrm{jk}}[n]$ for all the ranked pairs $\{\mathrm{x}_{\mathrm{j}}[n] \prec \mathrm{x}_{\mathrm{k}}[n]\}$ (7):

$$(\forall(j, k) \in I_r) \quad (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \Rightarrow \mathbf{r}_{jk}[n] = \mathbf{x}_k[n] - \mathbf{x}_j[n] \tag{28}$$

The differential vectors $r_{jk}[n]$ can be divided in the below sets R^+ and R^-:

$$R^+ = \{\mathbf{r}_{jk}[n]: j < k\} \textit{ and } R^- = \{\mathbf{r}_{jk}[n]: j > k\} \tag{29}$$

We can remark that one of the sets R^+ or R^- can be empty. The following *Lemma* has been proved [3].

Lemma 2. All the ranked relations ″$x_j[n] \prec x_k[n]$″ ($(j, k) \in I_r$) (7) can be preserved (8) by a linear model $y(x) = w'[n]^T x[n]$ (1) defined by a parameter vector $w'[n]$, if and only if the sets R^+ and R^- (29) are linearly separable (24).

We can infer from the *Lemma* 1 and the *Lemma* 2 that the linear separability of two sets constitutes a basis both for the interval regression models as well as for the ranked regression models.

5 *CPL* Penalty and Criterion Functions for Interval and Ranked Regression

The *augmented feature vectors* $\mathbf{z}_i^+[n+2]$ and $\mathbf{z}_j^-[n+2]$ (21) and the *augmented weight vector* $\mathbf{v}[n+2]$ (22) have been introduced for the case of the interval regression model. The family of linear inequalities (24) represents the concept of linear separability of the sets H^+ and H^- (25).

The convex and piecewise-linear (*CPL*) penalty functions $\varphi_{Hj}^+(\mathbf{v}[n+2])$ and $\varphi_{Hj}^-(\mathbf{v}[n+2])$ defined on the vectors (21) are linked to the expected inequalities (24).

$$(\forall \mathbf{z}_j^+[n+2] \neq \mathbf{0})$$

$$\textbf{\textit{if}} \quad \mathbf{v}[n+2]^T\mathbf{z}_j^+[n+2] < 1, \ \textbf{\textit{then}} \ \varphi_{Hj}^+(\mathbf{v}[n+2]) = 1 - \mathbf{v}[n+2]^T\mathbf{z}_j^+[n+2], \ \textbf{\textit{else}} \ \varphi_{Hj}^+(\mathbf{v}[n+2]) = 0 \tag{30}$$

$$(\forall \mathbf{z}_j^-[n+2] \neq \mathbf{0})$$

$$\textbf{\textit{if}} \quad \mathbf{v}[n+2]^T\mathbf{z}_j^-[n+2] > -1, \ \textbf{\textit{then}} \ \varphi_{Hj}^-(\mathbf{v}[n+2]) = 1 + \mathbf{v}[n+2]^T\mathbf{z}_j^-[n+2], \ \textbf{\textit{else}} \ \varphi_{Hj}^-(\mathbf{v}[n+2]) = 0 \tag{31}$$

The *CPL* criterion function $\Phi_H(\mathbf{v}[n+2])$ is defined as the weighted sum of the penalty functions $\varphi_{Hj}^+(\mathbf{v}[n+2])$ (30) and $\varphi_{Hj}^-(\mathbf{v}[n+2])$ (31) [8]:

$$\Phi_H(\mathbf{v}[n+2]) = \sum_j \beta_j \, \varphi_{Hj}^+(\mathbf{v}[n+2]) + \sum_j \beta_j \, \varphi_{Hj}^-(\mathbf{v}[n+2]) \tag{32}$$

where positive parameters β_j ($\beta_j \geq 0$) determine an *importance* of the particular vectors $\mathbf{z}_j^+[n+2]$ or $\mathbf{z}_i^-[n+2]$ (21).

