# Analysis and convergence of weighted dimensionality reduction methods

# Análisis y convergencia de métodos de reducción de dimensionalidad ponderados

*Juan Carlos Riaño Rojas.[1]\*, Flavio Augusto Prieto Ortiz.[2], Edgar Nelson Sánchez Camperos.[3], Carlos Daniel Acosta Medina[1], Germán Augusto Castellanos Domínguez[1]*

[1]Universidad Nacional de Colombia Sede Manizales. A.A. 127. Km 9 vía el aeropuerto campus la Nubia Manizales, Colombia

[2]Universidad Nacional de Colombia Sede Bogotá. A.A. 14490. Carrera 30 N° 45-03 Edificio IEI 406 Bogotá D.C., Colombia

[3]Centro de Investigaciones Avanzadas IPN, Cinvestav. Guadalajara. Av. Científica 1145 , colonia el Bajío, Zapopan , 45015, Jalisco, México

## Abstract

We propose to use a Fisher type discriminant objective function addressed to weighted principal component analysis (WPCA) and weighted regularized discriminant analysis (WRDA) for dimensionality reduction. Additionally, two different proofs for the convergence of the method are obtained. First one analytically, by using the completeness theorem, and second one algebraically, employing spectral decomposition. The objective function depends on two parameters $U$ matrix being the rotation and $D$ diagonal matrix weight of relevant features, respectively. These parameters are computed iteratively, in order to maximize the reduction. Relevant features were obtained by determining the eigenvector associated to the most weighted eigenvalue on the maximum value in $U$. Performance evaluation of the reduction methods was carried out on 70 benchmark databases. Results showed that weighted reduction methods presented the best behavior, PCA and PPCA lower than 17% while WPCA and WRDA higher than 45%. Particularly, WRDA method had the best performance in the 75% of the cases compared with the others studied here.

----- *Keywords:* PCA, PPCA, WPCA, WRDA, dimensionality reduction

\* AUTOR DE CORRESPONDENCIA falta

**Resumen**

En este trabajo se propone utilizar una función objetivo discriminante tipo de Fisher, para la reducción de la dimensionalidad, en el análisis de componentes principales ponderados (WPCA) y al análisis discriminante regularizado ponderados (WRDA). Además, se desarrollan dos pruebas de la convergencia del método. Primero analíticamente, usando el teorema de completitud, y una segunda prueba algebraica, empleando descomposición espectral. La función objetivo depende de dos parámetros: *U* matriz de rotaciones y *D* matriz pesos de características relevantes, respectivamente. Estos parámetros se calculan iterativamente, para maximizar la reducción. Las características relevantes fueron obtenidas determinando el vector propio asociado al valor propio con máximo valor en U. La evaluación del desempeño de los métodos de la reducción fue realizada sobre 70 bases de datos (benchmark). Los resultados mostraron que los métodos ponderados presentan un mejor comportamiento PCA y PPCA por debajo del 17% mientras que WPCA y WRDA por encima del 45%. Particularmente, el método WRDA tuvo el mejor funcionamiento en el 75% de los casos comparados con los otros estudiados en este trabajo.

*-----* *Palabras clave:* WPCA, WRDA, reducción de dimensión

## Introduction

The relevant information extraction from a data set with a great number of features has been considered as a big problem in machine learning and pattern recognition. These great sets appear normally in areas as bioinformatics and text recognition, where is common to find feature vectors with dimensions higher than $10^7$, but with a low number of relevant characteristics. Thus, the classification algorithm performance is limited and their computing time is high, reducing the application in real time tasks [1]. For solving this problem, irrelevant features that do not contribute to the extraction and selection process must be rejected, improving the classifiers performance. Traditionally, the dimensionality reduction has been developed by using linear techniques such as principal components analysis (PCA), probabilistic principal components analysis (PPCA) and factorial analysis [2-4]. Nevertheless, these linear techniques are not suitable for handling non linear complex data. For this reason, in recent years, a great number of non linear techniques for dimensionality reduction have been proposed several geometric methods for feature selection and dimensional

