

University of Groningen

Advanced methods for prototype-based classification

Schneider, Petra

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Schneider, P. (2010). *Advanced methods for prototype-based classification*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

7.1 Summary

The thesis presents different concepts to improve the performance of LVQ algorithms. One issue refers to metric learning in LVQ, i.e., the optimization of the employed distance measure for a specific application. A novel parameterized distance measure is proposed which can be adapted to the training data by means of any LVQ algorithm. Furthermore, three approaches to modify Robust Soft LVQ are introduced.

After providing the required background to Learning Vector Quantization in chapter 2, the novel distance measure is introduced in chapter 3. The Euclidean distance is extended by a full matrix of adaptive weight values. This approach generalizes the popular concept of relevance learning. The diagonal elements of the adaptive matrix correspond to an explicit feature weighting. The off-diagonal elements additionally weight combinations of features. Hence, correlations can be taken into account to evaluate the similarity between prototypes and feature vectors. Equivalently, this metric corresponds to the squared Euclidean distance after performing a linear transformation of data and prototypes to a new feature space. Full rank matrices can be learned as well as distance measures which are restricted to a smaller number of features. Moreover, global, as well as class-wise or prototype-specific metrics can be implemented. We derive the learning rules for matrix learning in Generalized LVQ and Robust Soft LVQ. Experiments on artificial and benchmark real life data sets illustrated the new technique in practice. The applications demonstrate that matrix learning is advantageous in two respects: The approach is clearly beneficial with respect to classification accuracy compared to the use of the Euclidean distance or the weighted Euclidean distance. This holds especially for local matrix adaptation in case of multi-class problems. Furthermore, the additional parameters improve the interpretability of the final model. Matrix parameters reveal the importance of the input dimensions for the diagonal elements and the importance of correlations for the off-diagonal elements. Additionally, the experiments

show that the two training algorithms display different learning dynamics and different characteristics of the resulting model.

In chapter 4, we present a regularization technique to influence the convergence behavior of matrix learning. The proposed method prevents the training algorithm from yielding low rank distance measures. As it is frequently observed, the update rules tend to select a single or very few directions in feature space. We propose a regularization term which favors balanced eigenvalue profiles of the relevance matrix. The technique can be applied in combination with any matrix learning scheme. We demonstrate the usefulness by means of matrix learning in GLVQ. Beside the desired effect on the convergence behavior, the technique turns out to be beneficial to prevent over-fitting effects and numerical instabilities during training.

Theoretical aspects of matrix learning in LVQ are investigated in chapter 5. In particular, we research the convergence behaviour in terms local matrix adaptation in LVQ1. Under simplifying model assumptions, it is shown that the updates lead to the selection of single directions in feature space. This direction is determined by the statistical properties of the data assigned to the respective prototype. Furthermore, we investigate the effect of the proposed regularization approach in this framework. We show that the method eases the tendency of the learning algorithm to strongly reduce the number of dimensions. The derivations are verified by practical experiments.

Chapter 6 presents three approaches to extend Robust Soft LVQ. The study refers to the original version of the algorithm based on the squared Euclidean distance. First, we propose to adapt the algorithm's hyperparameter in the course of training based on the gradient information of the RSLVQ cost function. Our experiments depict that hyperparameter learning makes the algorithm very robust with respect to the initial choice of the hyperparameter. Moreover, local hyperparameters can be learned for every prototype individually. Furthermore, we present an alternative decision rule for RSLVQ classifiers: Training the model parameters maximizes the underlying cost function which is defined in terms of a likelihood ratio. Against this background, we propose to assign a sample in the working phase to the class of highest likelihood ratio. Contrary to the distance based classification rule, this approach naturally follows the objective of RSLVQ training. The following practical examples show that the novel decision rule outperforms the distance based approach, if the method is combined with local, adaptive hyperparameters. Finally, we derive the generalization of the RSLVQ cost function with respect to vectorial class labels of the input data. We demonstrate on artificial data set that RSLVQ is a special case of the novel fuzzy algorithm.

