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.

Matrix Relevance LVQ in Steroid Metabolomics Based Classification of Adrenal Tumors

M. Biehl[1], P. Schneider[2], D.J. Smith[2], H. Stiekema[1], A.E. Taylor[2], B.A. Hughes[2], C.H.L. Shackleton[2], P.M. Stewart[2], and W. Arlt[2]

1- Univ. of Groningen - Johann Bernoulli Inst. for Math. and Computer Science
P.O. Box 407, 9700 AK Groningen - The Netherlands

2- University of Birmingham - Centre for Endocrinology, Diabetes, and Metabolism
Birmingham B15 2TT - United Kingdom

Abstract. We present a machine learning system for the differential diagnosis of benign adrenocortical adenoma (ACA) vs. malignant adrenocortical carcinoma (ACC). The data employed for the classification are urinary excretion values of 32 steroid metabolites. We apply prototype-based classification techniques to discriminate the classes, in particular, we use modifications of Generalized Learning Vector Quantization including matrix relevance learning. The obtained system achieves high sensitivity and specificity and outperforms previously used approaches for the detection of adrenal malignancy. Moreover, the method identifies a subset of most discriminative markers which facilitates its future use as a non-invasive high-throughput diagnostic tool.

1 Introduction

The adrenal glands are major components of the human endocrine system. They are located on top of the kidneys and produce and secrete steroid hormones. The latter regulate a wide variety of body functions such as metabolism, the immune system, and sexual development. Tumors of the adrenal cortex are common, but the classification of an adrenal mass as benign adrenocortical adenoma (ACA) or malignant carcinoma (ACC) is a major diagnostic challenge. Currently, diagnosis is based on criteria like tumor size and density as assessed by imaging techniques, which however lack satisfactory specificity. Hence, the identification of reliable diagnostic markers for ACC is of considerable interest [1, 2].

In this contribution, we present urinary steroid metabolomics and its analysis by means of machine learning techniques as a novel approach to this clinical problem. In particular, we apply recent modifications of Learning Vector Quantization (LVQ) [3] to provide a diagnostic test for the detection of adrenal malignancy. LVQ is a family of distance-based classification algorithms which are particularly attractive for complex real life applications. Since the model parameters are interpretable, LVQ systems facilitate new insight into the data and the underlying classification task. In addition, extensions of LVQ termed relevance learning provide a weighting of features with respect to their significance. This is a particularly advantageous aspect in the problem at hand, as the restriction to a reduced panel of markers would be highly desirable for the design of a high-throughput practical diagnosis tool [2].

We employ the framework of Generalized LVQ (GLVQ) [4, 5] in combination with relevance matrices, introduced as Generalized Matrix LVQ (GMLVQ) in

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.

[6]. For comparison we also apply a scheme restricted to diagonal matrices, equivalent to Generalized Relevance LVQ (GRLVQ) [5] as well as plain GLVQ [4]. We also compare with Fisher Linear Discriminant Analysis (LDA) [7].

We demonstrate that GMLVQ provides a highly sensitive and specific classifier and allows to identify a subset of most discriminative steroids. We obtain a promising biomarker tool for the diagnostic work-up of patients with adrenal tumors. The medical aspects and implications of this study have been discussed in greater detail in a recent publication [2]. Here, the focus is on the machine learning analysis which is outlined only briefly in [2].

2 The Data

Urine samples of 147 adrenal tumor patients were acquired within the *European Network for the Study of Adrenal Tumours* [1]. The study population consists of 102 ACA and 45 ACC samples. In addition, 88 samples from a healthy control cohort were included. The excretion values of 32 preselected steroid metabolites were quantified using gas chromatography/mass spectrometry (GC/MS). For a more detailed description of the medical and biochemical background, technical aspects, the study design, and the patient cohort we refer to [2].

