ijar
ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

A New Generalized Learning Vector Quantization Classifier Algorithm with Sliding-mode Optimized Training

Turgay Temel (PhD)
Department of Mechatronics Engineering, Bursa Technical University, Bursa, TURKEY
turgay.temel@btu.edu.tr / ttemel70@gmail.com

Manuscript Info

Manuscript History:

Received: 14 May 2015
Final Accepted: 26 June 2015
Published Online: July 2015

Key words:

Learning vector quantization; neural networks, machine learning; supervised learning; sliding-mode

***Corresponding Author**

Turgay Temel (PhD)

Abstract

A new Learning Vector Quantization classifier is proposed. The algorithm relies on a new training scheme for labeled sample vectors in feature space. Since weight or prototype vectors are conditioned to a well-known sliding-mode approach with use of a cost function to be minimized in terms of weight updates, new algorithm is called Optimized Generalized Learning Vector Quantization (OGLVQ). Consequently, weights are then associated to the proximity measure employed by conventional Generalized Learning Vector Quantization. New algorithm and some well-known predecessors are designed and tested for comparison with synthetic and publicly available datasets. From the experimental results, it is observed that the new classifier achieves faster training and is more successful and robust in generalizing labeled test samples picked from datasets studied than the counterparts it is compared to.

Introduction

An improved classifier algorithm should be able to successfully learn differing realizations of ensemble statistics given a dataset and exhibit robustness against noise or other undetermined conditions involved for stable and low-complexity learning/training. It is the general approach for learning process to reward against punishment for correct and incorrect classifications, respectively, for attaining a minimum misclassification even if data is not much available or prone to false alarm, [1]-[2]. Learning Vector Quantization (LVQ) and its variant algorithms have been known almost the most successful classifiers to achieve above objectives, [3]-[4]. They commonly employ Hebbian-based learning in training and nearest-neighborhood to relate the winner prototype or weight vector attributed to by a respective class label in feature space. However, learning with LVQ algorithms has a potential instability and bias problems in representing proximity measure between sample and weight vectors. Especially when a loser class has been incorrectly assigned during training, a non-adaptive proximity measure may lead to degraded generalization performance. Besides these problems, initialization of weight vectors has been a tackling issue for success of LVQ algorithm considered.

Hence, it is desirable to develop an adaptive training scheme that attains stable and consistent learning when dealing with varying statistical realizations of a given dataset. Most remedial attempts toward this objective have focused on devising an adaptive metric for representing dissimilarities in training: Generalized Learning Vector Quantization (GLVQ) algorithms implement training based on optimization of a decision-making cost function with use of gradient descent, [5], which in small dimensions allows classification performance almost independent of data representation. However, in higher domensions, weight dynamics will be prone to unstability and proximity metric between weight and sample vectors, which will loose its descriptive meaning, [6]-[7]. As a solution to above issues, may be to employ a kernel-based scaling for similarity of individual samples to prototype vectors, which is termed

Generalized Relevance Learning Vector Quantization (GRLVQ), [7]-[8], with a similar training phase description as GLVQ. Despite the fact that major (G/GR)LVQ schemes maintain descriptive nature in representing similarity between weight/prototype and data in learning. There has been no study up to now, which reveals appropriate stability constraints/conditions for expressing dynamics of weight vectors and their update with convergence and robustness to parametric changes given an arbitrary initial assignment.

This study proposes a new GLVQ classifier based on a new training algorithm associated to a sliding-mode converging rule for weight dynamics. Since new rule guarantees optimization of weight-update related cost function in overall, new classifier is called Optimally Generalized Learning Vector Quantization (OGVLQ). Proposed rule, then, is combined with the proximity measure exploited by previous (G/GR)LVQ schemes. New and some major (G)LVQ classifiers are designed and compared for synthetic and publicly available datasets in terms of statistical performance measures. Simulation results reveal that the new LVQ is faster in training and assigns correct classes more successfully with improved robustness against variation in learning parameter and randomness in initialization of weight vectors than the predecessors it is compared.

An Overview For (G/GR)LVQ Classification

Common operation of an LVQ classifier, which is depicted in Fig. 1, is to associate class labels which are known a priori to input samples to be classified. For a sample L-dimensional input column vector $\mathbf{x} = [x_1 \dots x_L]^T$ and given K classes with labels $C_{i=1,\dots,K}$, where the weight vector of the j-th class label C_j is $\mathbf{w}^j = \left[{w^j}^1 \dots {w^j}^L\right]^T$, the respective output class label which corresponds to the closest weight vector as the winner is assigned.