The vector $\mathbf{v}_H^*[n+2]$ constitutes the minimum of the criterion function $\Phi_H(\mathbf{v}[n+2])$:

$$(\forall \mathbf{v}[n+2]) \quad \Phi_H(\mathbf{v}[n+2]) \geq \Phi_H(\mathbf{v}_H^*[n+2]) = \Phi_H^* \geq 0 \tag{33}$$

where $\mathbf{v}_{\mathrm{H}}^*[n{+}2] = \mathbf{v}^*[n{+}2] = [\mathbf{w}^*[n]^{\mathrm{T}}, \theta^*, \beta^*]^{\mathrm{T}}$, and $\mathbf{w}^*[n] = [\mathrm{w}_1^*,\ldots., \mathrm{w}_{\mathrm{n}}^*]^{\mathrm{T}}$ (22).

The below theorem can be proved [8]:

***Theorem* 1.** The minimal value $\Phi_{\mathrm{H}}^* = \Phi_{\mathrm{H}}(\mathbf{v}_{\mathrm{H}}^*[n{+}2])$ (33) of the non-negative criterion function $\Phi_{\mathrm{H}}(\mathbf{v}[n{+}2])$ (32) is equal to zero ($\Phi_{\mathrm{H}}^* = 0$) and the sets H^+ and H^- (25) are linearly separable (24) if and only if there exists such weight vector $\mathbf{w}'[n]$ and the threshold θ', that the inequalities $y_{\mathrm{j}}^- < \mathbf{w}'[n]^{\mathrm{T}}\mathbf{x}_{\mathrm{i}}[n] + \theta' < y_{\mathrm{j}}^+$ (4) are fulfilled for each ranked pairs $\{\mathbf{x}_{\mathrm{j}}[n] \prec \mathbf{x}_{\mathrm{k}}[n]\}$ from the learning set C_3 (7).

Remark 1. If the minimal value $\Phi_{\mathrm{H}}^* = \Phi_{\mathrm{H}}(\mathbf{v}^*[n{+}2])$ (33) is equal to zero ($\Phi_{\mathrm{H}}^* = 0$) in the point $\mathbf{v}^*[n{+}2] = [\mathbf{w}^*[n]^{\mathrm{T}}, \theta^*, \beta^*]^{\mathrm{T}}$ with $\beta^* > 0$, then the optimal model $\hat{y} = (\mathbf{w}^*[n] \,/\, \beta^*)^{\mathrm{T}}\,\mathbf{x}[n] + \theta^*/\,\beta^*$ fulfils all the constraints (4):

$$(\forall j \in \{1,\ldots, m\}) \qquad y_{\mathrm{j}}^- < (\mathbf{w}^*[n] \,/\, \beta^*)^{\mathrm{T}}\,\mathbf{x}_{\mathrm{j}}[n] + \theta^*/\,\beta^* < y_{\mathrm{j}}^+ \tag{34}$$

If the minimal value Φ_{H}^* (42) is greater than zero ($\Phi_{\mathrm{H}}^* > 0$) in the point $\mathbf{v}^*[n{+}2]$, then the optimal model does not fulfil all the above inequalities.

In the case of the ranked models (*Definition* 2), the set of the expected linear inequalities (27) has been defined by using the differential vectors $\mathbf{r}_{\mathrm{ik}}[\mathrm{n}] = \mathbf{x}_{\mathrm{k}}[n] - \mathbf{x}_{\mathrm{j}}[n]$ (28) representing the ranked pairs $\{\mathbf{x}_{\mathrm{j}}[n] \prec \mathbf{x}_{\mathrm{k}}[n]\}$ (8). The positive set R^+ and the negative set R^- (29) has been defined on the basis of the lexicographical order of the indices j and k in the ranked pairs $\{\mathbf{x}_{\mathrm{j}}[n] \prec \mathrm{x}_{\mathrm{k}}[n]\}$ (7).

