Discovering Main Vertexical Planes in a Multivariate Data Space by Using CPL Functions

Leon Bobrowski

Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland
l.bobrowski@pb.edu.pl

Abstract. Data mining problems and tools are linked to the task of extracting important regularities (patterns) from multivariate data sets. In some cases, flat patterns can be located on vertexical planes in a multidimensional data space. Vertexical planes are linked to vertices in parameter space. Patterns located on vertexical planes can be discovered in large data sets through minimization of the convex and piecewise linear (*CPL*) criterion functions.

Keywords: high dimensional data sets, data mining, *CPL* criterion functions, main vertexical planes, flat patterns, *K-planes* clustering.

1 Introduction

Data mining tasks are aimed at discovering useful patterns in large data sets [1], [2], [3]. The term *patterns* stands for various types of regularities in an explored data set, such as decision rules, trends, or models of interactions. The extracted patterns are used in solving many practical problems linked, e.g. to medical diagnosis support, economic forecasting, marketing, fraud detection or to scientific discoveries.

Data sets can typically be represented as clouds of points in a multidimensional feature space. Clustering algorithms belong to the most powerful tools of data mining [2]. The *K-means* algorithms constitute the most popular and successful paradigm in clustering applications. The basic idea in the *K-means* algorithm is linked to partitioning of a given set C of m objects (points) into K subsets C_k centered around the class prototypes in the form of *central points*, which had been computed (defined) earlier. In the next step, the class prototypes are modified in accordance with the obtained subsets C_k. These steps are repeated until the central points are stabilized in the successive steps. The *K-means* algorithm has been generalized to the form of *K-means*, *K-planes* or *K-models* [4], [5]. In these approaches the concept of *central points* has been replaced by *central models*, e.g. in the form of planes in multivariate feature space.

Central planes can have the form of *vertexical planes* based on vertices in parameter space. *″Flat patterns″* located on the *main vertexical planes* can be discovered in large data sets through minimization of the convex and piecewise linear (*CPL*) criterion functions [6]. Analytical and computational properties of the method of main vertexical planes discovering through minimization of the *CPL* criterion functions are analyzed in the presented article.

P. Perner (Ed.): ICDM 2014, LNAI 8557, pp. 200–213, 2014.

2 Hyperplanes and Vertices in the Parameter Space

Let us assume that the data set C contains m feature vectors $\mathbf{x}_i[n] = [x_{j1},...,x_{jn}]^T$ belonging to a given n-dimensional feature space $F[n]$ ($\mathbf{x}_j[n] \in F[n]$):

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

Components x_{ji} of the feature vector $\mathbf{x}_j[n]$ can be treated as the numerical results of n standardized examinations of a given object O_j ($x_{ji} \in \{0,1\}$ or $x_{ji} \in R$).

We can assume without limitation that the feature space $F[n]$ is equal to the n-dimensional space of real numbers R^n ($F[n] = R^n$). Each feature vector $\mathbf{x}_j[n]$ can be treated as a point in the space R^n.

Each of m feature vector $\mathbf{x}_j[n]$ from the set C (1) defines the below hyperplane h_j in the parameter space R^n ($\mathbf{w}[n] \in R^n$):

$$(\forall \mathbf{x}_j[n] \in C) \qquad h_j = \{\mathbf{w}[n]: \mathbf{x}_j[n]^T\mathbf{w}[n] = 1\} \tag{2}$$

Each unit vector $\mathbf{e}_i[n] = [0,\ldots,1,\ldots,0]^T$ defines the below hyperplane h_i^0 in the parameter space R^n:

$$\begin{aligned}(\forall i \in \{1,\ldots,n\}) \qquad h_i^0 &= \{\mathbf{w}[n]: \mathbf{e}_i[n]^T\mathbf{w}[n] = 0\} = \\ &= \{\mathbf{w}[n]: w_i = 0\}\end{aligned} \tag{3}$$

Let us consider a set S_k of n linearly independent feature vectors $\mathbf{x}_j[n]$ ($j \in J_k$) and unit vectors $\mathbf{e}_i[n]$ ($i \in I_k$).

$$S_k = \{\mathbf{x}_j[n]: j \in J_k\} \cup \{\mathbf{e}_j[n]: i \in I_k\} \tag{4}$$

The k-th *vertex* $\mathbf{w}_k[n]$ in the parameter space R^n is the intersection point of n hyperplanes h_i (2) or h_i^0 (3) defined by the vectors $\mathbf{x}_i[n]$ ($j \in J_k$) and $\mathbf{e}_i[n]$ ($i \in I_k$) from the set S_k (4). The intersection point $\mathbf{w}_k[n]$ can be given by the below linear equations:

$$(\forall j \in J_k) \quad \mathbf{w}_k[n]^T\mathbf{x}_j[n] = 1 \tag{5}$$

and

$$(\forall i \in I_k) \quad \mathbf{w}_k[n]^T\mathbf{e}_i[n] = 0 \tag{6}$$

The equations (5) and (6) can be given in the below matrix form:

$$\mathbf{B}_k[n]\,\mathbf{w}_k[n] = \mathbf{1}'[n] = [1,\ldots,1,0,\ldots,0]^T \tag{7}$$

where $\mathbf{B}_k[n]$ is the square matrix, the k-th *basis* linked to the vertex $\mathbf{w}_k[n]$:

$$\mathbf{B}_k[n] = [\mathbf{x}_{j(1)}[n],...,\mathbf{x}_{j(n')}[n],\mathbf{e}_{i(n'+1)}[n],...,\mathbf{e}_{i(n)}[n]]^T \tag{8}$$

and

$$\mathbf{w}_k[n] = \mathbf{B}_k[n]^{-1}\mathbf{1}'[n] \tag{9}$$

The number of the subsets S_k (4) and the bases $\mathbf{B}_k[n]$ (8) could be very large. The same vertex $\mathbf{w}_k[n]$ can be determined by (9) more than one base $\mathbf{B}_k[n]$.

Definition 1: The *rank* r_k ($1 \le r_k \le n$) of the vertex $\mathbf{w}_k[n]$ (9) is defined as the number of the feature vectors $\mathbf{x}_j[n]$ ($j \in J_k$ (5)) in the base $\mathbf{B}_k[n]$ (8) linked (9) to this vertex.

It can be noted that the *rank* r_k of the vertex $\mathbf{w}_k[n] = [w_{k,1},\ldots,w_{k,n}]^T$ (9) is equal to the number of its nonzero components $w_{k,i}$ ($w_{k,i} \neq 0$).

