Classification of Mass-Spectrometric Data in Clinical Proteomics Using Learning Vector Quantization Methods

T. Villmann, F.-M. Schleif, M. Kostrzewa, A. Walch and B. Hammer

August 15, 2007

Abstract

In the present contribution we present two recently developed classification algorithms for analysis of mass-spectrometric data - the supervised neural gas and the fuzzy labeled self-organizing map. The algorithms are inherently regularizing, which is recommended, for these spectral data because of its high dimensionality and the sparseness for specific problems. The algorithms are both prototype based such that the principle of characteristic representants is realized. This leads to an easy interpretation of the generated classifcation model. Further, the fuzzy labeled self-organizing map, is able to process uncertainty in data, and classification results can be obtained as fuzzy decisions. Moreover, this fuzzy classifcation together with the property of topographic mapping offers the possibility of class similarity detection, which can be used for class visualization. We demonstrate the power of both methods for two exemplary examples: the classification of bacteria (listeria types) and neoplastic and non-neoplastic cell populations in breast cancer tissue sections.

1 Introduction

Exploration and analysis of mass spectrometric data in the field of clinical proteomics have become one of the key problems in computational proteomics. Thereby, the complexity of the mass spectrometric data is one difficult problem. Frequently, the data are given as huge-dimensional functional vectors with several thousands dimensions according to the resolution on the mass axis. Further, usually the number of samples is limited to a few data sets due to clinical restrictions. From am athematical point of view, the data space to be explored is sparsely filled. Thereby, the spectra may be overlaid by noise such that the contained signal is difficult to extract. A further problem arises for data analysis methods as consequence of the high dimensionality: the data can always be separated more or less independent of the separation criteria [31]. Thus, any method is confronted with the problem of the detection of the underlying regularities. Usually, this problem is overcome by cross-validation. Yet, the certainty of such an evaluation is diminished here, because of the humble number of data. Therefore, advanced methods for data analysis in mass spectrometry are required to be regularizing inherently, robust and to be able to deal with high-dimensional, sparse and noisy data.

In the following we will restrict us to classification problems. Thereby, we will concentrate to the following aspects: How we can achieve a good classification accuracy and how we can visualize classification results in an adequate manner. The latter problem is related to the problem of class similarity detection. Moreover, each classification result depends on the underlying similarity/dissimilarity measure for data.

Classification in traditional statistics is frequently realized by Fisher's discriminant analysis (FDA) or linear/quadratic discriminant analysis (LDA/QDA) [23]. FDA optimizes the inter-intra-class correlation ratio by weighting the data dimensions to obtain a good separation plane, i.e. it is based on a weighted Euclidean distance for data similarity. LDA/QDA tries to optimize the Bayes error of the classification by utilization of the (class dependent, QDA) covariance, i.e. the Mahalanobis distance between data is used inherently [9]. These classical statistic approaches are more and more supplemented by machine learning tools, which provide adaptive and robust methods for pattern recognition in complex data [1],[25],[26]. Thereby, machine learning algorithms comprise approaches like artificial neural networks (ANNs), evolutionary algorithms (EAs), decision trees (DTs), clustering, and other [4].

Beside the pure classification accuracy of a generated classification model, its interpretability plays an important role. Standard methods use (linear) principal component analysis (PCA) and Fisher's discriminant analysis or classical hierarchical clustering [7],[8]. Additionally, advanced preprocessing procedures, like denoising using wavelets or 'intelligent' peak picking heuristics including problem specific expert knowledge, are applied to improve the accuracy. Here, the flexibility of machine learning methods offers new ways which may result in better results [22]. For example, multilayer perceptron neural networks (MLPs) as universal function approximators offer, on the one hand side, greatest flexibility in learning and adaptation to achieve good classification results [**?**]. On the other hand, however, their decision scheme is more or less a 'black box', because all the information for the decision is distributed over the whole network. In contrast, prototype based classifiers realize the principle of '*characteristic representatives'* for data subsets or decision regions between them . Thus, the interpretation becomes easy. Examples for such tools are Support Vector Machines (SVM) [27], Kohonen's Learning Vector Quantization (LVQ), Self-Organizing Maps (SOMs) [17] and respective variants. New developments include the utilization of non-standard metrics (functional norms, scaled Euclidean metric) and task-dependent automatic metric adaptation (feature selection), fuzzy classification, and similarity based visualization of data. These properties offers new possibilities for analysis also of mass spectrometric data.

In the present paper we will give insights to two recently developed prototype based classifiers which fulfill the above requirements. The Supervised Neural Gas (SRNG) and the Fuzzy labeled SOM (FLSOM) are robust prototype based neural classifiers, which are inherently regularizing by neighborhood cooperativeness between prototypes and which are easy to interpret. Moreover, as we will explain FLSOM is able to detect class similarities and offers the possibility of fuzzy classification. Both algorithms share the flexibility of utilization of arbitrary data metrics, which may be adapted during the training process as well in dependence on the classification task to be learned [14].

The article is structured as follows. First, we shortly review both methods, classification based on LVQ and FLSOM, pointing out their different properties and abilities. Thereby, we will emphasize the ability of the usage of general, task adequate, similarity measures in both methods. Further, we will highlight the class similarity detection probability of the semi-supervised FLSOM, which can be used for adequate class visualization or clinical interpretation. The theoretic part is followed by two clinical example investigations. The first one is an investigation of mass spectrometric bacteria data to find an adequate classification. In the second application a classification of neoplastic and non-neoplatic cell populations in histological sections of breast cancer tissue is considered. For both applications we demonstrate the advanced abilities of the methods. Concluding remarks complete the paper.

2 Prototype based classifiers

Usually, spectrometric data in proteomic analysis are given as vectors $\mathbf{v} \in V \subseteq \mathbb{R}^D$. D is the data dimension which may be huge in this field. It depends on the spectral range and sampling resolution. Because the data represent spectra, or more general functions, they are called some times *functional data*. We remark that for functional data, the sequence of data dimensions is not independent.

In our consideration we assume, that there exist an underlying (unknown) data probability density P in V. Further, we assume for the training data that to each data vector $\mathbf{v}$ a unique class label $\mathbf{c}(\mathbf{v})$ exist. Prototype based classifiers distributes prototypes $\mathbf{w_r} \in \mathbb{R}^D$, $\mathbf{r} \in A$, as representations for classes in the data space V, whereby A is a given index set. The prototypes should represent class distributions in the data space and borders between different classes. For this purpose, each prototype has a class label $\mathbf{y_r}$.

Several approaches exist: the well-known LVQ family introduced by KOHONEN tries to minimize the Bayes classification error, but the adaptation dynamic is only a heuristic Hebbian like and does not perform a gradient descent on the misclassification error [17]. SVMs are based on structural risk optimization using a separation margin maximization approach [6]. Both methods are very powerful. In particular, SVMs frequently show superior results [27]. However, if new data become available for training a complete new learning has to be applied for SVM. Both methods have in common that they are not able to handle uncertainty in classification for training data (fuzzy class memberships). Further, the obtained classification model is crisp.

We now review two recently developed classification schemes, which both are inherently regularizing to address the above mentioned problem of noisy and sparse data in huge-dimensional data spaces. The first one is a generalization of Kohonen's LVQ scheme providing a gradient descent on a cost function. The second one extend the unsupervised SOM, such that a semi-supervised fuzzy classifier is obtained with excellent visualization abilities and the feature of class similarity detection. Both methods share the ability to proceed arbitrary (differentiable) may be parametrized data similarity measures, which itself can be in parallel subject of optimization with respect to the classification task.

