P

Jinwen Ma and Bin Cao

[illegible]
jwma@math.pku.edu.cn

Abstract. [illegible]

Clustering analysis is important not only in statistics but also in many aspects of practical applications. Analysis of clusters on a set of (unlabeled) sample data by means of mixture distribution, called mixture-model cluster analysis, is one of the most di cult problems in statistics, and its main task is to select a proper number of clusters in a sample data set. In the terminology of artificial neural networks, it is the problem of unsupervised classification on a sample data set where the number of clusters is unknown. Its importance and di culty has been noted by many researchers (e.g., [1]-[2]).

As an adaptive version of the classical k-means clustering, competitive learning (CL) has been developed from the field of neural networks and provided us a promising tool for unsupervised classification [3]. In the basic form of competitive learning, as each input sample is presented, the winning unit or neuron of the output layer of a CL network is activated and its corresponding weight vector is modified. While the rest of the units in the output layer are not activated and their corresponding weight vectors are not modified. This form of competitive learning is referred to as the simple or classical competitive learning (CCL) or winner-take-all (WTA) learning [4]. In the late of the 1980's, it was found that this simple learning mechanism has the dead unit (or under-utilized) problem. In order to solve it, the Frequency Sensitive Competitive Learning (FSCL) algorithm was developed under the light of conscience mechanism [5].

J. Wang et al. (Eds.): ISNN 2006, LNCS 3971, pp. 442–447, 2006.

However, these conventional competitive learning algorithms cannot be applied to solving the unsupervised classification problem on a set of sample data where the number of clusters is unknown. Xu, Krzyzak & Oja discussed this problem and showed that it is equivalent to the selection of an appropriate number of the units [6]. To tackle this problem, the Rival Penalized Competitive Learning (RPCL) algorithm was further proposed by adding a new mechanism into FSCL [6]. The basic idea is that for each input, not only the weight vector of the winner unit is modified (rewarded) to adapt to the input, but also the weight vector of its rival (the 2nd winner) is de-learned (penalized) by a smaller learning rate. When the learning and de-learning rates are appropriately selected, RPCL has the ability of automatically allocates an appropriate number of weight vectors for a sample data set, with the weight vectors of the extra units being pushed far away from the sample data. Actually, the RPCL algorithm is demonstrated well and has been applied to many fields such as clustering, vector quantization, and the training of RBF neural networks. Moreover, it has also been generalized to several versions for di erent types of sample data [7].

Via theoretical analysis, it has been recently proved that the RPCL algorithm can be considered as a special case of the distance sensitive RPCL (or DSRPCL) algorithms constructed by minimizing a cost function [8]. However, the distance between a sample point and a weight vector in the cost function is mainly analyzed and used in the form of the Euclidean distance, which actually limits the clusters to the spherical forms.

In this paper, we use the Mahalanobis distance instead of the Euclidean distance in the cost function and propose the Mahalanobis distance based rival penalized competitive learning (MDRPCL) algorithm that can be successful to determine the number of elliptical clusters in a data set and lead to a good classification result.

P

Given a set of sample data $\mathcal{S} = \{X^\mu\}_{\mu=1}^N$ with $X^\mu = [x_1^\mu, x_2^\mu, \cdots, x_d^\mu]^T \in \mathbb{R}^d$, the simple competitive learning or FSCL algorithm can be regarded as the adaptive versions of the well known k-mean algorithm, which essentially minimize the Mean Square Error (MSE) cost function as follows:

$$E_{MSE}(\mathbf{W}) = \frac{1}{2}\sum_{ij\mu} M_i^\mu (x_j^\mu - w_{ij})^2 = \frac{1}{2}\sum_{\mu} \| X^\mu - W_{c(\mu)} \|^2 \qquad (1)$$

where $\mathbf{W} = [W_1, W_2, \cdots, W_k]$, and each $W_i = [w_{i1}, w_{i2}, \cdots, w_{id}]^T \in \mathbb{R}^d$ is just the i-th weight vector in consideration. Moreover, we have

$$M_i^\mu = \begin{cases} 1 \text{ if } i = c(\mu) \text{ such that } \| X^\mu - W_{c(\mu)} \| = \min_j \| X^\mu - W_j \|, \\ 0 \text{ otherwise,} \end{cases}$$

with $c(\mu)$ denoting the index of the winner unit (or its weight vector) for the μ^{th} sample point.