All numerical steroid excretion values were log-transformed and subsequently rescaled by subtracting the corresponding mean values obtained in healthy controls and dividing by the respective standard deviations. A very small number of measurements were found to be *zero* within the sensitivity of the GC/MS analysis, these values were set to 10^{-10} before log-transformation. We have confirmed that the results presented in the following are not affected by the precise choice of this correction parameter within a range of sufficiently small values. Steroid excretion data contained a total of 56 missing values (out of 4704).

We obtain a total of 147 labeled feature vectors $\boldsymbol{\xi} \in \mathbb{R}^{32}$ representing the log-transformed and rescaled steroid excretion profiles of 102 patients with ACA and 45 patients with ACC. Each dimension of feature space corresponds to one of the 32 considered steroid markers, which can be grouped into Androgen metabolites and Androgen precursors (features 1-6), Mineralocorticoids and precursors thereof (7-13), Glucocorticoid precursors (14-19), and Glucocorticoid metabolites (20-32). For a detailed list of metabolites we point the interested reader to [2]. In the following we refer to markers by number 1-32 only.

3 Machine Learning Analysis

LVQ systems implement a classification of N-dim. feature vectors $\boldsymbol{\xi} \in \mathbb{R}^N$ in terms of prototypes $\left\{\boldsymbol{w}_j \in \mathbb{R}^N\right\}_{j=1}^K$, i.e. typical representatives of the classes $c(\mathbf{w_j}) \in \{1, 2 \ldots C\}$ in feature space. Together with a suitable distance measure $d^\Lambda(\boldsymbol{w}, \boldsymbol{\xi})$ they parameterize, e.g., a Nearest Prototype Classifier (NPC) which assigns any input $\boldsymbol{\xi}$ to the class represented by the closest prototype.

In general, the superscript Λ refers to a set of – possibly adaptive – parameters in the definition of the distance. The recently introduced GMLVQ [6] employs a full $N \times N$-dim. matrix Λ of adaptive parameters to define a general quadratic distance measure of the form

$$d^\Lambda(\boldsymbol{w}, \boldsymbol{\xi}) = (\boldsymbol{w} - \boldsymbol{\xi})^\top \Lambda (\boldsymbol{w} - \boldsymbol{\xi}) \quad \text{with } \Lambda = \Omega^\top \Omega \quad \text{and } \textstyle\sum_i \Lambda_{ii} = 1. \qquad (1)$$

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.

The parameterization of Λ in terms of $\Omega \in \mathbb{R}^{N\times N}$ ensures positive semi-definiteness and the normalization of the trace prevents numerical degeneration. The matrix Ω can be interpreted as parameterizing an arbitrary linear transformation of the original feature space, including rescaling of features and rotations.

Here we consider only the simplest setting with a single matrix Λ defining a global distance measure. For modifications using local distances or, e.g., rectangular matrices Ω we refer to [6, 8]. Note that missing values can be ignored when comparing distances of a given $\boldsymbol{\xi}$ from different prototypes. Hence, we refrain from imputing missing values in LVQ training and classification.

For a given set of example data, the training process is guided by a cost function introduced in the framework of GLVQ [4]:

$$E(\{\mathbf{w}^j\}, \Omega) = \sum_i \Phi(\mu_i) \quad \text{where} \quad \mu_i = \frac{d^\Lambda(\boldsymbol{w}_J, \boldsymbol{\xi^i}) - d^\Lambda(\boldsymbol{w}_K, \boldsymbol{\xi^i})}{d^\Lambda(\boldsymbol{w}_J, \boldsymbol{\xi^i}) + d^\Lambda(\boldsymbol{w}_K, \boldsymbol{\xi^i})}. \qquad (2)$$

Here, the sum is over all training examples. In general, Φ denotes a monotonic function, e.g. a sigmoidal or the identity $\Phi(x) = x$ which we employ throughout the following. For a given ξ^i, the index J (K) corresponds to the closest prototype which represents the correct class c^i (a class different from c^i). Hence, $\mu_i < 0$ indicates a correct classification and $|\mu_i|$ can be interpreted as its *margin*.

