Partially Supervised Approach in Signal Recognition

Catalina COCIANU[1], Luminita STATE[2], Doru CONSTANTIN[2], Corina SARARU[2]
[1]Academy of Economic Studies, Bucharest, Romania
[2]University of Pitesti, Pitesti, Romania,
ccocianu@ase.ro, radus@sunu.rnc.ro, cdomanid@yahoo.com

The paper focuses on the potential of principal directions based approaches in signal classification and recognition. In probabilistic models, the classes are represented in terms of multivariate density functions, and an object coming from a certain class is modeled as a random vector whose repartition has the density function corresponding to this class. In cases when there is no statistical information concerning the set of density functions corresponding to the classes involved in the recognition process, usually estimates based on the information extracted from available data are used instead. In the proposed methodology, the characteristics of a class are given by a set of eigen vectors of the sample covariance matrix. The overall dissimilarity of an object X with a given class C is computed as the disturbance of the structure of C, when X is allotted to C. A series of tests concerning the behavior of the proposed recognition algorithm are reported in the final section of the paper.
Keywords: *signal processing, classification, pattern recognition, compression/decompression*

1 Introduction

Machine recognition of patterns can be viewed as a two-fold task, comprising learning the invariant properties of a set of samples characterizing a class, and deciding that a new sample is a possible member of the class by noting that it has properties common to those of the set of samples.

A typical pattern recognition system consists of three phases. These are data acquisition, feature extraction and classification. In the data acquisition phase, depending on the environment within which the objects are to be classified, data are gathered using a set of sensors. These are then passed on to the feature extraction phase, where the dimensionality of the data is reduced by measuring and retaining only some characteristic features or properties. In a broader perspective, this stage significantly influences the entire recognition process. Finally, in the classification phase, the extracted features are passed on to the classifier that evaluates the incoming information and makes a final decision. This phase basically establishes a transformation between the features and the classes. Therefore pattern recognition can be described as a transformation from the measurement space M to the feature space F and finally to the decision space D.

The problem of classification basically establishes a transformation F→D between the features and the classes. In other words, it provides a partitioning of the feature space into regions, one region for each category of input. That is, it attempts to assign every data point in the entire feature space to one of the possible (say K) classes. Different forms of transformation can be a Bayesian rule of computing a posteriori class probability, nearest neighbor rule, linear discriminant functions, perceptron rule, nearest prototype rule, etc. Classifiers are usually, but not always, designed with labeled data, in which case these problems are sometimes referred to as supervised classification (where the parameters of a classifier function D are learned). Some common examples of the supervised pattern classification techniques are the nearest neighbor rule, Bayes' maximum likelihood classifier and the perceptron rule [4], [5], [10].

The process of feature selection aims to map a data space into a feature space that, in theory, has precisely the same dimension as the original data space. The mapping is designed such that the available data set is represented by a reduced number of effective

features, the most informative ones (that is by retaining most of the intrinsic information content of the data). Principal Component Analysis (PCA) has been broadly used in a large series of signal and image processing, pattern recognition and data analysis applications. Classical PCA is based on the second-order statistics of the data and, in particular, on the eigen-structure of the data covariance matrix and accordingly, the PCA neural models incorporate only cells with linear activation functions. More recently, several generalizations of the classical PCA models to non-Gaussian models, the Independent Component Analysis (ICA) and the Blind Source Separation techniques (BSS) have become a very attractive and promising framework in developing more efficient signal processing algorithms. [10]

The paper focuses on the potential of principal directions-based approaches in signal classification. The structure of a class is represented in terms of the estimates of its principal directions computed from data, the overall dissimilarity of a particular object with a given class being given by the "disturbance" of the structure, when the object is identified as a member of this class. A series of conclusions experimentally established are exposed in the final section of the paper.

2 The Proposed Methodology for Classification Purposes

According to the well known result established by Karhunen and Loeve, a set of principal directions corresponds to the maximum variability of the "cloud" from metric point of view, and they are also almost optimal from informational point of view, the principal directions being recommended by the maximum entropy principle as the most reliable characteristics of the repartition.

PCA allows the identification of a linear transform such that the axes of the resulted coordinate system correspond to the largest variability of the investigated signal. The signal features corresponding to the new coordinate system are uncorrelated, that is, in case of normal models these components are independent. The advantages of using principal components reside from the fact that the information contained by each band is maximum for the whole set of bits [3].

The principal directions of a class are given by a set of unit orthogonal eigen vectors of the covariance matrix.

When the available data is represented by a set of objects $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N$, belonging to a certain class C, the covariance matrix is estimated by the sample covariance matrix,

$$\hat{\mathbf{\Sigma}}_N = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}_N)(\mathbf{X}_i - \hat{\boldsymbol{\mu}}_N)^T \quad (1)$$

where $\hat{\boldsymbol{\mu}}_N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$.

Let us denote by $\lambda_1^N \geq \lambda_2^N \geq ... \geq \lambda_n^N$ the eigen values and by $\boldsymbol{\psi}_1^N, ..., \boldsymbol{\psi}_n^N$ a set of orthonormal eigen vectors of $\hat{\Sigma}_N$.