2.1 Classification by Supervised Neural Gas

As mentioned above, LVQ does not minimize the classification error by gradient descent prototype adaptation. Therefore SATO&YAMADA introduced a cost function based on a classification function μ such that the respective gradient descent is similar to the heuristic LVQ learning scheme preserving the Hebbian characteristic [21]. For a given data point $\mathbf{v}$ with class label $\mathbf{c}(\mathbf{v})$ the two best matching prototypes with respect to the data metric d, usually the quadratic Euclidian, are determined: $\mathbf{w}_{\mathbf{r}^+}$ has minimum distance $d^+ = d(\mathbf{v}, \mathbf{w}_{\mathbf{r}^+})$ and the class labels are identically: $\mathbf{y}_{\mathbf{r}^+} = \mathbf{c}(\mathbf{v})$. The other best prototype $\mathbf{w}_{\mathbf{r}^-}$ has has minimum distance $d^- = d(\mathbf{v}, \mathbf{w}_{\mathbf{r}^-})$ but the class labels are different: $\mathbf{y}_{\mathbf{r}^-} = \mathbf{c}(\mathbf{v})$. Then the classification function $\mu(\mathbf{v})$ is defined as

$$\mu(\mathbf{v}) = \frac{d^+ - d^-}{d^+ + d^-} \tag{1}$$

The value $d^+ - d^-$ yields the hypothesis margin of the classifier [5]. Then the *generalized* LVQ (GLVQ) is derived as gradient descent of the cost function

$$C_{GLVQ} = \sum_{\mathbf{v}} f(\mu(\mathbf{v})) \tag{2}$$

with respect to the prototypes. f is the sigmoid function

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{3}$$

In one learning step for a given data point, both $\mathbf{w}_{\mathbf{r}^+}$ and $\mathbf{w}_{\mathbf{r}^-}$ are adapted in parallel. Taking the derivative yields the updates

$$\triangle\mathbf{w}_{\mathbf{r}^+} = \epsilon^+ \cdot \operatorname{sgd}'_{\mu(\mathbf{v})} \cdot \xi^+ \cdot \frac{\partial[d(\mathbf{v}, \mathbf{w}_{\mathbf{r}^+})]}{\partial\mathbf{w}_{\mathbf{r}^+}}$$

and

$$\triangle\mathbf{w}_{\mathbf{r}^-} = -\epsilon^- \cdot \operatorname{sgd}'_{\mu(\mathbf{v})} \cdot \xi^- \cdot \frac{\partial[d(\mathbf{v}, \mathbf{w}_{\mathbf{r}^-})]}{\partial\mathbf{w}_{\mathbf{r}^-}}$$

where ϵ^+ and $\epsilon^- \in (0,1)$ are the learning rates. The logistic function $f(x)$ is evaluated at position $\mu(\mathbf{v})$, and we get

$$\xi^+ = \frac{2 \cdot d^-}{(d^+ + d^-)^2}$$

and

$$\xi^- = \frac{2 \cdot d^+}{(d^+ + d^-)^2}.$$

Yet, so far no inherently regularization is involved in the classification model. This feature can be included by combination of GLVQ with an unsupervised neural prototype vector quantizer - the neural gas (NG) [20]. NG realizes a Hebbian learning of prototypes together with regularization by neighborhood cooperativeness between prototypes. The level of cooperativeness is determined in dependence on the similarity of the prototypes to a given data vector: Let $\mathbf{W}$ be the set of prototypes and $L(\mathbf{v}, \mathbf{W})$ be the ordered list of prototype indices such that for each pair $\mathbf{r}_i, \mathbf{r}_k \in L$ with $i < k$ the relation $d\left(\mathbf{v}, \mathbf{w}_{\mathbf{r}_i}\right) \leq d\left(\mathbf{v}, \mathbf{w}_{\mathbf{r}_k}\right)$ holds. Then the position $i(\mathbf{r})$ denotes the rank of the competition of the prototypes to be the best matching for $\mathbf{v}$. The degree of cooperativeness is defined by

$$h_\sigma^{NG}(\mathbf{r}, \mathbf{v}, \mathbf{W}) = \exp\left(\frac{-i(\mathbf{r})}{2\sigma^2}\right)$$

with neighborhood range σ determining the regularization strength. High values σ lead to strong regularization whereas low values relax this restriction [20].

Including this regularization scheme into GLVQ the supervised neural gas (SNG) is obtained [12]. For this purpose, we modify the GLVQ cost function (2) to

$$C_{SNG} = \frac{1}{C\left(\sigma, N_{\mathbf{W}_{\mathbf{c}(\mathbf{v})}}\right)} \sum_{\mathbf{v}} h_\sigma^{NG}\left(\mathbf{r}, \mathbf{v}, \mathbf{W}_{\mathbf{c}(\mathbf{v})}\right) \cdot f\left(\mu^{\mathbf{r}}(\mathbf{v})\right)$$

whereby, $\mathbf{W}_{\mathbf{c}(\mathbf{v})}$ is the subset of all prototypes $\mathbf{w}_\mathbf{r}$ the class labels $\mathbf{y}_\mathbf{r}$ of which are equal to the class label $\mathbf{c}(\mathbf{v})$ of the data point. $C\left(\sigma, N_{\mathbf{W}_{\mathbf{c}(\mathbf{v})}}\right)$ is a constant depending on the cardinality $N_{\mathbf{W}_{\mathbf{c}(\mathbf{v})}}$ of the subset $\mathbf{W}_{\mathbf{c}(\mathbf{v})}$ and the regularization level σ. Derivation of this cost function leads to a similar adaptation scheme as for GLVQ. However, now a neighborhood cooperativeness is included between all prototypes of the correct class:

The update formulas for the prototypes can be obtained taking the derivative. For each $\mathbf{v}$, all prototypes $\mathbf{w}_\mathbf{r} \in \mathbf{W}_{\mathbf{c}(\mathbf{v})}$ are adapted by

$$\triangle \mathbf{w}_\mathbf{r} = \epsilon^+ \cdot \frac{\mathrm{sgd}'|_{\mu^\mathbf{r}(\mathbf{v})} \cdot \xi^+_\mathbf{r} \cdot h^{NG}_\sigma\left(\mathbf{r}, \mathbf{v}, \mathbf{W}_{\mathbf{c}(\mathbf{v})}\right)}{C\left(\sigma, N_{\mathbf{W}_{\mathbf{c}(\mathbf{v})}}\right)} \cdot \frac{\partial\left[d\left(\mathbf{v}, \mathbf{w}_\mathbf{r}\right)\right]}{\partial \mathbf{w}_\mathbf{r}}$$

and the closest wrong prototype is adapted by

$$\triangle \mathbf{w}_{\mathbf{r}^-} = -\epsilon^- \cdot \sum_{\mathbf{w}_\mathbf{r} \in \mathbf{W}_{\mathbf{c}(\mathbf{v})}} \frac{\mathrm{sgd}'|_{\mu^\mathbf{r}(\mathbf{v})} \cdot \xi^-_\mathbf{r} \cdot h^{NG}_\sigma\left(\mathbf{r}, \mathbf{v}, \mathbf{W}_{\mathbf{c}(\mathbf{v})}\right)}{C\left(\sigma, N_{\mathbf{W}_{\mathbf{c}(\mathbf{v})}}\right)} \cdot \frac{\partial\left[d\left(\mathbf{v}, \mathbf{w}_{\mathbf{r}^-}\right)\right]}{\partial \mathbf{w}_{\mathbf{r}^-}}$$

whereby ϵ^+ and $\epsilon^- \in (0,1)$ are learning rates and the logistic function is evaluated at position

$$\mu^\mathbf{r}(\mathbf{v}) = \frac{d_\mathbf{r} - d_{\mathbf{r}^-}}{d_\mathbf{r} + d_{\mathbf{r}^-}}.$$

The terms $\xi^+_\mathbf{r}$ and $\xi^-_\mathbf{r}$ are obtained as

$$\xi^+_\mathbf{r} = \frac{2 \cdot d_{\mathbf{r}^-}}{\left(d_\mathbf{r} + d_{\mathbf{r}^-}\right)^2}$$

and

$$\xi^-_\mathbf{r} = \frac{2 \cdot d_\mathbf{r}}{\left(d_\mathbf{r} + d_{\mathbf{r}^-}\right)^2}.$$

Note that the updates of GLVQ are recovered for vanishing regularization $\sigma \to 0$. We remark that SNG also optimizes the hypothesis margin because the cost function contains the term $d_\mathbf{r} - d_{\mathbf{r}^-}$.