7.2 Future work

This work could be extended in several directions. In particular, we consider the following topics for future research:

- Throughout the thesis, we apply the novel algorithms to benchmark data sets. Future work will address the application to more sophisticated data sets. Due to the interpretability and the easy handling of missing values, LVQ is especially attractive for interdisciplinary applications in the medical or biological domain. Relevance learning has already proven to be valuable for the analysis of such data, e.g. for biomarker detection. Matrix learning has the potential to further advance this work. Currently, we are working on a diagnosis system for adrenal cancer using the algorithms presented in chapter 3. The experiments show promising results which will be presented in a forthcoming publication.
- Future work will also combine the proposed extensions of Robust Soft LVQ (Sec. 6.2) with matrix learning in RSLVQ (Sec. 3.3.3)
- Several learning problems introduced in this thesis are formulated in terms of the optimization of a cost function, and the optimization is implemented in terms a stochastic gradient descent or ascent. Alternatively, advanced optimization algorithms could be used to train the model parameters, e.g. line search methods or the conjugate gradient algorithm (Bishop, 1995). Although these methods are computationally more expensive, the approach could eliminate the sensitivity of the learning process with respect to initialization and learning parameter. Advanced optimization techniques could make to learning dynamics more robust and reduce the required number of learning steps.
- Finally, continuing the work presented in Sec. 6.2.3, we plan to further generalize the RSLVQ cost function with respect to vectorial, adaptive class labels for the prototypes. This natural extension of Fuzzy RSLVQ allows to realize insecure classification results, which is highly desirable, e.g. in medical applications.

Samenvatting

Dit proefschrift beschrijft verschillende concepten voor het verbeteren van LVQ algoritmen. Een van de behandelde punten is het aanleren van een metriek binnen LVQ, dat wil zeggen, het optimaliseren van de gebruikte afstandsmaat voor een specifieke toepassing. Een nieuwe geparametriseerde afstandsmaat wordt voorgesteld, welke door ieder LVQ-algoritme aangepast kan worden aan de traindata. Ook zullen drie verschillende benaderingen om Robust Soft LVQ aan te passen geïntroduceerd worden.

Na het geven van de benodigde achtergrond van LVQ in hoofdstuk 2 zal een nieuwe afstandsmaat geïntroduceerd worden in hoofdstuk 3. De Euclidische afstand wordt uitgebreid met een volledige matrix van adaptieve gewichten. Deze aanpak generaliseert het populaire concept van het leren van relevantie. De diagonale elementen van de adaptieve matrix corresponderen met het expliciet wegen van features, terwijl de buitendiagonale elementen gewicht toevoegen van combinaties van features. Zodoende kunnen correlaties in acht worden genomen bij het evalueren van de gelijkheidswaarde tussen prototypen en featurevectoren. Deze metriek is equivalent aan de gekwadraterde Euclidische afstand na een lineaire transformatie van data en prototypen naar een nieuwe featureruimte. Zowel matrices met een volledige rang als afstandsmaten welke gelimiteerd zijn tot een klein aantal features kunnen aangeleerd worden. Daar bovenop kan zowel een globaal als een klassegewijs of prototype-specifiek metriek geïmplementeerd worden. We leiden de regels af voor het aanleren van matrices in Generalized LVQ en Robust Soft LVQ. Experimenten met kunstmatige datasets en reële referentiedatasets illustreren de nieuwe techniek in de praktijk. De toepassingen demonstreren dat het aanleren van de matrix is op twee punten profijtelijk: de aanpak heeft een duidelijke meerwaarde wanneer de accuraatheid van de classificatie in beschouwing wordt

genomen en wordt vergeleken met het gebruik van de Euclidische afstand of de gewogen Euclidische afstand. Dit geldt in het speciaal voor lokale-matrixadaptatie bij multi-klasseproblemen. Daarbovenop verbeteren de toegevoegde parameters de interpreteerbaarheid van het uiteindelijke model. Matrixparameters onthullen de belangrijke rol van de invoerdimensies op de diagonaalelementen en de belangrijke rol van correlaties op de buitendiagonale elementen. De experimenten leren ons alsmede dat de twee leeralgoritmen verschillen in leerdynamica en de karakteristieken van het resulterende model.

In hoofdstuk 4 presenteren we een regularisatietechniek om het convergentiegedrag van matrixleren te beïnvloeden. De voorgestelde methode voorkomt dat het leeralgoritme laagrangse afstandsmaten produceert. Zoals vaak wordt geobserveerd, hebben de aanpassingsregels de neiging om een enkele of een erg klein aantal richtingen in featureruimte te kiezen. We introduceren een regularisatieterm welke gebalanceerde eigenwaardeprofielen verkiest van de relevantiematrix. De techniek kan toegepast worden in combinatie met ieder matrixleerschema. We demonstreren haar bruikbaarheid door matrixleren met GLVQ. Naast het gewenste effect op het convergentiegedrag, blijkt de techniek profitabel om overspecialisatie-effecten en numerieke instabiliteiten te voorkomen tijdens het aanleren.