In GLVQ based training, all adaptive parameters are updated by stochastic gradient descent [4, 5, 6]. Upon presentation of a randomly selected example $\{\boldsymbol{\xi}^i, c^i\}$, the form of the updates at learning step t is

$$\boldsymbol{w}_{J,K}(t{+}1) = \boldsymbol{w}_{J,K}(t) - \eta_w \partial\Phi(\mu_i)/\partial \boldsymbol{w}_{J,K} \quad \text{and} \quad \Omega(t{+}1) = \Omega(t) - \eta_\Omega \partial\Phi/\partial\Omega \quad (3)$$

with $\boldsymbol{w}_{J,K}$ defined as above and the *learning rates* η_w and η_Ω. For a detailed derivation of the updates see [6]. The normalization of Λ or potential additional constraints are enforced explicitly after each update step.

Obviously, we recover Euclidean metrics from Eq. (1) by fixing Λ proportional to the N-dim. identity matrix: $\Lambda = I_N/N$. In this case, which is equivalent to plain GLVQ [4], the training prescription (3) applies only to the prototypes. The restriction to a diagonal matrix Λ represents a weighted Euclidean distance of the form $d^\Lambda(\boldsymbol{w}, \boldsymbol{\xi}) = \sum_j \Lambda_{jj} (w_j - \xi_j)^2$ and corresponds to the GRLVQ scheme introduced in [5].

We have analysed the available data in terms of the simplest system with only two prototypes, $\boldsymbol{w}_{1,2}$, representing classes 1 (ACA) and 2 (ACC), respectively. In order to evaluate the quality of the obtained classifier we select 93 ACA and 40 ACC examples randomly for training and retain the remaining samples for testing. All results reported here correspond to mean values over 1000 randomized splits of the data set, referred to as *runs* in the following.

For better interpretability of the emerging relevance matrix, a z-score transformation was performed with respect to the mean and standard deviation of features in the actual training set. Initially, prototypes $\boldsymbol{w}_j$ were set to the mean of a random selection of 50% of the training samples representing class $c(\boldsymbol{w}_j)$. The matrix Ω was initialized corresponding to $\Lambda = I_N/N$. For all results reported here, 100 epochs of stochastic gradient descent were performed at constant learning rates $\eta_w = \eta_\Omega = 10^{-3}$. Different settings were investigated,

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.

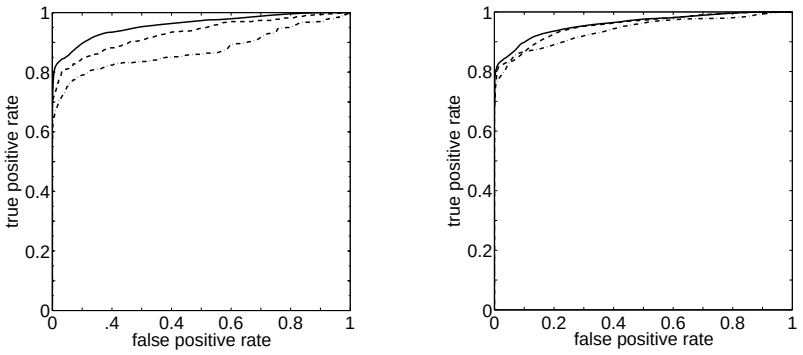

Fig. 1: Left panel: threshold-averaged ROC as obtained by GLVQ (chain line), GRLVQ (dashed), and GMRLVQ (solid). Right panel: ROC of GMRLVQ employing all 32 steroids (solid line, same as in the left panel) and restricted to the 9 (dashed) and 3 (chain) most discriminative markers in each run, respectively.

including schedules for decreasing $\eta_{w,\Omega}$ during training. Results were robust in terms of the considered performance measures.