If a new example $\mathbf{X}_{N+1}$ coming from the same class has to be included in the sample, the new estimate of the covariance matrix can be recomputed as,

$$\hat{\mathbf{\Sigma}}_{N+1} = \hat{\mathbf{\Sigma}}_N + \frac{1}{N+1}(\mathbf{X}_{N+1} - \hat{\boldsymbol{\mu}}_N)(\mathbf{X}_{N+1} - \hat{\mu}_N)^T - \frac{1}{N}\hat{\mathbf{\Sigma}}_N \quad (2)$$

Using first order approximations [11], the estimates of the eigen values and eigen vectors respectively are given by,

$$\lambda_i^{N+1} = \lambda_i^N + \left(\boldsymbol{\psi}_i^N\right)^T \Delta\hat{\mathbf{\Sigma}}_N \boldsymbol{\psi}_i^N = \left(\boldsymbol{\psi}_i^N\right)^T \hat{\mathbf{\Sigma}}_{N+1}\boldsymbol{\psi}_i^N \quad (3)$$

$$\boldsymbol{\psi}_i^{N+1} = \boldsymbol{\psi}_i^N + \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\left(\boldsymbol{\psi}_N^j\right)^T \Delta\hat{\mathbf{\Sigma}}_N \boldsymbol{\psi}_i^N}{\lambda_i^N - \lambda_j^N}\boldsymbol{\psi}_j^N \quad (4)$$

Let $\boldsymbol{\psi}_1^N, ..., \boldsymbol{\psi}_n^N$ be set of principal directions of the class C computed using $\hat{\mathbf{\Sigma}}_N$. When the example $\mathbf{X}_{N+1}$ is identified as a member of the class *C*, then the disturbance implied by extending C is expressed as,

$$D = \frac{1}{n}\sum_{k=1}^{n} d\left(\boldsymbol{\psi}_k^N, \boldsymbol{\psi}_k^{N+/}\right) (6)$$

where d is the Euclidian distance and $\boldsymbol{\psi}_1^{N+1},...,\boldsymbol{\psi}_n^{N+1}$ are the principal directions computed using $\hat{\boldsymbol{\Sigma}}_{N+1}$.

Let $H=\{C_1,C_2,...,C_M\}$ be a set of classes, where the class C_j contents N_j elements. The new object **X** is allotted to C_j, one of the classes for which

$$D=\frac{1}{n}\sum_{k=1}^{n} d\left(\boldsymbol{\psi}_{k,j}^{N_j},\boldsymbol{\psi}_{k,j}^{N_j+1}\right)=$$

$$\min_{1\le p\le M}\frac{1}{n}\sum_{k=1}^{n} d\left(\boldsymbol{\psi}_{k,p}^{N_p},\boldsymbol{\psi}_{k,p}^{N_p+1}\right) \quad (7)$$

In order to protect against misclassifications, due to insufficient "closeness" to any class, we implement this recognition technique using a threshold $T>0$ such that the example **X** is allotted to C_j only if relation (7) holds and $D<T$.

Briefly, the recognition procedure is described below [2].

Input: $H=\{C_1,C_2,...,C_M\}$

Repeat

i←1

Step 1: Let **X** be a new sample. Classify **X** according to (7)

Step 2: If $\exists j, 1\le j\le M$ such that **X** is allotted to C_j, then

2.1.re-compute the characteristics of C_j using (2), (3) and (4)

2.2. i++

Step 3: If i<PN goto Step 1

Else

3.1. For i=$\overline{1,M}$, compute the characteristics of class C_i using **M**.

3.2. goto Step 1.

Until the last new example was classified

Output: The new set $\{C_1,C_2,...,C_M\}\cup CR$

The performance of the proposed algorithm was evaluated mainly using the leaving one out method. Being given the complexity of the computation that has to be carried out in implementing this evaluation technique, instead of the exact eigen values and eigen vectors, the re-computation was mainly based on first order approximations derived by applying the perturbation theory. Also, in order to keep the computational complexity at a reasonable level, the recursive re-computation schemes established by the following lemma proved extremely useful.

Lemma. Let $\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_K$ be an n-dimensional Bernoullian sample. We denote by $\hat{\boldsymbol{\mu}}_N=\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$,

$\hat{\boldsymbol{\Sigma}}_N=\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_i-\hat{\boldsymbol{\mu}}_N)(\mathbf{X}_i-\hat{\boldsymbol{\mu}}_N)^T$, and let $\{\lambda_i^N\}_{1\le i\le n}$ be the eigen values and $\{\boldsymbol{\psi}_i^N\}_{1\le i\le n}$ a set of orthogonal unit eigen vectors of $\hat{\boldsymbol{\Sigma}}_N$, $2\le N\le K-1$. In case the eigen values of $\hat{\boldsymbol{\Sigma}}_{N+1}$ are pair wise distinct, the following first order approximations hold, [1]

$$\lambda_i^N=\lambda_i^{N+1}+\left(\boldsymbol{\psi}_i^{N+1}\right)^T\Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \quad (8)$$

$$\boldsymbol{\psi}_i^N=\boldsymbol{\psi}_i^{N+1}+\sum_{\substack{j=1\\ j\ne i}}^{n}\frac{\left(\boldsymbol{\psi}_j^{N+1}\right)^T\Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1}}{\lambda_i^{N+1}-\lambda_j^{N+1}}\boldsymbol{\psi}_j^{N+1} \quad (9)$$

where $\Delta\boldsymbol{\Sigma}_{N+1}=\hat{\boldsymbol{\Sigma}}_N-\hat{\boldsymbol{\Sigma}}_{N+1}$.

