A new variant of Conformal Map Approach method for computing the preimage in Input Space

JORGE ARROYO-HERNANDEZ
Universidad Nacional de Costa Rica
Department of Applied Mathematics
Avenue 1, Heredia
Costa Rica
jarroy@una.cr

Jose Pablo Alvarado-Moya
Instituto Tecnológico de Costa Rica
Electronics Engineering School
Cartago 30101
Costa Rica
palvarado@itcr.ac.cr

Abstract: This paper presents a complementary procedure to compute preimage in Conformal Map Approach method for dimensionality reduction. We show a mathematical construction that makes it possible to calculate the β parameter in this method. Also, we presented the results of implementation of this variant using artificial data base generator.

Key–Words: Kernel of Principal Components Analysis, preimage problem, de-noising, dimensionality reduction.

1 Introduction

Today, the growing volume of data generated by information and communication systems demand new data manipulation techniques in order to extract non-trivial information that implicitly resides in the data to obtain behavioral patterns immersed in themselves.

Due to this fact, the search for redundancy elimination mechanisms has been a topic widely studied in recent years in order to improve performance rates in the calculation of processes such as detection and pattern recognition implicit in high-dimensional data.

The success of these methods is the ability to search through hidden rules in a limited number of parameters to characterize them.

Different solutions have been published based on linear and nonlinear statistical and mathematical models. Nonlinear methods are distinguished by is calculated in another domain, called Feature Space through kernel functions.

This paper presents a variant of Conformal Map Approach Method. This consists shows mathematical construction to compute one parameter of formula the preimage in the Input Space.

2 Dimensionality reduction method

2.1 Kernel - PCA

Kernel Principal Components Analysis (KPCA) [1] is an unsupervised learning technique whose objective is calculate from the original data space to Feature Space F of low dimensions to determine the variability of data.

The data are mapped into Feature Space F via a kernel function defined by $\Phi : \mathbb{R}^n \to F, x \mapsto \Phi(x)$. Assuming the data set are centered $\left(\sum_{k=1}^{M} \Phi(x) = 0\right)$ compute the covariance matrix $\overline{C} = \frac{1}{M}\sum_{k=1}^{M} \Phi(x_i)\Phi(x_j)^T$. After this, the Eigenvalues $\lambda \geq 0$ and Eigenvector $V \in F - \{0\}$ are computed by the equation $\lambda V = \overline{C}V$. The solutions V lie in the span $\Phi(x_1), \Phi(x_2), \cdots, \Phi(x_M)$. As a result, the equations obtained are $\lambda(\Phi(x_k) \cdot V) = \Phi(x_k) \cdot \overline{C}V$, $V = \sum_{1=1}^{M} \Phi(x_i)$ and

$$\lambda \sum_{i=1}^{M} \alpha_i(\Phi(x_k)) \cdot \Phi(x_i) =$$

$$\frac{1}{M}\sum_{i=1}^{M}\lambda_i(\Phi(x_k)\sum_{i=1}^{M}\Phi(x_j))(\Phi(x_j)\cdot\Phi(x_i))$$

$\forall k = 1, ..., M$

Defining an $M \times M$ matrix K by

$$K_{ij} := (\Phi(x_i)\cdot\Phi(x_j))$$

the previous expression is re-expressed as $M\lambda\alpha = K^2\alpha$ where α represent the column vector with entries $\alpha_1,\cdots,\alpha_M$. K is a symmetric matrix, the equation is simplified $M\lambda K\alpha = K\alpha$ and gets solutions α. K is positive semidefinite. Therefore, the K's Eigenvalues will be nonnegative and to give the solutions of the last equation. Let $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M$ the Eigenvalues and $\alpha^1, \alpha^2, ... , \alpha^M$ with λ_p being the first nonzero. We normalize $\alpha^1, \alpha^2, ... , \alpha^M$ by requiring that the corresponding vector in F be normalized. Finally, the projections of Eigevector V^K with $p \leq k \leq M$ are $V \cdot \Phi(x) = \sum_{i=1}^{M}\alpha_i^k(\Phi(x_i)\cdot\Phi(x))$ may be called its nonlinear principal components corresponding to Φ.