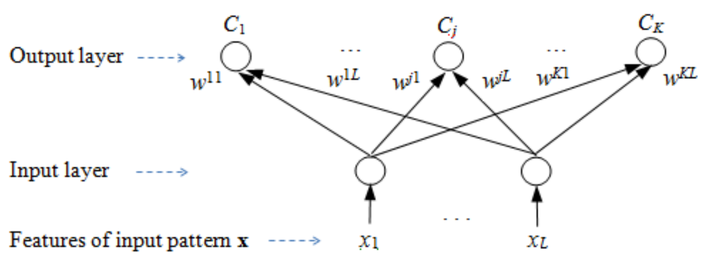


Fig. 1. Generalized operation of LVQ classifiers

In training phase of original LVQ, [9], with N_{train} samples per epoch, for the k-th input sample vector $\mathbf{x}_k = \mathbf{x}(k),\ k = 1 \dots N_{train}$, only the weight vector of the winning class label, $\mathbf{w}_k^* = \mathbf{w}^*(k)$, is adjusted by

$$\mathbf{w}_{k+1}^* = \mathbf{w}_k^* \pm \xi(\mathbf{x}_k - \mathbf{w}_k^*) \tag{1}$$

until a pre-specified convergence condition is met, e.g. $\mathbf{w}^j(k) = \mathbf{w}^j(k-1)$ for $j = 1, \dots, K$. The sign '±' is taken '+' if $\mathbf{x}_k$ has been correctly classified otherwise '-'. The winning output vector $\mathbf{w}^*$ is determined as the closest weight vector, i.e. $\mathbf{w}^* = \arg\min_{\mathbf{w}^j} d(\mathbf{x}_k, \mathbf{w}^j)$ where $d(\mathbf{x}_k, \mathbf{w}^j)$ is the squared Euclidean distance between $\mathbf{x}_k$ and $\mathbf{w}_j$. Other variants of original LVQ, e.g. LVQ2.1 and LVQ3, adjust weight vectors of some other classes, e.g. closest loser prototype of having the same class label as the sample, [10].

On the other hand, training of (G/GR)LVQ classifiers involves optimizing a cost function which relates correctly classified samples to particular class weight vectors while maximizing correct decision as a supervisory operation. Such a cost function is given by

$$E = \frac{1}{2}\sum\nolimits_{\forall k} f(\mu_k) \tag{2}$$

where the classifier function is $f(u) = \frac{1}{1+e^{-u}}$ for the proximity measure $\mu_k = \mu(\mathbf{x}_k) = \frac{d_k^+ - d_k^-}{d_k^+ + d_k^-}$. Dissimilarity measures $d_k^+ = d(\mathbf{x}_k, \mathbf{w}_k^* = \mathbf{w}^+) = \|\mathbf{x}_k - \mathbf{w}^+\|^2$ and $d_k^- = d(\mathbf{x}_k, \mathbf{w}_k^* = \mathbf{w}^-) = \|\mathbf{x}_k - \mathbf{w}^-\|^2$ are the squared distances of $\mathbf{x}_k$ to the closest prototype $\mathbf{w}^+$ with the same class label as $\mathbf{x}_k$ and the best matching prototype $\mathbf{w}^-$ with a class label different from that of $\mathbf{x}_k$, respectively. The weight update is then implemented as

$$\mathbf{w}_{k+1}^* \leftarrow \mathbf{w}_k^* \pm \underbrace{\xi(\mu_k)(\mathbf{x}_k - \mathbf{w}_k^*)}_{\Delta \mathbf{w}_k^*} \tag{3}$$

It is seen that (G/GR)LVQ algorithms adopt varying $\xi(\mu_k) = \frac{\partial f}{\partial \mu_k} \frac{d_k^{\pm}}{(d_k^+ + d_k^-)^2}$ to project spatial localization of weights with respect to sample $\mathbf{x}_k$ for improved accuracy, [11].

New GLVQ Classifier Description

If initial assignment for weights is made with improper choice and/or training samples have not been pruned while classes are overlapped, weights and their respective updates may lead to unstable non-convergent operation, [12]. Therefore, it is mandatory to devise an algorithm or rule toward a stable training, which involves a weight update dynamics against these issues. For such an objective, we consider minimizing the following cost function

