ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Integrating Class Relation Knowledge in Probabilistic Learning Vector Quantization

M. Kaden[1], R. Schubert[1], T. Geweniger[2], W. Hermann[3], and T. Villmann[1,4] *

1 - Saxon Institute for Computational Intelligence and Machine Learning (SICIM), Mittweida University of Applied Sciences, Saxony, Germany
[kaden1|villmann]@hs-mittweida.de

2 - West-Saxony University of Applied Sciences Zwickau, Germany

3 - SRO AG Spital Langenthal, Switzerland

4 - Technical University Freiberg, Saxony, Germany

Abstract. An interpretable approach to classification learning using cross-entropy loss is the Probabilistic Learning Vector Quantizer (PLVQ) as a robust prototype-based classifier. We propose a variant of the PLVQ, that allows the integration of domain knowledge. This strategy is becoming increasingly popular as a means of developing intelligent models that can enhance performance and gain acceptance from domain experts. In this paper, we put forth the idea of incorporating externally known class relations as supplementary information. We present theoretical aspects of the model and demonstrate its capabilities through numerical experiments.

1 Introduction

Learning of classification schemes is one of the most required tasks in machine learning. Frequently, classification learning is challenging due to problems like complex classification scenarios, high data dimensionality, and overlapping classes. One possibility to deal with the latter problem is to apply probabilistic classifiers such as Deep Multilayer Perceptron Models (DMLP) with cross-entropy loss [6]. Another strategy to deal with these difficult situations is to apply additional mathematical constraints for regularization. However, a more promising strategy would be to use additional knowledge about the data for regularization instead, if available [20]. Furthermore, an appropriate integration of external knowledge can contribute to a better explainability of the learned classifier model [4]. However, this aspect is limited, if the model in use does not constitute an interpretable approach such as DMPLs. Fortunately, interpretable classifiers for probabilistic decision making are known based on the learning vector quatization paradigm [15, 16]. In particular, Probabilistic Learning Vector Quantization (PLVQ) has been introduced to enable probabilistic class assignments for both training data as well as model decision using the Kullback-Leibler Divergence (KDL) for class assignment comparison [17]. Recently, a knowledge-informed learning vector quantizer has been established in the context of classification of gene expression data using external correlation information on

*M.K. is supported by the IAI-XPRESS and the DAIMLER project which both are parts of the interdisciplinary project Artificial Intelligence Meets Space (AIMS, No. 50WK2270E) funded by the German Space Agency (DLR).

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

gene expression behaviour [19]. The common path to knowledge-based learning is the availability of data-specific knowledge.[20, 4, 5].

In this paper, another type of external knowledge is used to improve classification learning: We assume that additional information about structural dependencies or relations between classes is available. We will show how this knowledge can be integrated in PLVQ using a respective Class-Relation-Knowledge-Graph (CRKG), which requires an appropriate redefinition of the KDL.

The paper is structured as follows: First, PLVQ is briefly reconsidered. Thereafter, we describe, how the class-knowledge can be integrated into PLVQ and how KDL has to be adjusted accordingly. Exemplary, numerical experiments validate the approach followed by concluding remarks.

2 Cross-Entropy Learning in PLVQ

Learning Vector Quantization (LVQ), as introduced by T. KOHONEN [12], assumes a set $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_N\}$ of prototypes $\mathbf{w}_k \in \mathbb{R}^n$ to represent and to classify data $\mathbf{x} \in \mathbb{R}^n$. For this purpose, each prototype is assigned to be responsible for a certain class by the class label $c(\mathbf{w}_k) \in \mathscr{C} = \{1, \ldots, C\}$. Each class is represented by at least one prototype. For the probabilistic variant (PLVQ, [17]) a training data set of observed pairs $(\mathbf{X}, \mathbf{T}) = \{\mathbf{x}_i, \mathbf{t}_i\}_{i=1}^{N_D}$ is supposed where $\mathbf{t}_i \in [0,1]^C$ provides the probabilistic *target* class information of the sample $\mathbf{x}_i$ with class components $t_{ic} \in [0,1]$ and $\sum_c t_{ic} = 1$. For unique mutually exclusive classification training data $t_{ic} \in \{0,1\}$ is required.