Proof Using the perturbation theory, we get, $\hat{\boldsymbol{\Sigma}}_N=\hat{\boldsymbol{\Sigma}}_{N+1}+\Delta\boldsymbol{\Sigma}_{N+1}$ and, $\boldsymbol{\psi}_i^N=\boldsymbol{\psi}_i^{N+1}+\Delta\boldsymbol{\psi}_i^{N+1}$, $\lambda_i^N=\lambda_i^{N+1}+\Delta\lambda_i^{N+1}$, $1\le i\le n$. Then,

$$\Delta\boldsymbol{\Sigma}_{N+1}=\frac{1}{N-1}\hat{\boldsymbol{\Sigma}}_{N+1}-\frac{N}{(N-1)(N+1)}(\mathbf{X}_{N+1}-\boldsymbol{\mu}_N)(\mathbf{X}_{N+1}-\boldsymbol{\mu}_N)^T \quad (10)$$

where $\boldsymbol{\mu}_N=\frac{(N+1)\boldsymbol{\mu}_{N+1}}{N}-\frac{\mathbf{X}_{N+1}}{N}$

$$\left(\hat{\boldsymbol{\Sigma}}_{N+1}+\Delta\boldsymbol{\Sigma}_{N+1}\right)\left(\boldsymbol{\psi}_i^{N+1}+\Delta\boldsymbol{\psi}_i^{N+1}\right)=$$
$$\left(\lambda_i^{N+1}+\Delta\lambda_i^{N+1}\right)\left(\boldsymbol{\psi}_i^{N+1}+\Delta\boldsymbol{\psi}_i^{N+1}\right) \quad (11)$$

Using first order approximations, from (11) we get,

$$\lambda_i^{N+1}\boldsymbol{\psi}_i^{N+1} + \hat{\boldsymbol{\Sigma}}_{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \cong \lambda_i^{N+1}\boldsymbol{\psi}_i^{N+1} + \lambda_i^{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \Delta\lambda_i^{N+1}\boldsymbol{\psi}_i^{N+1} \quad (12)$$

hence,

$$\left(\boldsymbol{\psi}_i^{N+1}\right)^T \hat{\boldsymbol{\Sigma}}_{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \cong \lambda_i^{N+1}\left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} + \Delta\lambda_i^{N+1}\left\|\boldsymbol{\psi}_i^{N+1}\right\|^2 \quad (13)$$

Using $\lambda_i^{N+1}\left(\boldsymbol{\psi}_i^{N+1}\right)^T = \left(\boldsymbol{\psi}_i^{N+1}\right)^T \hat{\boldsymbol{\Sigma}}_{N+1}$ we obtain ,

$$\lambda_i^{N+1}\left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} + \left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \cong \lambda_i^{N+1}\left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} + \Delta\lambda_i^{N+1} \quad (14)$$

hence $\Delta\lambda_i^{N+1} = \left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1}$ that is,

$$\lambda_i^{N} = \lambda_i^{N+1} + \left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \quad (15)$$

The first order approximations of the orthonormal eigen vectors of $\hat{\boldsymbol{\Sigma}}_N$ can be derived using the expansion of each vector $\Delta\boldsymbol{\psi}_i^{N+1}$ in the basis represented by the orthonormal eigen vectors of $\hat{\boldsymbol{\Sigma}}_{N+1}$,

$$\Delta\boldsymbol{\psi}_i^{N+1} = \sum_{j=1}^{n} b_{i,j}\boldsymbol{\psi}_j^{N+1} \quad (16)$$

Where

$$b_{i,j} = \left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} \quad (17)$$

Using the orthonormality, we get,

$$1 = \left\|\boldsymbol{\psi}_i^{N+1} + \Delta\boldsymbol{\psi}_i^{N+1}\right\|^2 \cong \left\|\boldsymbol{\psi}_i^{N+1}\right\|^2 + 2\left(\boldsymbol{\psi}_i^{N+1}\right)^T\left(\Delta\boldsymbol{\psi}_i^{N+1}\right) = = 1 + 2\left(\boldsymbol{\psi}_i^{N+1}\right)^T\left(\Delta\boldsymbol{\psi}_i^{N+1}\right) \quad (18)$$

that is

$$b_{i,i} = \left(\boldsymbol{\psi}_i^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} = 0 \quad (19)$$

Using (11), the approximation,

$$\hat{\boldsymbol{\Sigma}}_{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \lambda_i^{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \Delta\lambda_i^{N+1}\boldsymbol{\psi}_i^{N+1} \quad (20)$$

holds for each $1 \le i \le n$.

For $1 \le j \ne i \le n$, from (20) we obtain the following equations:

$$\left(\boldsymbol{\psi}_j^{N+1}\right)^T \hat{\boldsymbol{\Sigma}}_{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \cong \lambda_i^{N+1}\left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} + \Delta\lambda_i^{N+1}\left(\boldsymbol{\psi}_j^{N+1}\right)^T \boldsymbol{\psi}_i^{N+1} \quad (21)$$

$$\left(\boldsymbol{\psi}_j^{N+1}\right)^T \hat{\boldsymbol{\Sigma}}_{N+1}\Delta\boldsymbol{\psi}_i^{N+1} + \left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \cong \lambda_i^{N+1}\left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} \quad (22)$$

$$\lambda_j^{N+1}\left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} + \left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1} \cong \lambda_i^{N+1}\left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} \quad (23)$$

From (23) we get,

$$b_{i,j} = \left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\psi}_i^{N+1} = \frac{\left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1}}{\lambda_i^{N+1} - \lambda_j^{N+1}} \quad (24)$$

Consequently, the first order approximation of the eigen vectors of $\hat{\boldsymbol{\Sigma}}_N$ are,

$$\boldsymbol{\psi}_i^{N+1} + \Delta\boldsymbol{\psi}_i^{N+1} \cong \boldsymbol{\psi}_i^{N+1} + \sum_{\substack{j=1 \\ j\ne i}}^{n} \frac{\left(\boldsymbol{\psi}_j^{N+1}\right)^T \Delta\boldsymbol{\Sigma}_{N+1}\boldsymbol{\psi}_i^{N+1}}{\lambda_i^{N+1} - \lambda_j^{N+1}}\boldsymbol{\psi}_j^{N+1} \quad (25)$$

3 Experiments and Conclusive Remarks

Several tests on the proposed recognition procedure were performed on different classes of signals. The results proved very good performance in terms of the recognition error.

