ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Extending FSNPC to handle data points with fuzzy class assignments

Tina Geweniger[1] and Thomas Villmann[2]

1 - University of Leipzig - Working Group Computational Intelligence
Semmelweisstrasse 10, 04103 Leipzig - Germany

2 - University of Applied Sciences Mittweida
Department of Mathematics/Physics/Informatics, Computational Intelligence Group
Technikumplatz 17, 09648 Mittweida - Germany

Abstract. In this paper we present an advanced Nearest Prototype Classification to handle data points with unsharp class assignments. Therefore we extend the Soft Nearest Prototype Classification as proposed by Seo et al. and its further enhancement working with fuzzy labeled prototypes as introduced by Villmann et al. We adapt the cost function and derive appropriate update rules for the prototypes. We assess the performance on a toy data set and a real-world problem and compare the classification result with the results obtained by Fuzzy Robust Soft LVQ by means of Fuzzy Cohen's Kappa.

1 Introduction

Nearest prototype classification (NPC) [4] is a method to adapt a set of class dependent prototypes. They are positioned to optimize the classification of the data points according to their distances. A well known and widely used learning scheme using nearest prototype classification is the Learning Vector Quantization (LVQ) as introduced by Kohonen [4]. The original version has been the basis for a whole family of supervised learning algorithms like LVQ 2.1 [4], GLVQ [6], SLVQ [8], and RSLVQ [8] to name just a few. Seo et al. developed a soft version for the nearest prototype classification called SNPC which uses the Gaussian mixture approach to model soft assignments of the data points to their representing prototypes [7]. This method is a stochastic gradient descent on a cost function incorporating the probability density of the data points and can be interpreted as an annealed version of LVQ.

In NPC the prototypes realize only a crisp classification because of their unique class dependence. This has the disadvantage that overlapping classes can not be described appropriately. Therefore Villmann et al. established a new learning scheme called FSNPC which utilizes fuzzy prototype vectors [9]. The FSNPC is also based on the Gaussian mixture model.

Yet all these algorithms require crisp labeled training data. But there are various applications, for example in the medical or biological field like cancer identification based on tissue samples or the identification of barley grain tissue as described in the experimental section, where there is only a diffuse classification possible. And sometimes the training data can only be obtained by insecure methods like the manual subjective evaluation of specimen or other findings. Recently an approach called FRSLVQ for handling this type of fuzzy labeled data was introduced in [3]. This lack of alternative learning schemes motivated us to find another way to obtain prototypes for fuzzy data. Therefore we further extended the FSNPC to handle unsharp labeled data points implying a level of uncertainty within the data set itself. We developed an appropriate learning scheme based on a Gaussian mixture model and fuzzy prototypes.

We demonstrate the ability of the resulting learning scheme on a 2-dimensional toy data set consisting of two overlapping Gaussian distributions and on a real world problem classifying barley grain tissue samples. Assessing the performance of our algorithm requires a method to compare fuzzy classifiers with each other, which can be done using Fuzzy Cohen's Kappa [11].

2 FSNPC - a short review

Since our proposed method is an extension of the FSNPC as introduced by [9], we will give a short review of this algorithm. FSNPC itself is based on the Soft Nearest Prototype Classification by Seo et al. [7]. While SNPC works with crisp class information for the prototypes, FSNPC is working with fuzzy labeled prototypes.

Considering a set of N training data points $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ and a set of M initially unknown labeled prototypes $\mathcal{T} = \{(\boldsymbol{w}_j, \boldsymbol{c}_j)\}_{j=1}^M$ with $\boldsymbol{x}_i, \boldsymbol{w}_j \in \mathbb{R}^D$. y_i is the assigned class label for data point $\boldsymbol{x}_i$ with $y_i \in \mathcal{I}$ and $\boldsymbol{c}_j$ is the fuzzy label of prototype $\boldsymbol{w}_j$, which indicates the proportionate responsibility of $\boldsymbol{w}_j$ to all classes C with $\sum_{l=1}^C \boldsymbol{c}_j^l = 1$ and $\boldsymbol{c}_j^l \geq 0$. During training these labels have to be adjusted automatically.