Theoretische aspecten van het matrixleren in LVQ worden onderzocht in hoofdstuk 5. In het bijzonder onderzoeken we het convergentiegedrag in termen van lokale-matrixadaptatie in LVQ1. Bij gesimplificeerde modelaannames wordt aangetoond dat de aanpassingen er toe leiden dat een enkele richting in featureruimte gekozen wordt. Deze richting wordt bepaald door de statistische eigenschappen van de data welke toegekend is aan het respectievelijke prototype.

Hoofdstuk 6 presenteert drie verschillende werkwijzes om Robust Soft LVQ uit te breiden. De studie refereert naar de oorspronkelijke versie van het algoritme dat is gebaseerd op de gekwadraterde Euclidische afstand. Allereerst stellen we voor om tijdens het trainen van het algoritme, gebaseerd op de gradiëntinformatie van de RSLVQ-kostenfunctie, zijn hyperparameter te wijzigen. Onze experimenten bevestigen dat het leren van de hyperparameter zorgt voor een zeer robuust algoritme ten aanzien van de initiële keus van de hyperparameter. Ook kan voor ieder prototype individueel een lokale hyperparameter aangeleerd worden. Aansluitend presenteren we een alternatieve besluitregel voor RSLVQ-klassificatie: het trainen van de modelparameters maximaliseert de onderliggen kostenfunctie, welke gedefinieerd is door een waarschijnlijkheidsratio. Met deze achtergrond stellen we voor om een nieuw monster tijdens de werkfase toe te kennen aan de klasse met de hoogste waarschijnlijkheidsratio. In tegenstelling tot de afstandgebaseerde classificatieregels volgt deze aanpak van nature het doel van RSLVQ-training. De hieropvolgende praktische voorbeelden laten zien dat de nieuwe beslissingsregel beter presteert dan

de afstandgebaseerde aanpak, wanneer de methode gecombineerd wordt met lokale adaptieve hyperparameters. Tot slot leiden we de generalisatie van de RSLVQ-kostenfunctie af ten aanzien van de vectoriale klassebenoeming van de invoerdata. Op kunstmatige data demonstreren we dat RSLVQ een special situatie is van het nieuwe fuzzy algoritme.

Bibliography

- Arnonkijpanich, B., Hammer, B., Hasenfuss, A. and Lursinap, A.: 2008, Matrix learning for topographic neural maps, *International Conference on Artificial Neural Networks*, Prague, Czech Republic, pp. 572–582.
- Bibliography on the Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ)*: 2002, Neural Networks Research Centre, Helsinki University of Technology.
- Biehl, M., Breitling, R. and Li, Y.: 2007, Analysis of tiling microarray data by learning vector quantization and relevance learning, *International Conference on Intelligent Data Engineering and Automated Learning*, Springer LNCS, Birmingham, UK, pp. 880–889.
- Biehl, M., Ghosh, A. and Hammer, B.: 2007, Dynamics and generalization ability of LVQ algorithms, *Journal of Machine Learning Research* **8**, 323–360.
- Biehl, M., Hammer, B., Schleif, F.-M., Schneider, P. and Villmann, T.: 2009, Stationarity of relevance matrix learning vector quantization, *Technical Report MLR-01-2009*, University of Leipzig.
- Biehl, M., Hammer, B., Verleysen, M. and Villmann, T. (eds): 2009, *Similarity based clustering - recent developments and biomedical applications*, Vol. 5400 of *Lecture Notes in Artificial Intelligence*, Springer.
- Biehl, M., Pasma, P., Pijl, M. and Petkov, N.: 2006, Classification of boar sperm head images using learning vector quantization, in M. Verleysen (ed.), *European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 545–550.
- Biehl, M., Schneider, P., Hammer, B., Schleif, F.-M. and Villmann, T.: submitted, 2010, Stationarity of matrix updates in relevance learning vector quantization.
- Bishop, C. M.: 1995, *Neural Networks for Pattern Recognition*, 1 edn, Oxford University Press.
- Bishop, C. M.: 2007, *Pattern Recognition and Machine Learning*, 1 edn, Springer.