$$J = \left(\frac{1}{2}\right) \sum_k \|\Delta \mathbf{w}_k^*\|^2 \tag{4}$$

The constraint in (4) is met with decreased gradient of J by maintaining $\|\Delta \mathbf{w}_k^*\|^2 < \|\Delta \mathbf{w}_{k-1}^*\|^2$. We consider $\|\Delta \mathbf{w}_k^*\|^2 = \|\Delta \mathbf{w}_{k-1}^*\|^2 \pm \alpha_k (\|\mathbf{w}_k^*\|^2 - \|\mathbf{w}_{k-1}^*\|^2)$ with a decision-dependent parameter $|\alpha_k| < 1$ where $\Delta \mathbf{w}_{k-1}^* = \mathbf{w}_k^* - \mathbf{w}_{k-1}^*$ and $\Delta \mathbf{w}_k^* = \mathbf{w}_{k+1}^* - \mathbf{w}_k^*$. The sign '±' is taken '-' if $\mathbf{w}_k^*$ and $\mathbf{w}_{k-1}^*$ both refer to the same class label otherwise '+'. Thus, it is possible to relate $\mathbf{w}_k^*$ ($\mathbf{w}_{k-1}^*$) and $\Delta \mathbf{w}_k^*$ ($\Delta \mathbf{w}_{k-1}^*$) for stable dynamics. We consider the sliding-mode convergence rule

$$(\Delta \mathbf{w}_k^*)^T \mathbf{w}_k^* \leq -\eta \|\mathbf{w}_k^*\|^2 \tag{5}$$

similar to the approach in [13] where $\eta > 0$. By using $\mathbf{w}_{k-1}^* = \mathbf{w}_k^* - \Delta \mathbf{w}_{k-1}^*$ (5) can be rewritten as

$$(\Delta \mathbf{w}_k^*)^T \mathbf{w}_k^* \lesseqgtr \mp \eta \left\{ \frac{\|\Delta \mathbf{w}_k^*\|^2 - \|\Delta \mathbf{w}_{k-1}^*\|^2}{\alpha_k} \right\} \pm (\Delta \mathbf{w}_{k-1}^*)^T \mathbf{w}_{k-1}^* \tag{6}$$

where the operator sequence '$\leq, -, +$' ('>,+,-) is considered if $\mathbf{w}_k^*$and $\mathbf{w}_{k-1}^*$refer to the same (different) class label(s). Equating the right-hand side of (6) to 0 as a boundary for convergence and combining the resulting expression with that of (5) corresponding to (k-1)-th term will yield

$$\|\Delta \mathbf{w}_k^*\|^2 = \|\Delta \mathbf{w}_{k-1}^*\|^2 (1 \mp \alpha_k) \tag{7}$$

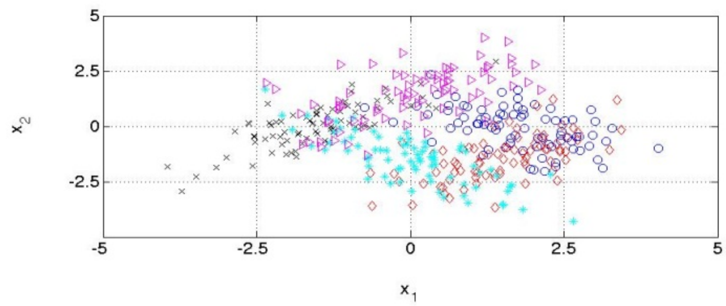
For convenience, it is appropriate to consider $\alpha_k = \mu_k$ for relating new learning algorithm with spatial saliency in decision making exploited by (G/GR)LVQ classifiers. Then, use of gradient for (7) leads to

$$(\Delta \mathbf{w}_k^*)^T = (\Delta \mathbf{w}_{k-1}^*)^T (1 \mp \alpha_k \mathbf{A}_k) \tag{8}$$

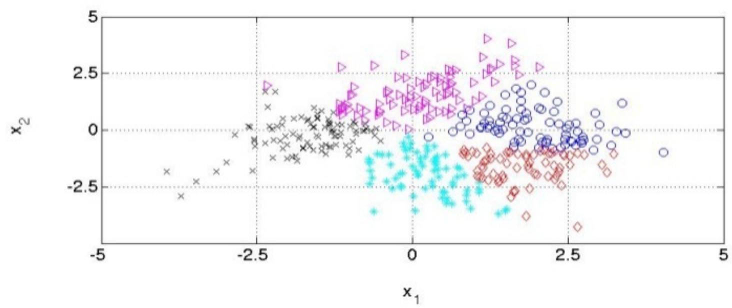
where $\mathbf{A}_k = \frac{\Delta \mathbf{w}_{k-1}^*}{(d_k^+ + d_k^-)^2} \left[d_k^+ \frac{\partial d_k^+}{\partial \mathbf{w}_k^+} - d_k^- \frac{\partial d_k^-}{\partial \mathbf{w}_k^-} \right]$ which is an L-by-L matrix and $\frac{\partial d_k^{\pm}}{\partial \mathbf{w}_k^{\pm}} = 2(\mathbf{x}_k - \mathbf{w}^{\pm})^T$ with sign previously given for (G/GR)LVQ. The sign '-' (+) in (8) refers to $\mathbf{w}_k^*$ and $\mathbf{w}_{k-1}^*$be same (different) class label(s).