Definition 2: The vertex $\mathbf{w}_k[n]$ of the rank r_k is degenerated when more than r_k hyperplanes h_j (2) pass (5) through it.

Let us note that the degenerated vertex $\mathbf{w}_k[n]$ can be defined (9) by at least two different matrices $\mathbf{B}_k[n]$ and $\mathbf{B}_{k'}[n]$ ($\mathbf{B}_k[n] \neq \mathbf{B}_{k'}[n]$ and $\mathbf{w}_k[n] = \mathbf{w}_{k'}[n]$).

Definition 3: The *degree of degeneration* of the vertex $\mathbf{w}_k[n]$ (9) of the rank r_k is defined as the number $d_k = m_k - r_k$, where m_k is the number of such feature vectors $\mathbf{x}_j[n]$ ($\mathbf{x}_j[n] \in C$) from the set C (1), which define the hyperplanes h_j (2) passing through this vertex ($\mathbf{w}_k[n]^T\mathbf{x}_j[n] = 1$).

The *degree of degeneration* d_k of the vertex $\mathbf{w}_k[n]$ (9) can be also seen as the number of different bases $\mathbf{B}_{k'}[n]$ (8) linked (9) to this vertex. The *degree of degeneration* of the vertex $\mathbf{w}_k[n]$ (9) can be defined also in a different way, for example as $d_k' = (m_k - r_k) / (m - r_k)$. It could be seen that $0 \le d_k' \le 1$.

3 Hyperplanes and Planes in the Feature Space

The hyperplanes $H(\mathbf{w}[n],\theta)$ in the feature space $F[n]$ are usually defined in the below manner [2]:

$$H(\mathbf{w}[n], \theta) = \{\mathbf{x}[n]: \mathbf{w}[n]^T\mathbf{x}[n] = \theta\} \tag{10}$$

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

The *weight vector* $\mathbf{w}[n]$ is perpendicular to $H(\mathbf{w}[n],\theta)$ and determines the orientation of this hyperplane. Changing the threshold θ causes a parallel displacement (shift) of the hyperplane $H(\mathbf{w}[n],\theta)$ (2). The dimension of the hyperplane $H(\mathbf{w}[n],\theta)$ (10) is equal to $n - 1$.

The vertex $\mathbf{w}_k[n]$ (9) of the rank r_k allows to define the $(r_k - 1)$ - dimensional *vertexical plane* $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ in the feature space $F[n]$ as the linear combination of r_k (5) feature vectors $\mathbf{x}_{j(i)}[n]$ belonging to the basis $\mathbf{B}_k[n]$ (8):

$$P_k(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n]) = \{\mathbf{x}[n]: \mathbf{x}[n] = \alpha_1\, \mathbf{x}_{j(1)}[n] + \ldots + \alpha_k\, \mathbf{x}_{j(rk)}[n]\} \tag{11}$$

where $j(i) \in J_k$ (5) and r_k parameters α_i ($\alpha_i \in R^1$) fulfill the below condition:

$$\alpha_1 + \ldots + \alpha_{rk} = 1 \tag{12}$$

If the vertex $\mathbf{w}_k[n]$ (9) has the rank $r_k = n$, then the vertexical plane (11) has the dimension equal to $(n - 1)$, similarly to the hyperplane $H(\mathbf{w}[n],\theta)$ (2). We can note that not every hyperplane $H(\mathbf{w}[n],\theta)$ (2) can be represented as $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(n)}[n])$ (11)

but the opposite statement is true. Every vertexical hyperplane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(n)}[n])$ (11) can be represented as $H(\mathbf{w}[n],\theta)$ (2).

Remark 1: The formula (11) without the condition (12) defines such an r_k - dimensional plane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ in the feature space $F[n]$ which passes through the point zero $\mathbf{0}[n]$ (*origin*).

The *line* $L_k(\mathbf{x}_{j(1)}[n],\mathbf{x}_{j(2)}[n])$ in the feature space $F[n]$ can be treated as the one-dimensional *plane* $P_k(\mathbf{x}_{j(1)}[n], \mathbf{x}_{j(2)}[n])$ (11) spanned by two different vectors $\mathbf{x}_{j(1)}[n]$ and $\mathbf{x}_{j(2)}[n]$ ($\mathbf{x}_{j(1)}[n] \neq \mathbf{x}_{j(2)}[n]$) by using only one parameter α ($\alpha \in R^1$):

$$L_k(\mathbf{x}_{j(1)}[n],\mathbf{x}_{j(2)}[n]) = \{\mathbf{x}[n]: \mathbf{x}[n] = \mathbf{x}_{j(1)}[n] + \alpha\,(\mathbf{x}_{j(2)}[n] - \mathbf{x}_{j(1)}[n])\} = \\ = \{\mathbf{x}[n]: \mathbf{x}[n] = (1-\alpha)\,\mathbf{x}_{j(1)}[n] + \alpha\,\mathbf{x}_{j(2)}[n]\} \tag{13}$$

One feature vector $\mathbf{x}_j[n]$ allows to define the line $L_0(\mathbf{x}[n])$ passing through the point $\mathbf{0}[n]$ (origin) of the feature space $F[n]$:

$$L_0(\mathbf{x}_{j(1)}[n]) = \{\mathbf{x}[\mathrm{n}]: \mathbf{x}[\mathrm{n}] = \alpha\,\mathbf{x}_{j(1)}[\mathrm{n}]\},\ \text{where } \alpha \in R^1 \tag{14}$$

Remark 2: If two feature vectors $\mathbf{x}_{i(1)}[n]$ and $\mathbf{x}_{i(2)}[n]$ are linearly dependent ($\mathbf{x}_{i(2)}[n]= c\ \mathbf{x}_{i(1)}[n]$, *where* $c \in R^1$), then the line $L(\mathbf{x}_{i(1)}[n],\mathbf{x}_{j(2)}[n])$ (13) passes through the origin, and the equation (13) can be reduced to (14).

Theorem 1: The feature vector $\mathbf{x}_i[n]$ defines hyperplane h_i (2) which passes through the vertex $\mathbf{w}_k[n]$ (9) of the rank r_k if and only if the vector $\mathbf{x}_i[n]$ is situated on the $(r_k\text{-}1)$-dimensional *vertexical plane* $P_k(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11), where $j(i)\in J_k$ (5).

