Multiple Training of Vector-Based Neural Networks to detect Density Centers in Input Space

David Sommer, Martin Golz
University of Applied Sciences Schmalkalden
Department of Computer Science, Postfach 182
D-98574 Schmalkalden, Germany
{sommer, golz}@informatik.fh-schmalkalden.de

ABSTRACT: In this paper we will propose a method for enhanced evaluation of classified input vectors. For an arbitrary noisy and labeled data set, which can have several maxima of the probability density function (density centers), an automatic distinction of input vectors into "good classifiable" and "bad classifiable" is done. In a first training and testing phase (filter phase) the goodness of each input vector is calculated by means of vector based artificial neural networks, typically Learning Vector Quantization. In a second phase (evaluation phase) only "good classifiable" input vectors are used to train a Self-Organizing Map (SOM). For an artificial and a real world data set with compact but overlapping probability density functions it is shown that the complexity of the SOM output space is reduced. In some applications it might be possible to enhance the generalization properties of the SOM classifyer by using only "good classifiable" input vectors during training.

KEYWORDS: Learning Vector Quantization, Self-Organizing Map

INTRODUCTION

Two basic assumptions of pattern recognition are compactness and separability of regions in the feature space. But many real world classification tasks suffer more or less from violations of these assumptions. One reason may be uncertainty of class membership. Other reasons one may give for are noisy processes. Automatically learning algorithms, like neural networks, are often overtaxed. A large number of neurons garantees high adaptivity to the training data set, but not necessarily to a high adaptivity to the test data set (generalization ability). For example, numerical simulations of a two-class problem in high dimensional input space, where data samples came from two Gaussian processes with identical parameters, resulted in very good reclassification rates (correct training set classifications) of 95 % and more. But classification rates (correct test set classifications) remained poor (50 %, the rate of arbitrarily classification). This has to be expected, because there is nothing to generalize in the data set.
The assumption of compact and separable regions is related to the concept of voronoi cells of nearest neighbour type classifiers, like vector based neural networks. After training a classifier should have voronoi cells containing input vectors of one class only. Otherwise, if this is not achievable, each Voronoi cell should have highest possible purity, that means the cell contains a large relative amount of input vectors with an uniform label.

METHOD

One way to improve the purity of Voronoi cells could be obtained by learning over multiple training. Before each training run the classifier is initialized randomly. After training the purity of each voronoi cell is calculated and assigned as weighting factors to their input vectors. Input vectors of cells with high purity are assigned to a high weight and input vectors of cells with low purity are assigned to a low weight. The weights for each input vector are stored for later visualization and further processing.
LVQ1 is a version of LVQ and belongs to the supervised learning; vector based neural networks [3]. With the principle of competition learning an adaptation of the prototype vectors to the distribution of the input vectors will be aimed. This paradigm is known as vector quantization. As similarity measure between input vector $\mathbf{x}$ and prototype vector $\mathbf{w}$ the euclidian distance was used. During training the prototype vector $\mathbf{w}_c$, which is closest to $\mathbf{x}$, is updated at iteration index t by:

$$\Delta \mathbf{w}_c(t) = \pm\, \eta(t)\, [\mathbf{x}(t) - \mathbf{w}_c(t)] \qquad (1)$$

If input vector **x** and winner neuron $\mathbf{w}_c$ belongs to the same class, than using the positive sign of Eq. (1); i.e. $\mathbf{w}_c$ decreases the distance to **x**. Otherwise $\mathbf{w}_c$ increases the distance to **x**. With increasing iteration index t the step size $\eta(t)$ becomes smaller, until finishing the training by a criterion [5].

In order to avoid dead neurons and a large variance of the classification rate due to random initialization a modified version of LVQ1 [1] was applied.

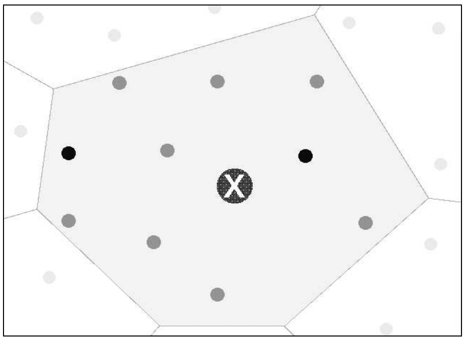


Figure 1: voronoi cell with prototype vector(cross) and 10 input vectors (small circles)

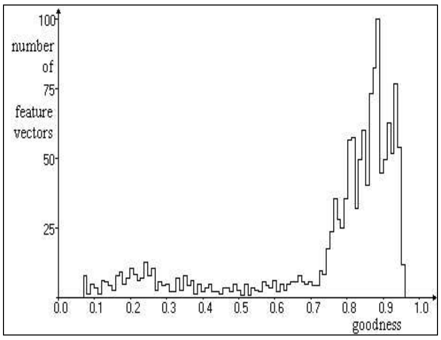


Figure2: histogram of input vectors weights from Fig. 3

After training the purity of each voronoi cells has to be calculated. In the case of two classes the purity g_i was estimated by:

$$g_i = n_i / (n_i + k_i) \qquad \textbf{(2)}$$

where n_i is the number of major input vectors of one class and k_i is the number of input vectors of the other class. The denominator of Eq. (2) is the size of the voronoi set. For the example (Fig. 1) one obtains: $n_i = 8$, $k_i = 2$, $g_i = 0.8$.

The g_i is assigned to each input vector of the major class in the voronoi cell of the prototype vector $\mathbf{w}_i$. The input vectors of the other class are assigned to $(1-g_i)$.

The training is repeated many times with new random initializations. The weights g_i, stored for every input vector, is averaged from training to training. Input vectors lying in compact regions of one class have a probability to reach high weights.