The final classification of unknown data points $\mathbf{v}$ is then realized by a winner take all mapping for both GLVQ and SNG:

$$\mathbf{v} \mapsto c(\mathbf{v}) = \mathbf{y}_\mathbf{r} \text{ such that } d(\mathbf{v}, \mathbf{w}_\mathbf{r}) \text{ is minimum.} \tag{4}$$

It was been demonstrated that SNG/GLVQ achieve excellent classification results [12], [34].

2.2 Semi-supervised fuzzy classification by fuzzy labeled SOM and class similarity detection

2.2.1 The fuzzy labeled SOM - FLSOM

We now turn to the more general task of fuzzy classification and classification visualization. For this purpose we assume that the number N_c of potential classes is known in advance. Then the class label $\mathbf{c}(\mathbf{v})$ is taken as a class membership vector $\mathbf{c}(\mathbf{v}) \in \mathbb{R}^{N_c}$ the elements $c_i(\mathbf{v}) \in [0,1]$ of which describe the fuzzy degree of class membership of the data vector $\mathbf{v}$ and sum up to $\sum_i c_i = 1$. Analogously, the prototype labels are taken as vectors $\mathbf{y}_\mathbf{r} \in \mathbb{R}^{N_c}$. In the following we will extend the unsupervised SOM model to deal with classification tasks.

SOMs are powerful models for unsupervised vector quantization [17]. In SOMs the index set A is a regular grid, usually a rectangular or hexagonal two-dimensional lattice. The indices $\mathbf{r}$ of the prototypes now indicate a location in the grid and, therefore, a natural metric $\|\cdot\|_A$ between them is induced. The mapping is like in SNG/GLVQ again a winner take all rule, which reads in the HESKES' variant of SOMs as

$$\mathbf{v} \mapsto s\left(\mathbf{v}\right) = \underset{\mathbf{r}\in A}{\operatorname{argmin}} \sum_{\mathbf{r}'\in A} h_\sigma(\mathbf{r},\mathbf{r}') \cdot d\left(\mathbf{v},\mathbf{w}_{\mathbf{r}'}\right) \tag{5}$$

with

$$h_\sigma(\mathbf{r},\mathbf{r}') = \exp\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|_A}{2\sigma^2}\right) \tag{6}$$

as neighborhood function [15]. It performs a topographic mapping of data under certain conditions [33], i.e. similar data points are mapped onto the same or onto neighbored grid locations[1]. The degree of topology preservation can be estimated by the *topographic product TP* [2]. TP-values nearby zero indicate adequate topographic mapping. For optimum results the lattice size and dimension can be dynamically adapted during learning [3]. This growing SOM (GSOM) generates a *non-linear* PCA of the data [32].

The learning in the Heskes-SOM follows a gradient descent on a cost function:

$$E_{\mathrm{SOM}} = \frac{1}{2C(\sigma)} \int P(\mathbf{v}) \sum_{\mathbf{r}} \delta_{\mathbf{r}}^{s(\mathbf{v})} \sum_{\mathbf{r}'} h_\sigma(\mathbf{r},\mathbf{r}')\xi(\ \mathbf{v},\mathbf{w}_{\mathbf{r}'})d\mathbf{v} \tag{7}$$

where $C\left(\sigma\right)$ is a constant, which we will drop in the following, and $\delta_{\mathbf{r}}^{\mathbf{r}'}$ is the usual Kronecker symbol checking the identity of $\mathbf{r}$ and $\mathbf{r}'$. All prototypes are adapted according to

$$\triangle\mathbf{w}_\mathbf{r} = -\epsilon h_\sigma\left(\mathbf{r}, s(\mathbf{v})\right) \frac{\partial\xi\left(\mathbf{v},\mathbf{w}_\mathbf{r}\right)}{\partial\mathbf{w}_\mathbf{r}} \tag{8}$$

with learning rate $\epsilon > 0$.

Now we extend the cost function of the SOM as defined in (7) to a cost function for semi-supervised fuzzy classification by

$$E_{\mathrm{FLSOM}} = \left(1-\beta\right) E_{\mathrm{SOM}} + \beta E_{\mathrm{FL}} \tag{9}$$

where E_{FL} measures the classification accuracy . The factor $\beta \in [0,1]$ is a factor balancing unsupervised and supervised learning. One can simply choose $\beta = 0.5$, for example. We choose

$$E_{\mathrm{FL}} = \frac{1}{2} \int P\left(\mathbf{v}\right) \sum_{\mathbf{r}} g_\gamma\left(\mathbf{v},\mathbf{w}_\mathbf{r}\right) \left(\mathbf{x}-\mathbf{y}_\mathbf{r}\right)^2 d\mathbf{v} \tag{10}$$

where $g_\gamma\left(\mathbf{v},\mathbf{w}_\mathbf{r}\right)$ is a Gaussian kernel describing a neighborhood range in the data space:

$$g_\gamma\left(\mathbf{v},\mathbf{w}_\mathbf{r}\right) = \exp\left(-\frac{d\left(\mathbf{v},\mathbf{w}_\mathbf{r}\right)}{2\gamma^2}\right). \tag{11}$$

This choice is based on the assumption that data points close to the prototype determine the corresponding label if the underlying classification is sufficiently smooth. Note that

[1] For a detailed discussion of topographic mapping and more general lattice structures we refer to [3],[33].

$g_\gamma(\mathbf{v},\mathbf{w_r})$ depends on the prototype locations, such that E_{FL} is influenced by both $\mathbf{w_r}$ and $\mathbf{y_r}$. Hence, prototype adaptation is now influenced by the classification task via the labels:

$$\frac{\partial E_{\mathrm{FLSOM}}}{\partial \mathbf{w_r}} = \frac{\partial E_{\mathrm{SOM}}}{\partial \mathbf{w_r}} + \frac{\partial E_{\mathrm{FL}}}{\partial \mathbf{w_r}} \tag{12}$$

which yields

$$\begin{aligned} \triangle \mathbf{w_r} &= -\epsilon(1-\beta)\cdot h_\sigma(\mathbf{r},s(\mathbf{v}))\frac{\partial d(\mathbf{v},\mathbf{w_r})}{\partial \mathbf{w_r}} \\ &\quad +\epsilon\beta\frac{1}{4\gamma^2}\cdot g_\gamma(\mathbf{v},\mathbf{w_r})\frac{\partial d(\mathbf{v},\mathbf{w_r})}{\partial \mathbf{w_r}}(\mathbf{x}-\mathbf{y_r})^2 . \end{aligned} \tag{13}$$

The label adaptation is only influenced by the second part E_{FL}. The derivative $\frac{\partial E_{\mathrm{FL}}}{\partial \mathbf{y_r}}$ yields

$$\triangle \mathbf{y_r} = \epsilon_l\beta\cdot g_\gamma(\mathbf{v},\mathbf{w_r})(\mathbf{x}-\mathbf{y_r}) \tag{14}$$

with learning rate $\epsilon_l > 0$. This label learning performs to a weighted average of the data fuzzy labels of those data close to the associated prototypes.