Proof: Each vector $\mathbf{x}_{i(i)}[n]$ belonging to the basis $\mathbf{B}_k[n]$ (8) fulfils the equation $\mathbf{w}_k[n]^T\mathbf{x}_{j(i)}[n] = 1$ (5). If the point $\mathbf{x}_i[n]$ is situated on the $(r_k - 1)$-dimensional vertexical plane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) then it satisfies the equation $\mathbf{x}_i[n] = \alpha_1\mathbf{x}_{j(1)}[n] +\ldots+ \alpha_k\mathbf{x}_{j(rk)}[n]$, where $\alpha_1 +\ldots+ \alpha_{rk} = 1$ (6). Therefore, the condition $\mathbf{w}_k[n]^T\mathbf{x}_j[n] = 1$ results.

Any feature vector $\mathbf{x}_j[n]$ can be represented as the linear combination of the basis vectors $\mathbf{x}_{j(i)}[n]$ ($j(i)\in J_k$) (5) and $\mathbf{e}_j[n]$ ($i\in I_k$) (6). The basis unit vectors $\mathbf{e}_j[n]$ fulfill the condition $\mathbf{w}_k[n]^T\mathbf{e}_i[n] = 0$ (6). Therefore, if the equation $\mathbf{w}_k[n]^T\mathbf{x}_j[n] = 1$ holds for some vector $\mathbf{x}_j[n]$, then the condition (6) implies that this vector is situated on the vertexical plane $P_k(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) in the feature space $F[n]$.

We are interested in discovering such a vertexical plane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) which would contain especially numerous feature vectors $\mathbf{x}_j[n]$ (1).

Definition 4: The *vertexical plane* $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) that includes m_k feature vectors $\mathbf{x}_j[n]$ from the data set C (1) is called the *main vertexical plane* if and only if the number m_k is large in comparison to the *rank* r_k (*Definition* 2) of the vertex $\mathbf{w}_k[n]$ (at least $m_k > r_k$).

Supposition I: Discovering *main vertexical planes* $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) in the feature space $F[n]$ can be based on the detection and examination of the vertices $\mathbf{w}_k[n]$ (9) with a high *degree of degeneration* r_k (*Definition* 3).

4 Convex and Piecewise Linear (*CPL*) Criterion Functions

Let us consider a convex and piecewise linear (*CPL*) penalty functions $\varphi_j(\mathbf{w})$ defined in the below manner on the feature vectors $\mathbf{x}_j[n]$ from the data set C (1) [6]:

$$(\forall \mathbf{x}_j[n] \in C\ (1))$$

$$\varphi_j(\mathbf{w}[n]) = \begin{cases} 1 - \mathbf{w}[n]^T\mathbf{x}_j[n] & \text{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] \le 1 \\ \mathbf{w}[n]^T\mathbf{x}_j[n] - 1 & \text{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] > 1 \end{cases} \tag{15}$$

The penalty functions $\varphi_j(\mathbf{w})$ are equal to the absolute values $|1 - \mathbf{w}[n]^T\mathbf{x}_j[n]|$.

The criterion function $\Phi_m(\mathbf{w}[n])$ is defined as the weighted sum of the penalty functions $\varphi_j(\mathbf{w}[n])$ defined by feature vectors $\mathbf{x}_j[n]$ from the subset C_m ($C_m \subset C$):

$$\Phi_m(\mathbf{w}[n]) = \sum_{i \in J_m} \alpha_j\ \varphi_j(\mathbf{w}[n]) \tag{16}$$

where $J_m = \{j\colon \mathbf{x}_i[n] \in C_m\}$ and the positive parameters α_i ($\alpha_i > 0$) in the below function $\Phi_m(\mathbf{w}[n])$ can be treated as the *prices* of particular feature vectors $\mathbf{x}_j[n]$. The standard choice of the parameters α_j values is one:

$$(\forall j \in J_m\ )\ \ \alpha_j = 1.0 \tag{17}$$

The criterion function $\Phi_m(\mathbf{w}[n])$ (16) is convex and piecewise linear as the sums of the *CPL* functions $\alpha_j\varphi_j(\mathbf{w}[n])$ (15). It can be proved that the minimal value of the function $\Phi_m(\mathbf{w}[n])$ can be found in one of the vertices $\mathbf{w}_k[n]$ (9) [7]:

$$(\exists \mathbf{w}_k^*[n])\quad (\forall \mathbf{w}[n])\quad \Phi_m(\mathbf{w}[n]) \ge \Phi_m(\mathbf{w}_k^*[n]) = \Phi_m^* \ge 0 \tag{18}$$

The basis exchange algorithms which are similar to the linear programming allow to find efficiently the minimum $\Phi_m(\mathbf{w}_k^*[n])$ of the criterion functions $\Phi_m(\mathbf{w}[n])$ (16) even in the case of large, multidimensional data subsets C_m ($C_m \subset C$) (1) [8].

Theorem 2: The minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) is equal to zero ($\Phi_m(\mathbf{w}_k^*[n]) = 0$), if and only if all the feature vectors $\mathbf{x}_j[n]$ from the subset C_m ($C_m \subset C$ (1)) are situated on a hyperplane $H(\mathbf{w}[n],\theta)$ (10) with $\theta \neq 0$.

Proof: Let us suppose that all the feature vectors $\mathbf{x}_j[n]$ from the subset C_m are situated on some hyperplane $H(\mathbf{w}'[n],\theta')$ (10) with $\theta' \neq 0$:

$$(\forall \mathbf{x}_j[n] \in C_m)\quad \mathbf{w}'[n]^T\mathbf{x}_j[n] = \theta' \tag{19}$$

From this

$$(\forall \mathbf{x}_j[n] \in C_m)\quad (\mathbf{w}'[n]\ /\ \theta')^T\mathbf{x}_j[n] = 1 \tag{20}$$

The above equations mean that functions $\varphi_j(\mathbf{w}'[n]\ /\ \theta')$ (15) are equal to zero in the point $(\mathbf{w}'[n]\ /\ \theta')$:

$$(\forall \mathbf{x}_j[n] \in C_m)\quad \varphi_j(\mathbf{w}'[n]\ /\ \theta') = 0 \tag{21}$$

so

$$\Phi_m(\mathbf{w}'[n] \,/\, \theta') = 0 \tag{22}$$

On the other hand, if the criterion function $\Phi_m(\mathbf{w}'[n])$ (16) is equal to zero in some point $\mathbf{w}'[n]$, then each of the penalty functions $\varphi_j(\mathbf{w}'[n])$ (15) has to be equal to zero:

$$(\forall \mathbf{x}_j[n] \in C_m) \quad \varphi_j(\mathbf{w}'[n]) = 0 \tag{23}$$

or

$$(\forall \mathbf{x}_j[n] \in C_m) \quad \mathbf{w}'[n]^T \mathbf{x}_j[n] = 1 \tag{24}$$

The above equations mean that each feature vector $\mathbf{x}_j[n]$ from the subset C_m is located on the hyperplane $H(\mathbf{w}'[n], 1)$ (10). ⇳

Taking into account that the minimal value (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) can be located in one of the vertices $\mathbf{w}_k[n]$ (9), the *Theorem* 2 can be reformulated in the below manner:

Theorem 2′: The minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) is equal to zero ($\Phi_m(\mathbf{w}_k^*[n]) = 0$), if and only if all the feature vectors $\mathbf{x}_j[n]$ from the subset C_m ($C_m \subset C$ (1)) are situated on some vertexical plane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(l)}[n])$ (11) (12) which does not pass through the point zero $\mathbf{0}[n]$ (*origin*).

