Flat Patterns Extraction with Collinearity Models
Leon Bobrowski\textsuperscript{1,2}, Paweł Zabielski\textsuperscript{1}
\textsuperscript{1}Faculty of Computer Science, Białystok University of Technology, Białystok, Poland
\textsuperscript{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 \( \mathbf{x}_j = \mathbf{x}_j[n] = [x_{j1},\ldots,x_{jm}]^T \) which represent the objects (patients) \( O_j \) and belong to a given \( n \)-dimensional feature space \( F[n] \) \( \{ \mathbf{x}_j \in F[n] \} \):

\[
C = \{ \mathbf{x}_j \mid j = 1,\ldots,m \}
\]

(1)

The feature space \( F[n] = \{ x_1,\ldots,x_n \} \) is composed of \( n \) features \( x_i (i \in I = \{ 1,\ldots,n \}) \). The \( j \)-th component \( x_{ij} (x_{ij} \in R \) or \( x_{ij} \in \{ 0,1 \}) \) 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 \( k \) \( (i \in k \subset I) \) 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 \( \mathbf{x}[n] = [x_1,\ldots,x_n]^T \) by neglecting these components \( x_i \) which represent features \( x_i \) with the indices \( i \) outside the set \( k \) \( (i \notin k) \). The regular hyperplane \( H_k(\mathbf{w}, \theta) \) in the \( k \)-th feature subspace \( F_k[n_k] \) is defined in the below manner:

\[
H_k(\mathbf{w}, \theta) = \{ \mathbf{x} : \mathbf{w}^T \mathbf{x} = \theta \}
\]

(2)

where \( \mathbf{x} = [x_1,\ldots,x_n]^T \) is the reduced feature vector \( (\in F_k[n_k]) \), \( \mathbf{w} = [w_1,\ldots,w_d]^T \) is the reduced weight vector \( (\in \mathbb{R}^k) \) and \( \theta \) is the threshold \( (\in \mathbb{R}) \).

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 \( \theta \) and the weights \( w_i \) are different from zero:

\[
(\theta \neq 0) \text{ and } (\forall i \in \{ 1,\ldots,n_k \} \text{ } w_i \neq 0)
\]

(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 \( J_k \) \( (j \in J_k \subset J = \{ 1,\ldots,m \}) \):

\[
C_k = C_k[n_k] = \{ \mathbf{x}_j \mid j \in J_k \}
\]

(4)

The \( k \)-th data subset \( C_k[n_k] \) \( (3) \) can be represented also as the matrix \( M[m_k \times n_k] \) with the \( m_k \) rows and \( n_k \) columns. The rows of the matrix \( M[m_k \times n_k] \) are constituted by particular feature vectors \( \mathbf{x}_j (j \in J_k) \).
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 $H_k(w, \theta)$ (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 $x_i = x_j[n_k]$ constitutes the collinear (flat) pattern $P_k$ if all elements $x_i$ of this subset are located on a regular hyperplane $H_k(w, \theta)$ (2) in the feature subspace $F_k[n_k]$:

\[
(\forall x_i \in C_k[n_k]) \quad w^T x_i = \theta
\] (5)

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

\[
S(w, \theta) = \{x: \theta - \varepsilon \leq (w || x ||)^T x \leq \theta + \varepsilon\}
\] (6)

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

**Figure 1.** An example of the $\varepsilon$ - layer $S(w, \theta)$ (6) in the two-dimensional ($n = 2$) feature subspace $F_k = \{x_2, x_1\}$.

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

\[
(\forall x_i \in C_k[n_k]) \quad \theta' - \varepsilon' \leq (w')^T x_i \leq \theta' + \varepsilon'
\] (7)

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

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

\[
(\forall x_i \in C_k[n_k]) \quad 1 - \varepsilon \leq w^T x_i \leq 1 + \varepsilon
\] (8)

where $w = w'/ \theta'$ and $\varepsilon = \varepsilon' / \theta'$.

3 **Dual hyperplanes and vertices in the parameter subspaces**

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

\[
(\forall x_i \in C_k[n_k]) \quad h_i = \{w: x_i^T w = 1\}
\] (9)

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

\[
S_k = \{x_{j(i)}: j(i) \in J_k\}
\] (10)

The hyperplanes $h_{j(i)}$ defined by the basis vectors $x_{j(i)}$ from the set $S_k$ (9) intersect at one point (vertex) $w_k$ determined the below equations:

\[
(\forall j(i) \in J_k) \quad x_{j(i)}^T w_k = 1
\] (11)

The above equations can be given in the matrix form:

\[
B_k^T w_k = 1
\] (12)

where $B_k = [x_{j(1)},...,x_{j(n_k)}]$ is the non-singular matrix called the $k$ -th basis and $1 = [1, 1, ..., 1]^T$.