Test 1. The evaluation of error using the leaving one out method.

Sequentially, one of the given examples is removed from the sample. The classifier is designed using the rest of 2NP-1 examples (that is the characteristics of the classes are computed in terms of the NP, NP-1 remaining examples) and the removed example is classified into one of resulted classes. The error is evaluated as $\frac{F}{2NP}$, where F is the number of misclassified examples.

Let $\{\psi_i^{1,NP}\}_{1\le i\le n}, \{\psi_i^{2,NP}\}_{1\le i\le n}$, $\{\lambda_i^{1,NP}\}_{1\le i\le n}, \{\lambda_i^{2,NP}\}_{1\le i\le n}$ be the characteristics of the classes and the corresponding eigen values at the initial moment and $\mu_{NP}^1, \mu_{NP}^2, \Sigma_{NP}^1, \Sigma_{NP}^2$ the sample means and the sample covariance matrices respectively. Let X be the removed example. In case X comes from the first class, the new characteristics are,

$$\psi_i^{1,NP-1} = \psi_i^{1,NP} + \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\left(\psi_j^{1,NP}\right)^T \Delta\Sigma_{NP}^1 \psi_i^{1,NP}}{\lambda_i^{1,NP} - \lambda_j^{1,NP}} \psi_j^{1,NP}$$

for the first class and remains unchanged for the second one, where

$$\Delta\Sigma_{NP}^1 = \Sigma_{NP-1}^1 - \Sigma_{NP}^1$$

$$\Sigma_{NP-1}^1 = \frac{NP-1}{NP-2}\Sigma_{NP} - \frac{NP-1}{(NP-2)NP}\left(X-\mu_{NP-1}^1\right)\left(X-\mu_{NP-1}^1\right)^T$$

$$\mu_{NP-1}^1 = \frac{NP\mu_{NP}^1}{NP-1} - \frac{X}{NP-1}$$

In case X comes from the second class, similar formulas are used.

The evaluation of the error is performed for $NP = 10,20,30,40,50,75,150$. Several tests were performed on samples generated from 3 repartitions, Gaussian, Rayleigh and geometric, each class corresponding to one of them. All tests reported to a surprising conclusion that is the misclassification error is near to 0.

a) The classes correspond to the Gaussian repartition and Rayleigh repartition respectively, NP=75, 150, n=50, e=50, where n is the data dimensionality and e is the number of epochs. The resulted values of empirical errors are 0.213 in case NP=75 and respectively 0.0327 in case NP=150. Several other tests confirmed the idea that the performance can be significantly improved by using larger size training sets.

The variation of the empirical error in terms of e is presented in Figure 1 and Figure 2. In Figure1 the test is performed for NP=75 and in Figure 2 the volume is NP=150

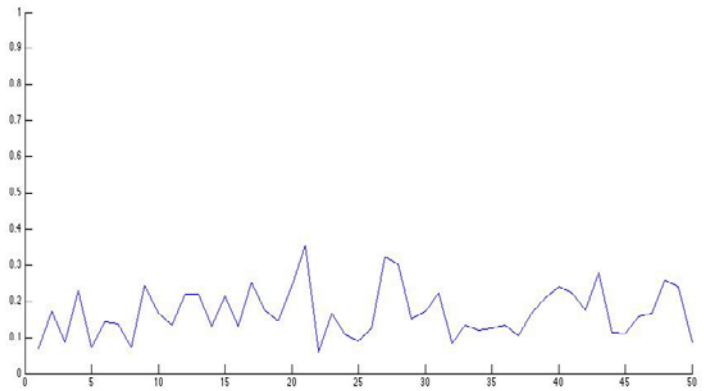

Fig. 1. The variation of empirical error for NP=75

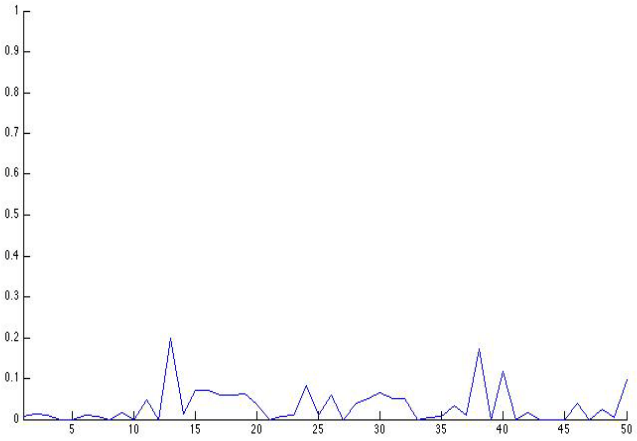

Fig. 2. The variation of empirical error for NP=150

b) The classes correspond to the geometric repartition and Rayleigh repartition respectively, NP=50, 150, n=50, e=50, where n is the data dimensionality and e is the number of epochs. The resulted values of empirical errors are 0.35 in case NP=50 and respectively 0.0112 in case NP=150. The variation of the empirical error in terms of e is presented in Figure 3 and Figure 4. In Figure3 the test is performed for NP=50 and in Figure 4 the volume is NP=150.

