Selection of High Risk Patients with Ranked Models Based on the CPL Criterion Functions*

Leon Bobrowski

Faculty of Computer Science, Białystok Technical University
Institute of Biocyberinetics and Biomedical Engineering, PAS, Warsaw, Poland
leon@ibib.waw.pl

Abstract. Important practical problems in computer support medical diagnosis are related to screening procedures. Identification of high risk patients can serve as an example of such a problem. The identification results should allow to select a patient in an objective manner for additional therapeutic treatment. The designing of the screening tools can be based on the minimisation of the convex and piecewise linear (CPL) criterion functions. Particularly ranked models can be designed in this manner for the purposes of screening procedures.

Keywords: screening procedures, convex and piecewise linear (CPL) criterion functions, ranked models.

1 Introduction

One of the most important groups of problems in computer aided medical diagnosis are those related to screening procedures. We are considering screening procedures which are aimed at selecting high risk patients. High risk patients should be possibly early directed to a special therapeutic treatment. For example, the success of cancer therapy depends on early detection and beginning of this disease therapy.

The screening procedures usually result from certain probabilistic prognostic models. The survival analysis methods give a theoretical framework for designing screening procedures [1], [2]. In particular, the Cox model is commonly used in survival analysis for selection of high risk patients [3].

However, constraints met in practical implementation of screening procedures indicate limitations of the probabilistic modeling in this context. In practice, the screening procedures are designed on the basis of available medical databases which rarely fit in the statistical principles of parameters estimation. First of all, the number of cases (patients) in medical databases is usually too low and the number of parameters (features) describing particular patients is too high for reliable estimation of selected prognostic model parameters. Secondly, the probabilistic assumptions linked to particular prognostic models are often unverifiable.

* This work was supported by the by the NCBiR project N R13 0014 04, and partially financed by the project S/WI/2/2010 from the Białystok University of Technology, and by the project 16/St/2010 from the Institute of Biocybernetics and Biomedical Engineering PAS.

P. Perner (Ed.): ICDM 2010, LNAI 6171, pp. 432–441, 2010.

For these reasons, designing of effective and reliable screening procedures is still related to open research and implementation problems. Here we are examining the possibility of using ranked modeling in designing screening procedures [4]. In particular, we are taking into account linear ranked models designed through the minimisation of the convex and piecewise linear (*CPL*) criterion functions [5], [6]. These criterion functions are defined on the survival analysis data sets. An important problem analysed here is feature selection aimed at reducing ranked models dimensionality [7].

2 Feature Vectors and Ranked Relations Originating from Survival Analysis

Let us assume that m patients O_j collected in a given medical database are represented as n-dimensional feature vectors $\mathbf{x}_j[n] = [x_{j1},......, x_{jn}]^T$ or as points in the n-dimensional feature space $F[n]$ ($\mathbf{x}_j[n] \in F[n]$, $j = 1,.....,m$). The component x_{ji} of the vector $\mathbf{x}_j[n]$ is the numerical value of the i-th feature x_i of the patient (*object*) O_j. For example, the components x_{ji} can be the numerical results of diagnostic examinations of the given patient O_j. The feature vectors $\mathbf{x}_j[n]$ can be of a mixed type and represent a type of measurement (for example ($x_{ji} \in \{0,1\}$, or $x_{ji} \in R^1$).

We are taking into consideration the learning data set C built from m feature vectors $\mathbf{x}_j[n]$ which represent particular patients O_j:

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

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_j can implicate the *ranked relation* "$\mathbf{x}_j[n] \prec \mathbf{x}_k[n]$" between adequate feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_k[n]$.

$$(O_j \text{ is } \textit{less risky}, \text{ then } O_k) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \tag{2}$$

The relation "$\mathbf{x}_j[n] \prec \mathbf{x}_k[n]$" between the feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_k[n]$ means that the pair $\{\mathbf{x}_j[n], \mathbf{x}_k[n]\}$ is *ranked*. The ranked relations between particular feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_k[n]$ should result from additional knowledge about the patients O_j and O_k. Such additional knowledge could result from an information about *survival time* T_j of particular patients O_j collected in the given database.

