Flat Patterns Extraction with Collinearity Models

Leon Bobrowski[1,2], Paweł Zabielski[1]

[1]Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland
[2]Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

l.bobrowski@pb.edu.pl, p.zabielski@pb.edu.pl

Abstract

The term *collinear* (*flat*) pattern means in this article, a set of a large number of feature vectors located on (or near) a plane in multidimensional feature space. Flat patterns extracted from large data set can provide a basis for modeling a local interactions in selected sets of features. Collinear patterns can be discovered in given data set through minimization of some kind of the convex and piecewise linear (*CPL*) criterion functions.

Keywords: data mining, flat patterns, CPL criterion functions, margins

1 Introduction

Data mining tools are used to extraction patterns from multivariate data sets (Hand and Smyth, 2001). The data sets considered in this article are assumed to be formed by the structuralized feature vectors of the same dimensionality and can be represented as the matrices. The word *pattern* means a data subset with a certain type of regularity. The overall goal of the data mining process is to obtain useful information on the basis of the extracted patterns.

The term collinear (*flat*) pattern means a subset of a large number of feature vectors located on and around selected hyperplanes in a certain feature subspace. Discovered collinear patterns can be used also for creating models of linear interaction between many selected features (genes).

Flat patterns can be discovered in data sets through minimization of a certain type of the convex and piecewise linear (*CPL*) criterion functions (Bobrowski, 2014). The basis exchange algorithms can be used for the *CPL* functions minimization. The role the margin in a special type of the *CPL* functions in the flat patterns discovering is examined in the presented paper. A special type of the *CPL* functions gives opportunity to discover the so called *layered patterns* in the feature space.

2 Data subsets in feature subspaces

Let consider the data set C composed of m feature vectors $\mathbf{x}_j = \mathbf{x}_j[n] = [x_{j,1},...,x_{j,n}]^T$ which represent the objects (patients) O_j and belong to a given n-dimensional feature space $F[n]$ ($\mathbf{x}_j \in F[n]$):

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

The feature space $F[n] = \{ x_1,\ldots, x_n\}$ is composed of n features x_i ($i\in I = \{1,...,n\}$). The i-th component $x_{j,i}$ ($x_{j,i} \in R$ or $x_{j,i} \in \{0,1\}$) of the feature vector $\mathbf{x}_j$ is the numerical value of the feature x_i measured on the j-th object O_j.

The k-th feature subspace $F_k[n_k]$ ($F_k[n_k] \subset F[n]$) is made of n_k such features x_i which have the indices i in the subset I_k ($i \in I_k \subset I$) and contains n_k - dimensional reduced vectors $\mathbf{x} = \mathbf{x}[n_k]$ ($\mathbf{x}[n_k] \in F_k[n_k]$). The reduced vectors $\mathbf{x}[n_k]$ are obtained from the feature vectors $\mathbf{x}[n] = [x_1,...,x_n]^T$ by neglecting these components x_i which represent features x_i with the indices i outside the set I_k ($i \notin I_k$). The regular hyperplane $H_k(\mathbf{w}, \theta)$ in the k-th feature subspace $F_k[n_k]$ is defined in the below manner:

$$H_k(\mathbf{w}, \theta) = \{\mathbf{x}: \mathbf{w}^T\mathbf{x} = \theta\} \tag{2}$$

where $\mathbf{x} = [x_1,...,x_{nk}]^T$ is the reduced feature vector ($\mathbf{x}\in F_k[n_k]$), $\mathbf{w} = [w_1,...,w_{nk}]^T$ is the reduced weight vector ($\mathbf{w} \in R^{nk}$) and θ is the threshold ($\theta \in R^1$).

Definition 1: The hyperplane $H_k(\mathbf{w},\theta)$ in the k-th feature subspace $F_k[n_k]$ is *regular* if and if the threshold θ and the weights $w_{,i}$ are different from zero:

$$(\theta \neq 0) \ \ and \ \ (\forall i\in\{1,\ldots, n_k\}) \ \ w_i \neq 0 \tag{3}$$

The k-th data subset $C_k[n_k]$ is constituted by such m_k reduced vectors $\mathbf{x}_j$ ($\mathbf{x}_j \in F_k[n_k]$) which have the indices j from the given subset J_k ($j \in J_k \subset J = \{1,..., m\}$):

$$C_k = C_k[n_k] = \{\mathbf{x}_j: j \in J_k\} \tag{4}$$

The k-th data subset $C_k[n_k]$ (3) can be represented also as the matrix $M[m_k * n_k]$ with the m_k rows and n_k columns. The rows of the matrix $M[m_k * n_k]$ are constituted by particular feature vectors $\mathbf{x}_j$ ($j \in J_k$). Similar representation of data sets is used in the biclustering methods. We pay attention to the data subsets $C_k[n_k]$ (3) with a *collinear* (*flat*) structure based on regular hyperplanes $H_k(\mathbf{w}, \theta)$ (2) in the feature subspace $F_k[n_k]$.