The below theorem characterizes the *invariance property* of the value $\Phi_k(\mathbf{w}_k^*[n])$:

Theorem 3: The minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) does not depend on linear, non-singular data transformations of the feature vectors $\mathbf{x}_j[n]$ from the subset C_m ($C_m \subset C$ (1)):

$$\Phi_m'(\mathbf{w}_k'[n]) = \Phi_m(\mathbf{w}_k^*[n]) \tag{25}$$

where $\Phi_m'(\mathbf{w}_k'[n])$ is the minimal value of the criterion functions $\Phi_m'(\mathbf{w}[n])$ (16) defined on the transformed feature vectors $\mathbf{x}_j'[n]$:

$$(\forall \mathbf{x}_j[n] \in C_m) \quad \mathbf{x}_j'[n] = \boldsymbol{A}[n]\, \mathbf{x}_j[n] \tag{26}$$

where $\boldsymbol{A}[n]$ is a non-singular matrix of dimension (n x n) ($\boldsymbol{A}^{-1}[n]$ exists).

Proof: The values $\varphi_i'(\mathbf{w}[n])$ of the penalty function $\varphi_i(\mathbf{w}[n])$ (15) in a point $\mathbf{w}'[n]$ are defined in the below manner on the transformed feature vectors $\mathbf{x}_j'[n]$ (26):

$$(\forall \mathbf{x}_j'[n] \in C_m) \quad \varphi_j'((\mathbf{w}'[n]) = |1 - \mathbf{w}'[n]^T \mathbf{x}_j'[n]| = |1 - \mathbf{w}'[n]^T \boldsymbol{A}[n] \mathbf{x}_j[n]| \tag{27}$$

If we take

$$\mathbf{w}_k'[n] = (\boldsymbol{A}[n]^T)^{-1}\, \mathbf{w}_k^*[n] \tag{28}$$

we obtain the below result

$$(\forall \mathbf{x}_j[n] \in C_m) \quad \varphi_j'(\mathbf{w}_k'[n]) = \varphi_j(\mathbf{w}_k^*[n]) \tag{29}$$

The above equations mean that the value $\Phi_m'(\mathbf{w}_k'[n])$ of the criterion functions $\Phi_m'(\mathbf{w}[n])$ (16) defined in the point $\mathbf{w}_k'[n]$ (28) on the transformed feature vectors

$\mathbf{x}_i'[n]$ (26) is equal to the minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) defined on the feature vectors $\mathbf{x}_j[n]$.

The minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) can be characterized by two below *monotonocity properties*:

i. The positive monotonocity property due to reduction of feature vectors $\mathbf{x}_i$[n]
Neglecting some feature vectors $\mathbf{x}_j[n]$ cannot result in an increase of the minimal value $\Phi_m(\mathbf{w}_m^*[n])$ (19) of the criterion function $\Phi_m(\mathbf{w}[n])$ (17):

$$(C_{m'} \subset C_m) \Rightarrow (\Phi_{m'}^* \leq \Phi_m^*) \tag{30}$$

where the symbol $\Phi_{m'}^*$ stands for the minimal value (18) of the criterion function $\Phi_{m'}(\mathbf{w}[n])$ (16) defined on the elements $\mathbf{x}_j[n]$ of the subset $C_{m'}$ ($\mathbf{x}_j[n] \in C_{m'}$).

The relation (30) can be justified by the remark that neglecting some feature vectors $\mathbf{x}_j[n]$ results in neglecting some non-negative components $\alpha_j\ \varphi_j(\mathbf{w}[n])$ (15) in the criterion function $\Phi_m(\mathbf{w}[n])$ (16).

ii. The negative monotonicity property due to reduction of features x_i
The reduction of the feature space $F[n]$ to $F'[n']$ by neglecting some features x_i cannot result in a decrease of the minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_k(\mathbf{w}[n])$ (16):

$$(F'[n'] \subset F[n]) \Rightarrow (\Phi_m' \geq \Phi_m^*) \tag{31}$$

where the symbol Φ_m' stands for the minimal value (18) of the criterion function $\Phi_m(\mathbf{w}[n'])$ (16) defined on the reduced vectors $\mathbf{x}_i'[n']$ ($\mathbf{x}_i'[n'] \in F'[n']$, $n' < n$). The relation (31) results from the fact that the neglecting some features x_i is equivalent to imposing additional constraints "$w_i = 0$" in the parameter space R^n..

5 Procedure of the Main Vertexical Planes Discovering

The feature vector $\mathbf{x}_i[n]$ is *included* in the vertexical plane $P(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) if the below equation holds:

$$\mathbf{x}_j[n] = \alpha_{j,1}\ \mathbf{x}_{j(1)}[n] + \ldots + \alpha_{j,rk}\ \mathbf{x}_{j(rk)}[n] \tag{32}$$

with the condition (12):

$$\alpha_{j,1} + \ldots + \alpha_{j,rk} = 1 \tag{33}$$

Definition 4: The *vertexical plane* $P_k(\mathbf{x}_{i(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) that includes m_k feature vectors $\mathbf{x}_j[n]$ from the data set C (1) is called the *main vertexical plane* if and only if the number m_k is a large in comparison with to the *rank* r_k (*Definition* 2) of the vertex $\mathbf{w}_k[n]$ (at least $m_k > r_k$).