Traditionally, the *survival analysis* data sets C_s have the below structure [1]:

$$C_s = \{\mathbf{x}_j[n], t_j , \delta_j\} \;\; (j = 1,.......,m) \tag{3}$$

where t_j is the *observed survival time* between the entry of the j-th O_j patient into the study and the end of the observation, δ_j is an indicator of failure of this patient ($\delta_j \in \{0,1\}$): $\delta_j = 1$ - means the end of observation in the event of interest (*failure*), $\delta_j = 0$ - means that the follow-up on the j-th patient ended before the event (*the right censored observation*). In this case ($\delta_j = 0$) information about survival time t_j is *not* *complete*. A great part of survival data set C_s can be censored. The survival analysis methods are based in a great part on the Cox model [2], [3]. The ranked models can be also used in the search for a solution of basic problems in survival analysis [4].

The *survival time* T_j can be defined in the below manner on the basis of the set C_s (3):

$$(\forall j = 1,.......,m) \ \textit{if} \ \delta_j = 1, \ \textit{then} \ T_j = t_j, \ \textit{and} \qquad (4)$$
$$\textit{if} \ \delta_j = 0, \ \textit{then} \ T_j > t_j$$

Assumption: If the survival time T_j (4) of the j-th patients O_j is longer then the survival time T_j of the j-th patients O_j, then the patients O_j was *less risky* (2) then the patients O_k:

$$(T_j > T_k) \Rightarrow (O_j \text{ is } \textit{less risky} \text{ than } O_k) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \qquad (5)$$

This implication can be expressed also by using the observed survival time t_j and t_k (3):

$$(t_j > t_k \text{ and } \delta_k = 1) \Rightarrow (O_j \text{ is } \textit{less risky} \text{ than } O_k) \Rightarrow (\mathbf{x}_j[n] \prec \mathbf{x}_k[n]) \qquad (6)$$

3 Linear Ranked Models

Let us consider such transformation of n-dimensional feature vectors $\mathbf{x}_j[n]$ on the ranked line $y = \mathbf{w}[n]^T\mathbf{x}_j[n]$, which preserves the ranked relations $''\mathbf{x}_j[n] \prec \mathbf{x}_k[n]''$ (2) as precisely as possible

$$y_j = y_j(\mathbf{w}[n]) = \mathbf{w}[n]^T\mathbf{x}_j[n] \qquad (7)$$

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

Definition 1: The relation $''\mathbf{x}_j[n] \prec \mathbf{x}_j[n]''$ (2) is fully preserved by the *ranked line* (7) if and only if the following implication holds:

$$(\forall (j, k)) \quad \mathbf{x}_j[n] \prec \mathbf{x}_k[n] \Rightarrow y_j(\mathbf{w}[n]) < y_k(\mathbf{w}[n]) \qquad (8)$$

The procedure of the ranked line designing can be based on the concept of positively and negatively oriented dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ [6].

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

$$(\forall\ (j, j') \in I^+) \quad \mathbf{x}_j[n] \prec \mathbf{x}_{j'}[n] \tag{9}$$

Definition 3: The ranked pair $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ $(j < j')$ of the feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_{j'}[n]$ constitutes the *negatively oriented dipole* $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ $(\forall (j, j') \in I^-)$, if and only if $\mathbf{x}_{j'}[n] \prec \mathbf{x}_j[n]$.

$$(\forall\ (j, j') \in I^-) \quad \mathbf{x}_{j'}[n] \prec \mathbf{x}_j[n] \tag{10}$$

Definition 4: The line $y(\mathbf{w}[n]) = \mathbf{w}[n]^T\mathbf{x}[n]$ (7) is fully consistent (*ranked*) with the dipoles $\{\mathbf{x}_j[n],\mathbf{x}_{j'}[n]\}$ orientations if and only if

$$\begin{aligned} &(\forall\ (j, j') \in I^+) \quad y_j(\mathbf{w}[n]) < y_{j'}(\mathbf{w}[n]) \ \ \textit{and} \\ &(\forall\ (j, j') \in I^-) \quad y_j(\mathbf{w}[n]) > y_{j'}(\mathbf{w}[n]), \text{ where } j < j' \end{aligned} \tag{11}$$

All the relations "$\mathbf{x}_j[n] \prec \mathbf{x}_{j'}[n]$" (2) are fully preserved (8) on the line (7) if and only if all the above inequalities are fulfilled.