3 Preimage problem

In the KPCA method, the direct compute of the preimage of a vector ψ in Feature Space is not always possible. The problem of preimagen [7] consists in finding a method that it considers the best approximation for $x \in \mathbb{R}^n$, suppose $x^* \in \mathbb{R}^n$, satisfying:

$$\Phi(x^*) \approx \psi \quad \text{if} \quad \psi \in F$$

According to [2] some solutions at the preimage problem such as Gradient Descent Techniques, Fixed-Point Iteration Method [3] and Learning The Preimage Map [6] all of them suffers several problems due to numerical instability, problems associated with kernel functions and minima local.

Another solution to the preimage problem is based on Conformal Map Approach method [2]. By virtue of Representer Theorem, each of de n coordinate functions can be written as a lineal expansion of availables images defined by $\Psi = \sum_{i=1}^{n}\alpha_i\Phi(x)$ for $l = 1, 2, \cdots n$ with unknow weigths to be determinated a new matrix Φ. Each coordinate can be obtained by a projection onto these coordinate functions: $\Psi_{x_i} = [\langle\psi_1, \Phi(x_i)\rangle\langle\psi_2, \Phi(x_i)\rangle\cdots\langle\psi_k, \Phi(x_i)\rangle]$. If the inner product is preserved betwen Input Space and Feature Space, the equation of coordinate systems $\Psi_{x_i}^T\Psi_{x_j} = x_i^T x_j$ can be resolved by minimizing the fitness the error of all pairs using:

$$\min_{\psi_1,\ldots\psi_n} = \sum_{i,j=1}^{n}|\Psi_{x_i}^T\Psi_{x_j} - x_i^T x_j|_F^2 + \eta\sum_{l=1}^{n}||\Psi_l||^2$$

where the second term is a parameter of regularization. In matrix form, the last equation is expressed by:

$$\min_{\Theta} = \frac{1}{2}||X^TX - K\Theta^T\Theta K||_F^2 + \eta Tr(\Theta^T\Theta K)$$

The solution is $\Theta^T\Theta = K^{-1}(X^TX - \eta K^{-1})K^{-1}$. The position of some preimage is determinated with $\psi = \sum_{i=1}^{n}\beta_i\Phi(x_i)$. Its coordinates are associated to the system of coordinate function on the Feature Space. Using a least square the solution is

$$x^* = (XX^T)^{-1}X(X^TX - \eta K^{-1})\beta$$

where $\beta = (\beta_1\beta_2\cdots\beta_n)$.

3.1 New method for computing β

We propose a new way of calculating the parameter β for the solution of preimage problem via CMA. The Eigenvalues are computed on Feature Space through of:

$$\psi = \sum_{j=1}^{n}\beta_j\phi(x_j) \qquad (1)$$

such as implicit representation de ψ of linear combination de β_j with $\phi(x_j)$ on Feature Space. Assuming that the ψ is an image of Eigenvector:

$$\psi = \sum_{i=1}^{m}\langle\phi(x^*), \psi_i\rangle\psi_i \qquad (2)$$

If both expressions (1) and (2) are combined, obtained:

$$
\begin{aligned}
\psi &= \sum_{i=1}^{m}\left(\left\langle \Phi(x^*), \sum_{j=1}^{n}\alpha_k^{(i)}\Phi(x_l)\right\rangle \sum_{j=1}^{n}\alpha_k^{(i)}\Phi(x_k)\right) \\
&= \sum_{i=1}^{m}\left(\left(\sum_{j=1}^{n}\left\langle \Phi(x^*), \alpha_k^{(i)}\Phi(x_l)\right\rangle\right) \sum_{j=1}^{n}\alpha_k^{(i)}\Phi(x_k)\right) \\
&= \sum_{k=1}^{n}\left(\sum_{i=1}^{m}\alpha_k^{(i)}\left(\sum_{l=1}^{n}\alpha_l^{(i)}\left\langle \Phi(x^*), \Phi(x_l)\right\rangle\right)\right)\Phi(x_k) \\
&= \sum_{k=1}^{n}\beta_k\Phi(x_k)
\end{aligned}
$$