The below multistage *Procedure Vertex* is proposed for discovering the main vertical plane $P_m(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) on the basis of the data set $C_m = C$ (1):

Procedure Vertex (34)

i. Find the minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) and the optimal vertex $\mathbf{w}_k^*[n]$ of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) defined on elements $\mathbf{x}_j[n$ of the subset C_m
ii. If $\Phi_m(\mathbf{w}_k^*[n]) = 0$, then the *Procedure Vertex* is **stopped** in the optimal vertex $\mathbf{w}_k^*[n]$, otherwise the next step is executed
iii. Find the vector $\mathbf{x}_{j'}[n]$ in the feature subset C_m with the highest value of the penalty function $\varphi_j(\mathbf{w}[n])$ (15) in the optimal vertex $\mathbf{w}_k^*[n]$ (18)

$$(\forall \mathbf{x}_j[n] \in C_m) \;\; \varphi_{j'}(\mathbf{w}_k^*[n]) \geq \varphi_j(\mathbf{w}_k^*[n]) \tag{35}$$

or with the parameters α_j (16) taking into account:

$$(\forall \mathbf{x}_j[n] \in C_m) \;\; \alpha_{j'}\varphi_{j'}(\mathbf{w}_k^*[n]) \geq \alpha_j \; \varphi_j(\mathbf{w}_k^*[n]) \tag{36}$$

iv. Remove the feature vector $\mathbf{x}_{j'}[n]$ from the subset C_m ($C_m \rightarrow C_m / \{ \mathbf{x}_{j'}[n] \}$) and go to the step i.

It can be proved that the *Procedure Vertex* is **stopped** in the optimal vertex $\mathbf{w}_k^*[n]$ (18) of the rank r_k after finite number of steps. This property is based on the *Theorem* 2 and on the *monotonocity property* (30).

Let the symbol C_m^* stand for optimal subset of feature vectors $\mathbf{x}_j[n]$ which is obtained when the *Procedure Vertex* is **stopped**.

Remark 3: The *degree of degeneration* d_k of the optimal vertex $\mathbf{w}_k^*[n]$ (18) (*Definition* 3) is equal to the difference between the number m_k of elements $\mathbf{x}_j[n]$ of the optimal subset C_m^* and the *rank* r_k of this vertex (*Definition* 2):

$$d_k = m_k - r_k \tag{37}$$

In accordance with the *Definition* 4, the *main vertexical plane* $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) based on the optimal vertex $\mathbf{w}_k^*[n]$ (18) of the *rank* r_k should contain especially numerous feature vectors $\mathbf{x}_j[n]$ (1) or, in other words, should have a high value of the degree of degeneration d_k (37).

There is no guarantee that the optimal vertex $\mathbf{w}_k^*[n]$ (18) resulting from the *Procedure Vertex* (34) will define the *main vertexical plane* $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) with the highest value of the *degree of degeneration* d_k (37). Modifications to the *Procedure Vertex* (34) could allow to find the plane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) with a higher degree of degeneration d_k (37). One of these modification would be to replace the rule (35) in the step iii. with the below rule:

iii'. Find such feature vector $\mathbf{x}_{j'}[n]$ ($\mathbf{x}_{j'}[n] \in C_m$) which when removed from the subset C_m causes the largest decrease $\Delta_j(\mathbf{w}_k^*[n])$ of the minimal value $\Phi_m(\mathbf{w}_k^*[n])$ (18) of the criterion function $\Phi_m(\mathbf{w}[n])$ (16):

$$(\forall \mathbf{x}_j[n] \in C_m) \;\; \Delta_{j'}(\mathbf{w}_k^*[n]) \geq \Delta_j(\mathbf{w}_k^*[n]) \tag{38}$$

The *Procedure Vertex* (34) may allow for discovering more than one vertexical plane $P_m(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) from a data set C (1) in subsequent cycles l. The optimal subset $C_{m(1)}{}^*$ is found during the first cycle (l = 1) of the *Procedure Vertex* (34), which began on the full data set C_1 = C (1). The second cycle (l = 2) (34) begins on the reduced data set $C_2 = C_1 / C_{m(1)}{}^*$ and allows to find the optimal subset $C_{m(2)}{}^*$. A possible third cycle (l = 3) begins on the data set set $C_3 = C_2 / C_{m(2)}{}^*$ and allows to find the optimal subset $C_{m(3)}{}^*$ and so on. Subsequent cycles l allow to generate the sequence of the K optimal subset $C_{m(l)}{}^*$ and the optimal vertices $\mathbf{w}_{k(l)}{}^*[n]$ (18):

$$(C_{m(1)}{}^*,\mathbf{w}_{k(1)}{}^*[n]), (C_{m(2)}{}^*,\mathbf{w}_{k(2)}{}^*[n]),\ldots,(C_{m(K)}{}^*,\mathbf{w}_{k(K)}{}^*[n]), \tag{39}$$

As a result, the sequence of the K vertexical planes $P_{m(l)}(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) can be generated on the basis of the optimal vertices $\mathbf{w}_{k(l)}{}^*[n]$ (39). The sequence of the K vertexical planes $P_{m(l)}(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) allows among others to divide the data sets C (1) into K subsets C(l) centered around K planes $P_{m(l)}(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11). Such procedure can be called as the *K – plane* clustering.

The sequence (39) is generated through gradual reduction of the data set C (1) by successive removing of the optimal subsets $C_{m(l)}{}^*$. The reduced subsets $C_{m(l)}{}^*$ may be enlarged in successive cycles l, which could improve generalization power of the proposed procedure of the *K – plane* clustering. For this purpose, the step ii. in the *Procedure Vertex* (34) can be replaced, by the below one with a small, positive parameter ε (ε > 0):

ii′. If $\Phi_{m(l)}(\mathbf{w}_k{}^*[n]) \le \varepsilon$, then the *Procedure Vertex* is **stopped** in the optimal vertex $\mathbf{w}_k{}^*[n]$, in the other case the next step is executed

The *Procedure Vertex* (34) with the parameter ε equal to zero (ε = 0) generates the optimal subset $C_{m(l)}{}^*$ constituted by feature vectors $\mathbf{x}_j[n]$ located precisely on the optimal plane $P_{m(l)}(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11). If the parameter ε becomes greater than zero (ε > 0), then the optimal subset $C_{m(l)}{}'$ may contain both the feature vectors $\mathbf{x}_j[n]$ located on the optimal plane $P_{m(l)}(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (11) as well as near this plane:

$$C_{m(l)}{}' = \{\mathbf{x}_j[n]: \mathbf{x}_j[n] \in C_{m(l)} \textit{ and } \Phi_{m(l)}(\mathbf{w}_k{}^*[n]) \le \varepsilon\} \tag{40}$$

where $\Phi_{m(l)}(\mathbf{w}_k{}^*[n])$ is the minimal value $\Phi_{m(l)}(\mathbf{w}_k{}^*[n])$ (18) of the criterion function $\Phi_{m(l)}(\mathbf{w}[n])$ (16) defined on elements $\mathbf{x}_j[n]$ of the subset $C_{m(l)}$ ($C_{m(l)} \subset C$ (1)).