The aim of FSNPC is to minimize the cost function

$$E(\mathcal{T}, \mathcal{S}) = \frac{1}{N} \sum_{k=1}^N lc((\boldsymbol{x}_k, y_k), \mathcal{T}) \tag{1}$$

where the local costs are given by

$$lc(\boldsymbol{x}_k, y_k) = \sum_{j=1}^M P(j|\boldsymbol{x}_k)(1 - \boldsymbol{c}_j^{y_k}). \tag{2}$$

Since the crisp class information for the prototypes, which is required for SNPC learning, is no longer available, a corresponding learning scheme has been derived by [9] as

$$\Delta \boldsymbol{w}_l = -\frac{\alpha_w}{2\sigma^2} P(l|\boldsymbol{x})(1 - \boldsymbol{c}_l - lc_t) \frac{\partial(\boldsymbol{x}, \boldsymbol{w}_l)}{\partial(\boldsymbol{w}_l)} \tag{3}$$

with learning rate α_w. The posterior probability

$$P(j|\boldsymbol{x}) = \frac{\exp\left(-d(\boldsymbol{x}, \boldsymbol{w}_j)/(2\sigma^2)\right)}{\sum_k \exp\left(-d(\boldsymbol{x}, \boldsymbol{w}_k)/(2\sigma^2)\right)} \tag{4}$$

indicates, that data point $\boldsymbol{x}$ was generated by the component j. For the special case of $\boldsymbol{c}_j$ resembling a crisp assignment to one class, equation (2) is equivalent

to the local cost function of the SNPC $lc(\boldsymbol{x}_k, y_k) = \sum_{j=1}^{M} P(j|\boldsymbol{x}_k)(1-\delta_{y_k,c_j})$, where c_j is is the assigned class label for prototype $\boldsymbol{w}_j$ with $c_j \in \mathcal{I}$ and the Kronecker symbol δ_{y_k,c_j} is one if $y_k = c_j$ and zero otherwise.
Parallely to adapting the prototypes, their fuzzy labels can be optimized by

$$\Delta \boldsymbol{c}_j = -\alpha_c P(j|\boldsymbol{x}_k) \tag{5}$$

followed by a subsequent normalization. α_c is the learning rate.

Once the prototypes are determined, new data points $\boldsymbol{x}$ can be classified according to

$$c = \texttt{arg}\max_{c'} \sum_{j:c_j=c'} P(j|\boldsymbol{x}) \tag{6}$$

As with the SNPC there is a window rule specifying the active region for the prototype update. By denoting $T = P(l|\boldsymbol{x}_t)(1-\boldsymbol{c}_l - lc_t)$ in equation (3) and rewriting it to $T_0 = (T_{lc} - T_{\boldsymbol{c}_l}) \cdot \Pi(\boldsymbol{c}_l)$ with $T_{lc} = lc(1-lc)$ and $T_{\boldsymbol{c}_l} = \boldsymbol{c}_l(1+\boldsymbol{c}_l)$, it can be shown that $-2 \leq T_0 \leq 0.25$ since $0 \leq T_{lc} \leq 0.25$ and $T_{\boldsymbol{c}_l} \leq 1$. The term $\Pi(\boldsymbol{c}_l)$ is given by

$$\Pi(\boldsymbol{c}_l) = \frac{\exp(-d(\boldsymbol{x}, \boldsymbol{w}_l)/(2\sigma^2)}{\sum_{l'}(1-\boldsymbol{c}_l - \boldsymbol{c}_{l'})/\exp(-d(\boldsymbol{x}, \boldsymbol{w}_{l'})/(2\sigma^2)}. \tag{7}$$

The absolute value of T_0 has to be significantly different from zero to have a valuable contribution in the update rule. This yields the window condition $0 \ll |T_0|$, which can be obtained by balancing the local loss lc and the value of the assignment variable $\boldsymbol{c}_l$.

3 FSNPC for fuzzy labeled data points

Now we extend the FSNPC to handle fuzzy labeled data points additionally to the fuzzy labeled prototypes. To adapt the cost function of the FSNPC to the fuzzy labeled training data we remodel the crisp class assignments y_i of the data points $\boldsymbol{x}_i$ to a C-dimensional possibilistic vector $\boldsymbol{y}_i$ with $\sum_{k=1}^{C} \boldsymbol{y}_i^k = 1$ and $\boldsymbol{y}_i^k \geq 0$ where C again is the number of classes. Hence, we consider a training data set $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$ with unsharp class assignments $\boldsymbol{y}_i$. The function of the local costs now takes the proportionate class assignments $\boldsymbol{y}$ of the training data points $\boldsymbol{x}$ into account and will therefore change to

$$lc(\boldsymbol{x}_k, \boldsymbol{y}_k) = \sum_{j=1}^{M} P(j|x_k)(\boldsymbol{y}_k - \boldsymbol{c}_j)^2 \tag{8}$$

where $(\boldsymbol{y}_k - \boldsymbol{c}_j)^2$ is used as an abbreviation for $(\boldsymbol{y}_k - \boldsymbol{c}_j)^T(\boldsymbol{y}_k - \boldsymbol{c}_j)$. Note that for the local cost yields $lc(\boldsymbol{x}_k, y_k) \leq 1$, because $\boldsymbol{y}_k$ and $\boldsymbol{c}_j$ are both possibilistic class assignments. The cost function $E(\mathcal{S}, \mathcal{T})$ (1), as the sum over the local costs of all the data points, will not change. The update rule for the prototypes, which aims to minimize the local costs, can again be derived by a stochastic gradient descent on the local costs (8) with respect to the prototypes $\boldsymbol{w}_j$ and is given as

$$\Delta \boldsymbol{w}_j = -\frac{\alpha}{2\sigma^2} P(j|\boldsymbol{x})((\boldsymbol{y} - \boldsymbol{c}_j)^2 - lc)\frac{\partial d(\boldsymbol{x}, \boldsymbol{w}_j)}{\partial \boldsymbol{w}_j} \tag{9}$$

where lc is used as an abbreviation for $lc(\boldsymbol{x}_k, \boldsymbol{y}_k)$.