PLVQ delivers after training a probabilistic class assignment vector $\mathbf{p}_W(\mathbf{x}) = (p_W(1|\mathbf{x}), \ldots, p_W(C|\mathbf{x}))$ with

$$p_W(c|\mathbf{x}) = \frac{P_W(\mathbf{x}, c)}{P_W(\mathbf{x})} \tag{1}$$

as the predicted class probability of the model such that $\sum_{c=1}^{C} p_W(c|\mathbf{x}) = 1$ is valid. Thereby, PLVQ considers

$$P_W(\mathbf{x}) = \sum_{j=1}^{N} p(\mathbf{x}|\mathbf{w}_j)\, p(\mathbf{w}_j) \tag{2}$$

as the probability density for $\mathbf{x}$ generated by the model with the prototypes $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$ as model parameters and $p(\mathbf{x}|\mathbf{w}_j)$ is the probability that $\mathbf{x}$ is generated by the jth model component determined by the prototype vector $\mathbf{w}_j$. The probabilities $p(\mathbf{w}_j)$ are the priors for the model components. Further, the model based estimator of the joint probability for $\mathbf{x}$ and an arbitrarily given but fixed class $c \in \mathscr{C}$ is

$$P_W(\mathbf{x}, c) = \sum_{j: c(\mathbf{w}_j) = c} p(\mathbf{x}|\mathbf{w}_j)\, p(\mathbf{w}_j) \tag{3}$$

as already suggested in [16]. The local loss to be optimized by PLVQ is the

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Kullback-Leibler Divergence

$$\begin{aligned} D_{KL}\left(\mathbf{t}_i||\mathbf{p}_W\left(\mathbf{x}_i\right)\right) &= \sum_c t_{ic} \cdot \log\left(\frac{t_{ic}}{p_W\left(c|\mathbf{x}_i\right)}\right) \\ &= H_S\left(\mathbf{t}_i\right) - Cr\left(\mathbf{t}_i, \mathbf{p}_W\left(\mathbf{x}_i\right)\right) \end{aligned} \tag{4}$$

where $H_S\left(\mathbf{t}_i\right)$ is the prototype-independent Shannon entropy of the target and

$$Cr\left(\mathbf{t}_i, \mathbf{p}_W\left(\mathbf{x}_i\right)\right) = \sum_c t_{ic} \cdot \log\left(p_W\left(c|\mathbf{x}\right)\right) \tag{5}$$

is the prototype dependent cross-entropy term [13, p. 221ff]. Yet, other divergences are possible [17]. Learning in PLVQ takes place by stochastic gradient descent learning (SGDL) with respect to the prototypes W determining the model class assignment $\mathbf{p}_W\left(\mathbf{x}_i\right)$. For this purpose, the probabilities $p\left(\mathbf{x}|\mathbf{w}_j\right)$ have to be defined, frequently taken as Gaussians $p_{\mathbf{\Omega}}\left(\mathbf{x}|\mathbf{w}_j\right) = \exp\left(-\left\|\mathbf{\Omega}\left(\mathbf{x}-\mathbf{w}_j\right)\right\|^2\right)$ using an adjustable mapping matrix $\mathbf{\Omega}$ [14, 1], which is also trained by SGDL. The prototype priors usually are chosen as $p\left(\mathbf{w}_j\right) = \frac{1}{N}$.