Definition 2: The data subset $C_k[n_k]$ (4) formed by a large number m_k of reduced vectors $\mathbf{x}_j = \mathbf{x}_j[n_k]$ constitutes the *collinear* (*flat*) *pattern* P_k if all elements $\mathbf{x}_j$ of this subset are located on a regular hyperplane $H_k(\mathbf{w}, \theta)$ (2) in the feature subspace $F_k[n_k]$:

$$(\forall \mathbf{x}_j \in C_k[n_k]) \quad \mathbf{w}^T\mathbf{x}_j = \theta \tag{5}$$

The ε - *layer* $S(\mathbf{w}, \theta)$ in the feature subspace $F_k[n_k]$ is defined on the regular hyperplane $H_k(\mathbf{w}, \theta)$ (2) in the below manner by using a small margin ε ($\varepsilon \geq 0$):

$$S(\mathbf{w}, \theta) = \{\mathbf{x}: \theta - \varepsilon \leq (\mathbf{w} / \| \mathbf{w} \|)^T \mathbf{x} \leq \theta + \varepsilon \tag{6}$$

where $\| \mathbf{w} \| = (\mathbf{w}^T \mathbf{w})^{1/2}$.

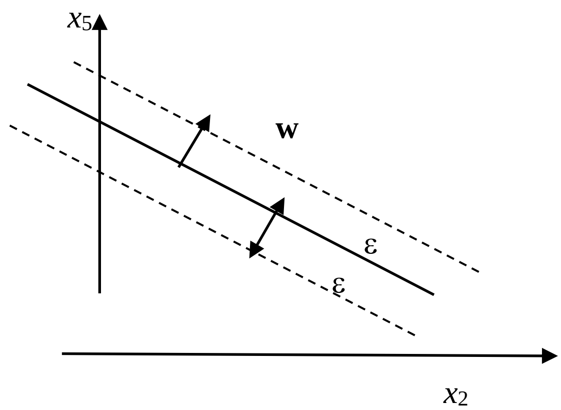

Figure 1. An example of the ε - *layer* $S(\mathbf{w}, \theta)$ (6) in the two-dimensional ($n_k = 2$) feature subspace $F_k = \{x_2, x_5\}$.

Definition 3: The data subset $C_k[n_k]$ (4) has the ε' - *collinear* structure with a margin ε' ($\varepsilon' > 0$) if it exists such weight vector $\mathbf{w}'$ and the threshold θ' that all elements $\mathbf{x}_j$ of this subset are located inside the layer $S(\mathbf{w}', \theta')$ (6):

$$(\forall \mathbf{x}_j \in C_k[n_k]) \quad \theta' - \varepsilon' \leq (\mathbf{w}')^T \mathbf{x}_j \leq \theta' + \varepsilon' \tag{7}$$

where $\| \mathbf{w}' \| = 1$ and $\theta' \neq 0$.

Because the threshold θ' is different from zero ($\theta' \neq 0$) the above inequalities can be given in the following form:

$$(\forall \mathbf{x}_j \in C_k[n_k]) \quad 1 - \varepsilon \leq \mathbf{w}^T \mathbf{x}_j \leq 1 + \varepsilon \tag{8}$$

where $\mathbf{w} = \mathbf{w}' / \theta'$ and $\varepsilon = \varepsilon' / \theta'$.

3 Dual hyperplanes and vertices in the parameter subspaces

Each of reduced feature vector $\mathbf{x}_j$ from the data subset $C_k[n_k]$ (4) defines the below dual hyperplane h_j in the n_k - dimensional parameter subspace R^{nk} ($\mathbf{w} \in R^{nk}$):

$$(\forall \mathbf{x}_j \in C_k[n_k]) \quad h_j = \{\mathbf{w}: \mathbf{x}_j^T\mathbf{w} = 1\} \tag{9}$$

Let consider the set $S_k = \{\mathbf{x}_{j(i)}\}$ of n_k linearly independent reduced feature vector $\mathbf{x}_j[n_k]$ from the subset $C_k[n_k]$ (4)

$$S_k = \{\mathbf{x}_{j(i)}: j(i) \in J_k\} \tag{10}$$

The hyperplanes $h_{j(i)}$ defined by the *basis* vectors $\mathbf{x}_{j(i)}$ from the set S_k (9) intersect at one point (*vertex*) $\mathbf{w}_k$ determined the below equations:

$$(\forall j(i) \in J_k) \quad \mathbf{x}_{j(i)}^T\mathbf{w}_k = 1 \tag{11}$$

The above equations can be given in the matrix form:

$$\mathbf{B}_k^T\mathbf{w}_k = \mathbf{1} \tag{12}$$

where $\mathbf{B}_k = [\mathbf{x}_{j(1)}, \ldots, \mathbf{x}_{j(nk)}]$ is the non-singular matrix called the k-th *basis* and $\mathbf{1} = [1,1,\ldots,1]^T$.

