ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Adaptive learning for complex-valued data

Kerstin Bunte[1], Frank-Michael Schleif[1], and Michael Biehl[2]

1- University of Bielefeld - CITEC Center of Excellence,
D-33501 Bielefeld - Germany

2- University of Groningen - Johann Bernoulli Institute for Mathematics and Computer Science, P.O. Box 407, 9700 AK Groningen, The Netherlands

Abstract. In this paper we propose a variant of the Generalized Matrix Learning Vector Quantization (GMLVQ) for dissimilarity learning on complex-valued data. Complex features can be encountered in various data domains, e.g. Fourier transformed mass spectrometry or image analysis data. Current approaches deal with complex inputs by ignoring the imaginary parts or concatenating real and imaginary parts in one real valued vector. In this contribution we propose a prototype based classification method, which allows to deal with complex-valued data directly. The algorithm is tested on a benchmark data set and for leaf recognition using Zernike moments. We observe that the complex version converges much faster than the original GMLVQ evaluated on the real parts only. The complex version has fewer free parameters than using a concatenated vector and is thus computationally more efficient than original GMLVQ.

1 Introduction

Machine learning methods which are able to deal with complex-valued (cv) data attracted substantial interest recently [6, 7, 3, 4]. Complex-valued signals occur in many disciplines and are of fundamental importance. Many modern life science measurement systems generate cv-data, like, e.g. fMRI or NMR spectra. Im many applications the imaginary part of the data is of significant importance.

Up to date most cv data is processed by either ignoring the complex nature of the signal, by considering real and imaginary part independently or by normalizing the data to remove the imaginary part [6, 5, 3].

Data are often converted to a real-valued representation for subsequent analysis. This however may lead to substantial errors or limits performance significantly [8]. In this contribution we propose an extension of Generalized Matrix Learning Vector Quantization (GMLVQ) for complex-valued data, taking real and imaginary parts of the signal into account explicitly. The method is detailed and evaluated on cv-benchmark data set and a real world data set using shape descriptors.

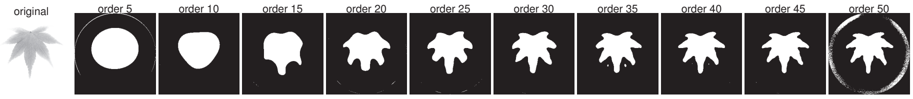

Fig. 1: Image reconstruction of Zernike-Moments on one example leave.

2 Methods

2.1 Complex-valued features

A widely used shape-descriptor[1] of a two-dimensional object is given by the Zernike-Moments [10]. Zernike-Moments are complex-valued moments based on radial Zernike-polynomials (V_{pq}):

$$Z_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{pq}(r,\theta)]^* f(r\cos\theta, r\sin\theta) r \; dr \; d\theta \quad \text{with} \tag{1}$$

$$V_{pq}(r,\theta) = R_{pq}(r) e^{jr\theta},\; R_{pq} = \sum_{s=0}^{(p-|q|)/2} (-1)^s \frac{(p-s)!}{s!(\frac{p+|q|}{2}-s)!(\frac{p-|q|}{2}-s)!} r^{p-2s} \quad ,$$

$p = 0, \ldots, \infty$, $|q| \leq p$ and $p-|q|$ even. Zernike moments constitute an orthogonal basis for the set of complex functions and can be adapted to be rotation, scale and translation invariant. Note that there are $\frac{1}{2}(n+1)(n+2)$ linear independent polynoms of degree $\leq n$, since $p - |q|$ is even. The Zernike-Moments of an image function f of order p and repetition q is the orthogonal projection to V_{pq}. For discrete data, using a coefficient representation (B_{pqk}) as given in [2], one obtains:

$$Z_{pq} = \frac{(p+1)}{\pi} \sum_{k=q}^{p} \sum_{j=0}^{l} \sum_{m=0}^{q} (-i)^m \binom{l}{j} \binom{q}{m} B_{pqk} m_{k-2j-q-m,2j+q-m} \;,\; l = \frac{k-q}{2}$$

and $m_{k-2j-q-m,2j+q-m}$ are the regular normalized moments [5]. The original image can be approximately reconstructed from a large set of moments as $f(x,y) = \lim_{N\to\infty} \sum_{p=0}^{N} \sum_q Z_{pq} V_{pq}(x,y)$where the second sum is over all $|q| \leq p$ with $p - |q|$ even. A sample of the obtained reconstructions using different parameters is shown in Fig. 1.