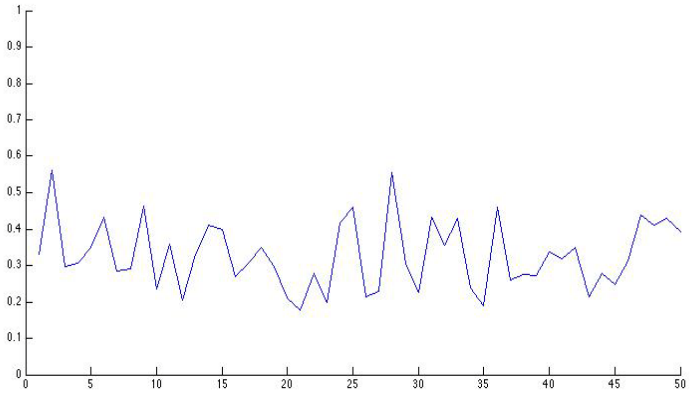

Fig. 3. The variation of empirical error for NP=50

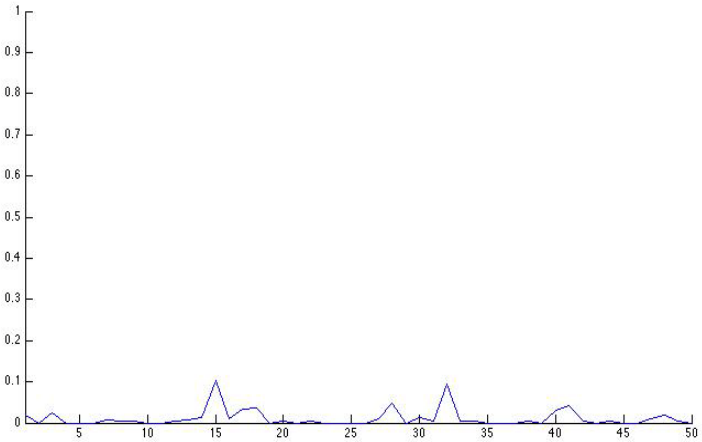

Fig. 4. The variation of empirical error for NP=150

c) The classes correspond to the Gaussian repartition, NP=50,150, n=50, e=50, where n is the data dimensionality and e is the number of epochs. The resulted values of empirical errors are 0.15 in case NP=50 and respectively 0.0261 in case NP=150. The variation of the empirical error in terms of e is presented in Figure 5 and Figure 6. In Figure5 the test is performed for NP=50 and in Figure 6 the volume is NP=150

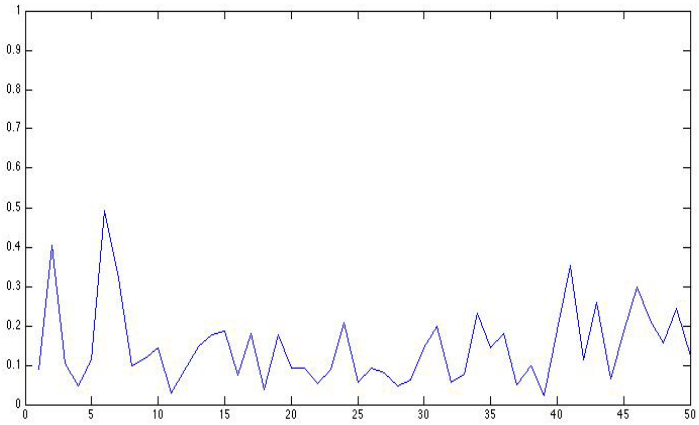

Fig. 5. The variation of empirical error for NP=50

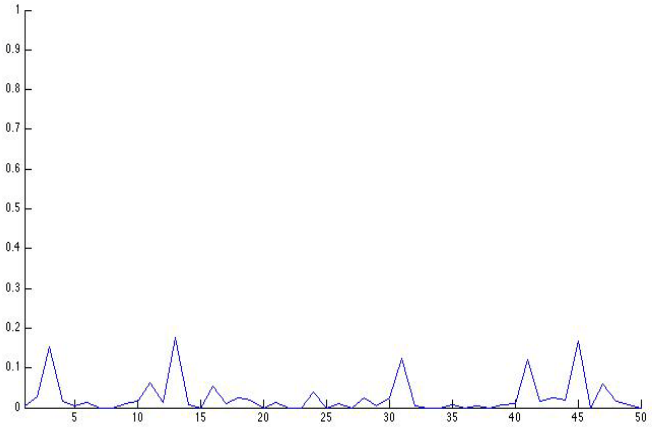

Fig. 6. The variation of empirical error for NP=150

Test 2. The evaluation of the error by counting the misclassified examples from a set of NC new test samples coming from the given classes of the same repartitions.

In this case, the learning is performed in an adaptive way, that is, each new classified example contributes to the new characteristics of the class the exampled is assigned to, the new characteristics being computed using first order approximations in terms of the previous ones. Besides, after each iteration, the characteristics of the new resulted classes are re-computed using an exact method **M**.