6 Modified *CPL* Criterion Functions with Feature Costs

The modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ includes additional *CPL* penalty functions $\phi_i(\mathbf{w}[n])$ in the form of the absolute values $|w_i|$:

$$(\forall i \in \{1,\ldots,n\})$$
$$\phi_i(\mathbf{w}[n]) = |w_i| = \begin{cases} -\mathbf{e}_i[n]^T\mathbf{w}[n] & \textit{if } \mathbf{e}_i[n]^T\mathbf{w}[n] < 0 \\ \mathrm{e}_i[n]^T\mathrm{w}[n] & \textit{if } \mathrm{e}_i[n]^T\mathrm{w}[n] \ge 0 \end{cases} \tag{41}$$

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

The modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ is a weighted sum of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) and the cost functions $\phi_i(\mathbf{w}[n])$ (41), where $i \in I = \{1,..., n\}$:

$$\Psi_{m\lambda}(\mathbf{w}[n]) = \Phi_m(\mathbf{w}[n]) + \lambda \sum_{i\in I} \gamma_i\, \phi_i(\mathbf{w}[n]) = \Phi_m(\mathbf{w}[n]) + \lambda \sum_{i\in I} \gamma_i\, |w_i| \tag{42}$$

where λ is the *cost level* ($\lambda \geq 0$), γ_i – is the *cost* of the feature x_i ($\gamma_i > 0$), typically $\gamma_i = 1$.

Similarly as the function $\Phi_m(\mathbf{w}[n])$ (16), the modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) is convex and piecewise linear (*CPL*). The basis exchange algorithms allow to find efficiently the minimum $\Psi_{m\lambda}(\mathbf{w}_{k\lambda}^*[n])$ of the criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) in one of the vertices $\mathbf{w}_k[n]$ (9) []:

$$(\exists \mathbf{w}_{k\lambda}^*[n])\ \ (\forall \mathbf{w}[n])\ \ \Psi_{m\lambda}(\mathbf{w}[n]) \geq \Psi_{m\lambda}(\mathbf{w}_{k\lambda}^*[n]) \tag{43}$$

The modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) is used in the *relaxed linear separability* (*RLS*) method of feature subset selection []. The reduction of unimportant features x_i in the cost sensitive manner is based in the *RLS* method on componets $w_{k,i}^*$ of the optimal wertex $\mathbf{w}_{k\lambda}^*[n] = [w_{k,1}^*,\ldots,w_{k,n}^*]^T$ (43) [9].

$$(w_{k,i}^* = 0) \Rightarrow (\text{the } i\text{-th feature } x_i \text{ is reduced}) \tag{44}$$

The reduction of the i-th feature x_i means that the feature vectors $\mathbf{x}_j[n]$ lose their k-th component $x_{j,i}$. Such components $x_{j,i}$ can be removed without changing the location of the optimal vertex $\mathbf{w}_k^*[n]$ (43) or the values of inner products $\mathbf{w}_k^*[n]^T\mathbf{x}_j[n]$ in the criterion functions $\Phi_m(\mathbf{w}[n])$ (16) or $\Psi_{m\lambda}(\mathbf{w}[n])$ (42).

The regularization component $\lambda\Sigma\, \gamma_I\, |w_i|$ used in the function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) is similar to the one used in the *Lasso* method [10]. The *Lasso* method was developed in the framework of the regression analysis for the model selection purposes. The main difference between the *Lasso* and the *RLS* methods of feature selection is in the types of the basic criterion functions. The basic criterion function used in the *Lasso* method is usually the *Last squares* type. The basic criterion function used in the *RLS* method is the *CPL* type. This difference affects, inter alia, the computational techniques used for minimizing the criterion functions.

7 Oriented Graph G_m Based on Polytopes in Parameter Space

Feature vectors $\mathbf{x}_j[n]$ from the subset C_m ($C_m \subset C$ (1)) define the hyperplanes h_j (2) in the parameter space R^n. Similarly, n unit vectors $\mathbf{e}_i[n]$ define the hyperplanes h_i^0 (3). The hyperplanes h_j (2) and h_i^0 (3) divide the parameter space R^n into the disjoined sets (*convex polytopes*) P_l with walls, vertices $\mathbf{w}_k[n]$ (9), and edges $l_{k,k'}$ which are characterized by the below properties:

- none of the hyperplanes h_j (2) or h_i^0 (3) intersect the set P_l (45)
- each *wall* of the polytope P_l is formed by one hyperplane h_j (2) or h_i^0 (3)
- each *wertex* $\mathbf{w}_k[n]$ (9) of the set P_l is the intersection point of at least n hyperplanes h_j (2) or h_i^0 (3)
- each *edge* $l_{k,k'}$ connects two neighboring vertices $\mathbf{w}_k[n]$ and $\mathbf{w}_{k'}[n]$ of the set P_l and can be defined as the intersection of $n - 1$ hyperplanes h_j (2) or h_i^0 (3)

We can remark that both the criterion function $\Phi_m(\mathbf{w}[n])$ (16) as well as the modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) are linear inside of each polytope P_l []. The gradient $\nabla\Phi_m(\mathbf{w}[n])$ of the criterion function $\Phi_m(\mathbf{w}[n])$ (16) inside selected polytope P_l is constant and can be given by the below expression ($\mathbf{w}[n] \in P_l$):

$$\nabla\Phi_m(\mathbf{w}[n]) = \sum_{j \in J_m} \alpha_j \, s_j(\mathbf{w}[n]) \, \mathbf{x}_j[n] \tag{46}$$

where

$$\begin{aligned} s_j(\mathbf{w}[n]) &= 1 \quad \textit{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] > 1 \quad \textit{and} \\ s_j(\mathbf{w}[n]) &= -1 \quad \textit{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] < 1 \end{aligned} \tag{47}$$

The gradient $\nabla\Psi_{m\lambda}(\mathbf{w}[n])$ of the modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) can be specified in a similar manner in a point $\mathbf{w}[n] = [w_1,\ldots,w_n]^T$ from the polytope P_l:

$$\nabla\Psi_{m\lambda}(\mathbf{w}[n]) = \nabla\Phi_m(\mathbf{w}[n]) + \lambda \sum_{i \in I} \gamma_i \, s_i^0(\mathbf{w}[n]) \, \mathbf{e}_i[n] \tag{48}$$

where

$$\begin{aligned} s_j^0(\mathbf{w}[n]) &= 1 \quad \textit{if} \quad w_i > 0 \quad \textit{and} \\ s_j^0(\mathbf{w}[n]) &= -1 \quad \textit{if} \quad w_i < 0 \end{aligned} \tag{49}$$

Remark 5: The gradient $\nabla\Psi_{m\lambda}(\mathbf{w}[n])$ of the modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) can be reduced to the gradient $\nabla\Phi_m(\mathbf{w}[n])$ (46) of the criterion function $\Phi_m(\mathbf{w}[n])$ inside each polytope P_l by reducing the cost level λ to zero.

