Input Data Preprocessing with Anisotropous Whitening Transformations

Fulgencio S. Buendía Buendía[1], Benjamin Ojeda Magaña[2], Antonio Vega-Corona[3] and Diego Andina[1]

[1] Universidad Politécnica de Madrid
Departamento de Señales, Sistemas y Radiocomunicaciones, E.T.S.I. Telecomunicación, Madrid, Spain
[2] Universidad de Guadalajara, México
[3] Universidad de Guanajuato
F.I.M.E.E., Guanajuato, México
diego@gc.ssr.upm.es

Abstract. In this paper the Anisotropous Whitening Transformation is proposed in order to perform input data preprocessing when multiple classes/distributions are managed in a pattern recognition problem. We present how to apply anisotropous transformations as an alternative to classical preprocessing techniques like PCA, Fisher or whitening ones. This transformation performs a clustering of the training patterns, in the way that neighbors samples tend to become together decreasing the influence when the distance grows. We summarize this paper in a brief description of the presented possibility of application for this transformation.

1 Introduction

The first step in a classification problem is the preprocessing of the data, most of the usual techniques applied to preprocess the input data in a pattern classification problem are linear, and always isotropous, that means that this transformation is equally applied to the whole data space [1]. There are multitude of different technics applied to handle the signals preprocessing as PCA, whitening, ICA, etc, depending on every certain problem, the proper one is selected. When the problem involves different distributions/classes, it is used to obtain the transformation to one of them and then apply it to all, but it usually doesn't produces satisfactory results. There are also procedures like the Fisher Discriminant that finds the axis transformation that maximizes the separability between classes. We propose the application of anisotropous transformations, in the sense that the influence for a transformation is maximum in the center of it and decreases with the distance, so it can be seen as an attractor self configured by the samples distribution. It is designed to handle the preprocessing in classification problems with different classes, and prepare the data to apply Radial Basis functions [2].

© A. Gelbukh, C. Yáñez Márquez, O. Camacho Nieto (Eds.)
Advances in Artificial Intelligence and Computer Science
Research on Computing Science 14, 2005, pp. 93-98

In chapter 3 this transformation is deeply explained. In chapter 5 we briefly conclude the paper remarking the most important aspects and possibilities of the AWT.

2 Whitening Transformations

The whitening transformation has been broadly studied and applied in different systems (Kalman filters, Spread Spectrum technics, etc)[3]; in essence, it consists in setting the covariance matrix of the distribution to the unity. In first place, the covariance matrix of the samples is obtained, and then its eigenvalues and eigenvectors. The eigenvectors represent the direction of the principal axes and the eigenvalues the variance of each of them; in this reference system the covariance matrix is diagonal and the diagonal elements are the eigenvalues. Being