The tests were performed for NP=50,100,150, NC=10,20,30,40,50, n=50, e=100, where n is the data dimensionality and e is the number of epochs.

a) The classes correspond to the Gaussian repartition and Rayleigh repartition respectively. The variation of the empirical error in terms of e is presented in Figure 7 and Figure 8. In Figure 7 the test is performed for NP=100 and in Figure 8 the volume is NP=150

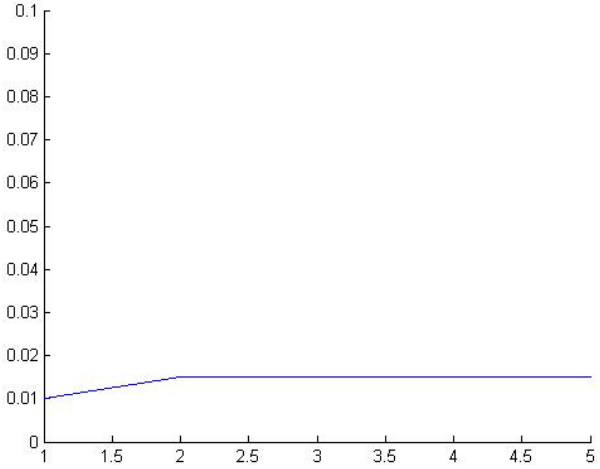

Fig. 7. The variation of empirical error for NP=100

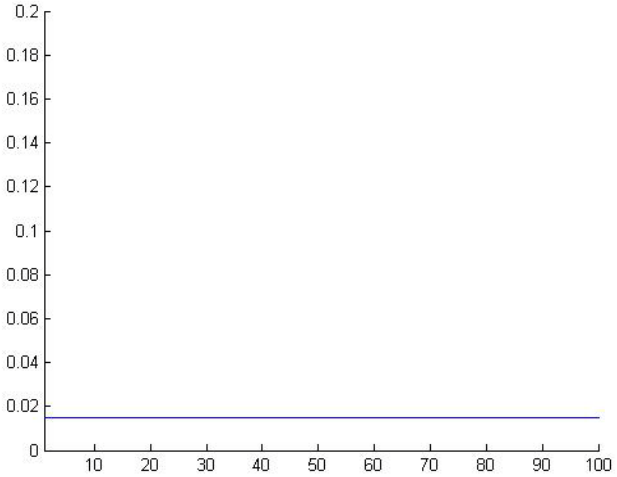

Fig. 8. The variation of empirical error for NP=150

b) The classes correspond to the geometric repartition and Rayleigh repartition respectively. The variation of the empirical error in terms of e is presented in Figure 9 and Figure 10. In Figure 9 the test is performed for NP=50 and in Figure 10 the volume is NP=150

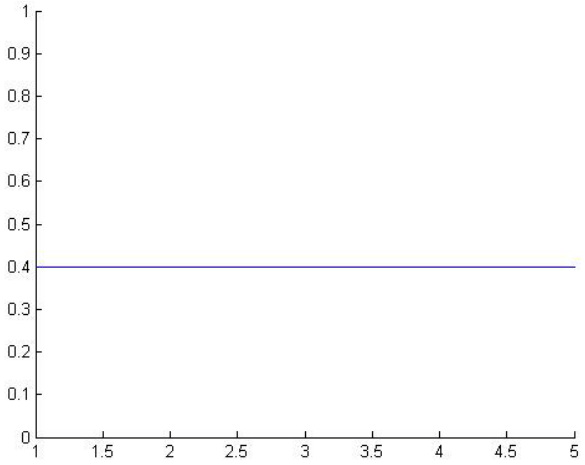

Fig. 9. The variation of empirical error for NP=50

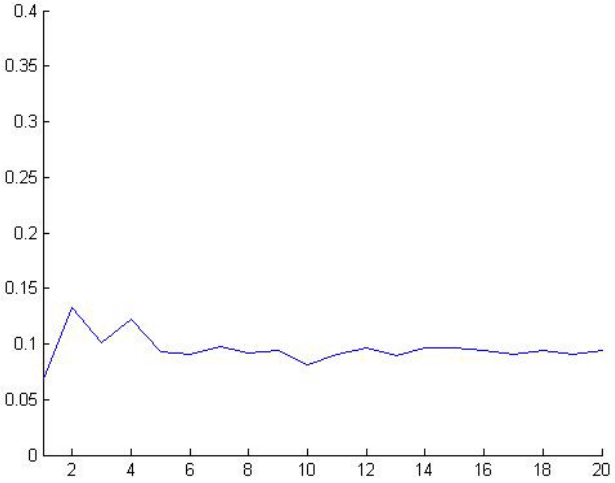

Fig. 10. The variation of empirical error for NP=150

c) The classes correspond to the Gaussian repartition. The variation of the empirical error in terms of e is presented in Figure 11 and Figure 12. In Figure 11 the test is performed for NP=50 and in Figure 12 the volume is NP=150