- Boehm, W. and Prautzsch, H.: 1993, *Numerical Methods*, Vieweg.
- Bojer, T., Hammer, B., Schunk, D. and von Toschanowitz, K. T.: 2001, Relevance determination in learning vector quantization, in M. Verleysen (ed.), *European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 271–276.
- Cover, T. and Hart, P.: 1967, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* **13**(1), 21–27.
- Crammer, K., Gilad-Bachrach, R., Navot, A. and Tishby, A.: 2003, Margin analysis of the lvq algorithm, *Advances in Neural Information Processing Systems*, Vol. 15, MIT Press, Cambridge, MA, USA, pp. 462–469.
- Darken, C., Chang, J., Z, J. C. and Moody, J.: 1992, Learning rate schedules for faster stochastic gradient search, *Neural Networks for Signal Processing 2 - Proceedings of the 1992 IEEE Workshop*, IEEE Press.
- Duda, R., Hart, P. and Stork, D.: 2000, *Pattern Classification*, second edn, Wiley-Interscience.
- Gath, I. and Geva, A. B.: 1989, Unsupervised optimal fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(7), 773–780.
- Ghosh, A., Biehl, M. and Hammer, B.: 2006, Performance analysis of lvq algorithms: a statistical physics approach, *Neural Networks* **19**(6), 817–829.
- Gustafson, E. and Kessel, W.: 1979, Fuzzy clustering with a fuzzy covariance matrix, *IEEE Conference on Decision and Control*, San Diego, CA, USA, pp. 761–766.
- Hammer, B., Schleif, F.-M. and Villmann, T.: 2005, On the generalization ability of prototype-based classifiers with local relevance determination, *Technical Report IfI-05-14*, Clausthal University of Technology.
- Hammer, B., Strickert, M. and Villmann, T.: 2004, Prototype based recognition of splice sites, *Bioinformatic using Computational Intelligence Paradigms*, Springer-Verlag, pp. 25–56.
- Hammer, B., Strickert, M. and Villmann, T.: 2005a, On the generalization ability of GRLVQ networks, *Neural Processing Letters* **21**(2), 109–120.
- Hammer, B., Strickert, M. and Villmann, T.: 2005b, Supervised neural gas with general similarity measure, *Neural Processing Letters* **21**(1), 21–44.
- Hammer, B. and Villmann, T.: 2002, Generalized relevance learning vector quantization, *Neural Networks* **15**(8-9), 1059–1068.
- Kaski, S.: 2001, Principle of learning metrics for exploratory data analysis, *Neural Networks for Signal Processing XI, Proceedings of the 2001 IEEE Signal Processing Society Workshop*, IEEE, pp. 53–62.
- Kietzmann, T. C., Lange, S. and Riedmiller, M.: 2008, Incremental grlvq: Learning relevant features for 3d object recognition, *Neurocomputing* **71**(13-15), 2868–2879.

- Kohonen, T.: 1986, Learning vector quantization for pattern recognition, *Technical Report TKK-F-A601*, Helsinki Univeristy of Technology, Espoo, Finland.
- Kohonen, T.: 1990, Improved versions of learning vector quantization, *International Joint Conference on Neural Networks*, Vol. 1, pp. 545–550.
- Kohonen, T.: 1997, *Self-Organizing Maps*, second edn, Springer, Berlin, Heidelberg.
- Kohonen, T.: 1998, Learning vector quantization, *The handbook of brain theory and neural networks*, MIT Press, Cambridge, MA, USA, pp. 537–540.
- Kusumoputro, B. and Budiarto, H.: 1999, Improvement of artificial odor discrimination system using fuzzy-lvq neural network, *International Conference on Computational Intelligence and Multimedia Applications*, IEEE Computer Society, Los Alamitos, CA, USA, pp. 474–478.
- Martinetz, T. and Schulten, K.: 1991, A "neural-gas" network learns topologies, *Artificial Neural Networks I*, 397–402.
- Mendenhall, M. and Merényi, E.: 2006, Generalized relevance learning vector quantization for classification driven feature extraction from hyperspectral data, *Proceedings of ASPRS 2006 Annual Conference and Technology Exhibition*, p. 8.
- Mwebaze, E., Schneider, P., Schleif, F.-M., Haase, S., Villmann, T. and Biehl, M.: 2010, Divergence based learning vector quantization, in M. Verleysen (ed.), *European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 247–252.
- Newman, D. J., Hettich, S., Blake, C. L. and Merz, C. J.: 1998, Uci repository of machine learning databases, <http://archive.ics.uci.edu/ml/>.
- Ong, C., A. Smola, A. and Williamson, R.: 2005, Learning the kernel with hyperkernels, *Journal of Machine Learning Research* **6**, 1043–1071.
- Perfetti, R. and Ricci, E.: 2006, Reduced complexity rbf classifiers with support vector centres and dynamic decay adjustment, *Neurocomputing* **69**(16-18), 2446–2450.
- Petersen, K. B. and Pedersen, M. S.: 2008, The matrix cookbook, <http://matrixcookbook.com>.
- Prudent, Y. and Ennaji, A.: 2005, A k nearest classifier design, *Electronic Letters on Computer Vision and Image Analysis* **5**(2), 58–71.
- Sato, A. and Yamada, K.: 1996, Generalized learning vector quantization, in M. C. M. D. S. Touretzky and M. E. Hasselmo (eds), *Advances in Neural Information Processing Systems*, Vol. 8, MIT Press, Cambridge, MA, USA, pp. 423–429.
- Sato, A. and Yamada, K.: 1998, An analysis of convergence in generalized lvq, in L. Niklasson, M. Bodén and T. Ziemke (eds), *Proceedings of the International Conference on Artificial Neural Networks*, Springer, pp. 170–176.