3 Integration of Class Relation Knowledge

As mentioned in the introduction, now we assume additional knowledge regarding the class relations is provided by a Class-Relation-Knowledge-Graph $\mathcal{K}$ as known from general knowledge graph approaches [5]. More specifically, we assume that class relations are given as similarity values $K_{ij} \in [0,1]$, $i,j = 1,\ldots,C$ with $K_{ii} = 1$. If $K_{ij} = 0$ is valid, no relational class knowledge between the classes i and j is available. These values are collected in the matrix $\mathbf{K} \in [0,1]^{C\times C}$. The class relational knowledge is included into the PLVQ by the modified targets $\boldsymbol{\tau}_i = \mathbf{K}\cdot\mathbf{t}_i$ for the sample $\mathbf{x}_i$. Yet, the new targets $\boldsymbol{\tau}_i$ are not longer probabilistic vectors, because the vector entries do not necessarily sum up to one. Instead, we have $\sum_c \tau_{ic} \leq C$ with still $\boldsymbol{\tau}_{ic} \in [0,1]$ being valid, i.e. $\boldsymbol{\tau}_i$ is a possibilistic vector. Hence, we have to replace the KLD in (4) by the *adjusted* KLD

$$D_{aKL}\left(\boldsymbol{\tau}_i||\mathbf{p}_W\left(\mathbf{x}_i\right)\right) = \sum_c \tau_{ic} \cdot \log\left(\frac{\tau_{ic}}{p_W\left(c|\mathbf{x}_i\right)}\right) - \left(\tau_{ic} - p_W\left(c|\mathbf{x}_i\right)\right)$$

proposed in [2, 18] for possibilistic vectors. Again, learning takes place as SGDL for D_{aKL} with respect to the prototypes contained in W as well as to adapt the mapping matrix $\mathbf{\Omega}$. Yet, the predicted class probabilities $\mathbf{p}_W\left(\mathbf{x}\right) = \left(p_W\left(1|\mathbf{x}\right),\ldots,p_W\left(C|\mathbf{x}\right)\right)$ remain to constitute a probabilistic vector.

We refer to this approach as *class-informed* PLVQ (CI-PLVQ).

4 Numerical Simulations

We tested the approach for two real world datasets denoted as *Wine red/white* and *Wilson*.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

	CI-PLVQ	PLVQ	GLVQ	CI-PLVQ	PLVQ	GLVQ
Wine red	**0.304**	0.383	0.410	**0.517**	0.614	0.825
Wine white	**0.405**	0.483	0.479	**0.727**	0.832	1.17
Wilson	**0.277**	0.292	0.404			

Table 1: The mean weighted test error for the different data sets and algorithms (left) and the mean absolute error for the ordinal regression problems (right).

Data Description The first data *Wine red/white* set is a classical problem of ordinal regression and the task is to predict the wine quality (levels from 3 (poor) to 9 (excellent)) using the information of overall 11 physicochemical features (input) and the sensoric classification (quality level) [3]. The data set is divided into 1599 data samples of Portuguese "Vinho Verde" red wines and 4898 of white wines. The related class relation knowledge matrix $\mathbf{K}^{\text{Wine}}$ is derived from the underlying regression problem with the matrix entries $[K]_{i,j}^{\text{Wine}} = \exp\left(-|y_i - y_j|\right)$ where $y_i \in \{3, \ldots, 9\}$ are the levels of the available classes.

The second data set *Wilson* comes from the medical diagnostic area with the task to classify patient neurological impairment profiles obtained from a ^{18}F-*Fluorodesoxyglucose-Positron-Emission-Tomography* ([^{18}F]FDG-PET,[7]). The patients suffer from Wilson disease, which is an autosomal-recessive disorder of copper metabolism leading to neurophysiological impairments. Thus, in the initial non-neurologic phase, impairments are negligible or at least not defacing, whereas later on (neurologic phase) the disturbances become severe [10] with a smooth transition between both phases. The non-neurologic phase can be divided into 3 sub-types (1 - pseudo-sclerotic (PS), 2 - pseudo-parkinsonic (PP), 3 - merged type) whereas the non-neurologic phase contains to sub-types (4 - hepatic type (HT), 5 - asymptomatic type (AT)); for neurological considerations, healthy volunteers (6 - volunteers (VT) can be seen as to be in the non-neurological phase [9]. The respective clinical class relations are collected in the knowledge matrix $\mathbf{K}^{\text{Wilson}}$ to be used as the domain knowledge in CI-PLVQ. The matrix $\mathbf{K}^{\text{Wilson}}$ is depicted in Fig. 1a.