reduction. Where they divide the methods into projective methods and methods that model the manifold on which the data lies. For projective methods, we review projection pursuit, principal component analysis (PCA), kernel PCA, probabilistic PCA, and oriented PCA; and for the manifold methods, we review multidimensional scaling (MDS), landmark MDS, Isomap, locally linear embedding, Laplacian eigenmaps and spectral clustering [5-7]. Nevertheless they lack of convergence analysis. Non linear reduction techniques have a good performance in complex artificial tasks; however, they do not overcome the traditional linear techniques in real word tasks using several databases without carrying out formal proofs of this fact [8]. In [9] an algebraic weighted variables approximation is presented. It is based on the Kernel matrix spectral properties. The main contribution consists in obtaining relevant variables using the weighted objective function, proving its convergence on employing strong hypothesis from the analysis fundaments, but being different to those used in this paper. An exhaustive review in extraction and selection features methods classification, grouping them in two classes, has been done in [10]. The first one contains the binary search methods, which

at the same time are catalogued as exhaustive search methods but forbidden because of its computational cost. These find the optimum of the objective function but provide unstable and not optimal results. The second group includes the weighted methods that multiply the features by continuous values, in order to employ mathematical analysis techniques, for optimizing the objective function. In reference [11-14] descriptive studies of the reduction and regression methods are observed. Some of them show applications in pattern classification for identifying faces, while others present applications for materials science. Although the two last references have the same abbreviation (WPCA) they refer to other aspects, employing "W" for window or whitened, being different to the methods considered in this paper, because they carried out a local reduction.

In this work, a complete study of two of these weighted methods is included, WPCA and WRDA. These methods were already introduced in [15,16] by using an EM algorithm and without including a theoretical study. The main advantage of these methods is the capability of combining in one step two tasks (features selection and extraction), returning the called relevant features. Here we present the methods as weighted rotations for maximizing the objective function $J_4$, which is the matrix traces ratio that represents the inter-classes and intra-classes dispersion. From this definition, we develop a convergence analysis. The main convergence result is obtained from two different points of view. First analytically by using the completeness theorem and second algebraically, employing spectral decomposition.

The convergence was validated employing artificial and real data for selecting relevant variables from a set of 70 geometric features (areas, perimeters, fractal dimension, curvatures, Hausdorff dimension among others), statistical (correlations, means, entropy and moments) needed for characterizing patterns that can separate two classes. In order to evaluate the performance of these methods, the ROC surfaces hyper-volume (hyper-surface) has to be calculated

by using the Monte Carlo method, additional to the error classification.

# Methodology

## *Relevance and variables selection using weighting*

The variables selection problem can be understood as choosing a subset of $p$ features, from the whole features set $c$, that allows obtaining a suitable performance in the classification process. This kind of search is handled by some evaluation function named the relevant set. On the other hand, the extraction techniques carry out a transformation of $c$ features space to a lower dimension space. In order to guarantee the optimum solution, these methods execute an exhaustive search, increasing the computational cost. For solving this problem, heuristic methodology has been proposed, but producing unstable behavior respect to the objective function. Other alternative consists in using weighting methods, although they are not the optimal solution, they are more stable and flexible, producing a suitable solution [17]. Some of these methods are described below.

## *Weighted probabilistic PCA*

It is a particular factorial analysis case. In this method, the original $X$ features are observed as a linear combination of $Z$ factors group, joined to a specific error $V$ and $C$ that represent charge coefficients that are modeled according to (1):

$$X = CZ + V; \ \ Z \approx N(0, I), \ \ V \approx N(0, I) \quad (1)$$

Where the random variables $Z$ can be assumed non dependent and identically distributed, with a unitary spherical Gaussian variance. Note that there is a difference between the probabilistic model PCA (PPCA) and PCA, where the random variable variance can be associated to the diagonal elements of Z. The model also considers the general perturbation matrix $V$, but in [18] a restriction of Gaussian variance $R = \varepsilon I$ (isotropic noise) is stated. The previous

model formulation is modified by introducing the weights on the original variables (features) and thus containing the weighted rotation. Let $D$ be a diagonal matrix that contains the $i$-th variable in the $d_{ii}$ element. If the new variable is assumed as $d_{ii}x_i$, a new data subset $y = Dx$ is generated, where the probabilistic model is defined in (2):

$$Y = XD = CZ + V \qquad (2)$$

where $Z$ and $V$ are distributed as in equation (1). From this definition $Y$ is normally distributed, with mean equal to zero and the covariance given by (3):