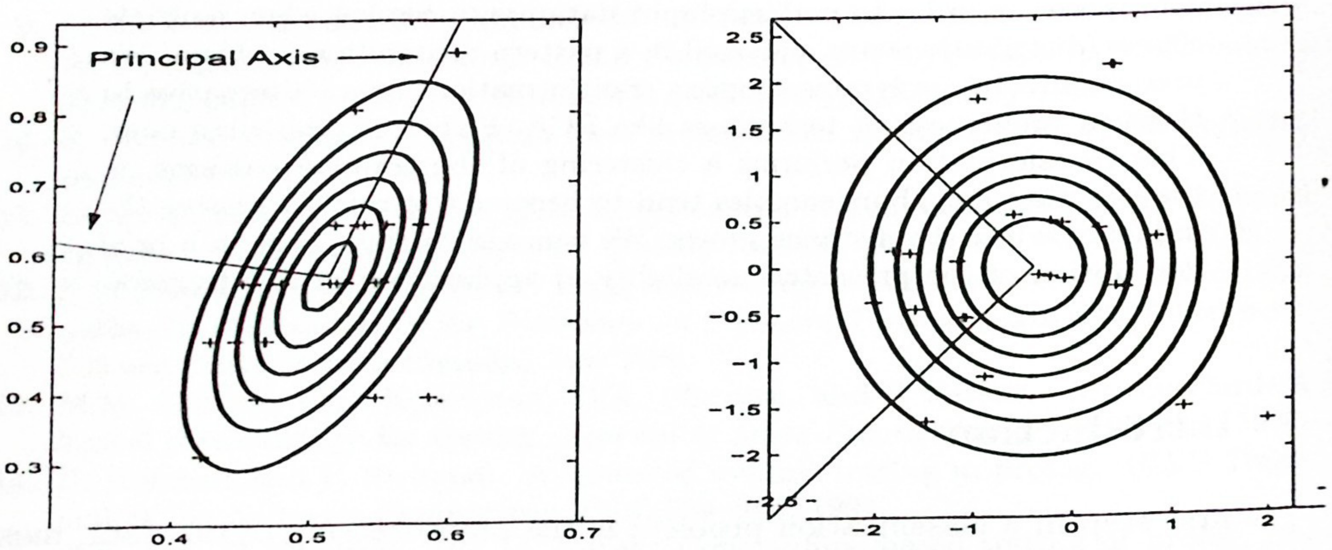

Fig. 1. Different Influence Functions for the whitening.

$\overline{\overline{C}}$ the covariance matrix, the whitening transformation is carried out obtaining the eigenvalues and eigenvectors of it, let be Φ the matrix whose columns are the eigenvectors and Λ the diagonal matrix of the eigenvalues, the whitening transformation is [1]:

$$\mathbf{A}_w = \mathbf{\Phi}\mathbf{\Lambda}^{-0.5} \tag{1}$$

3 Anisotropous Whitening Transformations

In order to be capable to apply this transformation to all the possible distributions in a set of samples, it is enough multiply the transformation by a function which influence over the space but the distribution were null. The transformation will be explained for a single distribution in $\mathbb{R}^2$,**A**. As has been explained in the previous section (2) first, the eigenvalues and eigenvectors are obtained.

$$\overline{\overline{\mathbf{C}}}_a \lambda_a = \overline{\overline{\mathbf{C}}}_a \overline{\mathbf{v}}_a \quad (2)$$

Then, it is subtracted the mean of the distribution, $\overline{m}_a$, to all the samples and then are multiplied by the matrix whose columns are the eigenvectors obtained $\overline{\overline{\mathbf{R}}} = (\overline{\mathbf{v}}_{a_1}, \overline{\mathbf{v}}_{a_2})$; these operations changes the reference system to the center of the principal axis of **A**.
To convert **A** into a circular distribution it is chosen its minimum eigenvalue (lets suppose λ_{a_2} corresponding to $\overline{\mathbf{v}}_{a_2}$), $\lambda_{a_1} = min(\lambda_{a_1}, \lambda_{a_2})$ and then set the other one equal to it. Taking into account that given a set of samples **X** and "a" a constant, $cov(a\mathbf{X}) = a^2 cov(\mathbf{X})$, the following matrix will set the variances of both axes equal:

$$\mathbf{C} \overline{\overline{=}} \begin{pmatrix} (\lambda_{a_2}/\lambda_{a_1})^{0.5} & 0 \\ 0 & 1 \end{pmatrix} \quad (3)$$

Once the samples have been compressed, we could came back to the original axes, just multiplying by the inverse of the rotating matrix and adding the mean vector, it would make **A** be almost whitened with the same mean as in the beginning. Note that we could set the covariance matrix of **A** to the identity $\overline{\overline{\mathbf{I}}}$ just applying this matrix instead of 3:

$$\overline{\overline{\mathbf{C_w}}} = \begin{pmatrix} (1/\lambda_{a_1})^{0.5} & 0 \\ 0 & (1/\lambda_{a_2})^{0.5} \end{pmatrix} \quad (4)$$

Until here the transformation made to an input vector $\overline{u}$ is:

$$\overline{v} = [(\overline{u} - \overline{m}_a)\overline{\overline{\mathbf{R}}}\overline{\overline{\mathbf{C}}}\overline{\overline{\mathbf{R}}}^{-1} + \overline{m}_a] \quad (5)$$

This transformation would modify equally both distributions **A** and **B**; so in order make that the transformation affects **A** without disturbing **B**, for every input sample $v = (x, y)$, (3) could be modified as:

$$\overline{\overline{\mathbf{C}}} = \begin{pmatrix} [1 - (\lambda_{a_2}/\lambda_{a_1})^{0.5} I(x,\sigma)] & 0 \\ 0 & 1 \end{pmatrix} \quad (6)$$

The a_{11} element of the matrix has been changed to

$$f(x,\sigma) = 1 - (\lambda_{a_2}/\lambda_{a_1})^{0.5} I(x,\sigma) \quad (7)$$

We will call this function, $I(x,\sigma)$, Influence Function (IF), that should accomplish the following conditions:

• $I(x,\sigma) = 1$ if $x = 0$

• $I(x,\sigma) \longrightarrow 0$ if $x \longrightarrow \infty$
• $I(x,\sigma)$ is C^1 and monotonous decressing

There are many functions that handle these requisites, the easiest is one minus an inverted gaussian, another example could be the following one:

$$I(x,\sigma) = 1 - e^{(\frac{-x^n * n}{\sigma^2})} \tag{8}$$

It is similar to a inverted gaussian but the n index. In figure(2) it can be seen the function with different values for the n index and his comparison to gaussian.