The $k$ -th vertex $w_k = [w_{k,1},...,w_{k,n_k}]^T$ (11) with the non-zero components $w_{ki}$ ($w_{ki} \neq 0$) allows to define the vertexical hyperplane $H_k(w_k, 1)$ in the feature subspace $F_k[n_k]$:

\[
H_k(w_k, 1) = \{x \in F_k[n_k]: (w_k)^T x = 1\}
\] (13)

The vertexical hyperplane $H_k(w_k, 1)$ (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 $x_i$ from the subset $C_k[n_k]$ (4) are situated on the hyperplane $H(w, \theta) = \{x: w^T x = 0\}$ with $\theta \neq 0$, if and only if each vector $x_i$ defines such dual hyperplane $h_i$ (8) which passes through the vertex $w_k$ (10).

The Remark 1 has been discussed in the paper.

4 **Penalty and criterion functions aimed at extraction of collinear patterns**

We consider convex and piecewise linear (CPL) penalty functions $\phi_\theta(w)$ defined on the $n_k$ - dimensional feature vectors $x_i$ from the $k$ -th data subset $C_k[n_k]$ (4):

DOI: 10.3384/ecp17142518 Proceedings of the 9th EUROSIM & the 57th SIMS September 12th-16th, 2016, Oulu, Finland 519
\[(\forall x_j \in C_k[n_k]) \quad \Phi_k(w) = \begin{cases} 1 - \varepsilon & \text{if } w^T x_j < 1 - \varepsilon \\ 0 & \text{if } 1 - \varepsilon \leq w^T x_j \leq 1 + \varepsilon \\ w^T x_j - 1 + \varepsilon & \text{if } w^T x_j > 1 + \varepsilon \end{cases} \]

where \(\varepsilon\) is a small, non-negative parameter (margin).

The non-negative function \(\phi_j(w)\) is equal to zero (\(\phi_j(w) = 0\)) if and only if the feature vector \(x_j\) is located in the layer \(S(w, \theta)\) (7) with \(\theta = 1\) (Fig. 2).

![Figure 2. The \(j\)-th penalty functions \(\phi_j(w)\) (8).](image)

The criterion function \(\Phi_k(w)\) is defined as the weighted sum of the penalty functions \(\phi_j(w)\) (8) linked to \(m_k\) feature vectors \(x_j\) constituting the subset \(C_k \subset C\) (1):

\[\Phi_k(w) = \sum_j \alpha_j \phi_j(w)\]  

(15)

where the positive parameters \(\alpha_j\) (\(\alpha_j > 0\)) are prices of particular feature vectors \(x_j\). The parameters \(\alpha_j\) may depend on the number \(m_k\) of the vectors \(x_j\) in the subset \(C_k\):

\[(\forall x_j \in C_k) \quad \alpha_j = 1 / m_k\]  

(16)

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

\[(\exists w_k^*) \quad (\forall w) \quad \Phi_k(w) \geq \Phi_k(w_k^*) = \Phi_k^* \geq 0\]  

(17)

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

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

\[(\forall x_i \in C_k[n_k]) \quad h_j^+ = \{ w : x_i^T w = 1 + \varepsilon \} \quad \text{and} \quad h_j^- = \{ w : x_i^T w = 1 - \varepsilon \}\]  

(18)

\[h_j^* = \{ w : x_i^T w = 1 \}\]

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

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

**Remark 2:** The minimal value \(\Phi_k(w_k^*)\) (17) of the criterion function \(\Phi_k(w)\) (15) determined on all elements \(x_j\) of the subset \(C_k\) (4) becomes equal to zero for a sufficiently high value of the parameter \(\varepsilon\).

For a given data subset \(C_k[n_k]\) (4) we can determine the minimum value \(\varepsilon_k\) of the parameter \(\varepsilon\) which allows to reset the minimal value \(\Phi_k(w_k^*)\) (17) of the criterion function \(\Phi_k(w)\) (15) determined on this subset:

\[\varepsilon_k = \min \{ \varepsilon : \Phi_k(w_k^*) = 0 \}\]  

(19)

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

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

The minimizing of the criterion function \(\Phi_k(w)\) (15) with parameter \(\varepsilon\) less than \(\varepsilon_k\) \((0 \leq \varepsilon < \varepsilon_k)\) 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 \(H_k(w_k, x)\) (13) in the \(n_k\) -dimensional feature subspace \(F_k[n_k]\) is defined by using the vertex \(w_k = [w_{k,1}, \ldots, w_{k,n_k}]^T\) with \(n_k\) non-zero components \(w_i\).
The vertex \( \mathbf{w}_k \) is linked to the \( k \)-th basis \( \mathbf{B}_k = [\mathbf{x}_{j(1)}, \ldots, \mathbf{x}_{j(\beta_k)}] \) constituted by \( \beta_k \) linearly independent basis vectors \( \mathbf{x}_{j(\beta_k)} \) of dimensionality \( \beta_k \).