$$D^{-1}CC^T D^{-1} + D^{-1}RD^{-1} \qquad (3)$$

The weights are found in $D$ and they are responsible of generating the noise for the $X$ variables. From (3) the EM (*Expectation - Maximization*) algorithm is obtained, in order to estimate the unknown variables state in the $E$-step and maximizing the total probability from the estimation of $Z$ and the observation of $Y$, in the $M$-step. $E$-step *and* $M$-step are observed in equation (4) and (5), respectively:

$$step-E: \quad Z^T = (C^T C)^{-1} C^T D X^T \qquad (4)$$

$$step-M: \quad C = DX^T Z(Z^T Z)^{-1} \qquad (5)$$

### Weighted regularized discriminant Analysis WRDA

RDA was proposed by Friedman [19] for being used in small samples, where data possess high dimensionality, trying to overcome the discriminant rule degradation. In this document, they were identified as the regularized linear discriminant analysis method. The aim of this technique is to find the lineal projection space where the dispersion between classes was the maximum value. One way consists in maximizing the ratio between projected classes in the dispersion matrix inter-classes $\Sigma_B$ and the dispersion matrix intra-class $\Sigma_W$, as is expressed in (6):

$$J = \frac{|W^T \Sigma_B W|}{|W^T \Sigma_W W|} \qquad (6)$$

where $W$ is the projection matrix, which dimension is defined by the number of linearly separated classes $k$. The aim is to maximize the previous objective function (6), under restriction $|W^T \Sigma_W W| = 1$. The solution is obtained employing the Lagrange multipliers, the solutions are $k$-1 eigenvectors generalized from $\Sigma_B$ and $\Sigma_W$, that correspond to the principal eigenvectors of $\Sigma_W^{-1} \Sigma_B$. The regularization is needed because for small samples size, $\Sigma_W$ cannot be inverted directly. Then, the solution would be reformulated as is expressed in the next equation, where $\Lambda$ is the eigenvalues matrix (a diagonal matrix) as is presented in (7):

$$(\Sigma_W + \delta I)^{-1} \Sigma_B W = W\Lambda \qquad (7)$$

After data weighting by obtaining $XD$, where $D$ be a diagonal matrix, the function to optimize is transformed in (8):

$$J_D = \frac{|W^T D\Sigma_B DW|}{|W^T D\Sigma_W DW|} \qquad (8)$$

### Weighted variables and relevance criteria

In previous subsections some weighted linear transformations were defined. Now, the interest is to project data in a $f$ dimension space. Such dimension depends on the chosen rotation criterion; for instance, there is a bi-class problem and the convergence of WRDA is required, the fixed dimension must be $f = 1$, in order to reach the convergence. For evaluating the weighted projection relevance at a fixed dimension, the measurement of separability is required. The parameter to be optimized is the weight diagonal matrix $D$, and the selected criterion is the inter-classes and intra-classes dispersion matrix traces coefficient, known as $J_4$ [20]. For projected and weighted data, this measurement is given by (9):

$$J_4(D,\Phi) = \frac{trace(\Phi^T D \Sigma_B D \Phi)}{trace(\Phi^T D \Sigma_W D \Phi)} \qquad (9)$$

where $\Phi$ can represent $U$ or $W$. The size of $\Phi$ is $c \times f$ and $f$ denotes the fixed dimension, corresponding to the number of projection vectors $\Phi$ such that: $\Phi = (\phi_1, \phi_2, ..., \phi_f)$.

Rewriting the $D$ matrix as a column vector $d$, and using the Hadamard matrix product (expressed as $\circ$), the traces of equation (9) can be rewritten as is presented in (10):

$$traza(\Phi^T D \Sigma_B D \Phi) =$$

$$\sum_{i=1}^{f} d^T (K \circ \phi_i \phi_i^T) d = d^T \left( \sum_{i=1}^{f} K \circ \phi_i \phi_i^T \right) d \quad (10).$$

$$(10)$$

Then, equation (9) is transformed in equation (11):

$$J_4(d) = \frac{d^T \left( \sum_{i=1}^{f} \Sigma_B \circ \phi_i \phi_i^T \right) d}{d^T \left( \sum_{i=1}^{f} \Sigma_W \circ \phi_i \phi_i^T \right) d} \qquad (11)$$