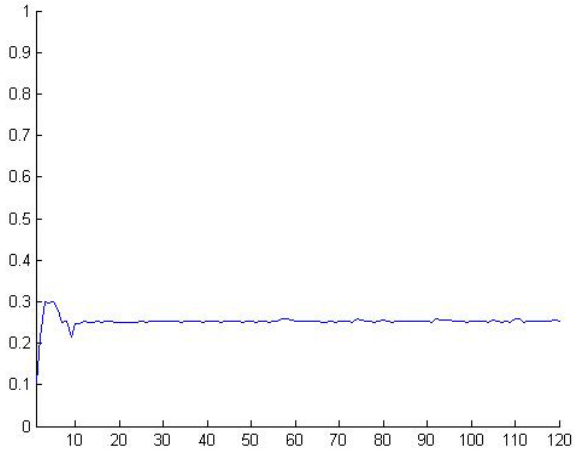

Fig. 11. The variation of empirical error for NP=50

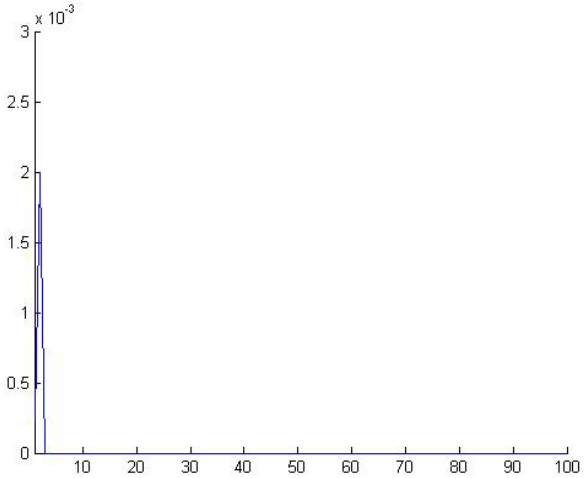

Fig. 12. The variation of empirical error for NP=150

The results of a test on a two-class problem in signal recognition are presented in Figure 13, Figure 14, and Figure 15.

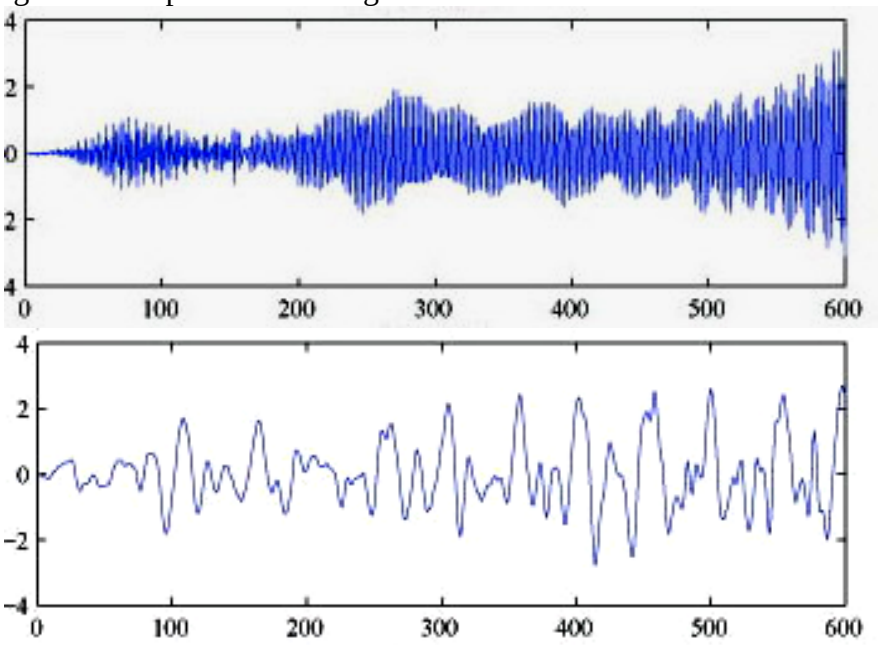

Fig. 13. The input signals

The samples are extracted from the signals depicted in Figure 13. In Figure 14 are represented the marginal probability distribution function (PDF) of each signal. The correct recognition of 20 new examples coming from these two classes using P1 failed in 2 cases.