The fuzzy prototype labels $\boldsymbol{c}_j$ can also be optimized by

$$\Delta \boldsymbol{c}_j = -2P(j|\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{c}_j), \tag{10}$$

again followed by subsequent normalization.

In complete analogy to the FSNPC a window rule for the active region for the prototype update can be derived by setting $T = P(j|\boldsymbol{x})((\boldsymbol{y} - \boldsymbol{c}_j)^2 - lc)$. Using the Gaussian form (4) for $P(j|\boldsymbol{x})$ the term T can be rewritten as $T = T_0 \cdot \Pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}_j, \boldsymbol{c}_j)$, with

$$\Pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}_j, \boldsymbol{c}_j) = \frac{\exp(-d(\boldsymbol{x}, \boldsymbol{w}_j)/(2\sigma^2))}{\sum_k ((\boldsymbol{y} - \boldsymbol{c}_j)^2 + (\boldsymbol{y} - \boldsymbol{c}_k)^2) \exp(-d(\boldsymbol{x}, \boldsymbol{w}_k)/(2\sigma^2))} \tag{11}$$

and $T_0 = ((\boldsymbol{y} - \boldsymbol{c}_j)^T (\boldsymbol{y} - \boldsymbol{c}_j))^2 - lc^2$. Obviously, $lc^2 \leq 1$ because $lc \leq 1$. Further, because all components of both $\boldsymbol{y}$ and $\boldsymbol{c}_j$ are less or equal to one and greater or equal to zero, $(\boldsymbol{y} - \boldsymbol{c}_j)^T (\boldsymbol{y} - \boldsymbol{c}_j) \leq K$, where K is a data dependend constant. Hence we have $-K^2 \leq T_0 \leq K^2 + 1$. The absolute value of T_0 has to be significantly different from zero to have a valuable contribution to the update. Therefore $|T_0| \gg 0$ defines a window rule as it is known from SNPC or LVQ.

4 Experiments

We applied our extended version of the FSNPC to two data sets with fuzzy labeled data points: an artificial and a real world data set.

In order to evaluate the classification accuracy, we compute Fuzzy Cohen's Kappa κ as introduced in [2]. This coefficient always lies in the interval $[-1; 1]$ and measures the agreement of two classifiers. The degree of classification agreement reaches from *slight* agreement with $0 < \kappa \leq 0.2$ over *fair*, *moderate*, and *substantial* up to *perfect* agreement with $0.8 < \kappa \leq 1.0$. Values beneath zero indicate a *poor* or *accidential* agreement only (see [5] for details).

As mentioned before there is a lack of classifiers based on fuzzy data assignments. For this reason, we can only assess the performance of our algorithm with respect to FRSLVQ. Comparison with classifiers based on crisp labeled data would require a modification on the data set itself and the interesting fuzzy aspect would get lost. Furthermore Fuzzy Cohen Kappa is not suitable for comparing fuzzy with crisp classifiers [11].

4.1 Artificial data set

The first experiment consists of a data set of two overlapping Gaussian clusters of equal variance in a two-dimensional space. We set the distribution's mean values to $\mu_1 = [-1; 0]$ and $\mu_2 = [1; 0]$. For the variance we choose different values. Each cluster consists of 500 samples. We define the class memberships $\boldsymbol{y}$ of sample $\boldsymbol{x}$ depending on the first component $\boldsymbol{x}(1)$. For the region between the means $-1 \leq \boldsymbol{x}(1) \leq 1$ we choose a linear relationship and a crisp class assignment for data points outside of the overlap with $\boldsymbol{x}(1) < -1$ and $\boldsymbol{x}(1) > 1$, respectively (figure 1, left).