The k-th vertex $\mathbf{w}_k = [w_{k,1},\ldots,w_{k,nk}]^T$ (11) with the non-zero components $w_{k,i}$ ($w_{k,i} \neq 0$) allows to define the *vertexical hyperplane* $H_k(\mathbf{w}_k, 1)$ in the feature subspace $F_k[n_k]$:

$$H_k(\mathbf{w}_k, 1) = \{\mathbf{x} \in F_k[n_k]: (\mathbf{w}_k)^T\mathbf{x} = 1\} \tag{13}$$

The vertexical hyperplane $H_k(\mathbf{w}_k, 1)$ (12) is defined in the k-th feature subspace $F_k[n_k]$ composed from n_k features x_i with the indices i belonging to the subset I_k ($i \in I_k$).

Remark 1: All feature vectors $\mathbf{x}_j$ from the subset $C_k[n_k]$ (4) are situated on the hyperplane $H(\mathbf{w},\theta) = \{\mathbf{x}: \mathbf{w}^T\mathbf{x} = \theta\}$ with $\theta \neq 0$, if and only if each vector $\mathbf{x}_j$ defines such dual hyperplane h_j (8) which passes through the vertex $\mathbf{w}_k$ (10).

The *Remark* 1 has been dicussed in the paper.

4 Penalty and criterion functions aimed at extraction of collinear patterns

We consider convex and piecewise linear (*CPL*) penalty functions $\varphi_j(\mathbf{w})$ defined on the n_k - dimensional feature vectors $\mathbf{x}_j$ from the k-th data subset $C_k[n_k]$ (4):

$(\forall \mathbf{x}_j \in C_k[n_k])$

$$\varphi_j(\mathbf{w}) = \begin{cases} 1 - \varepsilon - \mathbf{w}^T\mathbf{x}_j & \text{if} \quad \mathbf{w}^T\mathbf{x}_j < 1 - \varepsilon \\ 0 & \text{if} \quad 1 - \varepsilon \leq \mathbf{w}^T\mathbf{x}_j \leq 1 + \varepsilon \\ \mathbf{w}^T\mathbf{x}_j - 1 + \varepsilon & \text{if} \quad \mathbf{w}^T\mathbf{x}_j > 1 + \varepsilon \end{cases} \qquad (14)$$

where ε is a small, non-negative parameter (*margin*).

The non-negative function $\varphi_j(\mathbf{w})$ is equal to zero ($\varphi_j(\mathbf{w}) = 0$) if and only if the feature vector $\mathbf{x}_j$ is located in the layer $S(\mathbf{w}, \theta)$ (7) with $\theta = 1$ (Fig. 2)

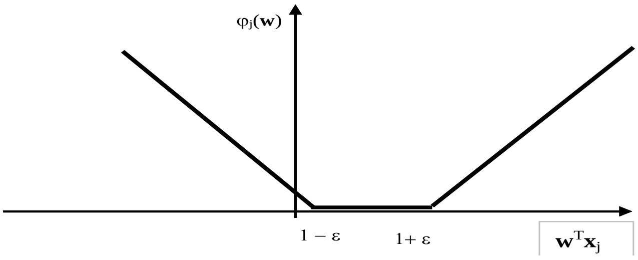

Figure 2. The *j*-th penalty functions $\varphi_j(\mathbf{w})$ (8).

The criterion function $\Phi_k(\mathbf{w})$ is defined as the weighted sum of the penalty functions $\varphi_j(\mathbf{w})$ (8) linked to m_k feature vectors $\mathbf{x}_j$ constituting the subset $C_k \subset C$ (1):

$$\Phi_k(\mathbf{w}) = \Sigma_j \, \alpha_j \, \varphi_j(\mathbf{w}) \qquad (15)$$

where the positive parameters α_j ($\alpha_j > 0$) are *prices* of particular feature vectors $\mathbf{x}_j$. The parameters α_j may depend on the number m_k of the vectors $\mathbf{x}_j$ in the subset C_k:

$$(\forall \mathbf{x}_j \in C_k) \qquad \alpha_j = 1 / m_k \qquad (16)$$

The criterion function $\Phi_k(\mathbf{w})$ (15) is convex and piecewise linear (*CPL*). It can be proved that the minimal value of the function $\Phi_k(\mathbf{w})$ can be found in one of the vertices $\mathbf{w}_k^*$ (11):

$$(\exists \, \mathbf{w}_k^*) \quad (\forall \mathbf{w}) \quad \Phi_k(\mathbf{w}) \geq \Phi_k(\mathbf{w}_k^*) = \Phi_k^* \geq 0 \qquad (17)$$

The basis exchange algorithms which are similar to the linear programming allow to find efficiently the optimal vertex $\mathbf{w}_k^*$ (19) constituting the minimal value $\Phi_k(\mathbf{w}_k^*)$ even in the case of large, multidimensional data subsets C_k (4) (Bobrowski, 2014).