Classification of unknown data is again obtained by the mapping rule (4), but now giving a fuzzy class membership vector response. Usually, the classification accuracy of FLSOM is slightly less then the accuracy of a pure (good) classifier, because the balancing parameter β cannot be set to the unit due to numerical stability reasons [36]. Hence, a remaining unsupervised ammount of data information may lead to reduced accuracy. However, this disadvantage is compensated by the feature of inherent class similarity detection and the visualization abilities of FLSOM [37].

2.2.2 Class visualization and class similarity detection

As mentioned above, unsupervised SOMs generate a topographic mapping from the data space onto the prototype grid A under specific conditions. If the classes are consistently determined with respect to the varying data in a classification problem, one can expect for the semi-supervised topographic FLSOM that the class labels $\mathbf{y_r}$ become ordered within the distribution over the grid structure of the lattice A. In this case the topological order of the prototypes should be transferred to the topological order of prototype labels, such that we have a smooth change between the fuzzy class label vectors within the neighbored of the considered grid locations. This is the consequence of the following fact: the neighborhood function $h_\sigma(\mathbf{r},\mathbf{s})$ of the usual SOM learning (8) forces the topological ordering of the prototypes. In FLSOM, this ordering is further influenced by the weighted classification error

$$ce(\mathbf{v},\mathbf{r}) = g_\gamma(\mathbf{v},\mathbf{w_r})(\mathbf{x}-\mathbf{y_r})^2 , \tag{15}$$

which contains the data space neighborhood $g_\gamma(\mathbf{v},\mathbf{w_r})$, eq. (11). Hence, the prototype ordering contains information of both data density and class distribution, whereby for high balancing value β the latter term becomes dominant. Otherwise, the data space neighborhood $g_\gamma(\mathbf{v},\mathbf{w_r})$ also triggers the label learning (14), which, of course, also dependents on the underlying learned prototype distribution and ordering. Thus, a consistent ordering of the labels is obtained in FLSOM [35].

As a consequence, the evaluation of the similarities between the prototype label vectors yields suggestions for the *similarity of classes*, i.e. similar classes are represented by prototypes in a local spatial area of the FLSOM lattice A. In case of overlapping class distributions the topographic processing leads to prototypes with unclear decision (labels), located between prototypes with clear vote. Further, if classes are not distinguish-able, there will exist prototypes responsive to those data, which have class label vectors containing approximately the same degree of fuzzy class membership for the respective classes.

The fuzzy class membership vectors allow an easy visualization of the classification using their similarity property. For this purpose, all label vectors $\mathbf{y}_\mathbf{r}$, $\mathbf{r} \in A$ are embedded into a color space preserving their similarities. This can be realized by multi-dimensional scaling (MDS), for example. Doing so, similar classes are coded by similar colors, which may be used for *class visualization* [37].

2.3 Classification task dependent metric adaptation

The dissimilarity measure $d(\mathbf{v}, \mathbf{w}_\mathbf{r})$ for the data space V is usually chosen as squared Euclidean metric in GLVQ, SNG and FLSOM. Thus the derivative $\frac{\partial d(\mathbf{v},\mathbf{w})}{\partial \mathbf{w}}$ simply becomes $-2(\mathbf{v} - \mathbf{w})$. Depending on the classification task, this choice could be not optimum. Therefore, more appropriate (differentiable) similarity measures can be plugged into these algorithms instead, reflecting the nature of data or structure of classification. For example, LEE&VERLEYSEN proposed a *functional metric* derived from the general *Minkowski-metric* for functional data paying attention to the spatial correlation between the components of functional vectors [18]. Other example, frequently used in biological and biochemical problems, are the Pearson correlation [28] and the Tanimoto kernel [29]. Due to the general formulation of the methods above, these metrics can easily plugged into the algorithms.

Yet, more flexibility can be obtained if $d(\mathbf{v}, \mathbf{w}_\mathbf{r})$ is a parametrized similarity measure. Then, the respective parameters may be also subject of optimization according to the given classification task [13],[12].

Generally, we consider a parametrized distance measure $d^\lambda(\mathbf{v}, \mathbf{w})$ with a parameter vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)$ with $\lambda_i \geq 0$ and normalization $\sum_{i=1}^M \lambda_i = 1$. Then, a classification task depending parameter optimization is achieved again by a gradient descent of the above cost functions but here with respect to these metric parameters.

One important example of a parametrized metric is the *scaled* squared Euclidean metric

$$d^\lambda(\mathbf{v}, \mathbf{w}) = \sum_i \lambda_i (v_i - w_i)^2 \tag{16}$$

(with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$). The derivative $\frac{\partial d^\lambda(\mathbf{v},\mathbf{w})}{\partial \mathbf{w}}$ becomes $= -2 \cdot \boldsymbol{\Lambda} \cdot (\mathbf{v} - \mathbf{w})$ with $\boldsymbol{\Lambda}$ is a diagonal matrix and its i-th diagonal entry is λ_i and $\frac{\partial d^\lambda(\mathbf{v},\mathbf{w})}{\partial \boldsymbol{\lambda}_i} = (v_i - w_i)^2$.

The parameter optimization of the *scaled* squared Euclidean metric allows a useful interpretation. The parameter λ_i weight the dimensions of the data space. Hence, optimization of these parameters in dependence on the classification problem leads to a ranking of the input dimensions according to their classification decision relevance. Therefore, metric parameter adaptation of the scaled Euclidean metric is called *relevance learning* [13]. This weighting is structurally similar to the weighting in FDA. In case of zero-valued λ_i relevance learning can also be seen as feature selection. The vector $\boldsymbol{\lambda}$ is called *relevance* profile. It can be used for advanced data investigation as it is shown in the applications.

3 Application of GRLVQ and FLSOM for clinical data sets

In this section we demonstrate the application of both, SNG and FLSOM, to classification of two spectrometric data sets in bioinformatics. The problems are characteristic for bioinformatic tasks and therefore exemplary:

1. identification of bacteria

2. breast cancer tissue slice classification

For both problems, the data metric for FLSOM was chosen as the squared scaled Euclidian metric (16).

3.1 Description of data and preprocessing

The data for both problems are based on mass-spectrometric profiles measured by linear MALDI-TOF MS devices from Bruker Daltonik, Bremen, Germany.

3.1.1 The bacteria data set

The bacteria samples are obtained from extracts from listeria cell cultures (original culture stems from the German Resource Center for Biological Material – DSMZ). The extracts, covered by a HCCA matrix, have been applied onto the MSP 96 target ground steel [19]. Profiling spectra were generated on a linear *Autoflex* MALDI-TOF MS. Details can be found in [16].