The gradient $\nabla\Psi_{m\lambda}(\mathbf{w}[n])$ (48) of the modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) allows for the below orientation of the edges $\boldsymbol{l}_{k,k'}$ connecting two neighboring vertices $\mathbf{w}_k[n]$ and $\mathbf{w}_{k'}[n]$ of particular polytopes P_l:

$$\begin{aligned} \boldsymbol{l}_{k,k'} &= \mathbf{w}_{k'}[n] - \mathbf{w}_k[n] \quad \textit{if} \quad \nabla\Psi_{m\lambda}(\mathbf{w}[n])^T(\mathbf{w}_{k'}[n] - \mathbf{w}_k[n]) < 0 \quad \textit{and} \\ \boldsymbol{l}_{k,k'} &= \mathbf{w}_k[n] - \mathbf{w}_{k'}[n] \quad \textit{if} \quad \nabla\Psi_{m\lambda}(\mathbf{w}[n])^T(\mathbf{w}_k[n] - \mathbf{w}_{k'}[n]) < 0 \end{aligned} \tag{50}$$

Remark 4: Each edge $\boldsymbol{l}_{k,k'}$ is oriented (50) decreasingly to the criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42). This means that the move between two vertices $\mathbf{w}_k[n]$ and $\mathbf{w}_{k'}[n]$ in accordance with the edge $\boldsymbol{l}_{k,k'}$ (50) always causes a decrease of the function $\Psi_{m\lambda}(\mathbf{w}[n])$.

The oriented graph G_m is defined on the basis of the set of the vertices $\{\mathbf{w}_k[n]\}$ (9) and the set of the oriented edges $\{\boldsymbol{l}_{k,k'}\}$ (50):

$$G_m = (\{\mathbf{w}_k[n]\}, \{\boldsymbol{l}_{k,k'}\}) \tag{51}$$

Remark 5: The oriented graph G_m (51) has no loops [8].

Any algorithm based on moving between vertices $\mathbf{w}_k[n]$ of the graph G (52) in accordance with the edges $\boldsymbol{l}_{k,k'}$ orientation (50) reaches the optimal vertex $\mathbf{w}_{k\lambda}^*[n]$ (43) after a finite number of steps. The oriented graph G_m (51) with the above properties has been used in the proof of the basis exchange algorithm convergence in a finite number of steps []. These types of algorithms was been used among others for the minimization of the criterion functions $\Phi_m(\mathbf{w}[n])$ (16) or $\Psi_{m\lambda}(\mathbf{w}[n])$ (42).

The modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) can be used for the purpose of reducing the *rank* r_k (*Definition* 1) of the optimal vertices $\mathbf{w}_{k\lambda}{}^*[n]$ (43). We can infer on the basis of the formula (42) that an increase of the the *cost level* λ causes an increase of the number of the components $w_{k,i}{}^*$ of the vector $\mathbf{w}_{k\lambda}{}^*[n]$ (43) equal to zero ($w_{k,i}{}^* = 0$). Therefore we can arbitrarily reduce the rank r_k of the optimal vertex $\mathbf{w}_{k\lambda}{}^*[n]$ (43) by choosing a sufficiently large value of the parameter λ (42).

Supposition II: Discovering *main vertexical planes* $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) (12) in the feature space $F[n]$ can be based on the detection and examination of vertices $\mathbf{w}_k[n]$ of the oriented graph G_m (51) with a large number of edges $\boldsymbol{l}_{k,k'}$.

Discovering *"flat patterns"* located on the *main vertexical planes* $P_k(\mathbf{x}_{i(1)}[n],\ldots, \mathbf{x}_{i(rk)}[n])$ (11) (12) of different ranks r_k allows to design different models of linear interactions between features x_i or objects $\mathbf{x}_j[n]$. Different linear models of interactions (relations) between features x_i or objects $\mathbf{x}_j[n]$ of a given *"flat pattern"* can be determined by using the base matrices $\mathbf{B}_k[n]$ (9) linked to vertices $\mathbf{w}_k[n]$ (9) of different ranks r_k.

8 An Example - A Toy Data Set in a Two-Dimensional Feature Space

Feature vectors $\mathbf{x}_j[n]$ situated on the main vertexical plane $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) (12) define the hyperplanes h_j (2) passing through the vertex $\mathbf{w}_k[n]$ (9) of the rank r_k, which is characterized by a high degree of degeneration d_k. Such vertex $\mathbf{w}_k[n]$ in the graph G_m (51) is characterized by a large number of the oriented edges $\{\boldsymbol{l}_{k,k'}\}$ (50).

To illustrate this property the artificial data sets shown in the *Table* 1 have been used. The *Table* 1 contains feature vectors $\mathbf{x}_j[2]$ situated along three lines (*Figure* 1). The resulting graph G_m (51) contains three degenerated vertices (*Figure* 2).

Table 1. The artificial data sets *Line* I, *Line* II and *Line* III of two-dimensional feature vectors $\mathbf{x}_j[2] = [x_{j,1}, x_{j,2}]^T$ ($\mathbf{x}_j[2] \in R^2$)

Number j	*Line* I $[x_{j,1}, x_{j,2}]$	*Line* II $[x_{j,1}, x_{j,2}]$	*Line* III $[x_{j,1}, x_{j,2}]$
1	[-1, 1]	[2, 1]	[3, -2]
2	[0, 1]	[0, -1]	[1, 0]
3	[1, 1]	[1, 0]	[0, 1]

The sets from the *Table* 1 are represented on the *Figure*1 and the *Figure*2.