and therefore

$$
\beta_k = \sum_{i=1}^{m}\alpha_k^{(i)}\left(\sum_{l=1}^{n}\alpha_l^{(i)}\left\langle \Phi(x^*), \Phi(x_l)\right\rangle\right)
$$

$$
\beta_k = \sum_{i=1}^{m}\alpha_k^{(i)}\left(\sum_{l=1}^{n}\alpha_l^{(i)}k(x^*, x_l)\right)
$$

where α_l is l-th entries of the i-th Eigenvector α.

4 Experimental results and discussion

In this section, we ilustrated the variant proposed of CMA method. The training sets used were data base artificially generated.

For this experiment the Input Space is $\mathbb{R}^2$. We use a set of $n = 500$ data point (x, y) whose are located randomly around of origin on Cartesian Plane and were painted red. (Figure 1).

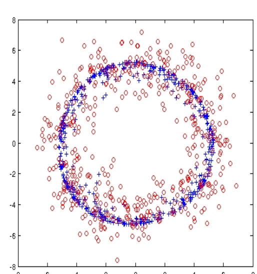

Figure 1: Conformal Map Approach with a new variant

The blue points are located were calculated using the variant of compute of preimage of CMA method show that method is success. Similarly, we prove thi variant using set of $n = 100$ data point (x, y) located on spiral and square form (Figure 2) show same results.

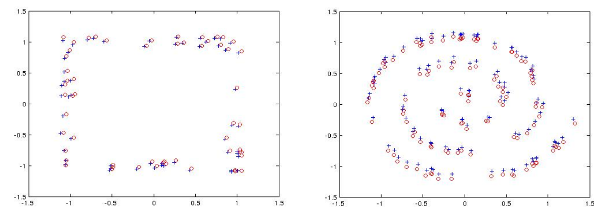

Figure 2: Conformal Map Approach with a new variant

The results shows that MCA method and the variant presents in this paper allows the reconstruction the preimage values in Input Space. The variant of the method give a complementary procedure of MCA for calculating the preimage in Input Space. Additionally, it is independent of the kernel function and is a closed approximation method, consecuently is more efficient.

Acknowledgements: The research was supported by Ministerio de Ciencia, Tecnología y Comunicaciones of the Costa Rica government (MICIT).

References:

[1] B. Scholkopf, A. Smola, and K. Muller, Nonlinear component analysis as a kernel eigenvalue problem. Technical Report, pp. 2–4, 1996,

[2] P. Honeine and C. Richard, Preimage problem in kernel-based machine learning, IEEE Signal Process, pp. 77–88, 2011.

[3] B. Scholkopf, S. Mika, C. Burges, P. Knirsch, K. Muller, G. Ratsch, and A. Smola, Input space versus feature space in kernel-based methods, IEEE TRANSACTIONS ON NEURAL NETWORKS, 1999

[4] T. F. Cox and M. A. A. Cox, Multidimensional Scaling, 2nd ed. (ser. Mono- graphs on Statistics and Applied Probability). London, Chapman and Hall, 2000.

[5] S. Mika, B. Scholkopf, A. Smola, K. Muller, M. Scholz, and G. Rtsch, Kernel PCA and denoising in feature spaces, in Proc. 1998 Conf. Advances in Neural Information Processing Systems II. Cambridge, MA: MIT Press, pp. 536-542, 1999.

[6] G. Bakir, J. Weston, and B. Scholkopf, Learning to find pre-images, in Neural Information Processing Systems 2003, S. Thrun, L. Saul, and B. Scholkopf, Eds. Cambridge, MA: MIT Press, 2004, vol. 16, pp. 449–456.

[7] J. Kwok and I. Tsang, The pre-image problem in kernel methods, Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, 2003

ISBN: 978-960-474-361-2