The sets R^+ and R^- (29) of the differential vectors $\mathrm{r}_{\mathrm{jk}}[n]$ are linearly separable, if and only if there exists such vector of parameters $\mathrm{w}'[n1]$, that all the below inequalities are fulfilled:

$$\begin{aligned} (\exists \mathbf{w}'[n]) \;\; (\forall \mathbf{r}_{\mathrm{jk}}[n] \in R^+) \quad & \mathbf{w}'[n]^{\mathrm{T}}\mathbf{r}_{\mathrm{jk}}[n] \geq 1 \\ \textbf{\textit{and}} \;\; (\forall \mathbf{r}_{\mathrm{jk}}[n] \in R^-) \quad & \mathbf{w}'[n]^{\mathrm{T}}\mathbf{r}_{\mathrm{jk}}[n] \leq -1 \end{aligned} \tag{35}$$

The below *CPL* penalty functions $\varphi_{\mathrm{jk}}^+(\mathbf{w}[n])$ and $\varphi_{\mathrm{jk}}^+(\mathbf{w}[n])$ are linked to the above inequalities:

$$\begin{gathered} (\forall \mathbf{r}_{\mathrm{jk}}[n] \in R^+) \\ \textbf{\textit{if}} \;\; \mathbf{w}[n]^{\mathrm{T}}\mathbf{r}_{\mathrm{jk}}[n] < 1, \textbf{\textit{ then }} \varphi_{\mathrm{jk}}^+(\mathbf{w}[n]) = 1 - \mathbf{w}[n]^{\mathrm{T}}\mathbf{r}_{\mathrm{jk}}[n], \;\; \textbf{\textit{else }} \varphi_{\mathrm{jk}}^+(\mathbf{w}[n]) = 0 \end{gathered} \tag{36}$$

$$\begin{gathered} (\forall \mathbf{r}_{\mathrm{jk}}[n] \in R^-) \\ \textbf{\textit{if}} \;\; \mathbf{w}[n]^{\mathrm{T}}\mathbf{r}_{\mathrm{jk}}[n] > -1, \textbf{\textit{ then }} \varphi_{\mathrm{jk}}^-(\mathbf{w}[n]) = 1 + \mathbf{w}[n]^{\mathrm{T}}\mathbf{r}_{\mathrm{jk}}[n], \;\; \textbf{\textit{else }} \varphi_{\mathrm{jk}}^-(\mathbf{w}[n]) = 0 \end{gathered} \tag{37}$$

The *CPL* criterion function $\Phi_{\mathrm{R}}(\mathbf{w}[n])$ is defined as the weighted sum of the penalty functions $\phi_{\mathrm{jk}}^+(\mathbf{w}[n])$ (36) and $\phi_{\mathrm{jk}}^-(\mathbf{w}[n])$ (37) [3]:

$$\Phi_{\mathrm{R}}(\mathbf{w}[n]) = \sum_{R^+} \gamma_{\mathrm{jk}}\, \varphi_{\mathrm{jk}}^+(\mathbf{w}[n]) + \sum_{R^-} \gamma_{\mathrm{jk}}\, \varphi_{\mathrm{jk}}^-(\mathbf{w}[n]) \tag{38}$$

where positive parameters γ_{jk} ($\gamma_{\mathrm{jk}} \geq 0$) determine an *importance* of particular ranked relations $\{\mathbf{x}_{\mathrm{j}}[n] \prec \mathbf{x}_{\mathrm{k}}[n]\}$ (7).

The optimal vector $\mathbf{w}_{\mathrm{R}}^*[n]$ constitutes the minimum of the function $\Phi_{\mathrm{R}}(\mathbf{w}[n])$:

$$(\forall \mathbf{w}[n])\ \ \Phi_R(\mathbf{w}[n]) \geq \ \Phi_R(\mathbf{w}_R{}^*[n]) \ = \Phi_R{}^* \geq 0 \tag{39}$$

The optimal vector $\mathbf{w}_R{}^*[n]$ (39) defines the ranked model:

$$\hat{y}_R = \mathbf{w}_R{}^*[n]^T\, \mathbf{x}[n]\ . \tag{40}$$

***Theorem* 2.** The minimal value $\Phi_R{}^* = \Phi_R(\mathbf{w}_R{}^*[n])$ (39) of the criterion function $\Phi_R(\mathbf{w}[n])$ (38) is equal to zero ($\Phi_R{}^* = 0$) and the sets R^+ and R^- (29) are linearly separable (33) if and only if there exists such weight vector $\mathbf{w}'[n]$, that the inequalities $\mathbf{w}'[n]^T\mathbf{x}_j[n] < \mathbf{w}'[n]^T\mathbf{x}_k[n]$ (8) are fulfilled for each ranked pair $\{\mathbf{x}_j[n] \prec \mathbf{x}_k[n]\}$ from the learning set C_3 (7) ([3], [9]).