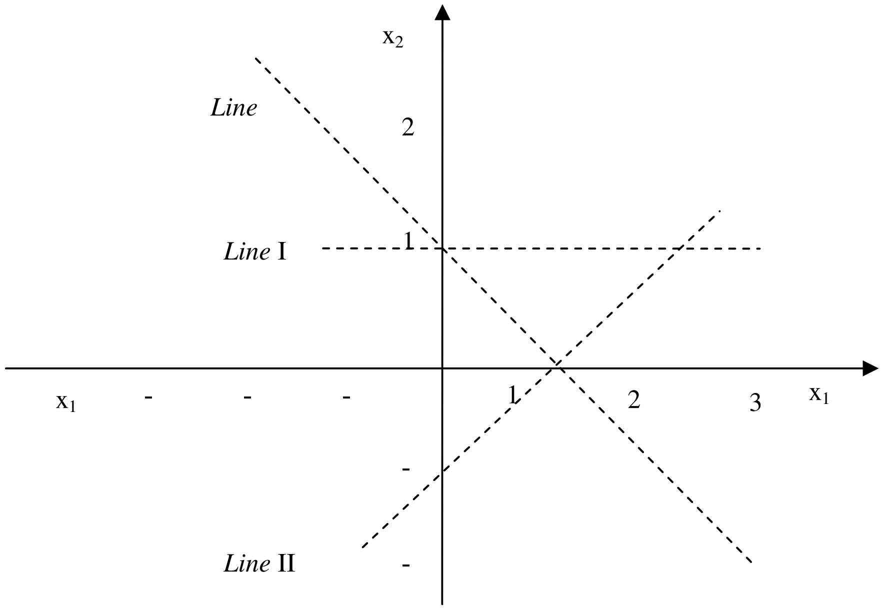

Fig. 1. Vertexical planes (lines) $P_m(\mathbf{x}_{j(1)}[2],\mathbf{x}_{j(2)}[2])$ (11) generated by the data sets *Line* I, *Line* II and *Line* III from the *Table* 1

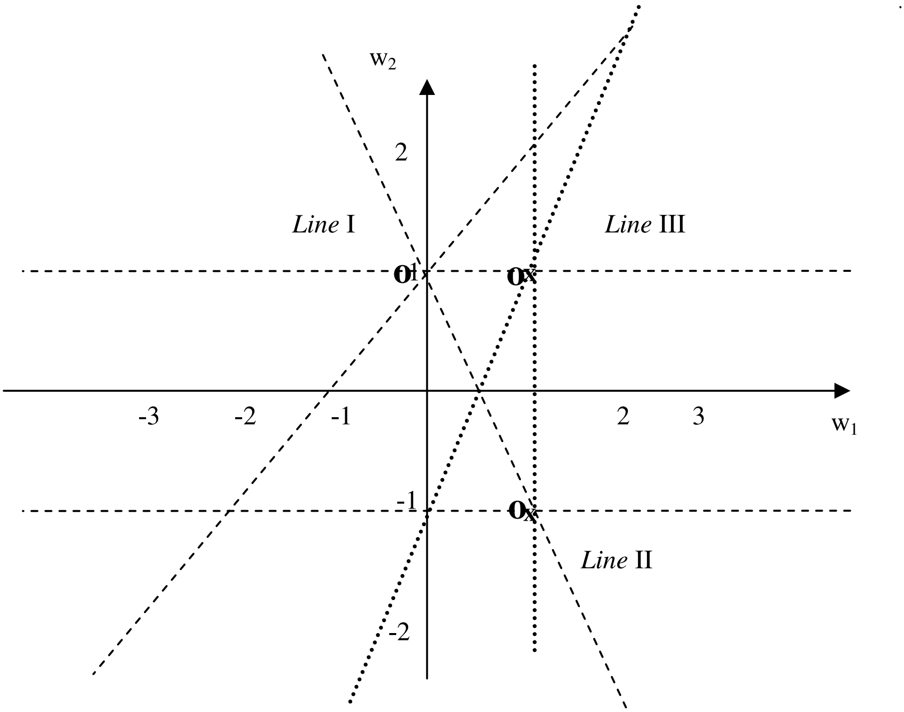

Fig. 2. Representation of the data sets *Line* I, *Line* II and *Line* III by the graph G_m (51)

9 Concluding Remarks

The main vertexical planes $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) (12) contain *"flat patterns"* in the feature space $F[n]$. The proposed procedure (34) of vertexical planes discovering from multivariate data set C (1) is based on the multiple minimization of the of the criterion function $\Phi_m(\mathbf{w}[n])$ (16).

The modified criterion function $\Psi_{m\lambda}(\mathbf{w}[n])$ (42) allows to reduce dimensionality $r_k - 1$ of the main vertexical planes $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) (12). Such planes with varied dimensionality could be useful, among others, in creating different degree models of interaction between features x_i or objects $\mathbf{x}_j[n]$.

The method of linking the main vertexical planes $P_k(\mathbf{x}_{j(1)}[n],\ldots, \mathbf{x}_{j(rk)}[n])$ (11) (12) in the feature space $F[n]$ with the degenerated vertices (9) of the oriented graphs G_m (51) is also described in the paper. This relationship which is based on common vertices $\mathbf{w}_k[n]$ (9) could create a bridge between data mining methods and graph methods. Such relationship could be used, for example, in modeling of social networks or in the decomposition of *mixture models* (e.g. [11]).

Acknowledgments. This work was supported by the project S/WI/2/2014 from the Białystok University of Technology, Poland.

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. Hand, D., Smyth, P., Mannila, H.: Principles of data mining. MIT Press, Cambridge (2001)
4. Bobrowski, L., Bezdek, J.C.: C-means clustering with the L_1 and L_∞ norms. IEEE Transactions on Systems Man and Cybernetics 21(3), 545–554 (1991a)
5. Bradley, P.S., Mangasarian, O.L.: K-plane clustering. Journal of Global Optimization 16, 23–32 (2000)
6. Bobrowski, L.: K-lines clustering with convex and piecewise linear (CPL) functions. In: MATHMOD, Vienna (2012)
7. Bobrowski, L.: Data mining based on convex and piecewise linear criterion functions. Technical University Białystok (2005) (in Polish)
8. Bobrowski, L.: Design of piecewise linear classifiers from formal neurons by some basis exchange technique. Pattern Recognition 24(9), 863–870 (1991)
9. Bobrowski, L., Łukaszuk, T.: Relaxed Linear Separability (*RLS)* Approach to Feature (Gene) Subset Selection. In: Xia, X. (ed.) Selected Works in Bioinformatics, pp. 103–118. INTECH (2011)
10. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58(1), 267–288 (1996)
11. Kwedlo, W.: A New Method for Random Initialization of the EM Algorithm for Multivariate Gaussian Mixture Learning. In: Burduk, R., Jackowski, K., Kurzynski, M., Wozniak, M., Zolnierek, A. (eds.) CORES 2013. AISC, vol. 226, pp. 85–94. Springer, Heidelberg (2013)