2.2 Complex-valued Learning Vector Quantization

We extend the concept of GMLVQ [9, 1] to be used on cv-data. GMLVQ is a prototype-based classification algorithm using the concept of adaptive similarity learning. The aim is to place the labeled prototypes $\boldsymbol{w}^j$ such that a high classification accuracy is achieved. We propose a variant designed for complex valued training data $\boldsymbol{x}^i \in \mathbb{C}^N$ and prototypes $\boldsymbol{w}^j \in \mathbb{C}^N$ accompanied with labels y^i and

[1] For an overview about the topic see [5].

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

$c(\boldsymbol{w}^{\boldsymbol{j}})$ in the following referred to as Complex GMLVQ (CGMLVQ). The cost function is defined by

$$E = \sum_{i=1}^{n} e^i = \sum_{i=1}^{n} \frac{d^J - d^K}{d^J + d^K} \ , \tag{2}$$

with d^J denoting the distance to the closest prototype with the same class label $c(\boldsymbol{w}^J) = y^i$ as $\boldsymbol{x}^i$ and d^K the distance of the closest wrong prototype $\boldsymbol{w}^K$. In the original GMLVQ formulation the distances are given by $d_o^j = (\boldsymbol{x}^i - \boldsymbol{w}^j)^\top \Omega^\top \Omega (\boldsymbol{x}^i - \boldsymbol{w}^j)$. We define a dissimilarity measure for complex values using a complex transformation matrix $\Omega \in \mathbb{C}^{M \times N}$

$$d^j = \sum |(\boldsymbol{x}^i - \boldsymbol{w}^j)\Omega^\top|^2 = (\boldsymbol{x}^i - \boldsymbol{w}^j)^{*\top} \Omega^{*\top} \Omega (\boldsymbol{x}^i - \boldsymbol{w}^j) \ . \tag{3}$$

where * denotes the complex conjugate. The learning is done using a stochastic gradient descent procedure presenting one example $\boldsymbol{x}^i$ at the time. We denote one sweep through the randomly shuffled training set as an epoch. Derivatives of the real-valued cost function and distances have to be taken with respect to the real and imaginary parts of the variables, respectively. Upon presentation of a single example $\boldsymbol{x}^i$ we obtain updates of the form $\boldsymbol{w} \to \boldsymbol{w} - \Delta \boldsymbol{w}$ and $\Omega \to \Omega - \Delta\Omega$

$$\Delta \boldsymbol{w} = \frac{\partial e^i}{\partial \Re(\boldsymbol{w})} + i \cdot \frac{\partial e^i}{\partial \Im(\boldsymbol{w})} \text{ and } \Delta\Omega = \frac{\partial e^i}{\partial \Re(\Omega)} + i \cdot \frac{\partial e^i}{\partial \Im(\Omega)} \ . \tag{4}$$

The derivatives are thus given by:

$$\gamma^J = \frac{2 \cdot d^K}{(d^J + d^K)^2} \text{ and } \gamma^K = \frac{-2 \cdot d^J}{(d^J + d^K)^2} \tag{5}$$

$$\Delta \boldsymbol{w}_n^L = \gamma^L \cdot \left(-2 \cdot \left[\sum_{m=1}^{M} [((\boldsymbol{x}^i - \boldsymbol{w}^L)\Omega^\top)^{*\top}]_m \cdot \Omega_{mn} \right]^* \right) \text{ with } L \in \{J, K\} \tag{6}$$

$$\begin{aligned} \Delta\Omega = &\gamma^J \cdot \left(2 \left(((\boldsymbol{x}^i - \boldsymbol{w}^J)\Omega^\top)^{*\top} \cdot (\boldsymbol{x}^i - \boldsymbol{w}^J) \right)^* \right) + \\ &\gamma^K \cdot \left(2 \left(((\boldsymbol{x}^i - \boldsymbol{w}^K)\Omega^\top)^{*\top} \cdot (\boldsymbol{x}^i - \boldsymbol{w}^K) \right)^* \right) \end{aligned} \tag{7}$$

After every iteration the matrix Ω is normalized auch that $\sqrt{\sum_{j=1}^{N} \Omega_{jj}^{*\top} \Omega_{jj}} = 1$.

3 Experiments

To evaluate CGMLVQ we use an artificial three dimensional benchmark and a real world data set for leaf recognition.