For the purpose of the minimization of the criterion function $\Phi_k(\mathbf{w})$ (15) with the penalty functions $\varphi_j(\mathbf{w})$ (14) it is useful to replace each dual hyperplane h_j (9) by the two hyperplanes h_j^+ and h_j^-:

$$(\forall \mathbf{x}_j \in C_k[n_k]) \quad h_j^+ = \{\mathbf{w}: \mathbf{x}_j^T\mathbf{w} = 1 + \varepsilon\} \ \ \textit{and} \qquad (18)$$

$$h_j^- = \{\mathbf{w}: \mathbf{x}_j^T\mathbf{w} = 1 - \varepsilon\}$$

Theorem 1: If all vectors $\mathbf{x}_j$ from the subset $C_k[n_k]$ (4) can be located inside some ε - *layer* $S(\mathbf{w}',\theta')$ with $\theta' \neq 0$ (5), then the minimal value $\Phi_k(\mathbf{w}_k^*)$ (16) of the criterion function $\Phi_k(\mathbf{w})$ (14) determined on this subset is equal to zero.

Proof: If the reduced vector $\mathbf{x}_j$ is located in the ε-*layer* $S(\mathbf{w}',\theta')$ with $\theta' \neq 0$ (6), then the inequalities (7) are fulfilled for $\mathbf{w} = \mathbf{w}' / \theta'$ and $\varepsilon = \varepsilon' / \theta'$. It means, that the penalty function $\varphi_j(\mathbf{w})$ (14) is equal to zero in the point $\mathbf{w} = \mathbf{w}' / \theta'$. If all elements $\mathbf{x}_j$ of the subset C_k (4) are located inside the *layer* $S(\mathbf{w}',\theta')$, then all the penalty function $\varphi_j(\mathbf{w})$ (13) are equal to zero. It means that the value $\Phi_k(\mathbf{w}_k^*)$ (16) of the criterion function $\Phi_k(\mathbf{w})$ (14) is equal to zero in the point $\mathbf{w} = \mathbf{w}' / \theta'$.

Remark 2: The minimal value $\Phi_k(\mathbf{w}_k^*)$ (17) of the criterion function $\Phi_k(\mathbf{w})$ (15) determined on all elements $\mathbf{x}_j$ of the subset C_k (4) becomes equal to zero for a sufficiently high value of the parameter ε.

For a given data subset $C_k[n_k]$ (4) we can determine the minimum value ε_k of the parameter ε which allows to reset the minimal value $\Phi_k(\mathbf{w}_k^*)$ (17) of the criterion function $\Phi_k(\mathbf{w})$ (15) determined on this subset:

$$\varepsilon_k = \min \{\varepsilon: \Phi_k(\mathbf{w}_k^*) = 0\} \qquad (19)$$

The minimal value ε_k of the parameter ε can be computed for data subset $C_k[n_k]$ (4) through multiple minimization of the criterion function $\Phi_k(\mathbf{w})$ (15) determined on this subset.

Definition 4: The *thickness* ρ_k of the data subset $C_k[n_k]$ (4) is defined to be equal twice the value of the parameter ε_k ($\rho_k = 2\varepsilon_k$) (19).

The minimizing of the criterion function $\Phi_k(\mathbf{w})$ (15) with parameter ε less than ε_k ($0 \leq \varepsilon < \varepsilon_k$) allows also to identify in the data subsets $C_k[n_k]$ (4) a part with the greatest collinearity.

5 Vertexical hyperplanes in feature subspaces

The vertexical hyperplane $H_k(\mathbf{w}_k,1)$ (13) in the n_k - dimensional feature subspace $F_k[n_k]$ is defined by using the vertex $\mathbf{w}_k = [w_{k,1},...,w_{k.nk}]^T$ with n_k non-zero components w_i

(4). The vertex $\mathbf{w}_k$ is linked to the k-th basis $\mathbf{B}_k = [\mathbf{x}_{j(1)},\ldots \mathbf{x}_{j(nk)}]$ constituted by n_k linearly independent basis vectors $\mathbf{x}_{j(i)}$ of dimensionality n_k.

The vertexical hyperplane $H_k(\mathbf{w}_k, 1)$ (13) can be represented also in a different manner by using the n_k basis vectors $\mathbf{x}_{j(i)}$ in the feature subspace $F_k[n_k]$:

$$H_k(\mathbf{w}_k, 1) = P_k(\mathbf{x}_{j(1)},\ldots,\mathbf{x}_{j(nk)}) = \\ = \{\mathbf{x}: \mathbf{x} = \alpha_1 \mathbf{x}_{j(1)} +\ldots+ \alpha_{nk}\mathbf{x}_{j(nk)}\} \tag{20}$$

where α_i are real numbers ($\alpha_i \in R^1$) which fulfills the below condition:

$$\alpha_1 +\ldots+ \alpha_{nk} = 1 \tag{21}$$

Remark 3: The dimensionality of the vertexical hyperplane $P_k(\mathbf{x}_{j(1)},\ldots,\mathbf{x}_{j(nk)})$ (20) is equal to n_{nk} - 1.

Theorem 2: The reduced feature vector $\mathbf{x}_j$ ($\mathbf{x}_j \in F_k[n_k]$) is situated on the *vertexical hyperplane* $P_k(\mathbf{x}_{j(1)}[n],\ldots,\mathbf{x}_{j(rk)}[n])$ (20), where $j(i) \in J_k$ (10) if and only if, the dual hyperplane h_j (9) passes through the vertex $\mathbf{w}_k$ (11) .