Remark 2. If the minimal value $\Phi_R{}^* = \Phi_R(\mathbf{w}_R{}^*[n])$ (39) is equal to zero ($\Phi_R{}^* = 0$), then the inequalities $\mathbf{w}_R{}^*[n]^T\mathbf{x}_j[n] < \mathbf{w}_R{}^*[n]^T\mathbf{x}_k[n]$ (8) are fulfilled for each ranked pair $\{\mathbf{x}_j[n] \prec \mathbf{x}_k[n]\}$ from the learning set C_3 (7). If the minimal value $\Phi_R{}^*$ (39) is greater than zero ($\Phi_R{}^* > 0$) in the point $\mathbf{w}_R{}^*[n]$, then the ranked model does not fulfil all the inequalities (8).

6 Relaxed Linear Separability (*RLS*) Method of Feature Selection for Prognostic Models

The feature selection process could mean a reduction as large amount of features x_i as possibly while assuring a high quality of the designed model (*Remark* 1).

For the purpose of feature selection in the interval regression the *CPL* criterion function $\Phi_H(\mathbf{v}[n+2])$ (32) has been modified by inclusion of *feature penalty functions* $\phi_i(\mathbf{v}[n+2])$ and the *costs* γ_i ($\gamma_i > 0$) related to particular features x_i [10]:

$$(\forall i \in \{1,\ldots..,n\}) \quad \phi_i(\mathbf{v}[n+2]) = |\mathbf{e}_i[n+2]^T\mathbf{v}[n+2]| = |w_i| \tag{41}$$

where $\mathbf{e}_i[n+2]$ are the unit vectors and $\mathbf{v}[n+2] = [\mathbf{w}[n]^T, \theta, \beta]^T$.

The modified *CPL* criterion function $\Psi_H(\mathbf{v}[n+1])$ has the below form [9]:

$$\Psi_H(\mathbf{v}[n+2]) \ = \ \Phi_H(\mathbf{v}[n+2]) + \ \lambda \sum_{i \in \{1,\ldots,n\}} \gamma_i\ \phi_i\ (\mathbf{v}[n+2]) \tag{42}$$

where λ ($\lambda \geq 0$) is the *cost level* and the *feature costs* γ_i are typically equal to one.

The criterion function $\Psi_H(\mathbf{v}[n+2)$ (42) similarly to the function $\Phi_H(\mathbf{v}[n+2])$ (32) is convex and piecewise-linear (*CPL*). The basis exchange algorithms allow to find efficiently the optimal vector of parameters (*vertex*) $\mathbf{v}_{H\lambda}{}^*[n+2]$ of the criterion function $\Psi_H(\mathbf{v}[n+2])$ (42) with different values of the cost level λ [10]:

$$(\exists \mathbf{v}_\lambda{}^*[n+ 2])\ \ (\forall \mathbf{v}[n+2])\ \ \Psi_H(\mathbf{v}[n+2]) \geq \Psi_H(\mathbf{v}_\lambda{}^*[n+2]) \tag{43}$$

where $\mathbf{v}_\lambda{}^*[n+2] = [\mathbf{w}_\lambda{}^*[n]^T, \theta_\lambda{}^*, \beta_\lambda{}^*]^T$, and $\mathbf{w}_\lambda{}^*[n] = [w_{\lambda 1}{}^*,\ldots., w_{\lambda n}{}^*]^T$ (22).