This function is essentially equal to the LDA (linear discriminant analysis) function, then, the solution of $d$ with norm $L_2$, will be chosen by the principal eigenvector given by (12):

$$\left( \sum_{i=1}^{f} \Sigma_W \circ \phi_i \phi_i^T \right)^{-1} \left( \sum_{i=1}^{f} \Sigma_B \circ \phi_i \phi_i^T \right) \qquad (12)$$

Note that this kind of description assumes that the $\Phi$ elements are static; this problem is overcome by overlapping $d$ and $\Phi$ calculation, until the convergence of both is reached [9]. Related to weight interpretability, in order to define the dispersion, positive values are generally required. Nevertheless, in the context of relevance function used in this work, negative values can be obtained, this is avoided taking the absolute value of $d$.

### Weighted reduction WPCA and WRDA

In [9] the convergence towards local maximum, using the power method applied to Q-α method with objective function similar to $J_4$ is shown. Such proof was carried out for a particular case, where the objective function is poor, which means that there is a subset of characteristics including a coherent cluster and a positive function, conditions that normally can not be demanded. For this reason, in this section the convergence objective function in WPCA is proved, but it requires using the lemma 1, which can be demonstrated from two perspectives: analytic and algebraic.

The next lemma guaranties the objective function has a maximum. Moreover, it is observed that any search method converges to the same limit.

*Lemma:* The objective function

$$J_4(W,D) = d^T \left( B \cdot \sum_{i=1}^{K} w_i w_i^T + \delta \ Id \right)^{-1} A \cdot \sum_{i=1}^{K} w_i w_i^T d$$

has a maximum.

*Proof (analytic version):* the objective function can be represented as: $J(W,D) = \langle d, \Gamma_W(d) \rangle$ where

$$\Gamma_w(d) = \left( B \cdot \sum_{i=1}^{K} w_i w_i^T + \delta \ Id \right)^{-1} A \cdot \sum_{i=1}^{K} w_i w_i^T d.$$

Let be $C = \left\{ \langle d, \Gamma_W(d) \rangle : \|d\| = 1 \right\}$ a set. The set is not empty since any eigenvector $d = \beta_i$ satisfies the condition in the $C$ set. Now, it is necessary to show that the set posses supremum. Taking a base of eigenvectors $\{\beta_1, \beta_2, ..., \beta_n\}$, associated to the transformation $\Gamma_W$, then, for any $d$ vector conformed by a linear combination of $\beta_i$, it is had that $d = \sum_{i=1}^{n} c_i \beta_i$, such that:

$$J(W,D) = \langle d, \Gamma_W(d) \rangle = \left\langle \sum_{i=1}^{n} c_i \beta_i, \Gamma_W \left( \sum_{i=1}^{n} c_i \beta_i \right) \right\rangle$$
$$= \sum_{i=1}^{n} \lambda_i c_i^2, \text{ where } \|d\|^2 = \left\| \sum_{i=1}^{n} c_i^2 \right\| > 1. \text{ Then } C \text{ is}$$

upper bounded, because of the supremum axiom $sup(C)$ exists. Considering $\|d\|^2 = \sum_{i=1}^{n} c_i^2 = 1$ and $\hat{\lambda} = \max_{i \leq i \leq n} \{\lambda_i\}$ it implies that $sup(C) = \hat{\lambda}$. For

this reason, if $d = \hat{\beta}$ the eigenvector associated to the greater eigenvalue $\hat{\lambda}$, maximizes the objective function.

*Proof 2 (algebraic version):* The objective function $J(w,D) = trace(W^T\,DADW)$ is maximized, under restrictions: $W^TW = Id$, $trace(W^T\,DBDW) = 1$ where $D$ be a diagonal matrix and $\|d\| = 1$. The original data is stored in the $X$ matrix and its covariance $X^TX$ is analyzed, written it as $X^TX = A + B$, being $A$ and $B$ symmetric and positive semi-defined matrixes, which can be substituted by Cholesky decomposition: $X^TX = A_1^T A_1 + B_1^T B_1$. Multiplying the left side by $W_1D$ and the right side by $DW$ and taken the traces, next expression is obtained:

$$J(W,D) = traza\left(W^T\,DA_1^T\,A_1DW\right) = traza\left(DA_1^T\,A_1DWW^T\right).$$

Using the $W$ orthogonality conditions, it can be written as $J(W,D) = traza\left(DA_1^T\,A_1D\right) = \|A_1D\|_F^2$.