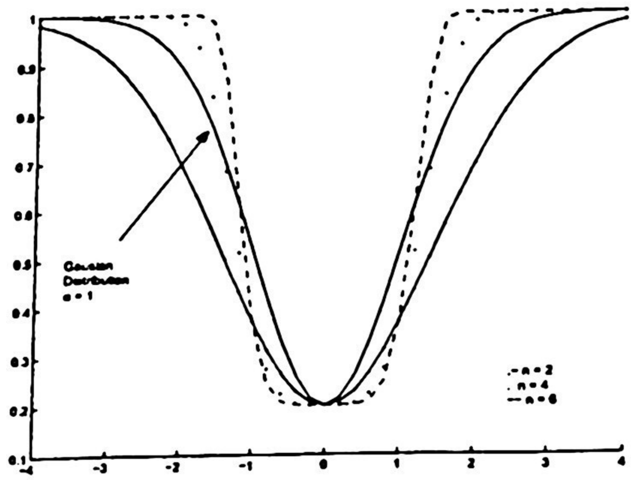

Fig. 2. Different Influence Functions for the whitening.

It can be observed that (8) is minimum in a ration of the function's center, then it raises and it tends to one when the distance raises. So if a set of samples are multiplied by this function, the samples near the center would tend the center however as well as they become distanced the influence of the function decreases. So actually this function works as an attractor to those samples near it, because the reference axis are in the center of the function; the influence of it will depend on the rest of the samples distribution. So, for a certain samples distribution the anisotropous Transformation (11), will affect to each sample individually depending on the distance to the mean of the set. If we have two different distributions in the input set, it could be possible to perform the whitening to each of the distributions without almost disturbing the other one, just as has been described for one. Lets suppose we have two distributions and **B** in $\mathbb{R}^2$. with covariances $\overline{\overline{\mathbf{C}}}_a$ and $\overline{\overline{\mathbf{C}}}_b$. The eigenvalues and eigenvectors are obtained for each distribution:

$$\overline{\overline{\mathbf{C}}}_a \lambda_a = \overline{\overline{\mathbf{C}}}_a \overline{\mathbf{v}}_a \tag{9}$$

$$\overline{\overline{\mathbf{C}}}_b \lambda_b = \overline{\overline{\mathbf{C}}}_b \overline{\mathbf{v}}_b \tag{10}$$