The problem of the ranked line designing can be linked to the concept of linear separability of two sets C^+ and C^- of the differential vectors $\mathbf{r}_{jj'}[n] = \mathbf{x}_{j'}[n] - \mathbf{x}_j[n]$ which are defined below:

$$\begin{aligned} &C^+ = \{\mathbf{r}_{jj'}[n] = (\mathbf{x}_{j'}[n] - \mathbf{x}_j[n]): (j, j') \in I^+\} \\ &C^- = \{\mathbf{r}_{jj'}[n] = (\mathbf{x}_{j'}[n] - \mathbf{x}_j[n]): \ (j, j') \in I^-\}, \text{ where } j < j' \end{aligned} \tag{12}$$

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

$$H(\mathbf{w}[n]) = \ \{\mathbf{x}[n]: \mathbf{w}[n]^T\mathbf{x}[n] = 0\} \tag{13}$$

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

Definition 5: The sets C^+ and C^- (12) are linearly separable with the threshold equal to zero if and only if there exists such a parameter vector $\mathbf{w}^*[n]$ that:

$$\begin{aligned} &(\forall\ (j, j') \in I^+) \quad \mathbf{w}^*[n]^T\, \mathbf{r}_{jj'}[n] > 0 \\ &(\forall\ (j, j') \in I^-) \quad \mathbf{w}^*[n]^T\, \mathbf{r}_{jj'}[n] < 0 \end{aligned} \tag{14}$$

The above inequalities can be represented in the following manner:

$$\begin{aligned} (\exists \mathbf{w}^*[n])\ (\forall\ (j, j') \in I^+) \quad &\mathbf{w}^*[n]^T\, \mathbf{r}_{jj'}[n] \ \geq \ 1 \\ (\forall\ (j, j') \in I^-) \quad &\mathbf{w}^*[n]^T\, \mathbf{r}_{jj'}[n] \ \leq -1 \end{aligned} \tag{15}$$

Remark 1: If the parameter vector $\mathbf{w}^*[n]$ linearly separates (14) the sets C^+ and C^- (12), then the line $y_j(\mathbf{w}^*[n]) = \mathbf{w}^*[n]^T\mathbf{x}_j[n]$ is fully consistent (11) with the dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ orientation.

4 *CPL* Criterion Functions

The separating hyperplane $H(\mathbf{w}[n])$ (13) could be designed through the minimisation of the convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w}[n])$ which is similar to the perceptron criterion function used in the theory of neural networks and pattern recognition [8], [9]. Let us introduce for this purpose the positive $\varphi_{jj'}^+(\mathbf{w}[n])$ and negative $\varphi_{jj'}^-(\mathbf{w}[n])$ penalty functions (Fig. 1):

$$(\forall\, (j, j') \in I^+) \quad \varphi_{jj'}^+(\mathbf{w}[n]) = \begin{cases} 1 - \mathbf{w}[n]^T\, \mathbf{r}_{jj'}[n] & \textit{if } \mathbf{w}[n]^T\mathbf{r}_{jj'}[n] < 1 \\ 0 & \textit{if } \mathbf{w}[n]^T\mathbf{r}_{jj'}[n] \geq 1 \end{cases} \tag{16}$$

$$\text{and} \quad (\forall\, (j, j') \in I^-) \quad \varphi_{jj'}^-(\mathbf{w}[n]) = \begin{cases} 1 + \mathbf{w}[n]^T\mathbf{r}_{jj'}[n] & \textit{if } \mathbf{w}[n]^T\mathbf{r}_{jj'}[n] > -1 \\ 0 & \textit{if } \mathbf{w}[n]^T\mathbf{r}_{jj'}[n] \leq -1 \end{cases} \tag{17}$$

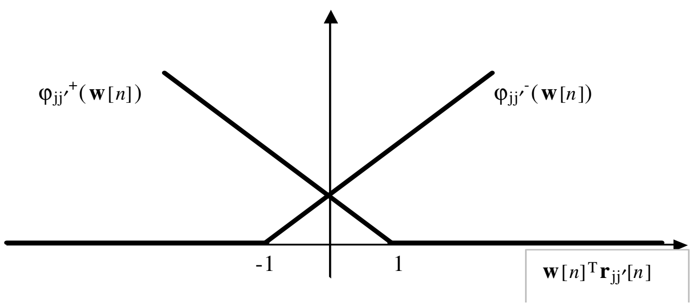

Fig. 1. The penalty functions $\varphi_{jj'}^+(\mathbf{w}[n])$ (16) and $\varphi_{jj'}^-(\mathbf{w}[n])$ (17)