The optimal vector $\mathbf{v}_\lambda^*[n+2]$ (43) allows to define both the interval regression model (4) as well as the below decision rule of the linear classifier which operates on elements of the sets H^+ or H^- (25) ($\mathbf{z}[n+2] = \mathbf{z}_j^+[n+2]$ or $\mathbf{z}[n+2] = \mathbf{z}_j^-[n+2]$ (21)).

$$\begin{aligned} &\textbf{\textit{if}}\ \ \mathbf{v}_\lambda^*[n+2]^T\mathbf{z}[n+2] \geq 0, \textbf{\textit{then}}\ \ \mathbf{z}[n+2]\ \textit{is allocated to the category}\ \ \omega^+ \\ &\qquad\qquad\qquad\qquad\qquad \textbf{\textit{else}}\ \ \mathbf{z}[n+2]\ \textit{is allocated to the category}\ \ \omega^- \end{aligned} \tag{44}$$

The element $\mathbf{z}[n+2] = \mathbf{z}_i^+[n+2]$ is wrongly classified by the rule (44) if it is allocated to the category ω^-. Similarly, the element $\mathbf{z}[n+2] = \mathbf{z}_j^-[n+2]$ is wrongly classified if it is allocated to the category ω^+.

The quality of the linear classifier (44) can be evaluated by using the error estimator (*apparent error rate*) $e_a(\mathbf{v}_\lambda^*[n+2])$ as the fraction of wrongly classified elements $\mathbf{z}[n+2]$ of the sets H^+ and H^- (25):

$$e_a(\mathbf{v}_\lambda^*[n+2]) = m_a(\mathbf{v}_\lambda^*[n+2]) \,/\, m_H \tag{45}$$

where m_H is the number of all elements $\mathbf{z}[n+2]$ of the sets H^+ and H^- (25), and $m_a(\mathbf{v}_\lambda^*[n+2])$ is the number of such elements $\mathbf{z}[n+2]$ which are wrongly allocated by the rule (44).

The parameters $\mathbf{v}_\lambda^*[n+2]$ of the linear classifier (44) are estimated from the sets H^+ and H^- (25) through minimization of the *CPL* criterion function $\Psi_H(\mathbf{v}[n+2])$ (42) defined on all elements $\mathbf{z}[n+2]$ of these sets. It is known that if the same vectors $\mathbf{z}[n+2]$ are used for classifier designing and classifier evaluation, then the evaluation results are too optimistic (*biased*).

For the purpose of the bias reduction of the apparent error rate estimator $e_a(\mathbf{v}_\lambda^*[n+2])$ (45), the cross validation procedures are applied [2]. The term *p-fold cross validation* means that data sets H^+ and H^- (25) have been divided into p parts P_i, where $i = 1,\ldots, p$ (for example $p = 10$). The vectors $\mathbf{z}[n+2]$ contained in $p - 1$ parts P_i are used for the definition of the criterion function $\Psi_H(\mathbf{v}[n+2])$ (42) and for finding (43) the parameters $\mathbf{v}_\lambda^*[n+2]$. The remaining vectors $\mathbf{z}[n+2]$ are used as the *test set* (one part $P_{i'}$) for computing (evaluation) of the error rate $e_{i'}(\mathbf{v}_\lambda^*[n+2])$ (45). Such evaluation is repeated p times, and each time different part $P_{i'}$ is used as the test set. The *cross-validation error rate* $e_{CVE}(\mathbf{v}_\lambda^*[n+2])$ (45) is estimated in the cross validation procedure as the mean value of the error rates $e_{i'}(\ \mathbf{v}_\lambda^*[n+2])$ evaluated on various parts (test sets) $P_{i'}$. The cross validation procedure uses different vectors $\mathbf{z}[n+2]$ for the classifier designing and evaluation. In result, the bias of the error rate estimation (45) can be reduced.

For the purpose of feature selection in the interval regression the *CPL* criterion function $\Phi_R(\mathbf{w}[n])$ (38) has been modified in a similar manner to (42):

$$\Psi_R(\mathbf{w}[n]) = \Phi_R(\mathbf{w}[n]) + \lambda \sum_{i \in \{1,\ldots,n\}} \gamma_i\, \phi_i(\mathbf{w}[n]) = \Phi_R(\mathbf{w}[n]) + \lambda \sum_{i \in \{1,\ldots,n\}} \gamma_i\, |w_i| \tag{46}$$

The minimization of the *CPL* criterion function $\Psi_R(\mathbf{w}[n])$ (45) with the cost level λ allows to find the optimal vector of parameters $\mathbf{w}_\lambda^*[n]$:

$$(\exists \mathbf{w}_\lambda^*[n])\ \ (\forall \mathbf{w}[n])\ \ \Psi_R(\mathbf{w}[n]) \geq \Psi_R(\mathbf{w}_\lambda^*[n]) \tag{47}$$