Finally it can be expressed as $J(W,D) = \left\|[d_1 A_1(1,:), d_2 A_1(2,:), \cdots, d_n A_1(n,:)]^T\right\|_2^2$. Its matrix representation is given by:

$$J(W,D) = \left\|\begin{bmatrix} A_1(:,1) & 0 & \cdots & 0 \\ 0 & A_1(:,2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_1(:,n) \end{bmatrix}\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}\right\|_2^2.$$

Then, $trace\,(W^T\,DB_1^T\,B_1DW) = 1$ implies that $\sum d_i^2 B_1(:,i)^T B_1(:,i) = 1$, and it can be transformed in $\sum \tilde{d}_i = 1$ where $\tilde{d}_i = d_i y_i$, such that $y_i = B_1(:,i)^T B_1(:,i)$. Therefore $D$ can be expressed in a matrix form as:

$$J(W,D) = \left\|\begin{bmatrix} 1/y_1 & 0 & \cdots & 0 \\ 0 & 1/y_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/y_n \end{bmatrix}\begin{bmatrix} \tilde{d}_1 \\ \tilde{d}_2 \\ \vdots \\ \tilde{d}_n \end{bmatrix}\right\|_2^2.$$

The last expression can be transformed in the next function for maximizing: $J(W,D) = \|M\tilde{d}\|_2^2$, joined to $\|\tilde{d}\| = 1$, which has as a maximum $\|M\|$ as is evident.

Following, the WPCA algorithm and its convergence is presented.

### *WPCA algorithm*

The iterative nature and the convergence of EM and PCA probabilistic parameters estimation is used for obtaining the WPCA algorithm, which is described as follow, employing $r$ as the iteration index:

i  Normalize each characteristics vector for obtaining zero mean and the one Euclidean norm ($\|x\|_2 = 1$).

ii  Start with some orthogonal set of vectors $U^{(0)}$.

iii  Calculate $D^{(r)}$ from the solution given in equation (12) and weighted data.

iv  Calculate the *step-E* and *step-M*, from equation (4) and (5) respectively. Normalize the $C$ columns de $C$ for obtaining $\|C(:,i)\|_2 = 1$.

v  If $\|C^{(r)} - C^{(r-1)}\|_2 > \varepsilon$, return to numeral iii.

vi  Orthonormalize the subspace obtained, finding its singular values decomposition (SVD), as follow: $\text{SVD}(C^T DX^T XDC) = ASA^T$, $C_{final} = A^T C$ where $A$, $S$ the elements obtained from the decomposition SVD.

### *WPCA convergence*

As was stated in the last section, weighting the characteristics by integrating with *EM* method can guarantee the convergence of steps $D$, and it is possible to ensure the relevant features obtained. From equation (4) the relationship (13) was reach in the algorithm:

$$D^{(r)}X = C^{(r)}Z^{(r)} + V^{(r)} \qquad (13)$$

Again, $r$ corresponds to the iteration. As *EM* is applied (increasing $r$) the perturbation $V^{(r)}$ decreases, due to it is approximated to the most discriminant axes. That is, if $r \to \infty$, then $\|V^{(r)}\| \to 0$, guarantying the convergence.

*Theorem 1:* If $C(r) \to \hat{C}$ y $Z^{(r)} \to \hat{Z}$ then $D^{(r)} \to \hat{D}$.