3.1 Artificial data

We construct an artificial data set in the following way: 1000 samples are drawn from a three-dimensional uniform distribution with values between -1 and 1

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

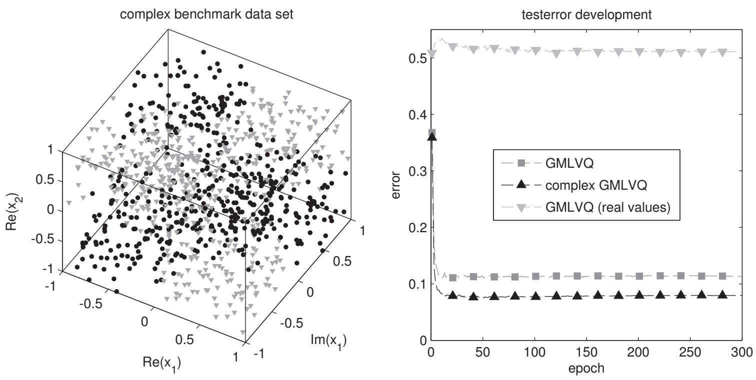

Fig. 2: Left panel: visualization of the three information containing dimensions of the benchmark data set. Right panel: error development on the test data.

$D = \mathcal{U}(-1, 1) \in \mathbb{R}^3$. Whenever the product of the three dimensions is positive the sample is assigned to class 1 and to class 2, otherwise. This results in a checkerboard-like volume as shown in Fig. 2. The first two dimensions build the real and imaginary part of the first complex dimension. The real and imaginary part of the second dimension equal to dimension three. This two-dimensional complex-valued data set is embedded into 10 dimensions as described in Eq. (8):

$$X = \{D_1 + i\cdot D_2,\ D_3 + i\cdot D_3,\ \{\mathcal{U}(-0.5, 0.5) + i\cdot\mathcal{U}(-0.5, 0.5)\} \in \mathbb{R}^8\} \in \mathbb{C}^{10} \quad . \tag{8}$$

For the evaluation we perform a ten-fold cross-validation using 10% of the data as a test set and $t = 300$ epochs. All algorithms are run with 4 prototypes per class initialized near the class centers with small random deviation. The matrix Ω is initialized as the identity matrix. The mean test error development per epoch for all methods is shown in Fig. 2 in the right panel. It can be seen that GMLVQ on the real values only, fails to learn the structure. This was expected, because the class structure is built by a correlation of the real and imaginary parts of the first two dimensions. Also the GMLVQ on the concatenated real and imaginary parts, which takes into account pairwise correlations of the real and imaginary parts, cannot achieve the quality of the CGMLVQ. In the complex version it is possible to take pairwise correlations of the complex dimensions into account. Inspection of GMLVQ and CGMLVQ reveals that GMLVQ has twice the number of free parameters compared to CGMLVQ.

3.2 Leaf image classification

The Flavia-Dataset from [11] is a real world data consisting of 1907 images of different leaves in 32 classes such as *ginkgo, peach, oleander* and others (see Fig. 3). In the original article [11] a training error of 98% and a test error of $\approx 90\%$ is reported. This was achieved using specific features and a probabilistic neural network model but the evaluation was done on a simple training (1800) / test (107) split rather a full cross-validation.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

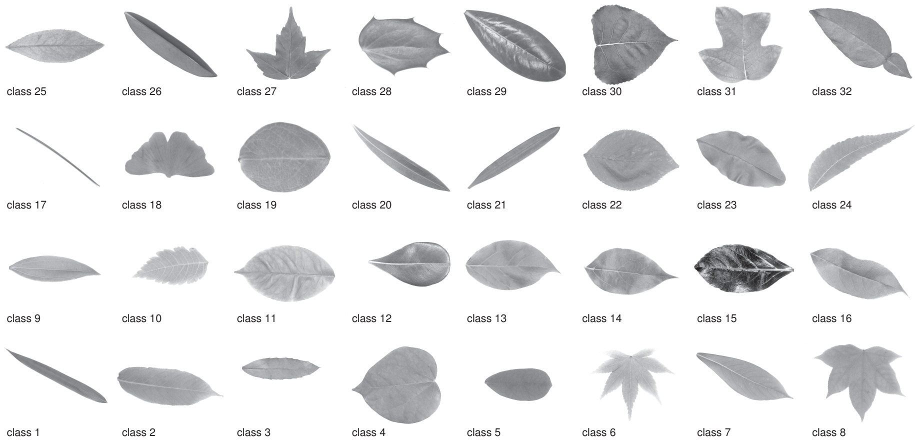

Fig. 3: Examples of the green channel of 32 leave classes used for the experiment.

It can be expected that the samples can be sufficiently represented using the shape information of the leaf. This motivates the use of Zernike-Moments (1) which are known to provide optimal, complex-valued shape descriptors. We generated Zernike-Moments from the green-channel of the images, up to order 20 resulting in 121 complex-valued features per sample. The obtained feature vectors have been normalized to zero mean and unit variance. We compare results of CGMLVQ with GMLVQ on the real-valued features only and by concatenating real and imaginary part.