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

$$\Phi(\mathbf{w}[n]) = \sum_{(j,j')\in I^+} \alpha_{jj'}\, \varphi_{jj'}^+(\mathbf{w}[n]) + \sum_{(j,j')\in I^-} \alpha_{jj'}\, \varphi_{jj'}^-(\mathbf{w}[n]) \tag{18}$$

where $\alpha_{jj'}$ ($\alpha_{jj'} > 0$) is a nonnegative parameter (*price*) related to the dipole $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ ($j < j'$)

The criterion function $\Phi(\mathbf{w}[n])$ (18) is the convex and piecewise linear (*CPL*) function as the sum of such type of the penalty functions $\phi_{jj'}^{+}(\mathbf{w}[n])$ (16) and $\phi_{jj'}^{-}(\mathbf{w}[n])$ (17). The basis exchange algorithms, similarly to linear programming, allow to find a minimum of such functions efficiently, even in the case of large, multidimensional data sets C^+ and C^- (12) [10]:

$$\Phi^* = \Phi(\mathbf{w}^*[n]) = min\ \Phi(\mathbf{w}[n]) \geq 0 \tag{19}$$

The optimal parameter vector $\mathbf{w}^*[n]$ and the minimal value Φ^* of the criterion function $\Phi(\mathbf{w}[n])$ (18) can be applied to a variety of data ranking problems. In particular, the below *ranked model* can be designed this way.

$$y(\mathbf{w}^*[n]) = \mathbf{w}^*[n]^T\mathbf{x}[n] \tag{20}$$

Lemma 1: The minimal value Φ^* (19) of the criterion function $\Phi(\mathbf{w}[n])$ (18) is non-negative and equal to zero if and only if there exists such a vector $\mathbf{w}^*[n]$ that the ranking of the points $y_j = \mathbf{w}^*[n]^T\mathbf{x}_j[n]$ on the line (7) are fully consistent (11) with the dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ orientations.

The proof of this *Lemma* can be found in the earlier paper [6].

The modified criterion function $\Psi_\chi(\mathbf{w}[n])$ which includes additional *CPL* penalty functions in the form of the absolute values $|w_i|$ multiplied by the *feature costs* γ_i has been introduced for the purpose of feature selection [7].

$$\Psi_\lambda(\mathbf{w}[n]) = \Phi(\mathbf{w}[n]) + \lambda \sum_{i \in I} \gamma_i |w_i| \tag{21}$$

where λ ($\lambda \geq 0$) is the *cost level*, and $I = \{1,\ldots\ldots,n\}$.

The criterion function $\Psi_\lambda(\mathbf{w}[n])$ (21), similarly to the function $\Phi(\mathbf{w}[n])$ (18) is convex and piecewise-linear (*CPL*). The basis exchange algorithms allow to find efficiently the optimal vector of parameters (*vertex*) $\mathbf{w}_\lambda[n]$ of the function $\Psi_\lambda(\mathbf{w}[n])$ with different values of parameter λ [10]:

$$(\exists(\mathbf{w}_\lambda[n]))\ (\forall \mathbf{w}[n])\ \ \Psi_\lambda(\mathbf{w}[n]) \geq \Psi_\lambda(\mathbf{w}_\lambda[n]) = \hat{\Psi_\lambda} \tag{22}$$

The parameters $\mathbf{w}_\lambda[n] = [w_{\lambda 1},\ldots.., w_{\lambda n}]^T$ (22) define the optimal separating hyperplane $H(\mathbf{w}_\lambda[n])$ (13). Such features x_i which have the weights $w_{\lambda i}$ equal to zero ($w_{\lambda i} = 0$) in the optimal vector $\mathbf{w}_\lambda[n]$ (22) can be reduced without changing the location of the optimal separating hyperplane $H(\mathbf{w}_\lambda[n])$ (13). As a result, the below rule of the feature reduction based on the components $w_{\lambda i}$ of the optimal vector of parameters $\mathbf{w}_\lambda[n] = [w_{\lambda 1},\ldots.., w_{\lambda n}]^T$ (22) has been proposed [7]:

$$(w_{\lambda i.} = 0) \Rightarrow (\text{the feature } x_i \text{ is reduced}) \tag{23}$$