However, $E_{MSE}(\mathbf{W})$ cannot be used for detecting a correct number k because it decreases monotonically as the number of the units increases. Thus, it does not apply to the RPCL algorithm. Recently, Ma and Wang [8] constructed the following cost function to describe the mechanism of the RPCL algorithm:

$$E(\mathbf{W}) = E_1(\mathbf{W}) + E_2(\mathbf{W}) \tag{2}$$

$$E_1(\mathbf{W})\frac{1}{2}\sum_{\mu} \| X^{\mu} - W_{c(\mu)} \|^2, \quad E_2(\mathbf{W}) = \frac{1}{P}\sum_{\mu, i\neq c(\mu)} \| X^{\mu} - W_i \|^{-P},$$

where $\mathbf{W} = vec[W_1, W_2, \cdots, W_n]$ and P is a positive number.

This new cost function is composed of the square mean error E_1 and the model selection term E_2. Actually, the minimization of E_2 can allocate the proper number of weight vectors to the sample data, while the minimization of E_1 makes the classification of the sample data possible. It is shown by theoretical analysis and experiment in [8] that the minimization of this cost function can lead to a a correct number of weight vectors located around the centers of the clusters in the sample data, respectively, with the other weight vectors being driven away far from the sample data. However, the discussions in [8] are mainly based on the Euclidean distance for the cost function, which essentially assumes that the clusters in the sample data are spherical. To relax this constraint, we can use the Mahalanobis distance instead of the Euclidean one in the cost function. That is, the distance $\| W_i - X^{\mu} \|$ is given by

$$\| X^{\mu} - W_i \|^2 = (X^{\mu} - W_i)^T \Sigma_i^{-1} (X^{\mu} - W_i), \tag{3}$$

where Σ_i is the covariance matrix of the cluster i assumed positive definite. Thus, the cost function given in Eq.(2) becomes a function of the parameters $(W_1, \Sigma_1), \cdots, (W_n, \Sigma_n)$.

With the above preparations, we now construct the Mahalanobis distance based rival penalized competitive learning algorithm as follows. First we get the derivatives of E with respect to W_j:

$$\begin{aligned}\frac{\partial E}{\partial W_j} = &-\sum_{\mu} \delta^j_{c(\mu)} \Sigma_j^{-1}(X^{\mu} - W_j) \\ &+ \sum_{\mu, j}(1 - \delta^j_{c(\mu)}) ||X^{\mu} - W_j||^{-P-2} \Sigma_j^{-1}(X^{\mu} - W_j)\end{aligned}$$

where δ^i_j is the Kronecker number.

Then, the MDRPCL algorithm for the update of the weight vectors can be constructed as a gradient descent algorithm of the cost function as follows.

$$W_j^{(t)} = W_j^{(t-1)} + \triangle W_j, \tag{4}$$

where$\triangle W_j = -\eta \frac{\partial E}{\partial W}$ and $\eta > 0$ is the learning rate which is generally selected as a small positive constant number.