When applying the classifier to the test set after training, we modify the NPC scheme by introducing a threshold θ. The modified system assigns $\boldsymbol{\xi}$ to class 1 (ACA) if $d^{\Lambda}(\boldsymbol{w}_1, \boldsymbol{\xi}) \leq d^{\Lambda}(\boldsymbol{w}_2, \boldsymbol{\xi}) - \theta$ and to class 2 (ACC), else. Hence, θ controls a bias towards one of the classes and by considering a sufficiently large range of θ we obtain the full Receiver Operator Characteristics (ROC) of a given LVQ classifier [9]. In the ROC curve, the true positive rate for detecting ACC (sensitivity, $SENS$) is plotted vs. the false positive rate (1-specificity, $1 - SPEC$). A particular *working point* can be selected by the domain expert according to problem specific requirements. ROC curves displayed here were obtained as threshold averages [9] over 1000 randomized training runs. We consider the area under the ROC curve (AUC) as a quality measure when comparing different classifiers [9]. In addition we provide results for the example working point with $SENS = SPEC$.

Along these lines we present our findings corresponding to GMLVQ, GRLVQ and simple GLVQ in the next section. In addition we report results obtained by LDA [7] in the implementation of van der Maaten's toolbox for dimensionality reduction [10]. When applying LDA, missing values were replaced by the class conditional means which should theoretically give a performance advantage compared with the LVQ approaches.

We employ a heuristic scheme to identify most significant steroid markers by means of the obtained relevance matrices. Their diagonal elements Λ_{ii} can be interpreted as to quantify the *importance* of features in the classification. Note that applying this heuristics in GMLVQ goes beyond a simple univariate approach since Λ_{ii} accumulates the weights of pairs of original features in the projections $\Omega\,\boldsymbol{\xi}$: $\Lambda_{ii} = \sum_j \Omega_{ij}^{\top}\Omega_{ji} = \sum_j \Omega_{ji}^2$.

The practical realization of steroid metabolomics as a diagnostic tool would

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.

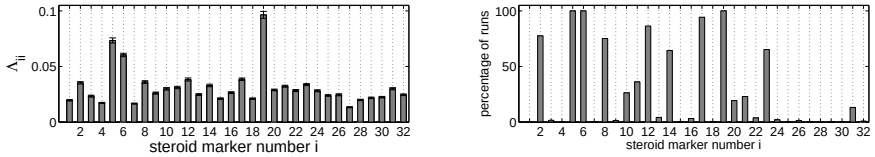

Fig. 2: Diagonal relevances Λ_{ii}. Left: Mean and standard deviation obtained over 1000 runs of the training process. Right: Fraction of runs in which a metabolite was among the nine most relevant markers.

greatly benefit from the consideration of a reduced panel of markers since rapid profiling of up to 10 steroids is technically feasible. In order to identify such a panel, we determined after each GMLVQ run the set of 9 or 3 markers with the largest diagonal relevances Λ_{ii}. We then repeated the GMLVQ training using only the individually selected markers in the given training set. This way, we obtained an estimate of the achievable performance when making use of restricted sets of 9 or 3 features, only. For comparison we also applied the LDA analysis to the same feature sets.

4 Results

The threshold-averaged ROC with respect to test set performance using all 32 steroid markers are displayed in Figure 1 (left panel). We observe that performance increases significantly with the complexity of the adaptive distance measure. Plain GLVQ yields an area under the ROC of $AUC \approx 0.873$ and $SENS = SPEC$ at approximately 0.82. Introducing adaptive diagonal relevances (GRLVQ) in the distance measure improves these results to $AUC \approx 0.928$ and $SENS = SPEC \approx 0.86$. The use of a full matrix of relevances further enhances the performance: We obtained an ROC with $AUC \approx 0.965$ and $SENS = SPEC \approx 0.90$. These findings demonstrate that matrix relevance learning increases the flexibility and power of LVQ systems significantly. Applying LDA to the full panel of 32 steroid markers, we observed strong overfitting effects. With respect to the test set performance, LDA achieved an ROC with $AUC \approx 0.925$ and $SENS = SPEC \approx 0.871$, only.