*Proof:* Given the Eq. (13) for iterations $r$ and $r + 1$, the subtraction produces: $(D^{(r+1)} - D^{(r)})X = (C^{(r+1)}Z^{(r+1)} - C^{(r)}Z^{(r)}) + (V^{(r+1)} - V^{(r)})$. Applying to the last relationship, any type of norm follows: $\left\|(D^{(r+1)} - D^{(r)})X\right\| \leq \left\|C^{(r+1)}Z^{(r+1)} - C^{(r)}Z^{(r)} + V^{(r+1)} - V^{(r)}\right\|$. It is known that if $r \to \infty$, then $\|V^{(r)}\| \to 0$, then: $\left\|(D^{(r+1)} - D^{(r)})X\right\| \leq \left\|C^{(r+1)}Z^{(r+1)} - C^{(r)}Z^{(r)}\right\|$. Adding and subtracting $C^{(r+1)}Z^{(r)}$, to the right side, it is obtained: $\left\|(D^{(r+1)} - D^{(r)})X\right\| \leq \left\|C^{(r+1)}(Z^{(r+1)} - Z^{(r)}) + (C^{(r+1)} - C^{(r)})Z^{(r)}\right\|$. Applying the triangular inequality and the multiplicative property, the expression: $\left\|(D^{(r+1)} - D^{(r)})\right\|\|X\| \leq \left\|C^{(r+1)}\right\|\left\|(Z^{(r+1)} - Z^{(r)})\right\| + \left\|(C^{(r+1)} - C^{(r)})\right\|\left\|Z^{(r)}\right\|$ is reached, when $r \to \infty$, then $\left\|(D^{(r+1)} - D^{(r)})\right\|\|X\| \leq \left\|\hat{C}\right\|\left\|(Z^{(r+1)} - Z^{(r)})\right\| + \left\|(C^{(r+1)} - C^{(r)})\right\|\|\hat{Z}\|$. By hypothesis, $X$ is the original data matrix, then $\|X\| > 0$. In a norm space, if a sequence converges, is a sequence of Cauchy. If $r \to \infty$, then $\left\|(C^{(r+1)} - C^{(r)}\right\| \to 0$, $\left\|Z^{(r+1)} - Z^{(r)}\right\| \to 0$, and $\left\|D^{(r+1)} - D^{(r)}\right\| \to 0$. Thus the weight convergence working in a Banach space is had.

### WRDA Algorithm

For this algorithm, errors produced by the rotation of weighted data are not important, since not only the function of each rotation but also the weighting function have similar directions. The algorithm is described as follow:

Fix the $k - 1$ dimension, being $k$ the number of classes.

Normalize each vector of feature for having media zero and Euclidian norm one.

Start with some orthogonal set of vectors $W^{(0)}$.

Calculate $d^{(r)}$ from the solution of equation (12), weighting the data.

Calculate the $W^{(r)}$ from equation (8) y (9).

If $\left\|W^{(r)} - W^{(r-1)}\right\|_2 > \varepsilon$, return to numeral iii. $\varepsilon$ is the fixed error in the process.

Its objective function is precisely the observed in the lemma 1.

## Results and analysis

By using a support vectors machine (SVM) classifier and evaluating its performance using two different approaches: employing ROC curves and hyper-surfaces and using the classification error, the dimensionality reduction PCA, PPCA, WPCA y WRDA techniques behavior was studied. Real data generated from geometric features as area, perimeters, orientations, dispersion, centroids and different statistical moments were used, obtaining 70 features applied to 50 capillary images of people without lupus erythematosus and 50 capillary images of ill people. In figure 1, data projection on the principal plane is shown. It is observed that in WPCA and WRDA methods, the data clouds belonging to healthy people class (circle) versus lupus erythematosus ill people class (cross), are more compacted. They vary between -0.1 and 0.1 on the horizontal axis and take values between -0.04 and 0.04 on the vertical axis. While in PCA and PPCA methods, the clouds are enlarged and the horizontal axis varies between -0.5 and -0.5 and the vertical axis varies between -0.3 and 0.2.