Listeria is a bacterial genus containing six species. This species consist of *listeria monocytogenes*, *listeria innocua*, *listeria ivanovii*, *listeria seeligeri*, *listeria welshimeri* and *listeria grayi*. Listeria occur very common in nature environments and are also present in water, plants, food and the bowel of humans. The identification of listeria is therefore an important problem in biology. Listeria are known to be the bacteria responsible for listeriosis, a rare but lethal food-borne infection that has a devastating mortality rate of 25% [30]. Listeria, also has a particularly high occurrence rate in newborns because of its ability to infect the fetus by penetrating the endothelial layer of the placenta [30]. Thereby Listeria monocytogenes is considered to be pathogenic for humans. Listeria in food are relatively rare but due to the increasing industrial production of food with many processing steps, the risk of a listerial contamination is increasing, which rises the needs for improved product safety and quality control. The diagnosis of listeria at an early stage is important for therapeutic approaches on humans. The expression of a infection caused by listeria may delay upto 8 weeks. To identify whether a listeria infection is present, the blood or matter is taken from the patient and a cultivation is tried. This, however, fails in part and, hence, the disease can not be diagnosed in time.

In the available data set all six listeria types are present. Thereby, for the listeria grayi a subgroup of *listeria grayi murrei* can be identified and for listeria ivanovii a distinction into the subgroups *listeria ivanovii ssp ivanovii* and *listeria ivanovii ssp londoniensis* can be made. Thus, the data set consists of 109 profile spectra in 8 classes with at least 6 samples for each class. The spectral range is between 2kDa and 20kDa. The obtained spectra have been smoothed, baseline corrected and, aligned following the standardized preprocessing tool BIOTYPER$^{\mathrm{TM}}$ 1.1 from Bruker Daltonik, Bremen, Germany. The involved peak picking generates a peak list vector for each spectrum. All peak list vectors are aggregated such that finally a data matrix with 937 intensity components and 109 rows is achieved as data base. Thereby, a peak shift tolerance of 300ppm was used. A classification tree obtained by hierarchical clustering using the BIOTYPER$^{\mathrm{TM}}$ 1.1 software yields a class separation tree as depicted in Fig.1. This tree can serve for comparison for FLSOM class similarity detection.

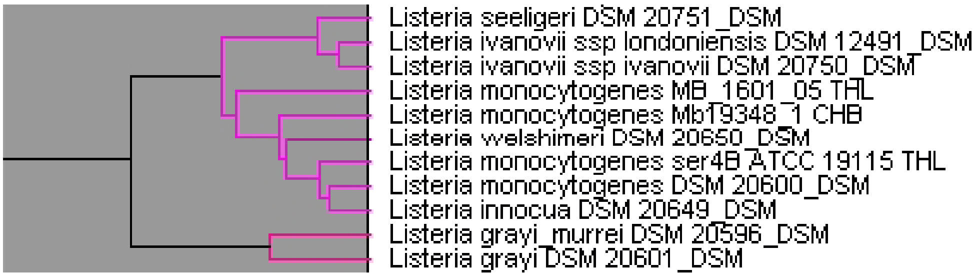
Listeria seeligeri DSM 20751_DSM
Listeria ivanovii ssp londoniensis DSM 12491_DSM
Listeria ivanovii ssp ivanovii DSM 20750_DSM
Listeria monocytogenes MB_1601_05 THL
Listeria monocytogenes Mb19348_1 CHB
Listeria welshimeri DSM 20650_DSM
Listeria monocytogenes ser4B ATCC 19115 THL
Listeria monocytogenes DSM 20600_DSM
Listeria innocua DSM 20649_DSM
Listeria grayi_murrei DSM 20596_DSM
Listeria grayi DSM 20601_DSM

Figure 1: Separation tree of different listeria types obtained by BIOTYPER$^{\mathrm{TM}}$ 1.1 software based on hierarchical clustering.

3.1.2 The breast cancer tissue data set

The breast cancer tissues are collected at the Institute of Pathology in Neuherberg, GSF-National Research Center for Environment an Health, Germany. The generic measurement procedure can be summarized as follows. The frozen tissue is cut using a cryomicrotome in sections of 12μm and transferred to a conductive slide, washed in ethanol and coated by matrix. Reference sections are used for histological staining (Hematoxylin&Eosin, immunohistochemical staining for HER2) and histomorphological classification. The slices are subsequently measured in a *Ultraflex II* MALDI-TOF and subsequently visualized using the FlexImaging$^{\mathrm{TM}}$ tool provided by Bruker Daltonik, Bremen, Germany [11].

The breast cancer tissue slices are manually labeled by a clinical expert (pathologist). In this exemplary study the slice of one patient is used. Four different spatial regions of the slice are marked according to histomorphologically classfied cell types: connective tissue, inflammation, and two moorphologically distinct tumor cell populations (tumor-type-1, tumor-type-2). From the whole slice 687 spectral record are generated, 438 of them are labeled with at least 51 records per region.

These profiles were preprocessed (baseline correction, alignment, peak picking and, peak feature extraction by means of maximum intensities) according to the CLINPROTOOLS$^{\mathrm{TM}}$ 2.1 from Bruker Daltonik, Bremen, Germany, see [10]. Finally, each preprocessed data record is a 70-dimensional data vector.

3.2 Application of the methods and interpretation of the results

Both data sets are analyzed by SNG and FLSOM using the scaled quadratic Euclidean metric as data similarity measure. For comparison we also applied SVM with different kernels and LDA based on linear PCA. Additionally, for the bacteria data set a *class dependence tree* provided as standard solution of the BIOTYPER$^{\mathrm{TM}}$ 1.1 tool based on hierarchical clustering is also available for comparison [19].

LDA	SVM1	SVM2	SVM3	SNG	FLSOM
61.5%	34.9%	< 10%	96.3%	97.8%	73.4%

Table 1: Classification accuracies for the different classifiers for the listeria data set. SVM_1 is a linear SVM, SVM_2 uses a radial basis function kernel and SVM_3 uses a Tanimoto-distance-kernel.LDA is based on linear PCA suggesting 5 principal components to be sufficient. For FLSOM majority vote is taken to obtain the crisp classification.

In case of FLSOM application we further investigate the detected class similarities and provide visualization results. The optimum FLSOM lattice size was estimated by a GSOM using standard Euclidean metric for data.

3.2.1 Results for the bacteria data set

First, we applied SNG with 3 prototypes per class and the neighborhood range σ for regularization is slowly decreased to zero during the adaptation process. For FLSOM a two-dimensional 12 × 4 lattice structure is suggested by the GSOM. Due to the large number of prototypes for FLSOM in comparison to SNG and paying attention to the sparse data the final regularization parameter for FLSOM is set to $\sigma = 0.4$, which yields non-vanishing regularization. The balancing parameter was set to $\beta = 0.05$ in the beginning and increasung up to the final value of $\beta = 0.85$. The topology preservation of the FLSOM is preserved, as the topographic product value $TP = -0.0066$ indicates. The 5-fold cross-validated classification accuracy results are collected in Tab. 1. Both, SNG and FLSOM show very good performance in comparison to the other algorithms. In particular, we remark that SVM heavily depends on the used kernel, as it is also known from other applications [24]. The slightly decreased accuracy of the FLSOM is the consequence of the balancing parameter $\beta < 1.0$, which is necessary for stability reasons of the algorithm as described before. However, this disadvantage is compensated by the class similarity detection feature provided by FLSOM [35]. These results are now under deeper consideration. The fuzzy class label vectors $\mathbf{y_r}$ of the prototypes are depicted in Fig. 2a) according to their distribution with respect to the FLSOM lattice. This distribution of the prototype labels suggests the following interpretation of FLSOM-detected class similarities, which should also be compared to the above given separation tree obtained by the $\text{BIOTYPER}^{\text{TM}}$ 1.1 software depicted in Fig 1: The listeria of grayi types (class 1 & 2) should not be distinguished according to their proteom finger print. The class 8 (listeria welshimeri) is clearly isolated from each other. Classes 4, 5 and 7 (listeria ivanovii ssp ivanovii, listeria seeligeri and listeria ivanovii ssp londoniensis) are very similar. Further there is a class similarity between the classes 3 and 4 (listeria innocua and listeria ivanovii ssp ivanovii). Although this similarity in the tree classification can not be ruled out, the tree visualization suggests a stronger separation. This 'separation' would be disappear, if the respective branch would be rotated. Thus, the FLSOM label distribution is more adequate. Further, FLSOM detected a similarity between the classes 6 and 4 (listeria monocytogenes and listeria ivanovii ssp ivanovii), which is also not easily detectable in the tree visualization. The class 3 (listeria innocua) shows multiple similarities to several other species based on the proteom finger print. This sharing property can not be reflected adequately in the tree classification. However, an expert biologist independently suggested a similarity between both types[2].