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

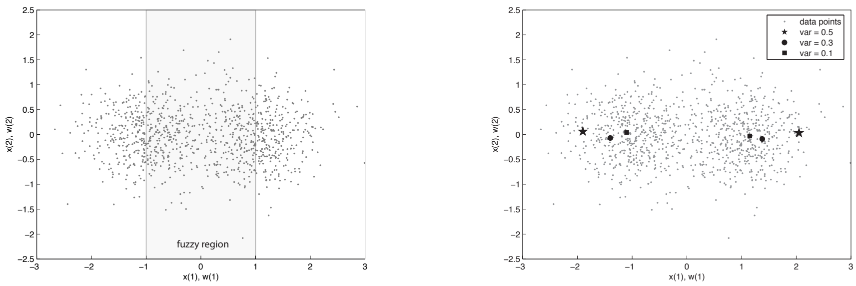

Fig. 1: Artificial data. Left: fuzzy region between the centers of the Gaussians. Right: visualization of the relationship between the variance and the final prototype position - with increasing variance the prototypes move away from the decision boundary.

We perform several runs with a varying number of prototypes and different settings for the variance. The initial fuzzy prototype labeling is set to around 50% for each class. In each training step the prototypes and their fuzzy class assigments are updated. The latter allowed the protoypes to switch their affiliation to the classes. After successful training classification agreement of $\kappa \approx 0.94$ (FRSLVQ $\kappa \approx 0.83$) was reached for the trained data. Comparing the agreement of the classification of untrained data points with the expected results gives $\kappa \approx 0.87$ (FRSLVQ $\kappa \approx 0.78$). All these indicate substantial to perfect agreement in terms of Fuzzy Cohen's Kappa. Comparing the classification results obtained by FRSLVQ and our extended FSNPC version we obtain $\kappa = 0.63$.

Further can be observed, that contrary to the FSNPC, where the prototypes get positioned near the decision boundary, our extended version behaves differently. The prototypes tend to move into regions of the data space, where there is a higher degree of classification agreement within the data set (figure 1, right). The more crisp the labels of the data points the more attractive is this region to the prototypes. We already observed this particular behavior by comparing the RSLVQ and its fuzzy version FRSLVQ [3].

4.2 Real-world problem

In our second experiment we used a dataset consisting of a series of transverse sections of barley grains at different developmental stages which also has been used in [1], [3], and [10]. The tissue samples can be classified into 11 different types like nuclear epidermis, transfer cell, and chlorophyll layer. The classification of these samples was done manually and especially for border tissue there was no distinct type identification possible. Therefore about 50% of the 4418 data points have fuzzy class assignments. Each data point is described by means of 144 features. For training and testing the dataset was randomly split into two groups of 3800 and 618 data points, respectively.

The agreement between the classification of FSNPC and FRSLVQ is averaging $\kappa \approx 0.72$, which implies a substantial agreement. Comparing the classification result of our method to the original data, gives a Fuzzy Cohen Kappa value of $\kappa \approx 0.64$. This also indicates a substantial agreement. This lower κ-value is due to the characteristic of the data set, which contains a great amount of crisp labeled data points, which is disadvantageous for the calculation of Fuzzy Cohen Kappa [11].

5 Conclusion

We introduced an extension for the FSNPC to handle fuzzy labeled data points. The rules for the prototype update as well as the update of their fuzzy class assignments were derived by a stochastic gradient descent on the cost function, which incorporates the fuzzy data assignments. We applied this extended FSNPC to two exemplary settings showing a substantial agreement with the expected classification. Due to a lack of other fuzzy classifiers besides FRSLVQ there were no further comparison in terms of accuracy possible. The characteristic behavior, that the prototypes of fuzzy labeled data sets tend to move away from the decision boundary, needs to be further investigated in future studies.

References

[1] C. Brüß, F. Bollenbeck, F.-M. Schleif, W. Weschke, T. Villmann, and U. Seiffert. Fuzzy image segmentation with fuzzy labeled neural gas. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2006)*, pages 563–568, Brussels, Belgium, 2006. d-side publications.

[2] W. Dou, Y. Ren, Q. Wu, S. Ruan, Y. Chen, and D. B. A.-M. Constans. Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, 70:726–734, 2007.

[3] T. Geweniger, P.Schneider, F.-M. Schleif, M-Biehl, and T. Villmann. Extending rslvq to handle data points with uncertain class assignments. Machine Learning Report 2, University of Leipzig Working Group Computational Intelligence, November 2009.

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

[5] L. Sachs. *Angewandte Statistik*. Springer Verlag, 7-th edition, 1992.

[6] A. S. Sato and K. Yamada. Generalized learning vector quantization. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 423–429. MIT Press, 1995.

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

[8] S. Seo and K. Obermayer. Soft learning vector quantization. *Neural Computation*, 15(7), July 2003.

[9] T. Villmann, F.-M. Schleif, and B. Hammer. Prototype-based fuzzy classification with local relevance for proteomics. *Neurocomputing*, 69(16–18):2425–2428, October 2006.

[10] 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.

[11] D. Zühlke, T. Geweniger, U. Heimann, and T. Villmann. Fuzzy Fleiss-Kappa for comparison of fuzzy classifiers. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2009)*, pages 269–274, Evere, Belgium, 2009. d-side publications.