Experiments

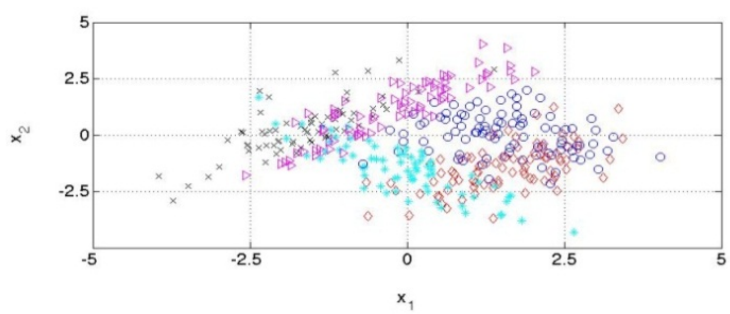
New classifier, and its previously cited original LVQ, LVQ2.1, and GLVQ counterparts were designed and compared on statistical basis in training complexity, and generalisation capability with two sets of 100 experiments for $\xi = 0.05$ and $\xi = 0.1$, respectively. At each experiment of the first set, 100 random vectors $\mathbf{x}=[x_1\ x_2]$ were populated from each of 2D (bivariate) normal densities: C_1: N([0 1.5], [1 0.5;0.5 1]), C_2: N([1.5 0], [1 -0.5;-0.5 1]), C_3: N(-1.5 0], [1 0.75;0.75 1]), C_4: N([1.5 -1.5], [1 0.5;0.5 1]), and C_5: N([0 -1.5], [1 -0.75;-0.75 1]) where the first vector in parentheses is the mean vector while the second vector/matrix is the covariance matrix with rows separated by semicolon. At each experiment, 20 (80) samples were set aside per class in forming a training (testing) set as union, i.e. N_{train}= 100 (N_{test}= 400). As an illustrative example, scatters for a testing sample dataset and respective classification results for original and new LVQ classifiers, respectively, for $\xi = 0.05$ are shown in Fig. 2.



(a)



(b)



(c)

Fig. 2. Example scatter of (a) a testing dataset, and classification results with (b) original LVQ, and (c) proposed algorithms for $\xi = 0.05$.

In the second set of 100 experiments, Character Trajectories Dataset (http://archive.ics.uci.edu/ml/machine-learning-databases/character-trajectories/) was used. The dataset consists of 3-dimensional 2858 labelled samples of pen tip segment trajectories for the 20 single pen-down characters, e.g. ‘a’, ‘e’, ‘w’. The feature vectors are composed of respective coordinates x, y, and pen tip force. At each experiment, 15 (60) training (testing) random samples from each of randomly selected 5 characters were drawn in forming a training (testing) dataset as a union, i.e. $N_{train} = 75$ ($N_{test} = 300$). With these datasets, relevant statistical simulation results for the algorithms examined are summarized in Table I. In table, '-' is used to separate the respective classification results for either dataset where the first group refers to those of combined normal densities while the latter quantities correspond to those for Character Trajectory dataset.

Method	Classification success, %, Average / Std. deviation		# of training epochs, Average / Std. deviation	
	$\xi = 0.05$	$\xi = 0.1$	$\xi = 0.05$	$\xi = 0.1$
OGLVQ	84.5 / 2.7 - 80.3 / 3.2	81.1 / 3.0 - 83.5 / 3.0	36.3 / 3.2 - 31.1 / 2.9	35.2 / 3.4 - 36.9 / 3.5
LVQ	46.1 / 5.1 - 49.2 / 4.9	51.5 / 5.0 - 55.7 / 4.7	59.2 / 6.2 - 51.6 / 5.1	51.1 / 5.6 - 46.5 / 4.8
LVQ2.1	63.5 / 4.5 - 68.3 / 4.6	67.4 / 4.4 - 64.3 / 4.8	50.5 / 5.7 - 44.6 / 5.1	47.5 / 5.3 - 42.4 / 5.0
GLVQ	77.8 / 3.9 - 80.4 / 3.6	71.8 / 4.2 - 77.3 / 4.0	40.6 / 5.2 - 48.8 / 4.9	43.7 / 5.2 - 45.9 / 4.6

Table I. Some statistical performance measures of new (OGLVQ) and some other LVQ algorithms for 2D normal densities and character trajectory dataset.