The minimal value (22) of the *CPL* criterion function $\Psi_\lambda(\mathbf{w}[n])$ (21) represents an optimal balance between linear separability of the sets C^+ and C^- (12) and features costs determined by the parameters λ and γ_i. We can remark that a sufficiently increased value of the *cost level* λ in the minimized function $\Psi_\lambda(\mathbf{w}[n])$ (21) results in an increase number of the reduced features x_i (23). The dimensionality of the feature $F[n]$ can be reduced arbitrarily by a successive increase of the parameter λ in the criterion function $\Psi_\lambda(\mathbf{w}[n])$ (21). Such method of feature selection has been named *relaxed linear separability* [7].

The feature selection procedure is an important part of designing ranked models (20). The feature selection process is aimed at reducing the maximal number of unimportant features x_i. The *reduced ranked model* $y(\mathbf{w}_\lambda'[n'])$ can be defined by using the optimal vector of parameters $\mathbf{w}_\lambda[n]$ (22) and the rule (23):

$$y(\mathbf{w}_\lambda'[n']) = \mathbf{w}_\lambda'[n']^T\mathbf{x}[n'] \tag{24}$$

where $\mathbf{w}_\lambda'[n']$ is such vector of parameters which is obtained from the optimal vector $\mathbf{w}_\lambda[n]$ (22) by reducing components $w_{\lambda i}$ equal to zero ($w_{\lambda i.} = 0$), and $\mathbf{x}[n']$ is the reduced feature vector (23) obtained in the same way as $\mathbf{w}_\lambda'[n']$.

The dimensionality reduction of the prognostic model (24), which is based on the relaxed linear separability method allows, among others, to enhance *risk factors* x_i influencing given disease.

5 Selection of High Risk Patients

The reduced ranked model (24) can be used in selection of high risk patients O_j. These models define transformations of the multidimensional feature vectors $\mathbf{x}_j[n']$ (1) on the points y_j (7) which represent particular patients O_j on the ranked line. As a result, the *ranked sequence* of patients $O_{j(i)}$ can be obtained:

$$O_{j(1)},\ O_{j(2)}, \ldots\ldots, O_{j(m)}, \text{ where} \tag{25}$$
$$y_{j(1)} \geq y_{j(2)} \geq \ldots\ldots \geq y_{j(m)}$$

The patients $O_{j(i)}$ with the largest values $y_{j(i)}$ which are situated on the top of the above sequence can be treated as *high risk patients*. The patients $O_{j(i)}$ which are situated at the beginning of this sequence can be treated as a *low risk*.

The models (20) or (24) can be applied not only to the patients O_j represented by the feature vectors $\mathbf{x}_j[n]$ from the set C (1). Let us assume that a new patient O_a is represented by the feature vector $\mathbf{x}_a[n]$. The model (24) allows to compute the point $y_a = \mathbf{w}_\lambda'[n']^T\mathbf{x}_a[n']$ on the ranked line. If the new patient O_a is located at the top of the ranked sequence (25),) next to the high risk patients $O_{j(i)}$, then it can be also treated as a *high risk*. In other cases, O_a should not be treated as a *high risk* patient.

The transformation (24) can be treated as a *prognostic model* of a given disease ω_k development. Such model can represent a *main trend* in the disease ω_k development. The model is designed on the basis of information contained in the *survival analysis* data sets C_s (3).

A very important problem in practice is quality evaluation of the prognostic models (24). One of the possibilities is to use the *ranked error rate* $e_r(\mathbf{w}_\lambda'[n'])$ the model (24) evaluation.

$$e_r(\mathbf{w}_\lambda'[n']) = m_r'(\mathbf{w}_\lambda'[n']) / m_r \tag{26}$$

where m_r is the number of positive (9) and negative (10) dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$, where $j < j'$. $m_r'(\mathbf{w}_\lambda'[n'])$ is the number of such dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$, which are wrongly oriented (not consistent with the rule (11)) on the line $y(\mathbf{w}_\lambda'[n'])$ (24).