The optimal vector $\mathbf{w}_\lambda^*[n]$ (46) defines both the ranked model $\hat{y}_R = \mathbf{w}_\lambda^*[n]^T\mathbf{x}[n]$ (40) as well as the below decision rule of the linear classifier which operates on elements $\mathbf{r}[n]$ of the sets R^+ or R^- (29) ($\mathbf{r}[n] = \mathbf{r}_{jk}[n]$).

$$\begin{aligned} &\textbf{\textit{if}}\ \ \mathbf{w}_\lambda^*[n]^T\mathbf{r}[n] \geq 0, \textbf{\textit{then}}\ \ \mathbf{r}[n]\ \textit{is allocated to the category}\ \ \omega^+, \\ &\qquad\qquad\qquad\qquad \textbf{\textit{else}}\ \mathbf{r}[n]\ \textit{is allocated to the category}\ \ \omega^- \end{aligned} \tag{48}$$

The quality of the linear classifier (47) can be evaluated by using the error estimator (*apparent error rate*) $e_a(\mathbf{w}_\lambda^*[n])$ as the fraction of wrongly classified elements $\mathbf{r}[n]$ of the sets R^+ and R^- (29):

$$e_a(\mathbf{w}_\lambda^*[n]) = m_a(\mathbf{w}_\lambda^*[n]) \,/\, m_R \tag{49}$$

where m_R is the number of all elements $\mathbf{r}[n]$ of the sets R^+ and R^- (29), and $m_a(\mathbf{w}_\lambda^*[n])$ is the number of such elements $\mathbf{r}[n]$ which are wrongly allocated by the rule (47).

We can remark, that such features x_i which have the weights $w_{\lambda i}^*$ equal to zero ($w_{\lambda i}^* = 0$) in the optimal vector $\mathbf{v}_\lambda^*[n+2]$ (43) can be reduced without changing the decision rule (44). The weights $w_{\lambda i}^*$ equal to zero ($w_{\lambda i}^* = 0$) does not change also the decision rule (47. The below feature reduction rule has been proposed basing on this property [10]:

$$(w_{\lambda i}^* \ = \ 0) \ \Rightarrow \text{(the feature } x_i \text{ is reduced)} \tag{50}$$

In accordance with the *relaxed linear separability* (*RLS*) method of feature subsets selection, a successive increase of the *cost level* λ in the minimized criterion function $\Psi_H(\mathbf{v}[n+2)$ (42) or the criterion function $\Psi_R(\mathbf{w}[n])$ (45) reduces more weights $w_{\lambda i}^*$ to zero ($w_{\lambda i}^* = 0$) and, in result, reduces additional features x_i (49). In this way, the less important features x_i are eliminated and the descending sequence of feature subspaces $F_k[n_k]$ ($n_k > n_{k+1}$) is generated. Each feature subspace $F_k[n_k]$ in the below sequence can be linked to some value λ_k of the cost level λ in the criterion function $\Psi_H(\mathbf{v}[n+2)$ (42) or the criterion function $\Psi_R(\mathbf{w}[n])$ (45):

$$F[n] \supset F_1[n_1] \supset \ldots \supset F_k[n_k], \ \textit{where}\ \ 0 \leq \lambda_0 < \ \lambda_1 < \ldots < \lambda_k \tag{51}$$

Particular feature subspaces $F_k[n_k]$ in the sequence (50) can be evaluated by using the cross-validation error rate (*CVE*) of the optimal linear classifier (44) or (47) designed in a given subspace $F_k[n_k]$ [10]. Such subspace $F_{k\,i}^*[n_k]$ which is characterized by the lowest cross-validation error rate (*CVE*) is treated as the optimal subspace in accordance with the *RLS* approach.