From above table it is seen that the proposed GLVQ classifier (OGLVQ) has improved and more consistent statistical training and classification performance despite variation in learning parameter with arbitrarily initialized weight vectors than its predecessors studied for both datasets. Original LVQ algorithm has poor classification for the datasets since it yields Voronoi-like regions in feature space. On the other hand, GLVQ counterpart that is a major competitor as expected proves relatively successful classification performance. However, new GLVQ still outperforms it in both statistical terms of training and generalization. Moreover, classification success of the new algorithm exhibits robustness against learning rate parameter variations while other counterparts yield considerable dependency on the parameter.

Conclusions

A new Learning Vector Quantization classifier called Optimized Generalized Learning Vector Quantization (OGLVQ) is presented, which relies on a new training scheme. Dynamics of weight/prototype vectors is related to sliding-mode approach for assuring a respective cost function to be bounded and minimized in terms of weight updates with an adaptive manner. Resulting weights are then associated to the proximity metric employed by conventional Generalized Learning Vector Quantization for attaining optimum correct decision-making while maintaining a convergent training. New algorithm and some well-known predecessors are tested for comparison with synthetic and publicly available datasets. From the experimental results, it is seen that the new classifier proves faster training and is more successful and robust in generalizing unlabeled test samples picked from datasets studied than the counterparts it is compared.

References

[1] A. Abackhaus, and U. Seiffert, "Classification in high-dimensional spectral data: Accuracy vs. Spectraldata: Accuracy vs. interpretability vs. model size," Neurocomputing, vol. 131, no. 5, pp. 15-22, 2014. doi:10.1016/j.neucom.2013.09.048

[2] T. Temel, "A new digital cochlea model neuro-spike representation of auditory signals and its application to classification of bat-like biosonar echoes," Neural Network World, vol. 20, no.2, pp. 223-239, 2010.

[3] D. Nova, and P. Estévez, "A review of learning vector quantization classifiers," Neural Computing and Applications, vol. 25, no. 3-4, pp. 511-524, 2013. doi: 10.1007/s00521-013-1535-3

[4] G. R. Lloyd, R. G. Brereton, R. Faria, and J. C. Duncan, “Learning Vector Quantization for MulticlassClassification: Application to Characterization of Plastics,” J. Chem. Inf. Model. vol. 47, no. 4, pp. 1553-1563, 2007. doi:10.1021/ci700019

[5] A. Sato, and K. Yamada, “Generalized learning vector quantization,” Advances in Neural Information Processing Systems, pp. 423-429, 1996.

[6] B. Hammer, M. Strickert, and T. Villmann, “On the generalization ability of GRLVQ networks,” Neural Processing Letters, vol. 21, no. 2, pp. 109-120, 2005.

[7] M. Kästner, B. Hammer, M. Biehl, and T. Villmann, “Functional relevance learning in generalized learning vector quantization,” Neurocomputing, vol. 90, pp. 85-95, 2012.

[8] A. Cataron, and R. Andonie, “Energy Generalized LVQ with Relevance Factors,” Proc. IEEE Int. Joint Conf. Neural Networks, IJCNN, pp. 1421-1426, 2004. doi:10.1109/IJCNN.2004.1380159

[9] T. Kohonen, “Self-Organized Formation of Topologically Correct Feature Maps,” Biological Cybernetics, vol. 43, no. 1, pp. 59-69, 1982. doi:10.1007/bf00337288

[10] T. Temel, and B. Karlik, “An improved odor recognition system using learning vector quantization with a new discriminant analysis,” Neural Network World, vol. 17, no. 4, pp. 287-294, 2007.

[11] M. Kaden, M. Lange, D. Nebel, M. Riedel, T. Geweniger, and T. Villmann, “Aspects in ClassificationLearning-Review of Recent, Developments in Learning Vector Quantization,” Foundations of Computing and Decision Sciences, vol. 39, no. 1, pp. 79-105, 2014. doi:10.2478/fcds-2014-0006

[12] A. Boubezoul, S. Paris, and M. Ouladsine, “Application of the cross entropy method to the GLVQ algorithm,” Pattern Recognition, vol. 41, pp. 3173-3178, 2008.

[13] S. Janardhanan, and B. Bandyopadhyay, “On Discretization of Continuous-Time Terminal Sliding Mode,” IEEE Trans. Automatic Control, vol. 51, no. 9, pp. 1532-1536, 2006. doi: 10.1109/TAC.2006.880805