In figure 2, the reduction methods performance is shown. In the left curve, the highest point indicates that WRDA obtained the greatest efficiency percent, while the right curve shows that WRDA has the lowest classification error compared with the other methods.
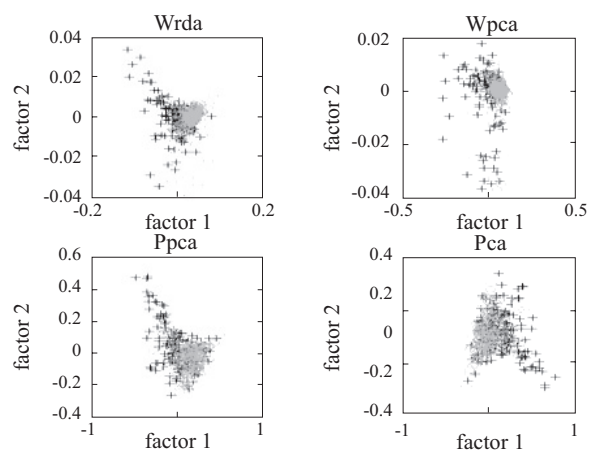
**Figure 1** Projection of reduction features methods PCA, PPCA WPCA and WRDA applied to the set of feature extraction from capillary images
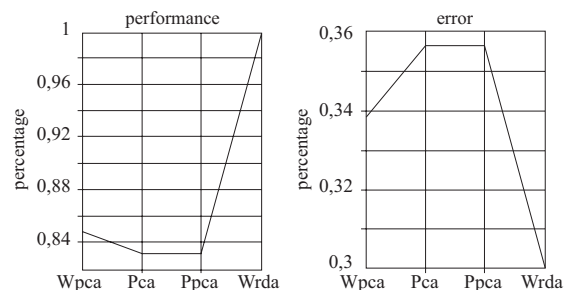


**Figure 2** Performance evaluation of reduction methods for data related with 100 capillary images of persons with lupus erytematosus and persons without lupus

The figure 3 shows the reduction methods performance in the classification applied to 70 databases discharged from the website [21]. From the figure, it is observed that the WRDA method presented the best performance compared to the others methods.
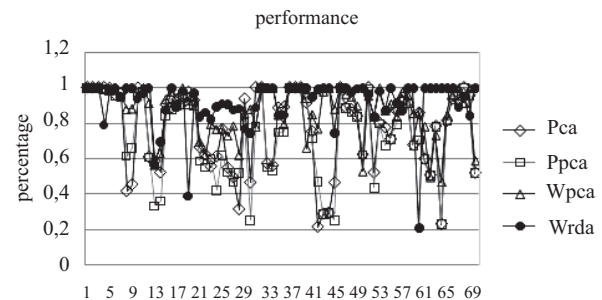


**Figure 3** Reduction methods performance applied to 70 databases

In table 1, the performance behavior and classifier errors of reduction methods applied to 70 databases were summarized. The values were obtained by using the mean, median, standard deviation and minima. In row "number of match" a best performance for WRDA was observed in 53 of the 70 databases, while PCA and PPCA were the worst.

**Table 1** Reduction methods evaluation

| | Performance evaluation in percent of the weighted dimensionally reduction methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **PCA** | | **PPCA** | | **WPCA** | | **WRDA** | |
| | *Efficiency* | *Error* | *Efficiency* | *Error* | *Efficiency* | *Error* | *Efficiency* | *Error* |
| Mean | 0.7265 | 0.0013 | 0.7341 | 0.0013 | 0.8678 | 0.0008 | 0.9146 | 0.0008 |
| Deviation | 0.2393 | 0.0012 | 0.2304 | 0.0013 | 0.1427 | 0.0006 | 0.1402 | 0.0008 |
| Median | 0.7956 | 0.0010 | 0.7874 | 0.0010 | 0.9109 | 0.0007 | 0.9729 | 0.0006 |
| Minimum | 0.2101 | 0 | 0.2283 | 0 | 0.4658 | 0 | 0.2047 | 0 |
| No. Mach | 11 | | 12 | | 30 | | 53 | |

## Conclusions

Proofs of WPCA and WRDA reduction methods were carried out from two points of view: algebraic and geometric. Such proofs were relatively weaker than those known for other weighted methods. Results of weighted reduction methods PPCA, WPCA and WRDA performance were present. These results indicate that WRDA method has a best performance in 75% of databases, being the first compared to the other methods. In the literature, no rigorous proofs of the WPCA method and its variants are reported. For this reason, great part of the document was dedicated to carry out formalization tests, one analytic and other algebraic.

Capillary images showed great complexity for extracting relevant features due to the signal noise ratio. The feature reduction methods implemented in this paper. It was confirmed by the errors generated in the order of 30%. It can be observed since the classes were highly overlapped. Nevertheless, the method with the best performance was WRDA, reducing the classification error in spite of the classes overlapping.