7 Concluding Remarks

Designing linear prognostic models (1) on the basis of the interval learning set C_2 (3) or the ranked learning set C_3 (7) has been considered in the paper. It was pointed out, that designing the interval prognostic model (4) can be based on exploration of the linear separability of the data sets H^+ and H^- (25). Similarly, designing the ranked prognostic model (8) can be based on the exploration of the linear separability of the data sets R^+

and R^- (29). The linear separability of the data sets H^+ and H^- (25) appears if and only if the minimal values $\Phi_H(\mathbf{v}_H^*[n+2])$ (33) of the *CPL* criterion function $\Phi_H(\mathbf{v}[n+2])$ (32) is equal to zero. The linear separability of the data sets R^+ and R^- (29) appears if and only if the minimal value $\Phi_R(\mathbf{w}_R^*[n])$ (39) of the *CPL* criterion function $\Phi_R(\mathbf{w}[n])$ (38) is equal to zero. It can be assumed, that the vector $\mathbf{v}_H^*[n+2]$ (33) defines the optimal interval model (34) both in the case of linearly separable data sets H^+ and H^- (25), as well as in the case when these sets are not linearly separable ($\Phi_H(\mathbf{v}_H^*[n+2]) > 0$). Similarly, the vector $\mathbf{w}_R^*[n]$ (40) defines the optimal ranked model (40) both in the case of linearly separable data sets R^+ and R^- (35), as well as in the case when these sets are not linearly separable ($\Phi_R(\mathbf{w}_R^*[n]) > 0$).

Exploration of the linear separability can be carried out through minimization of the convex and piecewise-linear (*CPL*) criterion functions defined on a given pair of data sets. The minimal value and the optimal vector of particular *CPL* criterion functions can be computed efficiently even in the case of large high-dimensional data sets by applying the basis exchange algorithms, which are similar to the linear programming [10].

The designing process based on the linear separability allows to apply the relaxed linear separability (*RLS*) method of feature subset selection to the interval prognostic models (34) or to the ranked prognostic models (40) [10]. This possibility indicates practical significance as it allows to identify the most influential input patterns. For example, the identification of such subset of genes of a given patient which increase the risk of a cancer disease could be performed by using the methods described in the paper.

More generally, choosing a subset of variables is a crucial step in designing prognostic models. It is particularly important, when the number n of variables (*features*) X_i is high in comparison to the number m of objects O_j. Typically such situation occurs in the case of bioinformatics data sets.

Acknowledgment. This work was supported by the by the NCBiR project N R13 0014 04, and partially financed by the project S/WI/2/2011 from the Białystok University of Technology, and by the project 16/St/2011 from the Institute of Biocybernetics and Biomedical Engineering PAS.

References

1. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis. Prentice-Hall, Inc., Englewood Cliffs (1991)
2. Duda, O.R., Hart, P.E., Stork, D.G.: Pattern Classification. J. Wiley, New York (2001)
3. Bobrowski, L.: Ranked linear models and sequential patterns recognition. Pattern Analysis & Applications 12(1), 1–7 (2009)
4. Gomez, G., Espinal, A., Lagakos, S.: Inference for a linear regression model with an interval-censored covariate. Statistics in Medicine 22, 409–425 (2003)
5. Klein, J.P., Moeschberger, M.L.: Survival Analysis, Techniques for Censored and Truncated Data. Springer, NY (1997)

6. Bobrowski, L.: Interval Uncertainty in CPL Models for Computer Aided Prognosis. In: Hippe, Z.S., Kulikowski, J.L., Mroczek, T. (eds.) Human-Computer Systems Interaction. Backgrounds and Applications. Advances in Soft Computing, vol. 2. Springer, Heidelberg (in the press, 2011)
7. Bobrowski, L.: Selection of high risk patients with ranked models based on the CPL criterion functions. In: Perner, P. (ed.) ICDM 2010. LNCS, vol. 6171, pp. 432–441. Springer, Heidelberg (2010)
8. Bobrowski, L.: Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych, Data mining based on convex and piecewise linear criterion functions, Technical University Białystok (2005) (in Polish)
9. Bobrowski, L.: Design of piecewise linear classifiers from formal neurons by some basis exchange technique. Pattern Recognition 24(9), 863–870 (1991)
10. Bobrowski, L., Łukaszuk, T.: Feature selection based on relaxed linear separability. Biocybernetics and Biomedical Engineering 29(2), 43–59 (2009)