[2] Personal communication with Dr. Thomas Maier, BRUKER Daltonik Leipzig, Germany.

LDA	SVM1	SVM2	SVM3	SNG	FLSOM
59.8%	62.8%	42.7%	84.2%	80.4%	72.4%

Table 2: Classification accuracies for the different classifiers for the breast cancer data set. SVM_1 is a linear SVM, SVM_2 uses a radial basis function kernel and SVM_3 uses a Tanimoto-distance-kernel.LDA is based on linear PCA suggesting 8 principal components to be sufficient. For FLSOM majority vote is taken to obtain the crisp classification.

Thus summerizing, the FLSOM provides detailed class similarity description together with comparable classification accuracy, whereas SNG achieves best accuracy.

3.2.2 Results for the breast cancer tissue data set

Again, we applied SNG with 3 prototypes per class and the neighborhood range σ for regularization also slowly decreasing to zero during the adaptation process. For FLSOM a two-dimensional 15×5 lattice structure is suggested here by the GSOM. As for the listeria data set the final regularization parameter for FLSOM is set to $\sigma = 0.4$, which yields non-vanishing regularization. The balancing parameter setting was the same as for the bacteria data set. The topographic product value $TP = 0.0001$ ensures the topographic mapping of the FLSOM. The 5-fold cross-validated classification accuracy results are collected in Tab. 2.

The obtained accuracies for SNG and FLSOM show high levels with a slightly decreased value for FLSOM, whereas SNG achieves the overall best result.

For class similarity investigation we consider the distribution of the label vectors within the FLSOM-grid, which is depicted in Fig. 2. The connective-tissue class is well separated. Further, we have in the distribution plane an overlapping region between the both tumor classes, which indicates a similarity between them. The FLSOM detects a clear distinction between tumor-1-class and the connective tissue class according to the spatial distribution of the respective labels in the FLSOM grid, whereas small overlapping between tumor-2-class and the connective-tissue class occurs. The inflammation class shows similarity to type-2-tumor. Using an MDS color embedding of all label vectors $\mathbf{y_r}$ of the FLSOM into the RGB-color space, the FLSOM classification of the tissue can be easily visualized, see Fig 3. A high agreement between original manually labeled tissue and obtained coloring based on the classification and class similarity detection generated by the FLSOM can be observed.

Further information can be obtained by consideration of the learned relevance profile of the problem specific scaled Euclidian metric. It is depicted in Fig.4. The highest relevance for class separation can be assigned to the 4971Da-peak in the original proteomic spectra. Recoloring of the original tissue according to the 4971Da-intensities shows that this peak mainly separates the connective tissue class from the other classes, see Fig 5. Analogously, the other relevance peaks could be evaluated. In this way, a detailed analysis of the information contained in the FLSOM model can be obtained.

4 Concluding remarks

In this article two recently developed prototype based methods for classification, SNG and FLSOM, are reviewed in the light of the analysis of mass spectrometric data in bioninformatics. Both approaches are adaptive machine learning approaches and allow easy retraining, if new data become available. They are both inherently regularizing, such that they are able to handle sparse, high-dimensional and noisy data. As demonstrated for two exemplary problems in classification of proteomic spectra (bacteria and breast cancer tissue), the generated classification models show good performance compared to other machine learning and statistical methods.

Additionally, FLSOM provides the possibility of processing uncertain class information for training data (fuzzy) and returns a fuzzy classification scheme. Moreover, FLSOM provides a class similarity detection based on the fuzzy labels, which give the possibility of deeper class analysis offering more information than simple classification trees. The fuzzy classification can further be used for class dependent data visualization whereby similar class information is encoded by similar colors such that an easy interpretation can be made.

References

[1] P. Baldi and S. Brunak. *Bioinformatics – The Machine Learning Approach.* The MIT Press, 1998.

[2] H.-U. Bauer and K. R. Pawelzik. Quantifying the neighborhood preservation of Self-Organizing Feature Maps. *IEEE Trans. on Neural Networks*, 3(4):570–579, 1992.

[3] H.-U. Bauer and T. Villmann. Growing a Hypercubical Output Space in a Self–Organizing Feature Map. *IEEE Transactions on Neural Networks*, 8(2):218–226, 1997.

[4] C. Bishop. *Pattern Recognition and Machine Learning.* Springer Science+Business Media, LLC, New York, NY, 2006.

[5] K. Crammer, R. Gilad-Bachrach, A.Navot, and A.Tishby. Margin analysis of the LVQ algorithm. In *Proc. NIPS 2002*, http://www-2.cs.cmu.edu/Groups/NIPS/NIPS2002/NIPS2002preproceedings/index.html, 2002.

[6] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge University Press, 2000.

[7] M. de Noo, A. Deelder, M. van der Werff, A. Özalp, and B. Martens. MALDI-TOF serum protein profiling for detection of breast cancer. *Onkologie*, 29:501–506, 2006.

[8] M. de Noo, B. Martens, A. Özalp, M. Bladergroen, M. van der Werff, C. van de Velde A. Deelder, and R. Tollenaar. Detecting of colorectal cancer using MALDI-TOF serum protein profiling. *European Journal of Cancer*, 42:1068–1076, 2006.

[9] R. Duda and P. Hart. *Pattern Classification and Scene Analysis.* Wiley, New York, 1973.

[10] M. Gerhard, S.-O. Deininger, and F.-M. Schleif. Statistical classification and visualization of MALDI-imaging data. In P. Kokol, M. Zorman, V. Podgerelec, M. Verlic, and D. Micetic-Turk, editors, *Proceedings of the 20th IEEE Symposium on Computer-based Medical Systems*, pages 403–405. IEEE Press, 2007.

[11] M. Groseclose, M. Andersson, W. Hardesty, and R. Caprioli. Identifications of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry. *Journal of Mass Spectrometry*, 42:254–262, 2007.

[12] B. Hammer, M. Strickert, and T. Villmann. Supervised neural gas with general similarity measure. *Neural Processing Letters*, 21(1):21–44, 2005.

[13] B. Hammer and T. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15(8-9):1059–1068, 2002.

[14] B. Hammer and T. Villmann. Classification using non-standard metrics. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2005)*, pages 303–316, Brussels, Belgium, 2005. d-side publications.

[15] T. Heskes. Energy functions for self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 303–316. Elsevier, Amsterdam, 1999.

[16] R. Ketterlinus, S.-Y. Hsieh, S.-H. Teng, H. Lee, and W. Pusch. Fishing for biomarkers: analyzing mass spectrometry data with the new clinprotools software. *Bio techniques*, 38(6):37–40, 2005.