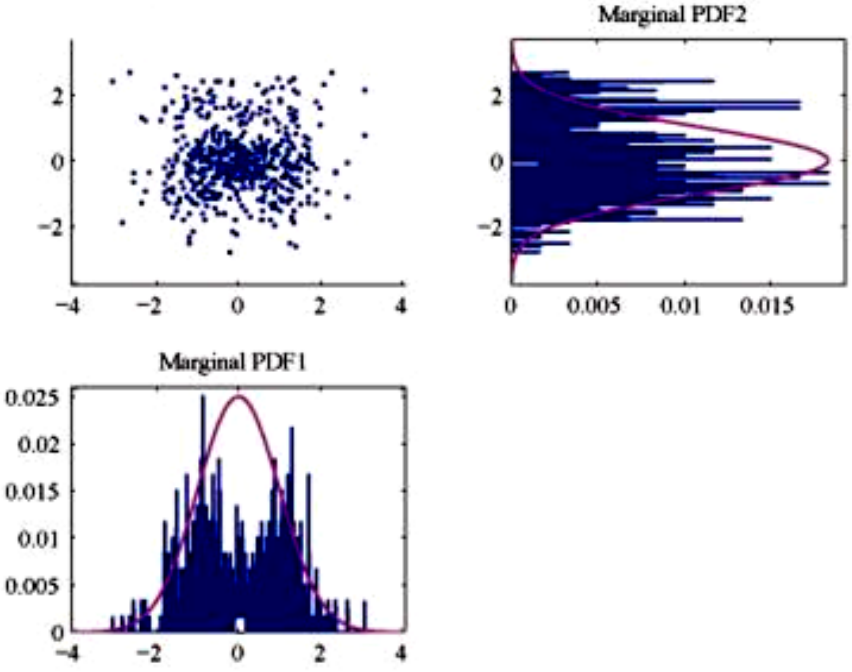

Fig. 14. The marginal PDF of the input signals

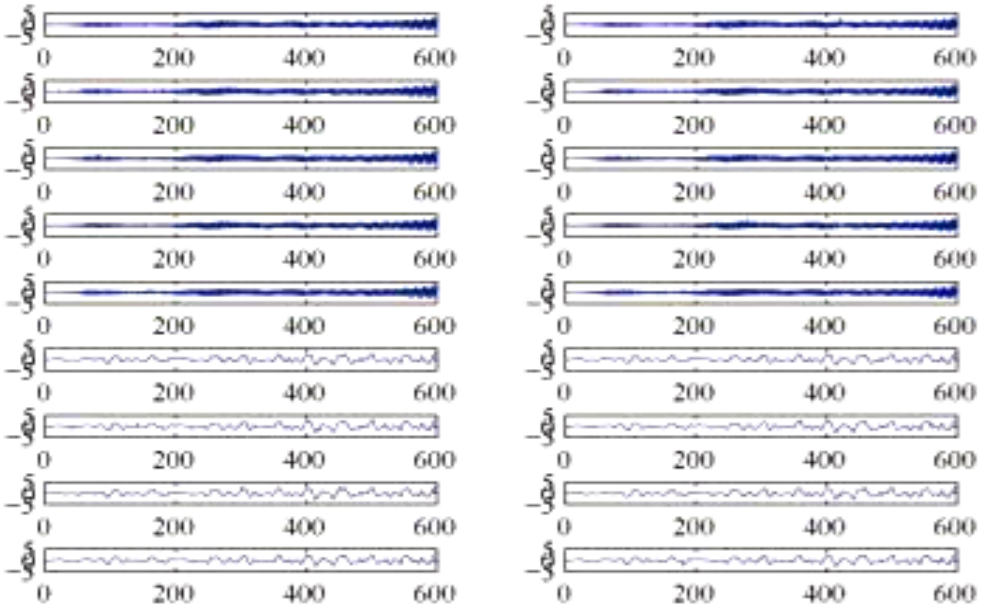

Fig. 15. A set of correct recognized examples using the proposed algorithm

The correctly recognized examples are presented in Figure 15. The performance was improved significantly when the volume of the initial samples increases. The values of the resulted empirical mean error are less than 0.05 (more than 95% new examples are correctly recognized).

References

[1] C. Cocianu, L. State, P. Vlamos, D. Constantin and C. Sararu, "PCA Supervised and Unsupervised Classifiers in Signal Processing", in *Proceedings of PRIS 2009*, Milan, Italy, 2009.

[2] C. Cocianu, L. State, I. Rosca, and P. Vlamos, "A New Adaptive Classification Scheme Based on Skeleton Information", in *Proceedings of SIGMAP 2007*, Barcelona, Spain, pp. 128-136, 2007.

[3] K.I. Diamantaras and S.Y. Kung, *Principal Component Neural Networks: theory and applications*, John Wiley &Sons, 1996.

[4] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, London, 1982.

[5] M. Friedman and A. Kandel, *Introduction to Pattern Recognition, statistical, structural, neural and fuzzy logic approaches*, World Scientific, Singapore, 1999.

[6] J. Goldberger, S. Roweis, G. Hinton and R. Salakhutdinov, "Neighbourhood Component Analysis", in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, pp. 324-341, 2004.

[7] A. D. Gordon, *Classification*, Chapman & Hall/CRC, 2nd Edition, 1999.

[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Springer, 2001.

[9] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*. John Wiley &Sons, 2001.

[10] D. T. Larose, *Data Mining. Methods and Models*. Wiley-Interscience, A John Wiley and Sons, Inc Publication, Hoboken, New Jersey, 2006.

[11] L. State, C. Cocianu, P. Vlamos and V. Stefanescu, "PCA-Based Data Mining Probabilistic and Fuzzy Approaches with Applications in Pattern Recognition", in *Proceedings of ICSOFT 2006*, Portugal, pp. 55-60, 2006.

[12] L. State, C. Cocianu and P. Vlamos, "An unsupervised skeleton based method to

discover the structure of the class system", in *Proceedings of the IEEE International Conference on Research Challenges in Information Science, RCIS 2008,*, Marrakech, Morocco, IEEE Computer Society Press, pp. 169-178, 2008.

Catalina-Lucia COCIANU, Professor, PhD, currently working with Academy of Economic Studies, Faculty of Cybernetics, Statistics and Informatics, Department of Informatics in Economy. Competence areas: statistical pattern recognition, digital image processing. Research in the fields of pattern recognition, data mining, signal processing. Author of 12 books and more than 70 papers published in national and international journals.

Luminita STATE, Professor, PhD, currently working with University of Pitesti, Department of Mathematics and Computer Science. Competence areas: artificial intelligence, machine learning, statistical pattern recognition, digital image processing. Research in the fields of machine learning, pattern recognition, neural computation. Author of 15 books and more than 120 papers published in national and international journals.

Doru CONSTANTIN, Lecturer, PhD, currently working with University of Pitesti, Department of Mathematics and Computer Science. Competence areas: pattern recognition, artificial intelligence, digital signal processing, programming logic. Research in the fields of pattern recognition, independent component analysis and neural computation. Author of more than 20 papers published in national and international journals.

Corina SARARU, Assistant, PhD, currently working with University of Pitesti, Department of Mathematics and Computer Science. Competence areas: digital image processing, pattern recognition, dynamical systems. Research in the fields of image processing and pattern recognition. Author of 6 papers published in national and international journals.