- Schneider, P., Biehl, M. and Hammer, B.: 2007, Relevance matrices in lvq, in M. Verleysen (ed.), *European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 37–42.
- Schneider, P., Biehl, M. and Hammer, B.: 2009a, Adaptive relevance matrices in learning vector quantization, *Neural Computation* **21**(12), 3532–3561.
- Schneider, P., Biehl, M. and Hammer, B.: 2009b, Distance learning in discriminative vector quantization, *Neural Computation* **21**(10), 2942–2969.
- Schneider, P., Bunte, K., Stiekema, H., Hammer, B., Villmann, T. and Biehl, M.: 2010, Regularization in matrix relevance learning, *IEEE Transactions on Neural Networks* **21**(5), 831–840.
- Schneider, P., Schleif, F.-M., Villmann, T. and Biehl, M.: 2008, Generalized matrix learning vector quantizer for the analysis of spectral data, in M. Verleysen (ed.), *European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 451–456.
- Seo, S., Bode, M. and Obermayer, K.: 2003, Soft nearest prototype classification, *IEEE Transactions on Neural Networks* **14**(2), 390–398.
- Seo, S. and Obermayer, K.: 2003, Soft learning vector quantization, *Neural Computation* **15**(7), 1589–1604.
- Seo, S. and Obermayer, K.: 2006, Dynamic hyper parameter scaling method for lvq algorithms, *International Joint Conference on Neural Networks*, Vancouver, CA.
- Shalev-Schwartz, S., Singer, Y. and Ng, A.: 2004, Online and batch learning of pseudo-metrics, *Proceedings of the 21st International Conference on Machine Learning*, ACM, New York, USA, p. 94.
- Strickert, M., Witzel, K., Mock, H.-P., Schleif, F.-M. and Villmann, T.: 2007, Supervised attribute relevance determination for protein identification in stress experiments, *Machine Learning in Systems Biology*, pp. 81–86.
- Tamura, H. and Tanno, K.: 2008, Midpoint-validation method for support vector machine classification, *IEICE - Transactions on Information Systems* **E91-D**(7), 2095–2098.
- Thiel, C., Sonntag, B. and Schwenker, F.: 2008, Experiments with supervised fuzzy lvq, in L. Prevost, S. Marinai and F. Schwenker (eds), *Artificial Neural Networks in Pattern Recognition*, Vol. 5064 of *Lecture Notes in Computer Science*, Springer, pp. 125–132.
- Tsang, I. W., Kocsor, A. and Kwok, J. T.: 2006, Diversified svm ensembles for large data sets, *Machine Learning: ECML 2006*, Vol. 4212 of *Lecture Notes in Computer Science*, Springer, pp. 792–800.
- Weinberger, K., Blitzer, J. and Saul, L.: 2006, Distance metric learning for large margin nearest neighbor classification, in Y. Weiss, B. Schölkopf and J. Platt (eds), *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, pp. 1473–1480.
- Wu, K.-L. and Yang, M.-S.: 2003, A fuzzy-soft learning vector quantization, *Neurocomputing* **55**(3-4), 681–697.