The proof of a similar theorem can be found in the paper (Bobrowski, 2014).

Definition 5: The vertexical hyperplane $H_k(\mathbf{w}_k,1)$ (13) supports the *flat pattern* P_k if a large number m_k of the reduced vectors $\mathbf{x}_j$ are located on this hyperplane.

Definition 6: The vertexical hyperplane $H_k(\mathbf{w}_k,1)$ (13) supports the ε - *flat pattern* P_k' if a large number m_k' of the reduced vectors $\mathbf{x}_j$ are located in the ε - *layer* $S(\mathbf{w}_k,1)$ (6) around this hyperplane.

Definition 7: The *rank* r_k of the *flat patterns* P_k or P_k' is equal to the number n_k ($r_k = n_k$) of the basis vectors $\mathbf{x}_{j(i)}$ in the k-th base $\mathbf{B}_k = [\mathbf{x}_{j(1)},\ldots, \mathbf{x}_{j(nk)}]$ (12).

Definition 8: The dimensionality of of the *flat patterns* P_k or P_k' is equal to r_k - 1.

Example 1: The vertexical hyperplane $H_k(\mathbf{w}_k, 1)$ (13) in the feature subspace $F_k[2] = \{x_{i(1)}, x_{i(2)}\}$ represented as the line $l_k(\mathbf{x}_{j(1)},\mathbf{x}_{j(2)})$ spanned (19) by two basis vectors $\mathbf{x}_{j(1)}$ and $\mathbf{x}_{j(2)}$:

$$l_k(\mathbf{x}_{j(1)}, \mathbf{x}_{j(2)}) = \{\mathbf{x}: \mathbf{x} = \alpha\, \mathbf{x}_{j(1)} + (1 - \alpha)\, \mathbf{x}_{j(2)}\} \tag{22}$$

where $\alpha \in R^1$.

The *rank* r_k of the *flat patterns* P_k or P_k' supported by the line $l_k(\mathbf{x}_{j(1)},\mathbf{x}_{j(2)})$ (21) is equal 2 ($r_k = 2$).

Example 2: The vertexical hyperplane $H_k(\mathbf{w}_k, 1)$ (13) in the feature subspace $F_k[3] = \{x_{i(1)}, x_{i(2)}, x_{i(3)}\}$ represented as the plane $P_k(\mathbf{x}_{j(1)},\mathbf{x}_{j(2)},\mathbf{x}_{j(3)})$ (19) spanned by three basis vectors $\mathbf{x}_{j(i)}$:

$$P_k(\mathbf{x}_{j(1)},\mathbf{x}_{j(2)},\mathbf{x}_{j(3)}) = \{\mathbf{x}: \mathbf{x} = \alpha_1 \mathbf{x}_{j(1)}+\alpha_2 \mathbf{x}_{j(2)}+\alpha_2 \mathbf{x}_{j(2}\} \tag{23}$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $\alpha_i \in R^1$.

The *rank* r_k of the *flat patterns* P_k or P_k' supported by the plane $P_k(\mathbf{x}_{j(1)},\mathbf{x}_{j(2)},\mathbf{x}_{j(3)})$ (22) is equal 3 ($r_k = 3$).

The *flat patterns* P_k or P_k' can be extracted from the data set C (1) trough minimization of the criterion functions $\Phi_k(\mathbf{w})$ (15).

6 Properies of the criterion functions $\Phi_k(\mathbf{w})$

The criterion function $\Phi(\mathbf{w})$ is defined as the weighted sum (15) of the penalty functions $\varphi_j(\mathbf{w})$ (8) on the basis of m feature vectors $\mathbf{x}_j = \mathbf{x}_j[n]$ ($\mathbf{x}_j \in F[n]$) constituting the data set C (1). The criterion function $\Phi_k(\mathbf{w})$ is defined (15) on the basis of m_k reduced vectors $\mathbf{x}_j$ ($\mathbf{x}_j \in F_k[n_k]$) from the subset $C_k[n_k]$ (4). In accordance with the *Definition* 1, the data subset $C_k[n_k]$ (4) is the k-th the *flat pattern* P_k if all elements $\mathbf{x}_j$ of this subset can be located on the hyperplane $H_k(\mathbf{w}_k,1)$ (12). We can infer from the *Theorem* 2, that a large number m_k of the vectors $\mathbf{x}_j$ located on the vertexical hyperplane $H_k(\mathbf{w}_k,1)$ (12) causes passing m_k dual hyperplanes h_j (9) through the vertex $\mathbf{w}_k$ (11). In result, the vertex $\mathbf{w}_k$ (11) becomes highly degenerated. The minimization of the criterion functions $\Phi_k(\mathbf{w})$ (15) allows to discover highly degenerated vertices $\mathbf{w}_k$ (11) and, in result, to extract flat patterns P_k.