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

\[
H_k(\mathbf{w}_k, 1) = P_k(\mathbf{x}_{j(1)}, \ldots, \mathbf{x}_{j(\beta_k)}) = \{ \mathbf{x}: \mathbf{x} = \alpha_1 \mathbf{x}_{j(1)} + \cdots + \alpha_{\beta_k} \mathbf{x}_{j(\beta_k)} \}
\]

(20)

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

\[
\alpha_1 + \cdots + \alpha_{\beta_k} = 1
\]

(21)

Remark 3: The dimensionality of the vertex hyperplane \( P_k(\mathbf{x}_{j(1)}, \ldots, \mathbf{x}_{j(\beta_k)}) \) (20) is equal to \( n_k - 1 \).

Theorem 2: The reduced feature vector \( \mathbf{x}_j \) (\( \mathbf{x}_j \in F_k[n_k] \)) is situated on the vertex hyperplane \( P_k(\mathbf{x}_{j(1)}, \ldots, \mathbf{x}_{j(\beta_k)}) \) (20), where \( j(i) \in L(10) \) 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 vertex hyperplane \( H_k(\mathbf{w}_k, 1) \) (13) supports the flat patterns \( P_k \) if a large number \( m_k \) of the reduced vectors \( \mathbf{x}_j \) are located on this hyperplane.

Definition 6: The vertex hyperplane \( H_k(\mathbf{w}_k, 1) \) (13) supports the \( \varepsilon \)-flat patterns \( P_k \) if a large number \( m_k \) of the reduced vectors \( \mathbf{x}_j \) are located on the vertex hyperplane \( P_k \) (13).

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

Definition 8: The dimensionality of the flat patterns \( P_k \) or \( P_k' \) is equal to \( n_k - 1 \).

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

\[
h_k(\mathbf{x}_{j(1)}, \mathbf{x}_{j(2)}) = \{ \mathbf{x}: \mathbf{x} = \alpha \mathbf{x}_{j(1)} + (1 - \alpha) \mathbf{x}_{j(2)} \}
\]

(22)

where \( \alpha \in R^1 \).

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

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

\[
P_k(\mathbf{x}_{j(1)}, \mathbf{x}_{j(2)}, \mathbf{x}_{j(3)}) = \{ \mathbf{x}: \mathbf{x} = \alpha_1 \mathbf{x}_{j(1)} + \alpha_2 \mathbf{x}_{j(2)} + \alpha_3 \mathbf{x}_{j(3)} \}
\]

(23)

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

The rank \( n_k \) 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 \( (n_k = 3) \).

The flat patterns \( P_k \) or \( P_k' \) can be extracted from the data set \( C(1) \) through minimization of the criterion functions \( \Phi_k(\mathbf{w}) \) (15).
The implication (22) can be proved by the fact that omission of certain feature vectors \( x_i \) results in omission of certain non-negative components \( a_i \varphi_j(w) \) (14) in the criterion function \( \Phi_k(w) \) (15).

**ii. The negative monotonicity due to reduction of features \( x_i \)**

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

\[
(F_k[n_k] \subset F_i[n_i]) \Rightarrow (\Phi_k^* \geq \Phi_i^*)
\]  
(25)

where the symbol \( \Phi_k^* \) stands for the minimal value (17) of the criterion function \( \Phi_k(w) \) (15) defined on the reduced vectors \( x'_i \) (\( x'_i \in F_k[n_k] \), \( n_k < n_i \)). 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 \( \Phi_k(w^*_k)(17) \) of the criterion function \( \Phi_k(w) \) (15) defined on reduced feature vectors \( x_i \) from the subset \( C_k \) (4) does not depend on linear, non-singular transformations of the feature vectors \( x_i \) from this subset:

\[
\Phi_k^*(w^*_k) = \Phi_k(w^*_k)
\]  
(26)

where \( \Phi_k^*(w^*_k) \) is the minimal value of the criterion functions \( \Phi_k^*(w) \) (15) defined on the transformed feature vectors \( x'_i[n] \):

\[
(\forall x_i \in C_k) \quad x'_i = A x_i
\]  
(27)

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

**Proof:** The values \( \varphi_j^*(w[n]) \) of the penalty function \( \varphi_j\varphi(w[n]) \) (15) in a point \( w^*[n] \) are defined in the below manner on the transformed feature vectors \( x'_i[n] \) (26):

\[
(\forall x_i \in C_k) \quad \varphi_j^*(w') = |1 - (w')^TAx_i| = |1 - (w')^T_{x_i}|(28)
\]