The neurological impairments profile for each patient/proband consists of a 11-dimensional vector with the normalized glucose consumption in different brain regions (*frontal lobe,parietal lobe, temporal lobe, occipital lobe, ant. cingulum, post cingulum, putamen, caput nuclei caudati, cerebellum, midbrain, thalamic area*), which are measured by [^{18}F]FDG-PET [8]. A detailed data description can be found in [7, 11].

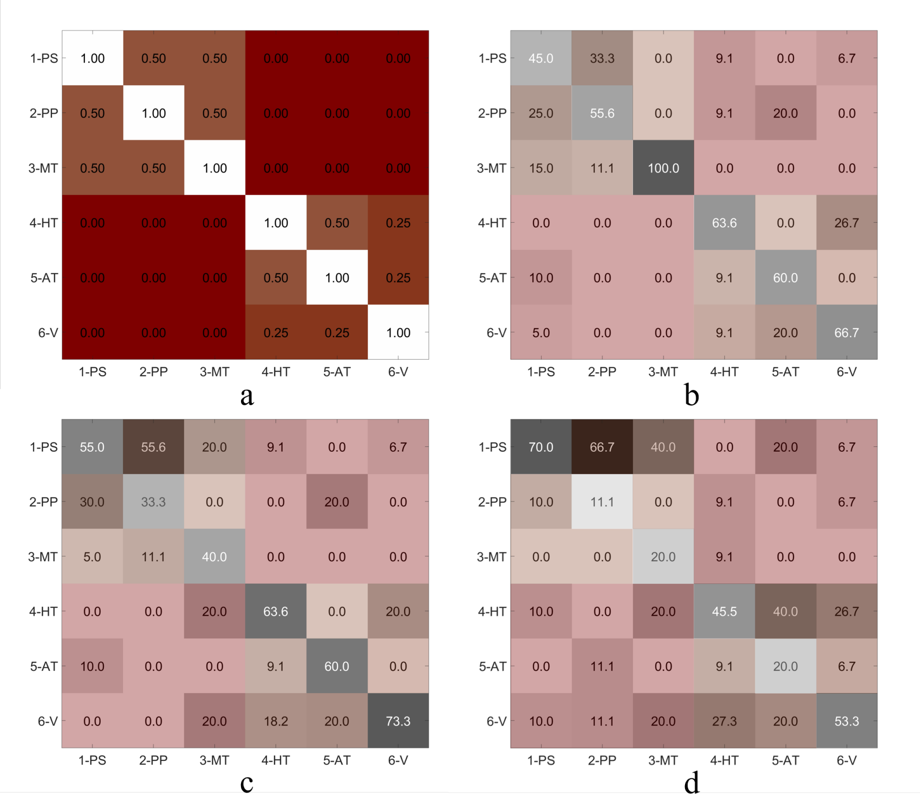
Data Analysis and Results To compare the results and obtain a meaningful performance metric, the integrated class relation information has to be taken into account for result evaluation. Thus we apply the weighted error

$$wErr(\mathbf{C}, \mathbf{K}) = \|\mathbf{C} \circ \mathbf{K}\|$$

in this paper, where $\mathbf{C}$ is the confusion matrix, $\mathbf{K}$ refers to the class relation matrix in use, and $\circ$ denotes the Hardamard product.

The results for both datasets are depicted in Tab. 1. We observe a moderately improved performance for the proposed CI-PLVQ if we compare with standard

Fig. 1: Wilson Disease data: (a) - knowledge matrix $\mathbf{K}^{\mathrm{Wilson}}$, (b-d) resulted relative confusion matrices by CI-PLVQ(b), PLVQ(c), and GLVQ(d). The axes are the Wilson-subtypes, see text.



PLVQ and GLVQ. Hence, the class relation knowledge contributes to better predictions. This statement is underlined by the evaluation of the corresponding relative confusion matrices as visualized in Fig. 1b-d: The CI-PLVQ matrix fits best to the class relations matrix.