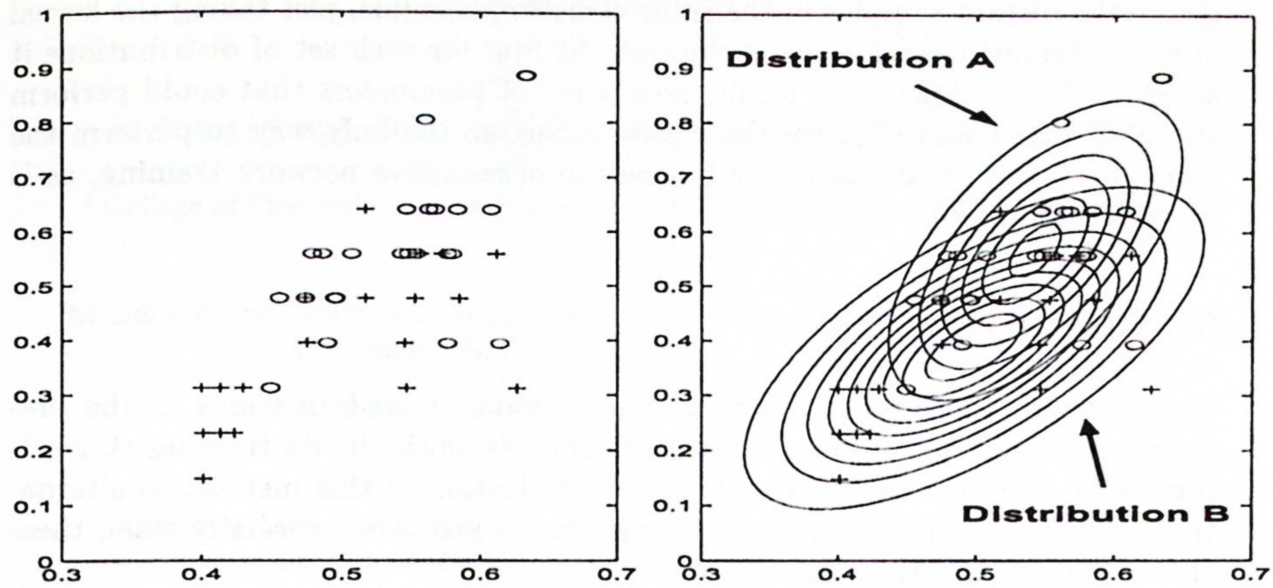

Fig. 3. Two different classes to be classified, the left figure shows the samples and the right figure shows the covariance curves.

Solving the systems above, it will be obtained $\overline{\overline{\mathbf{G_A}}}$, $\overline{\overline{\mathbf{G_B}}}$, $\overline{\overline{\mathbf{C_A}}}$ and $\overline{\overline{\mathbf{C_B}}}$

$$\overline{\overline{\mathbf{C_A}}} = \begin{pmatrix} [1-(\frac{\lambda_{a_2}}{\lambda_{a_1}})^{0.5} I(x,\sigma_A)] & 0 \\ 0 & 1 \end{pmatrix} \overline{\overline{\mathbf{C_B}}} = \begin{pmatrix} [1-(\frac{\lambda_{b_2}}{\lambda_{b_1}})^{0.5} I(x,\sigma_B)] & 0 \\ 0 & 1 \end{pmatrix} \quad (11)$$

,and then from 5

$$\overline{v_A} = [(\overline{u} - \overline{m}_a)\overline{\overline{\mathbf{R_A}}}\,\overline{\overline{\mathbf{C_A}\mathbf{R_A}}}^{-1} + \overline{m}_a] \quad (12)$$

$$\overline{v_B} = [(\overline{u} - \overline{m}_b)\overline{\overline{\mathbf{R_B}}}\,\overline{\overline{\mathbf{C_B}\mathbf{R_B}}}^{-1} + \overline{m}_b] \quad (13)$$

Applying both transformations separately to the input samples $\mathbf{U}$ there will be obtained the sets $\mathbf{V}_a$ and $\mathbf{V}_b$; the output set $\mathbf{V}$ will be the arithmetic media of them. In each transformation, all the samples are modified but $\mathbf{V}_a$ will affect mainly to those samples of U belonging to $\mathbf{A}$ and equally $\mathbf{V}_b$ will mainly affect to $\mathbf{B}$.

4 Improving the application fitting parameters

Applying this transformation once, usually will make the distributions to be properly whitened, the distant samples are not as influenced as the ones in the distribution's centers. Making this filtering iteratively, enhances its effect, until a stationary situation is reached. The final distributions will be determined by the width of the IF applied and the relation of each sample with the original distribution. The simplest way to make the distributions to be whitened, is to ignore the distant samples in the transformation calculus, just taking the kernel of the distribution and filtering the rest. At first for each set of distributions is not posible to know if it would exist a set of parameters that could perform an AWT that could improve the classification, so the only way to perform the selection of the parameters may be the use of recursive network training, as described in [4], [5].

5 Conclusions

In this paper the idea of applying Anisotropous Transformations in the preprocessing stage for systems such as Neural Networks in its training stage, presented. We focus in the capability of application of this method as alternative for preprocessing inputs in multiple classes problems, specially when these classes present different distributions.

References

1. Duda R. O., Hart P.E. and Stork D.G. *Pattern Classification.* John Wiley and Sons, Inc. (2001). ISBN 0-471-05669-3
2. Bishop C. M. *Neural Networks for Pattern Recognition.* Clarendon Press,Oxford (1996). ISBN 0-19-853849-9196
3. Tood K.M. and Wynn C.S. *Mathematical Methods and Algorithms.* Prentice Hall Inc., Upper Saddle River, New Jersey (2000). ISBN 0-201-36186-8
4. Buendía F.S., Barrón-Adame J.M, Vega-Corona A. and Andina D. Improving GRNNs in CAD systems. *Lecture Notes in Computer Science*, **3195** (2004) 160–16
5. Cao, L.J. and Tay, F.E.H. *Support vector machine with adaptive parameters financial time series forecasting. IEEE Transactions on Neural Networks*, **14** (6) (2003) 1506 - 1518