The folowing properies of the criterion functions $\Phi_k(\mathbf{w})$ (15). can be useful in flat patterns extraction from the data set C (1). The minimal value $\Phi_k(\mathbf{w}_k^*)$ (18) of the criterion function $\Phi_k(\mathbf{w})$ (14) can be characterized by two below *monotonocity properties* (Bobrowski, 2014):

i. The positive monotonocity due to reduction of feature vectors $\mathbf{x}_j$

Neglecting some feature vectors $\mathbf{x}_j$ the data set C (1) cannot result in an increase of the minimal value $\Phi_k(\mathbf{w}_k^*)$ (17) of the criterion function $\Phi_k(\mathbf{w})$ (15):

$$(C_{k'} \subset C_k) \Rightarrow (\Phi_{k'}^* \leq \Phi_k^*) \tag{24}$$

where the symbol Φ_k^* stands for the minimal value (18) of the criterion function $\Phi_k(\mathbf{w})$ (14) defined on the elements $\mathbf{x}_j$ of the subset C_k ($\mathbf{x}_j \in C_k$).

The implication (22) can be proved by the fact that omission of certain feature vectors $\mathbf{x}_j$ results in omission of certain non-negative components $\alpha_j\ \varphi_j(\mathbf{w})$ (14) in the criterion function $\Phi_k(\mathbf{w})$ (15).

ii. The negative monotonicity due to reduction of features x_i

The reduction of the feature space $F_k[n_k]$ to $F_{k'}[n_{k'}]$ by neglecting some features x_i cannot result in a decrease of the minimal value $\Phi_k(\mathbf{w}_k^*)$ (17) of the criterion function $\Phi_k(\mathbf{w})$ (15):

$$(F_{k'}[n_{k'}] \subset F_k[n_k]) \Rightarrow (\Phi_{k'}^* \geq \Phi_k^*) \tag{25}$$

where the symbol Φ_k^* stands for the minimal value (17) of the criterion function $\Phi_k(\mathbf{w})$ (15) defined on the reduced vectors $\mathbf{x}_j'$ ($\mathbf{x}_j' \in F_{k'}[n_{k'}]$, $n_{k'} < n_k$). The implication (25) results from the fact that the omission of certain features x_i is equivalent to imposing an additional constraint $''w_i = 0''$ during the minimization (17) in the parameter space R^{nk}.

Theorem 3: The minimal value $\Phi_k(\mathbf{w}_k^*)$ (17) of the criterion function $\Phi_k(\mathbf{w})$ (15) defined on reduced feature vectors $\mathbf{x}_j$ from the subset C_k (4) does not depend on linear, non-singular transformations of the feature vectors $\mathbf{x}_j$ from this subset:

$$\Phi_k'(\mathbf{w}_k') = \Phi_k(\mathbf{w}_k^*) \tag{26}$$

where $\Phi_k'(\mathbf{w}_k')$ is the minimal value of the criterion functions $\Phi_k'(\mathbf{w})$ (15) defined on the transformed feature vectors $\mathbf{x}_j'[n]$:

$$(\forall \mathbf{x}_j \in C_k) \quad \mathbf{x}_j' = \boldsymbol{A}\, \mathbf{x}_j \tag{27}$$

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

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

$$(\forall \mathbf{x}_j \in C_k) \quad \varphi_j'(\mathbf{w}') = |1 - (\mathbf{w}')^T \mathbf{x}_j'| = |1 - (\mathbf{w}')^T \boldsymbol{A}\mathbf{x}_j| \tag{28}$$

If we take (17)

$$\mathbf{w}' = (\boldsymbol{A}^T)^{-1}\, \mathbf{w}_k^* \tag{29}$$

we obtain the below result

$$(\forall \mathbf{x}_j \in C_k) \quad \varphi_j'(\mathbf{w}') = \varphi_j(\mathbf{w}_k^*) \tag{30}$$

The above equation mean that the value $\Phi_k'(\mathbf{w}')$ of the criterion functions $\Phi_k'(\mathbf{w})$ (15) defined in the point $\mathbf{w}'$ (29) on the transformed feature vectors $\mathbf{x}_j'$ (26) is equal to the minimal value $\Phi_k(\mathbf{w}_k^*)$ (17) of the criterion function $\Phi_k(\mathbf{w})$ (15) defined on the feature vectors $\mathbf{x}_j$ ($\mathbf{x}_j \in C_k[n_k]$ (4)).