[17] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[18] J. Lee and M. Verleysen. Generalization of the l_p norm for time series and its application to self-organizing maps. In M. Cottrell, editor, *Proc. of Workshop on Self-Organizing Maps (WSOM) 2005*, pages 733–740, Paris, Sorbonne, 2005.

[19] T. Maier and M. Kostrzewa. Fast and reliable MALDI-TOF MS-based microorganism identification. *Chemistry Today*, 25(2):68–71, 2007.

[20] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten. 'Neural-gas' network for vector quantization and its application to time-series prediction. *IEEE Trans. on Neural Networks*, 4(4):558–569, 1993.

[21] A. Sato and K. Yamada. Generalized learning vector quantization. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8. Proceedings of the 1995 Conference*, pages 423–9. MIT Press, Cambridge, MA, USA, 1996.

[22] F.-M. Schleif, T. Elssner, M. Kostrzewa, T. Villmann, and B. Hammer. Analysis and visualization of proteomic data by fuzzy labeled self-organizing maps. In D. Lee, B. Nutter, S. Antani, S. Mitra, and J. Archibald, editors, *19th IEEE International Symposium on Computer- based Medical Systems Salt Lake City (CBMS)*, pages 919–924. IEEE Computer Society Press, Los Alamitos, 2006. 0769525171.

[23] J. Schürmann. *Pattern Classification.* J. Wiley and Sons Inc., New York, 1996.

[24] B. Schölkopf and A. Smola. *Learning with Kernels.* MIT Press, 2002.

[25] U. Seiffert, B. Hammer, S. Kaski, and T. Villmann. Neural networks and machine learning in bioinformatics - theory and applications. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2006)*, pages 521–532, Brussels, Belgium, 2006. d-side publications.

[26] U. Seiffert, L. C. Jain, and P. Schweizer. *Bioinformatics using Computational Intelligence Paradigms.* Springer-Verlag, 2004.

[27] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis and Discovery.* Cambridge University Press, 2004.

[28] M. Strickert, U. Seiffert, N. Sreenivasulu, W. Weschke, T. Villmann, and B. Hammer. Generalized relevance LVQ (GRLVQ) with correlation measures for gene expression analysis. *Neurocomputing*, 69(6–7):651–659, March 2006. ISSN: 0925-2312.

[29] S. Swamidass and P. Baldi. Bounds and algorithms for fast exact searches of chemical fingerprints in linear and sublinear time. *Journal of Chemical Information and Modeling*, 47(2):302–317, 2007.

[30] K. Todar. Todar's online textbook of bacteriology – listeria monocytogenes and listerisis. Univsersity of Wisconsin-Madion Department of Biology, http://textbookofbacteriology.net/Listeria.html, 2003.

[31] M. Verleysen and D. François. The curse of dimensionality in data mining and time series prediction. In J. Cabestany, A. Prieto, and F. S. Hernández, editors, *Computational Intelligence and Bioinspired Systems, Proceedings of the 8th International Work-Conference on Artificial Neural Networks 2005 (IWANN), Barcelona.*

[32] T. Villmann and H.-U. Bauer. Applications of the growing self-organizing map. *Neurocomputing*, 21(1-3):91–100, 1998.

[33] T. Villmann, R. Der, M. Herrmann, and T. Martinetz. Topology Preservation in Self–Organizing Feature Maps: Exact Definition and Measurement. *IEEE Transactions on Neural Networks*, 8(2):256–266, 1997.

[34] T. Villmann, F.-M. Schleif, and B. Hammer. Comparison of relevance learning vector quantization with other metric adaptive classification methods. *Neural Networks*, 19:610–622, 2006.

[35] T. Villmann, F.-M. Schleif, E. Merényi, M. Strickert, and B. Hammer. Class imaging of hyperspectral satellite remote sensing data using flsom. In H. Ritter, editor, *Proc. Workshop on Self-Organizing Maps WSOM*, page in press. Bielefeld, Germany, 2007.

[36] T. Villmann, U. Seiffert, F.-M. Schleif, C. Brüß, T.Geweniger, and B. Hammer. Fuzzy labeled self-organizing map with label-adjusted prototypes. In F. Schwenker and S. Marinai, editors, *Proceedings of Conference Artificial Neural Networks in Pattern Recognition (ANNPR) 2006, Ulm, Germany*, LNAI 4087, pages 46–56. Springer Verlag, 2006.

[37] T. Villmann, M. Strickert, C. Brüß, F.-M. Schleif, and U. Seiffert. Visualization of fuzzy information in fuzzy-classification for image segmentation using MDS. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2007)*, pages 103–108, Brussels, Belgium, 2007. d-side publications.