We used the same experimental setup as for the artificial data but now with one prototype per class. The evolution of the training and test error is shown in Fig. 4. In this case the real values contain already a lot of useful information for the classification task. However, the complex version converges much faster taking into account the nature of the data explicitly. GMLVQ on the concatenated vector of real and imaginary parts can achieve a better classification in this example, but with the additional costs of twice as many free parameters than CMGLVQ.

4 Conclusions

In this contribution we propose a variant of the Generalized Matrix Learning Vector Quantization (GMLVQ) for complex valued data called Complex GM-LVQ (CGMLQ). Many applications which naturally deal with complex values may benefit from methods which take into account the nature of the data rather than ignoring or modifying the structure to fit into conventional techniques. Concatenating the real and imaginary parts facilitates the use of all the information. However, CGMLVQ reduces the number of free parameters. If the data is expected to have just pairwise correlations between real and imaginary parts the concatenation is recommended, but if correlations between complex dimensions are expected, CGMLVQ bears the capability to outperform the naive

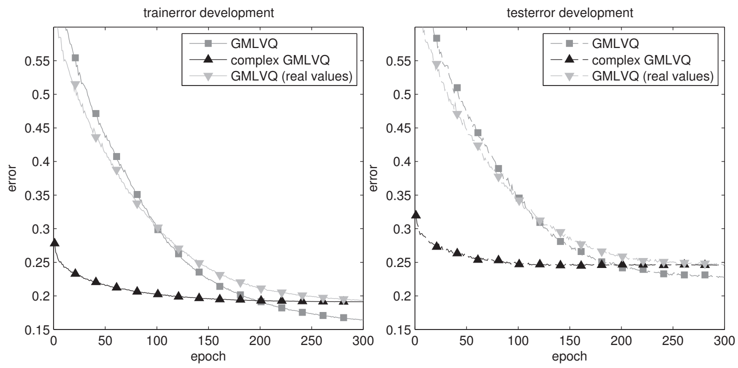

Fig. 4: Comparison of the training and test error development during training for GMLVQ using only the real values and CGMLVQ.

approach. Further work should address the use of alternative dissimilarities (for example localized matrices) and the exploration of more complex data domains.

Acknowledgment

This work was supported by the "German Science Foundation (DFG)" under grant number HA-2719/4-1. Further, financial support from the Cluster of Excellence 277 Cognitive Interaction Technology funded in the framework of the German Excellence Initiative is gratefully acknowledged.

References

[1] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl. Limited rank matrix learning – discriminative dimension reduction and visualization. *Neural Networks*, 26(4):159–173, February 2012.

[2] R. Courant and D. Hilbert. *Methods of Mathematics in Physics, Vol I.* Interscience, New York, 1953.

[3] A. Hirose. *Complex-Valued Neural Networks*, volume 32 of *Studies in Computational Intelligence*. Springer, 2006.

[4] A. Hirose and S. Yoshida. Comparison of complex- and real-valued feedforward neural networks in their generalization ability. In B.-L. Lu, L. Zhang, and J. T. Kwok, editors, *International Conference on Neural Information Processing ICONIP (1)*, volume 7062 of *Lecture Notes in Computer Science*, pages 526–531. Springer, 2011.

[5] A. K. Jain, R. P. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:4–37, 2000.

[6] L. Li and B. A. Prakash. Time series clustering: Complex is simpler! In *International Conference on Machine Learning (ICML)*. Omnipress, 2011. In press.

[7] X.-L. Li and T. Adali. Complex independent component analysis by entropy bound minimization. *IEEE Trans. on Circuits and Systems I*, 57(7):1417–1430, 2010.

[8] F.-M. Schleif, T. Riemer, U. Börner, and L. S.-H. M. Cross. Genetic algorithm for shift-uncertainty correction in 1-D NMR based metabolite identifications and quantifications. *Bioinformatics*, 27(4):524–533, 2011.

[9] P. Schneider, M. Biehl, and B. Hammer. Adaptive relevance matrices in learning vector quantization. *Neural Computation*, 21(12):3532–3561, 2009.

[10] M. R. Teague. Image analysis via the general theory of moments. *Journal of Optical Society of America*, 70:920–930, 1980.

[11] S. G. Wu, F. S. Bao, E. Y. Xu, Y. Wang, Y.-F. Chang, and Q.-L. Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. *Proceedings of CoRR*, abs/0707.4289, 2007.