7 Procedure of flat pattrerns extraction

The *collinear* (*flat*) patterns P_k (*Def.* 2) can be extracted from the data set C (1) through multiple minimization of the criterion functions $\Phi_k(\mathbf{w})$ (15). The procedure *Vertex* can be used for this purpose (Bobrowski, 2014). The basic form of this procedure is given below with using the counter l:

Procedure Vertex

i. $l = 1$; $C_l = C$ (1); (31)

ii. Define the criterion function $\Phi_l(\mathbf{w})$ (15) on all elements $\mathbf{x}_j$ of the data set C_l and find the optimal vertex $\mathbf{w}_l^*$ (11) which constitutes the minimal value $\Phi_k(\mathbf{w}_l^*)$ (17) of this function.

iii. If $\Phi_l(\mathbf{w}_l^*) = 0$, then the procedure is **stopped** in the optimal vertex $\mathbf{w}_l^*$, otherwise the next step is executed

iv. Find the vector $\mathbf{x}_{j'}$ in the feature subset C_l with the highest value of the penalty function $\varphi_j(\mathbf{w})$ (14) in the optimal vertex $\mathbf{w}_l^*$ (18):

$$(\forall \mathbf{x}_j \in C_l) \quad \varphi_{j'}(\mathbf{w}_l^*) \geq \varphi_j(\mathbf{w}_l^*) \tag{32}$$

or with an additional emphasis on the parameters α_j (15):

$$(\forall \mathbf{x}_j \in C_l) \quad \alpha_{j'}\ \varphi_{j'}(\mathbf{w}_l^*) \geq \alpha_j\ \varphi_j(\mathbf{w}_l^*) \tag{33}$$

v. Remove the feature vector $\mathbf{x}_{j'}$ from the subset C_l:

$$C_l \rightarrow C_l / \mathbf{x}_{j'} \tag{34}$$

vi. Increase the counter k:

$$l \rightarrow l + 1 \tag{35}$$

vii. Go to the step *ii*.

The resulting set C_k^* (4) of feature vectors $\mathbf{x}_j$, the set J_k^* (4) of these vectors indices j, and the optimal vertex $\mathbf{w}_k^*$ (11) can be created as a result of the *Vertex* procedure:

$$C_k^* = C_l \ (34) = \{\mathbf{x}_j : j \in J_k^*\} \quad \textit{and} \qquad \mathbf{w}_k^* = \mathbf{w}_l^* \tag{36}$$

It can be proved that the *Procedure Vertex* is **stopped** in some vertex $\mathbf{w}_k^*$ (17) of after finite number of steps l. The vertex $\mathbf{w}_k^*$ resulting from the procedure fulfils the below condition (17):

$$(\forall \mathbf{w}) \quad \Phi_k(\mathbf{w}) \geq \Phi_k(\mathbf{w}_k^*) = 0 \tag{37}$$

where the criterion function $\Phi_k(\mathbf{w})$ (15) is determined on all elements $\mathbf{x}_j$ of such reduced data subset C_k which results from the *Vertex* procedure.

The vertex $\mathbf{w}_k^* = [w_{k,1}^*,...,w_{k,nk}^*]^T$ (36) obtained from the procedure *Vertex* (31) should be *regularized* before using it in the definition of the *vertexical hyperplane* $H_k(\mathbf{w}_k^*, 1)$ (13). The regularization process means in this case the neglecting of such components $w_{k,i}^*$ in the vector $\mathbf{w}_k^*$ which are equal to zero ($w_i = 0$) (*Def.* 1). The regularization means additionally the neglecting of such features x_i and components $x_{j,i}$ of the feature vectors $\mathbf{x}_j = [x_{j,1},...,x_{j,n}]^T$ from the reduced data subset C_k which are linked to weights $w_{k,i}^*$ equal to zero ($w_{k,i}^* = 0$):

$$(\forall i \in \{1,...,n\}) \ (\forall j \in J_k(36)) \quad (38)$$

if ($w_{k,i}^* = 0$), ***then*** (the i-th feature x_i and the i-th component $x_{j,i}$ of the j-th feature vector $\mathbf{x}_j$ are neglected)

Remark 4: The reduction of feature vectors $\mathbf{x}_j$ in the set C (1) in accordance with the procedure *Vertex* combined with the reduction of features x_i in accordance with the rule (38) leads in a finite number of steps l to the extraction of the collinear data subset $C_k[n_k]$ (5) composed of m_k reduced vectors $\mathbf{x}_j$ ($\mathbf{x}_j \in F_k[n_k]$) which fulfill the equation (5).

The *Remark* 4 can justify directly on the basis of the description of the procedure *Vertex* and the rule (38).

Remark 5: If the number m_k of elements $\mathbf{x}_j$ of the final subset $C_k[n_k]$ obtained in result of the procedure *Vertex* and the rule (38) is a large enough, than this subset constitutes the flat pattern P_k (5) (*Def.* 2).

The *Procedure Vertex* (31) gives possibility for discovering and extraction more than one flat pattern P_k (5) from a given data set C (1). For this purpose the data set data set C (1) can be reduced in subsequent cycles k of the below procedure:

During the first cycle (k = 1), the *Procedure Vertex* (31) is activated on the data set C_1 equal to the full data set C (1) and ends with the set C_1^* (36).

The initial data set $C_1 = C$ (1) is reduced by the final set C_1^* (36) after the first cycle:

$$C_2 = C_1 \ / \ C_1^* = C \ / \ C_1^* \quad (39)$$

The second cycle (k = 2) is activated on the data set C_2 and ends with the set C_2^*:

$$C_3 = C_2 \ / \ C_2^* \quad (40)$$

The third cycle (k = 3) is activated on the set C_3 and so on.