We analysed in greater detail the relevance matrices obtained by GMLVQ. Figure 2 (left panel) displays the diagonal elements as observed on average over the 1000 randomized runs. Note that the Λ_{ii} corresponding to markers 5,6, and 19 are particularly large. While this could be expected from medical insight and experience [2], the significance of other markers was less obvious beforehand. Note that, for instance, markers 8 and 12 display very low predictive power according to a univariate analysis, see [2]. Accordingly, we found low relevances in GRLVQ where Λ is restricted to diagonal form. However, matrix relevance learning shows that the combination of markers 8 and 12 is indeed discriminative and, consequently, plays an important role in the GMLVQ classifier.

In each run, we determined the subset of 9 most significant markers as indicated by the largest Λ_{ii}. Figure 2 (right panel) displays the percentage of runs

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.

in which a particular marker was included in this subset. An analogous analysis with respect to the 3 leading markers shows that markers (5,6,19) were selected in more than 95% of the runs.

We repeated the GMLVQ analysis for each training set, restricting the system to the individually selected subset of features. Figure 1 (right panel) shows the corresponding ROC curves. The performance of the resulting classifiers was only slightly inferior compared to using the full panel: We obtained $AUC \approx 0.960$, $SENS = SPEC \approx 0.88$ for 9 steroid markers and $AUC \approx 0.942, SENS = SPEC \approx 0.87$ when using only the 3 leading figures. Reassuringly, LDA yielded comparable results when using 9 features ($AUC \approx 0.957$) or 3 features only ($AUC \approx 0.93$).

5 Conclusion

Our results show that urinary steroid profiling in combination with the machine learning analysis provides a promising diagnostic tool for the differentiation of benign and malignant adrenocortical tumors. Our novel approach, steroid metabolomics, provides a highly specific and sensitive classification and relevance learning allows for a reduction of the panel of markers in view of a practical high-throughput tool.

Our study has the limitation of being retrospective. Prospective validation with respect to novel patient data will be essential for establishing the diagnostic tool. Extensions of the approach will include the monitoring of patients after surgery or under treatment. The potential identification of tumor subtypes by more complex LVQ systems will also be addressed in a forthcoming project.

Acknowledgment: This work was supported by the Medical Research Council UK (Strategic Biomarker Grant G0801473).

References

[1] The European Network for the Study of Adrenal Tumours (ENS@T), url: www.ensat.org

[2] W. Arlt, M. Biehl, A.E. Taylor, S. Hahner, R. Libé, B.A. Hughes, P. Schneider, D.J. Smith, H. Stiekema, N. Krone, E. Porfiri, G. Opocher, J. Bertherat, F. Mantero, B. Allolio, M. Terzolo, P. Nightingale, C.H.L. Shackleton, X. Bertagna, M. Fassnacht, and P.M. Stewart, Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors, *J. Clinical Endocrinology & Metabolism*, 96: 3775-3784, 2011.

[3] T. Kohonen, *Self Organizing Maps*, Springer, Berlin, 2001.

[4] A.S. Sato and K. Yamada, Generalized Learning Vector Quantization, *Advances in Neural Information Processing Systems*, 8:423–429, 1996.

[5] B. Hammer and T. Villmann, Generalized Relevance Learning Vector Quantization, *Neural Networks*, 15:1059–1068, 2002.

[6] P. Schneider, M. Biehl, and B. Hammer, Adaptive Relevance Matrices in Learning Vector Quantization, *Neural Computation*, 21:3532–3561, 2009.

[7] P. McCullagh, J.A. Nelder, *Generalized Linear Models*, Chapman & Hall/CRC, 1989.

[8] 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*, in press.

[9] T. Fawcett, An introduction to ROC analysis, Pattern Rec. Lett., 27:861-874, 2006.

[10] L.J.P. van der Maaten, Matlab Toolbox for Dimensionality Reduction (v0.7b). url: http://hompepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html