If the same dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ are used for the model (24) designing and for the model evaluation, then the error rate estimator $e_r(\mathbf{w}_\lambda'[n'])$ (26) is *positively biased* [11]. The error rate estimator $e_r(\mathbf{w}_\lambda'[n'])$ (26) is called the *apparent error* (*AE*). The *cross-validation* techniques are commonly used in model evaluation because these techniques allow to reduce the *bias* of the error estimation.

In accordance with the cross-validation *p – folds* procedure, the set of all the dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ (9), (10) is divided in the p near equal parts P_i (for example p = 10). During one step the model is designed on the dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$ belonging to $p - 1$ (*learning*) parts P_i and evaluated on the elements of one (*testing*) part $P_{i'}$. Each part P_i serves once as the testing part $P_{i'}$ during successive p steps. The *cross-validation error rate* (*CVE*) $e_{CVE}(\mathbf{w}_\lambda'[n'])$ is obtained as the mean value of the error rates $e_r(\mathbf{w}_\lambda'[n'])$ (26) evaluated on the p testing parts $P_{i'}$.

The *leave one method* is the particular case of the *p – folds* cross-validation, when p is equal to the number m_r ($p = m_r$) of the positively (9) and negatively (10) dipoles $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$. In this case, each testing part $P_{i'}$ contains exactly one element $\{\mathbf{x}_j[n], \mathbf{x}_{j'}[n]\}$.

The *relaxed linear separability* method of feature selection can be used during designing the reduced ranked models $y(\mathbf{w}_\lambda'[n'])$ (24) [7]. In accordance with this method, a successive increase of the *cost level* λ in the minimized function $\Psi_\lambda(\mathbf{w}[n])$ (21) causes a reduction of additional features x_i (23). 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 (21):

$$\begin{gathered} F[n] \supset F_1[n_1] \supset F_2[n_2] \supset \ldots\ldots\ldots \supset F_k[n_k], \text{ where} \\ 0 \le \lambda_0 < \lambda_1 < \lambda_2 < \ldots\ldots < \lambda_k \end{gathered} \tag{27}$$

Particular feature subspaces $F_k[n_k]$ (27) have been evaluated by using the cross-validation error rate (*CVE*) $e_{CVE}(\mathbf{w}_\lambda'[n_k])$ (26). The below figure shown an example of experimental results. The evaluation results of descending sequence (27) of feature subspaces $F_k[n_k]$ is shown on this figure:

The above Figure illustrates the feature reduction process which begins from the dimensionality n = 100. In this case, the data set C (1) contained about m = 200 feature vectors $\mathbf{x}_j[n]$. It can be seen on the Figure 1, that it exists such feature subspaces $F_k[n_k]$ of dimensionality n_k = 30 or more, which allow to obtain the fully consistent (11) ranked lines $y(\mathbf{w}[n_k])$ (7).