The above procedure should be stopped after extraction of an adequate number K of the flat patterns P_k (5). The stop criterion should take into account that the numbers m_k of elements $\mathbf{x}_j$ in the final subsets C_k^* (36) can not be too small.

8 Examples of experimental results

The computational pro/cedures described in this paper are currently being implemented. The first results of the calculations are shown in this paragraph.

Two synthetic data sets D_1 and D_2, has been created for the purpose of the computational experiments. The set D_1 contained m_1 = 100 two-dimensional feature vectors $\mathbf{x}_j$ ($\mathbf{x}_j \in R^2$). The set D_2 contained m_2 = 100 three-dimensional feature vectors $\mathbf{x}_j$ ($\mathbf{x}_j \in R^3$). The data sets D_1 and D_2 were ***collinear***. It means in this case, that elements $\mathbf{x}_j$ of each set D_k (k = 1, 2) has been located on the vertexical line $l_k(\mathbf{x}_{j(1)}, \mathbf{x}_{j(2)})$ (22) defined by two basic feature vectors $\mathbf{x}_{j(1)}$ and $\mathbf{x}_{j(2)}$ contained in the basis $\mathbf{B}_k$ (12):

The basic feature vectors $\mathbf{x}_{j(1)}$ and $\mathbf{x}_{j(2)}$ (25) were pre-selected as:

$$P_1: \ \mathbf{x}_{j(1)} = [1,0]^T \ \textit{and} \ \mathbf{x}_{j(2)} = [0,1]^T \ \textit{and} \quad (41)$$
$$P_2: \ \mathbf{x}_{j(1)} = [1,1,0]^T \ \textit{and} \ \mathbf{x}_{j(2)} = [0,1,1]^T$$

The computational experiments were carried out both on the collinear data set P_1 with added *outliers*. as well as on the set P_2 without *outliers*. The term *outliers* means here such additional feature vectors $\mathbf{x}_j$ which were not located on the vertexical line $l_k(\mathbf{x}_{j(1)}, \mathbf{x}_{j(2)})$ (22). The outlier feature vectors $\mathbf{x}_j$ were generated in accordance with the normal distribution $N_2(\mathbf{0}, \boldsymbol{I})$ with the unit covariance matrix $\boldsymbol{I}$.

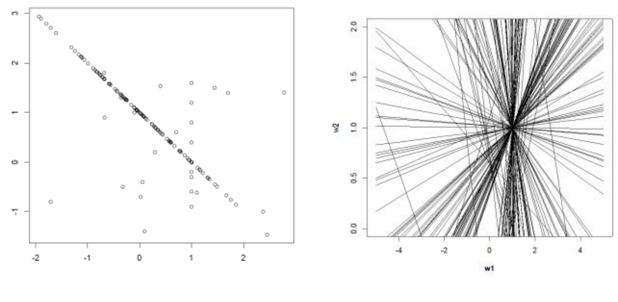

Figure 3. Representations of the collinear pattern P_1 with added outliers in the two-dimensional feature space (*left*) and in the parameter space (*right*).

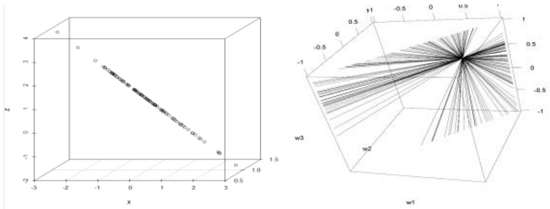

Figure 4. Representations of the collinear pattern P_2 (41) without outliers in the three-dimensional feature space (*left*) and in the parameter space (*right*).

The computational experiments allowed to extract the flat patterns P_1 and P_1 (41) from the data sets given in the feature space.

9 Concluding remarks

Collinear patterns P_k (*Def.* 2) can be discovered in large, high-dimensional data sets *C* (1) through minimization of the convex and piecewise linear (*CPL*) criterion functions $\Phi_k(\mathbf{w})$ (12).

Discovering collinear patterns P_k can be linked to a search for degenerated vertices (9) in the parameter space.

The proposed by us method of discovering collinear patterns on the basis of the CPL functions can be compared with the methods based on the Hough transformation used in computer vision for detection lines and curves in pictures (Duda and Hart, 1972; Ballard, 1981).

Acknowledgments: The present study was supported by a grant S/WI/2/2013 from Bialystok University of Technology and founded from the resources for research by Ministry of Science and Higher Education.

References

D. H. Ballard. Generalizing the *Hough Transform* to Detect Arbitrary Shapes, *Pattern Recognition*, 13(2):111-122, 1981

L. Bobrowski. Discovering main vertexical planes in a multivariate data space by using *CPL* functions, *ICDM* 2014, Ed. Perner P., Springer Verlag, Berlin 2014

O. R. Duda and P. E. Hart. Use of the Hough Transformation to Detect Lines and Curves in Pictures, *Communications of Association for Computing Machinery*, 15(1):11-15, 1972.

D. Hand, P. Smyth, and H. Mannila. *Principles of data mining*, MIT Press, Cambridge (2001)