If we take (17)

\[
w' = (A^T)^{-1}w^*_k
\]  
(29)

we obtain the below result

\[
(\forall x_i \in C_k) \quad \varphi_j^*(w') = \varphi_j(w^*_k)
\]  
(30)

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

### 7 Procedure of flat patterns 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 \( \Phi_k(w) \) (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**

1. \( l = 1; \quad C_1 = C \) (1);

2. Define the criterion function \( \Phi_k(w) \) (15) on all elements \( x_i \) of the data set \( C \) and find the optimal vertex \( w^*_1 \) (11) which constitutes the minimal value \( \Phi_k(w^*_1)(17) \) of this function.

3. If \( \Phi_k(w^*_1) = 0 \), then the procedure is stopped in the optimal vertex \( w^*_1 \), otherwise the next step is executed

4. Find the vector \( x_i \) in the feature subset \( C \) with the highest value of the penalty function \( \varphi_j(w) \) (14) in the optimal vertex \( w^*_1 \) (18): \( \Phi_k(w^*_1) \)

\[
(\forall x_i \in C_1) \quad \varphi_j(w^*_1) \geq \varphi_j(w^*_1)
\]  
(32)

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

\[
(\forall x_i \in C_1) \quad a_j \varphi_j(w^*_1) \geq a_j \varphi_j(w^*_1)
\]  
(33)

5. Remove the feature vector \( x_i \) from the subset \( C \):

\[
C_i \rightarrow C_i / x_i
\]  
(34)

6. Increase the counter \( k \):

\[
l \rightarrow l + 1
\]  
(35)

7. Go to the step ii.

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

\[
C_k^* = C \quad J_k^* \quad w^*_k = w^*_k
\]  
(36)

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

\[
(\forall w) \quad \Phi_k(w) \geq \Phi_k(w^*_k) = 0
\]  
(37)

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

The vertex \( w^*_k = [w^*_{k,1} \ldots w^*_{k,n_c}]^T \) (36) obtained from the procedure **Vertex** (31) should be **regularized** before...
using it in the definition of the **vertexical hyperplane** $H_k(w_k, 1)$ (13). The regularization process means in this case the neglecting of such components $w_{ki}^*$ in the vector $w_k^*$ which are equal to zero ($w_i = 0$) (Def. 1). The regularization means additionally the neglecting of such features $x_i$ and components $x_{ij}$ of the feature vectors $x_i = [x_{i1}, ..., x_{in}]^T$ from the reduced data subset $C_k$ which are linked to weights $w_{ki}^*$ equal to zero ($w_{ki}^* = 0$):

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

if $(w_{ki}^* = 0)$, then (the $i$-th feature $x_i$ and the $i$-th component $x_{ij}$ of the $j$-th feature vector $x_j$ are neglected)

**Remark 4:** The reduction of feature vectors $x_j$ 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 $x_j$ ($x_j \in P_k[n_k]$) 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 $x_j$ of the final subset $C_k[n_k]$ obtained in result of the procedure **Vertex** and the rule (38) is a large enough, then 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 $C_1 = C (1) \text{ is reduced by the final set } C_1^* (36) \text{ after the first cycle:}$

$$C_2 = C_1 / C_1^* = C / C_1^*$$

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

$$C_3 = C_2 / C_2^*$$

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 $x_j$ in the final subsets $C_k^*$ (36) can not be too small.

**8 Examples of experimental results**

The computational procedures 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 $x_j$ ($x_j \in R^2$). The set $D_2$ contained $m_2 = 100$ three-dimensional feature vectors $x_j$ ($x_j \in R^3$). The data sets $D_1$ and $D_2$ were collinear. It means in this case, that elements $x_i$ of each set $D_k (k = 1, 2)$ has been located on the vertexical line $l(x_{j(1)}, x_{j(2)})$ (22) defined by two basic feature vectors $x_{j(1)}$ and $x_{j(2)}$ contained in the basis $B_k$ (12):

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

$$P_1: \quad x_{j(1)} = [1,0]^T \text{ and } x_{j(2)} = [0,1]^T$$

$$P_2: \quad x_{j(1)} = [1,1,0]^T \text{ and } x_{j(2)} = [0,1,1]^T$$

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 $x_i$ which were not located on the vertexical line $l(x_{j(1)}, x_{j(2)})$ (22). The outlier feature vectors $x_i$ were generated in accordance with the normal distribution $N(0, I)$ with the unit covariance matrix $I$.

**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 $\Phi_k(w)$ (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