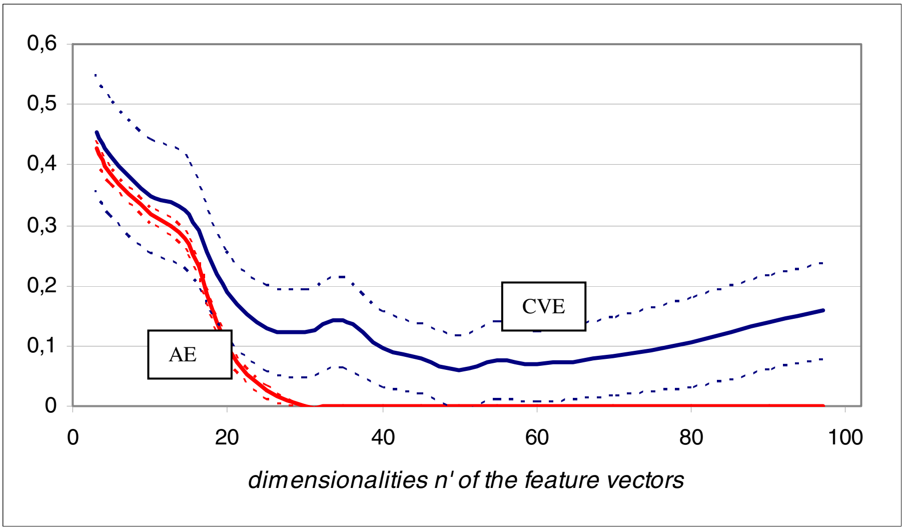

Fig. 2. The apparent error (*AE*) (26) and the cross-validation error (*CVE*) in different feature subspaces $F_k[n_k]$ of the sequence (27). The upper solid line represents the cross-validation error (*CVE*), the lower solid line represents the apparent error (*AE*) (26). The broken line represents the standard deviation.

In accordance with the relaxed linear separability method, the process of successive reduction of the less important features x_i should be stopped at the dimensionality $n_k' \approx 50$, where the cross-validation error rate $e_{CVE}(\mathbf{w}_\lambda'[n'])$ (26) reaches its lowest value. The feature subspace $F_k[n_k']$ is treated as the optimal one in accordance with this method of feature selection [7].

6 Concluding Remarks

Ranked modeling has been applied here to designing linear prognostic models $y(\mathbf{w}_\lambda'[n'])$ on the basis of the survival analysis data set C_s (3). The described designing process is based on multiple minimization of the convex and piecewise linear (*CPL*) criterion function $\Psi_\lambda(\mathbf{w}[n])$ (21) defined on the data set C_s (3). The basis exchange algorithms allow to carry out such multiple minimization efficiently.

The process of ranked models designing described here includes feature selection stage, which is based on the relaxed linear separability method. This method of feature selection is linked with the evaluation of prognostic models $y(\mathbf{w}_\lambda'[n'])$ (24) by using the cross-validation techniques.

The linear ranked models designing has been discussed here in the context of the screening procedures aimed at selecting high risk patients. High risk patients should be possibly early detected for the purpose of special therapeutic treatment. The proposed solution can be applied also to other areas. For example, the bankruptcy in economy or the mechanical reliability problems can be analyzed and solved in a similar manner. The patients O_j could be replaced by sets of dynamical objects or events E_j.

One of the important problems in ranked modeling could be feature decomposition of nonlinear family of ranked relations (9), (10) into a set of linear families. The single linear model which does not fit well to all ranked relations should be replaced in this case by a family of well fitted local ranked models.

References

[1] Marubini, E., Valsecchi, M.G.: Analysing Survival Data from Clinical Trials and Observational Studies. John Wiley & Sons, Chichester (1995)
[2] Klein, J.P., Moeschberger, M.L.: Survival Analysis, Techniques for Censored and Truncated Data. Springer, NY (1997)
[3] Cox, D.R.: Regression Model and Life Tables (with Discussion). Journal of the Royal Statistical Society B, 187–220 (1972)
[4] Bobrowski, L.: Ranked modeling of risk on the basis of survival data. In: ICSMRA 2007 - International Conference on Statistical Methods for Risk Analysis, Lisbon (2007)
[5] Bobrowski, L., Łukaszuk, T., Wasyluk, H.: Ranked modeling of causal se-quences of diseases for the purpose of early diagnosis. In: Kącki, E., Rudnicki, M., Stempczyńska, J. (eds.) Computers in Medical Activity. Advances in Intelli-gence and Soft Computing, vol. 65, pp. 23–31. Springer, Heidelberg (2009)
[6] Bobrowski, L.: Ranked linear models and sequential patterns recognition. Pattern Analysis & Applications 12(1), 1–7 (2009)
[7] Bobrowski, L., Łukaszuk, T.: Feature selection based on relaxed linear separabilty. Biocybernetics and Biomedial Engineering 29(2), 43–59 (2009)
[8] Duda, O.R., Hart, P.E., Stork, D.G.: Pattern Classification. J. Wiley, New York (2001)
[9] Bobrowski, L.: Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych (Data mining based on convex and piecewise linear (CPL) criterion func-tions), Technical University Białystok (in Polish) (2005)
[10] Bobrowski, L., Niemiro, W.: A method of synthesis of linear discriminant function in the case of nonseparabilty. Pattern Recognition 17, 205–210 (1984)
[11] Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis. Prentice- Hall, Inc., Englewood Cliffs (1991)