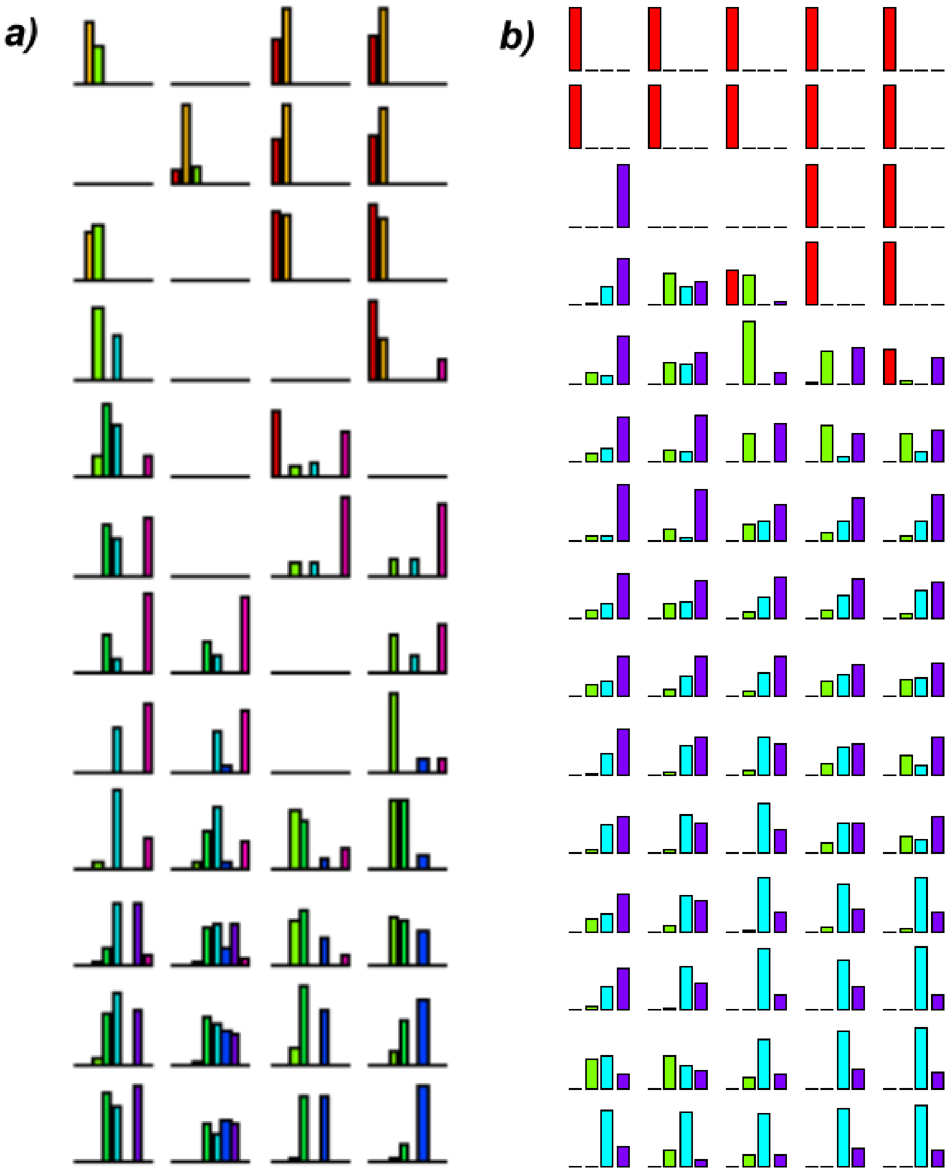

Figure 2: Distribution of the class responsibilities within the FLSOM-lattices for a) the listeria classification problem and b) the breast cancer problem. The label vectors $\mathbf{y}_\mathbf{r}$ are depicted as barplots arranged according to the FLSOM-grid structure. Each barplot refers to a label vector, whereby the height of the bars within is according to the probability that the prototype is responsible for the respective class (left class 1 – right class 8). The coloring of the bars is only for better visualization and does not contain any information. Flat lines show 'dead' prototypes, i.e. which did not won the competing process for the available data.

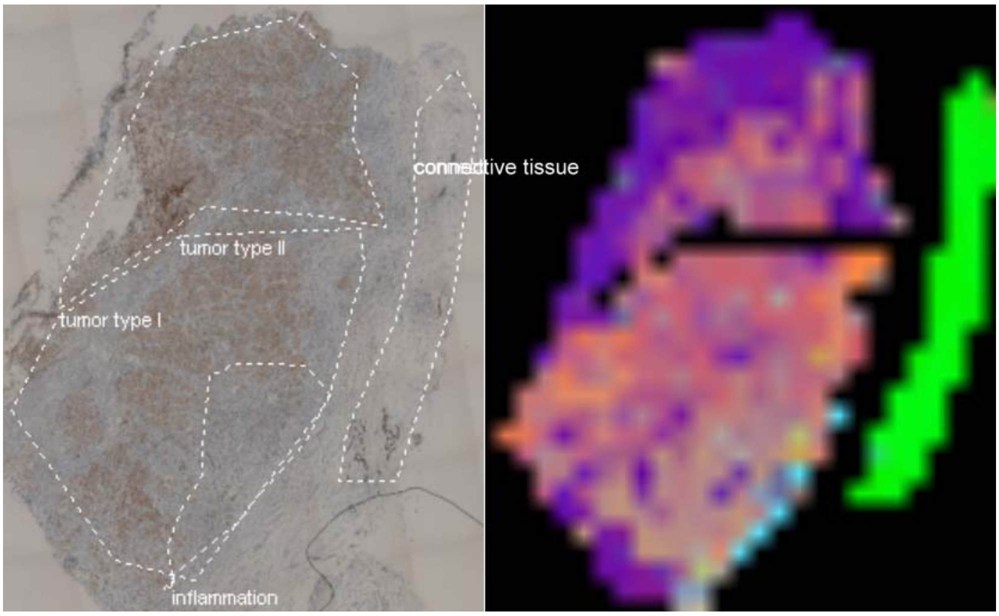

Figure 3: Breast cancer tissue section with manually labeled areas used for classification training is depicted left. Right hand, the classification obtained by the FLSOM classifier is plotted using an MDS RGB-color embedding of the FLSOM-label vectors $\mathbf{y_r}$. Thereby, similar colors represent similar class properties as detected by FLSOM (black - not used for classification). One clearly see the fine agreement with the manual labelling.

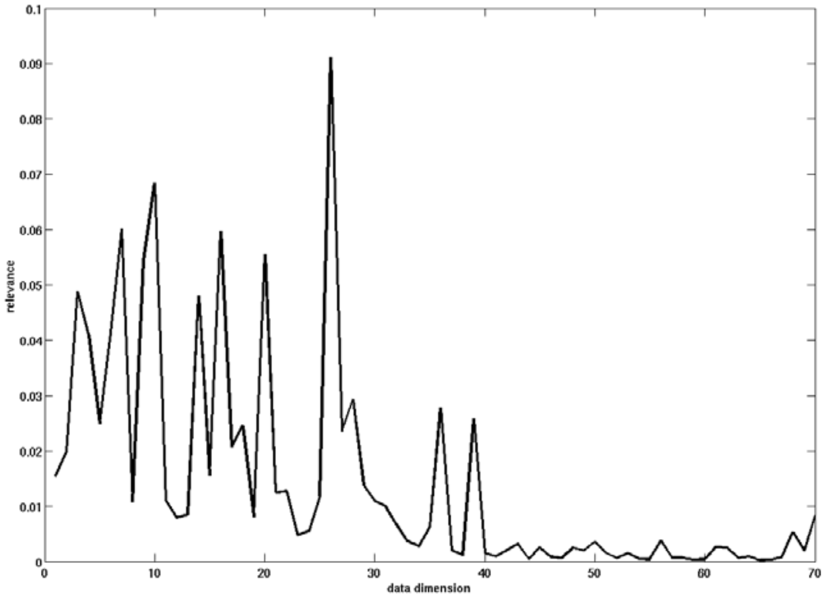

Figure 4: Relevance profile of the scaled Euclidian metric for the breast cancer problem. The highest relevance peak can be assigned to the data dimension 26 which is assigned to the 4971Da-peak in the original proteomic spectrum.

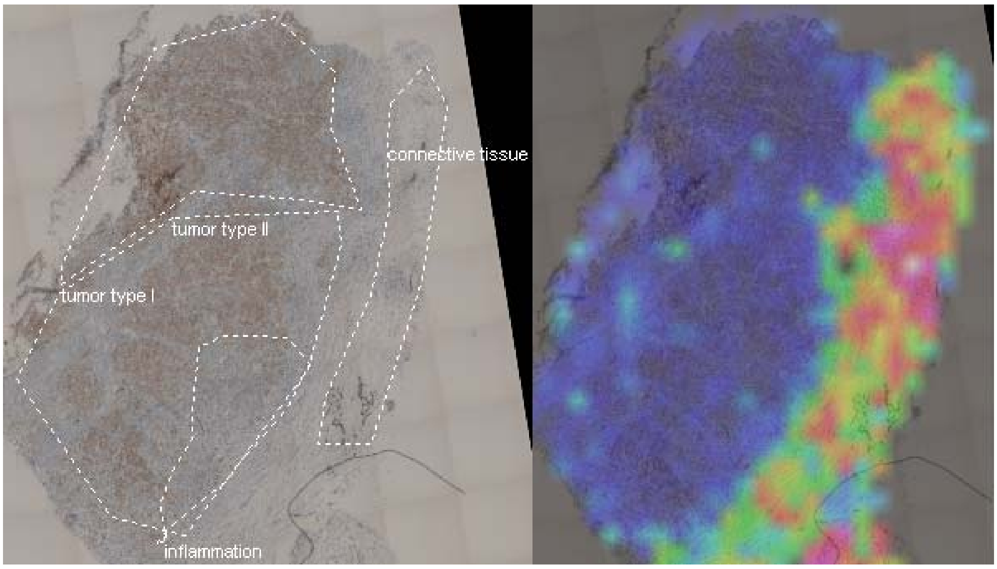

Figure 5: Recoloring of the original tissue according to the 4971Da-intensities. High intensities are colored red, low values are coded by blue colors.