5 Conclusions

In this contribution, we examine the potential for integrating external class relation knowledge into the probabilistic variant of LVQ (PLVQ), a prototype-based classifier. The relational information serves as additional knowledge for the classifier training. The respective loss function is the cross-entropy loss based on an adjusted Kullback-Leibler divergence to deal with possibilistic vectors. The theoretical framework is provided, and the behaviour of the approach is illustrated for two data sets. One of these is the classification of Wilson disease patients based on [^{18}F]FDG-PET-analysis. Here, with the help of the CI-PLVQ, fewer data points are misclassified between the two stages of neurological and non-neurological/volunteers in contrast to the PLVQ/GLVQ.

References

[1] M. Biehl, B. Hammer, and T. Villmann. Prototype-based models in machine learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(2):92–111, 2016.

[2] A. Cichocki, R. Zdunek, A. Phan, and S.-I. Amari. *Nonnegative Matrix and Tensor Factorizations*. Wiley, Chichester, 2009.

[3] C. A. A. F. M. T. Cortez, Paulo and J. Reis. Wine Quality. UCI Machine Learning Repository, 2009. DOI: https://doi.org/10.24432/C56S3T.

[4] T. Dash, S. Chitlangia, A. Ahuja, and A. Srinivasan. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Nature Scientifi Reports*, 12(1040):1–15, 2022.

[5] G. Futia and A. Vetrò. On the integration of knowledge graphs into deep learning models for a more comprehensible AI – Three challenges for future research. *Information*, 11(122):1–10, 2020.

[6] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.

[7] W. Hermann, H. Barthel, S. Hesse, F. Grahmann, H.-J. Kühn, A. Wagner, and T. Villmann. Comparison of clinical types of Wilson's disease and glucose metabolism in extrapyramidal motor brain regions. *Journal of Neurology*, 249(7):896–901, 2002.

[8] W. Hermann, S. Hesse, and H. Barthel. Nuklearmedizinische Diagnostik beim Morbus Wilson. *Der Nervenarzt*, 94:327–334, 2022.

[9] W. Hermann, T. Villmann, F. Grahmann, H. Kühn, and A. Wagner. Investigation of fine motoric disturbances in Wilson's disease. *Neurological Sciences*, 23(6):279–285, 2003.

[10] W. Hermann, T. Villmann, and A. Wagner. Elektrophysiologisches Schädigungsprofil von Patienten mit einem Morbus Wilson'. *Der Nervenarzt*, 74(10):881–887, 2003.

[11] M. Kaden, W. Hermann, and T. Villmann. Optimization of general statistical accuracy measures for classification based on learning vector quantization. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2014)*, pages 47–52, Louvain-La-Neuve, Belgium, 2014. i6doc.com.

[12] T. Kohonen. Learning Vector Quantization. *Neural Networks*, 1(Supplement 1):303, 1988.

[13] J. Principe. *Information Theoretic Learning*. Springer, Heidelberg, 2010.

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

[15] S. Seo, M. Bode, and K. Obermayer. Soft nearest prototype classification. *IEEE Transaction on Neural Networks*, 14:390–398, 2003.

[16] S. Seo and K. Obermayer. Soft learning vector quantization. *Neural Computation*, 15:1589–1604, 2003.

[17] A. Villmann, M. Kaden, S. Saralajew, and T. Villmann. Probabilistic learning vector quantization with cross-entropy for probabilistic class assignments in classification learning. In *Proceedings of the 17th International Conference on Artificial Intelligence and Soft Computing - ICAISC*, LNCS 10841, pages 736–749. Springer International Publishing, Switzerland, 2018.

[18] T. Villmann and S. Haase. Divergence based vector quantization. *Neural Computation*, 23(5):1343–1392, 2011.

[19] J. Voigt, S. Saralajew, M. Kaden, L. Reuss, and T. Villmann. Biologically-informed shallow classification learning integrating pathway knowledge. In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC'2024) – Bioinformatics Workshop*, volume 1, pages 357–367. SCITEPRESS – Science and Technology Publications, Lda., 2024.

[20] L. von Rüden, S. Mayer, K. B. B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker. Informed machine learning – A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):614–633, 2023.