## Acknowledgments

## References

1. B. A. Olshausen, D. J. Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature*. Vol. 381. 1996. pp. 607-609.

2. C. M. Bishop. *Pattern recognition and machine learning*. Ed. Springer. New York. 2006. pp. 1-738.

3. M. E. Tipping, C. M. Bishop. "Mixtures of probabilistic principal component analyzers". *Neural Computation*. Vol. 11. 1999. pp. 443-482.

4. A. Sharma, K. K. Paliwal, G. C. Onwubolu. "Pattern Classification: An Improvement Using Combination of VQ and PCA Based Techniques". *American Journal of Applied Sciences*. Vol. 2. 2005. pp. 1445-1455.

5. J. C. Burges. "Geometric Methods for Feature Selection and Dimensional Reduction" in: O. Maimon, L. Rokach. (editors). *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*.. Ed. Springer. New York. 2006. pp. 59-91.

6. L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, D. D. Lee. "Spectral methods for dimensionality reduction". In: O Chapelle, B. Schölkopf, A. Zien. (editores). *Semisupervised Learning*. Ed. MIT Press. Cambridge. Massachusetts. USA. 2006. pp. 279-294.

7. J. Venna. *Dimensionality reduction for visual exploration of similarity structures*. PhD thesis. Helsinki University of Technology. Helsinki. 2007. pp. 11-32.

8. L. V. Maaten, E. Postma, J. V. Herik. "Dimensionality Reduction: A Comparative Review". *Elsevier Journal of Machine Learning Research*. Vol. 10. 2009. pp. 1-41.

9. L. Wolf, A. Shashua. "Feature Selection for Unsupervised and Supervised Inference: The Emergence of Sparsity in a Weight-Based Approach". *Journal of Machine Learning Research*. Vol. 6. 2005. pp. 1855-1887.

10. A. Blum, P. Langley. "Selection of relevant features and examples in machine learning". *Artificial Intelligence*. Vol. 97. 1997. pp. 245-271.

11. M. A. Turk, A. P. Pentland. "Face Recognition Using Eigenfaces". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Maui. Hawaii. USA. Vol. 1. 1991. pp. 586-591.

12. G. Balcerowska, R. Siuda. "Inelastic background subtraction from a set of angle-dependent XPS spectra using PCA and polynomial approximation". *Vacuum*. Vol. 54. 1999. pp.195-199.

13. W. Deng, J. Hu, J. Guo, W. Cai, D. Feng. "Robust, accurate and efficient face recognition froma single training image: A uniform pursuit approach". *Pattern Recognition*. Vol. 43. 2010. pp. 1748–1762.

14. Q. Zhao, H. Lu, D. Zhang. "A fast evolutionary pursuit algorithm based on linearly combining vectors". *Pattern Recognition*. Vol. 39. 2006. pp. 310-312.

15. L. G. Sánchez-Giraldo, F. Martínez-Tabares, G. Castellanos-Domínguez. "Functional Feature Selection by Weighted Projections in Pathological Voice Detection". *Lecture Notes in Computer Science*. Vol. 5856. 2009. pp. 329-336.

16. L. G. Sánchez-Giraldo, G. Castellanos-Domínguez. "Weighted feature extraction with a functional data extension". *Neurocomputing*. Vol. 73. 2010. pp. 1760-1773.

17. D. Skocaj, A. Leonardis. "Weighted and robust incremental method for subspace learning". *Proceedings of the Ninth IEEE International Conference on Computer Vision*. Vol. 2. 2003. *pp*. 1494-1501.

18. M. Tipping, C. Bishop. "Probabilistic principal component analysis". *Journal of the Royal Statistical Society. Series B*. Vol. 61. 1999. pp. 611-622.

19. J. H. Friedman. "Regularized discriminant analysis". *Journal of the American Statistical Association*. Vol. 84. 1989. pp. 165-175.

20. A. R. Webb. *Statistical Pattern Recognition*. 2ª. ed. Ed. John Willey and Sons. London. 2002. pp. 305-360.

21. C. Lai. W. J. Lee. M. Loog. P. Paclik. D. Tax. URL:http://ict.ewi.tudelft.nl/~davidt/occ/index.html. Consultada el 1 de Julio de 2009.