TWO-DIMENSIONAL EXAMPLE

Two overlapping classes, each distributed in two regions, were generated in two-dimensional space (Fig. 3). All 2,000 input vectors were applied to multiple training of LVQ1 networks with 20 neurons. The number of training runs was T = 100.

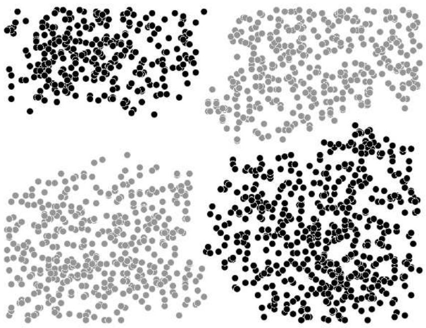


Figure3: Gaussian mixture data of 2 classes (class1: grey, class2: black)

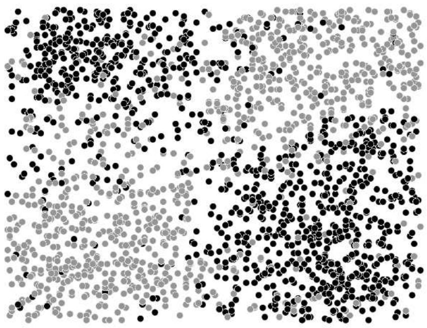


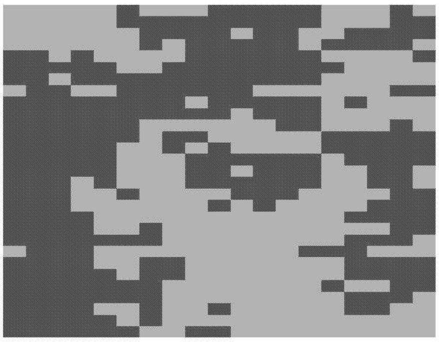
Figure 4: Filtered data from Fig. 3

The calculated histogram (Fig. 2) shows high input vector weights for the most input vectors. The simulation shows that singular input vectors of one class in a region of the other class (e.g. black dots in the upper right and the low left light gray regions in Fig. 3) were assigned to very low weights.

Introducing an arbitrary threshold leads to Fig. 4, where about 20 percent of all input vectors having weights lower than 0.65 were blanked. Without the blanked input vectors the voronoi cells of the LVQ network reach fully purity.

MULTIDIMENSIONAL EXAMPLE

The following example comes from a motoric ability test. 21 normal subjects aged between 18 and 32 years (17males, 4 females) took part in a test to hold the balance on the left leg, while standing on a double movable base plate. The resulting two-dimensional swinging movement x(t) and y(t) was measured and digitized with a sampling rate of 90 Hz. The test was done without and under the influence of alcohol [4].
Segments of the length of 5 sec were extracted and the spectral power densities were estimated using discrete fourier transform. Further features were estimated by the empirical covariance matrix of all samples and by calculation of the entropy leading to learning set of 810 classified 36-dimensional input vectors.
The multidimensional learning set was processed in the same way by multiple LVQ1 training, but visualization is out of reach. Therefore the Self-Organizing Map (SOM) as a dimensionality reducing neural networks was applied [2]. For the calibration of the SOM the class information of input vectors were used.
The separability of two classes on the 20 x 30 map (Fig. 5) is complicate. There might be an interclass region with input vectors of both classes.
After applying the multiple LVQ1 training procedure with 100 trainings and 20 neurons the vectors were weighted and setting the threshold to 0.5, leading to blanking of 160 input vectors. The resulting map shows well separable regions in the output space (Fig. 6).
By choosing the number of neurons for the LVQ1 training one can control the number of voronoi cells and therefore the complexity of the interclass border. A high number of voronoi cells lead to a high purity and high weights of the input vectors. But the extension of the interclass region is also controlled by the weights and by the threshold.



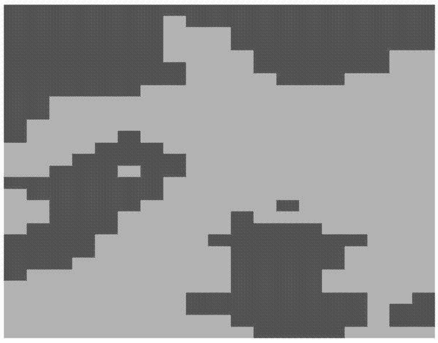


Figure 5: 20x30 SOM for the non-filtered feature vectors of BIOSWING data

Fig. 1. 20 x 30 SOM for the filtered feature vectors of BIOSWING data

dark gray: ‘with alcohol class’; light gray: ‘without alcohol class’

The proposed method as a combination of filtering and visualization could be used for removing of problematic feature vectors. Unproblematic feature vectors are assigned to high weights and are unchanged for following analyses. The method can simplify classifiers to represent fewer compact regions, whose validity has to estimate separately.

REFERENCES

[1] Golz, M; Sommer, D; Lembcke, T; Kurella, B; Classification of the Prestimulus-EEG of K-complexes using Competitive Learning Networks; Proc. Eufit’98, Aachen, 1998, 1767-1771.

[2] Kohonen, T.; Self-organized formation of topologically correct feature maps; Biol. Cybernet, 43 (1982), 59-69.

[3] Kohonen, T.; The Self-Organizing Map; Proc. of the IEEE, Vol. 78, No. 9,1990, 1464-1480.

[4] Recknagel, S.; Applikation neuroinformatischer Methoden für die Klassifikation von Daten aus einem motorischen Leistungstest; University of Applied Sciences, Schmalkalden, 1999 (in german).

[5] Zell, A.; Simulation Neuronaler Netze; Addison-Wesley, 1994 (german).