Instead of the batch mode, we can get the following adaptive algorithm for the update of the weight vectors: at current step t, we random choose a sample X^μ, and have

$$\frac{\partial E_1}{\partial W_j} = \begin{cases} -\Sigma_j^{-1}(X^\mu - W_j), & \text{if } j = c(\mu); \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

$$\frac{\partial E_2}{\partial W_j} = \begin{cases} 0, & \text{if } j = c(\mu); \\ ||X^\mu - W_i||^{-P-2}\Sigma_j^{-1}(X^\mu - W_j), & \text{otherwise,} \end{cases} \tag{6}$$

Meanwhile, we use the following update rule to modify the covariance matrix Σ_j at each step in the batch mode:

$$\Sigma_j^{(t)} = \begin{cases} (1-\eta')\Sigma_j^{(t-1)} + \eta' \sum_\mu (X^\mu - W_{c(\mu)})(X^\mu - W_{c(\mu)})^T, & \text{if } j = c(\mu); \\ \Sigma_j^{(t-1)}, & \text{otherwise,} \end{cases} \tag{7}$$

and in the adaptive mode:

$$\Sigma_j^{(t)} = \begin{cases} (1-\eta')\Sigma_j^{(t-1)} + \eta'(X^\mu - W_{c(\mu)})(X^\mu - W_{c(\mu)})^T, & \text{if } j = c(\mu); \\ \Sigma_j^{(t-1)}, & \text{otherwise.} \end{cases} \tag{8}$$

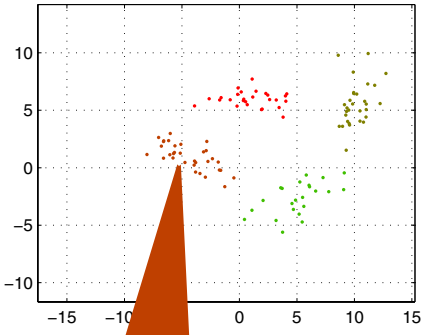
where $\eta' > 0$ is a small positive constant number. This update rule was given by Xu [7] for the extension of the RPCL algorithm to the data sets of elliptical clusters. Although it does not follow the gradient learning rule of the cost function exactly, it can maintain a good stability on the convergence of the covariance matrices. Actually, the exact gradient learning rule often leads to a degenerate covariance matrix such that the algorithm cannot be convergent.

Since the MDRPCL algorithm is a type of gradient descent algorithm, it may be possible to trap in a local minimum so that it may lead to a wrong clustering result. In order to overcome this problem, we can apply an simulated annealing mechanism to the MDRPCL algorithm in the same way as the ASRPCL algorithm given in [8]. We will use this simulated annealing method to make the classification of the wine data in the next section.

In this section, some simulated experiments are carried out to demonstrate the performance of the MDRPCL. Moreover, the MDRPCL algorithm is applied to the classification of the wine data.

3.1 On Simulated Data Sets

We conducted two experiments on the set of samples drawn from a mixture of four bivariate Gaussians densities (i.e., $d = 2$), which can observed from the backgrounds of Fig.1(a)&(b), respectively. Clearly, all the four Gaussian distributions in each case were selected to be elliptical, with a certain degree


10
5
0
−5
−10
−15
−10
0
5
10
15

It is shown by the experiments that the simulated annealing MDRPCL algorithm can detect the three classes in the wine data set with a classification accuracy of 100% or 99.64%(there is only one error) which is a rather good result for the unsupervised learning methods.

In this paper, we have investigated the function of the Mahalanobis distance in the special cost function associated with the RPCL algorithm and proposed a Mahalabis distance based RPCL algorithm. By relaxing the Euclidean distance into the Mahalanobis one, the proposed RPCL algorithm is able to be applied to a sample data set of elliptical clusters. It is demonstrated by the experiments that the MDRPCL algorithm can be successful to determine the number of elliptical clusters in a simulated or real data set and lead to a good classification result.

This work was supported by the Natural Science Foundation of China for Project 60471054.

1. p s , . ., r b s , .A.: A p d r b o d for s r A p s s. o - p o p s p s s 3(1974) 1-2
2. r p, .A.: s r b o p rob s p s r p. p . p p (d .): ss - o p d s r p. A d r ss, or (1977) 45-72
3. sr b d , ., p, .A.: o d p s p or p o p: A . r p s. o p. 36 (1988) 957-971
4. b - s p, .: ro o p p, A d d so p- s , d p A (1990)
5. A b , ., r s b p r , A. ., b p, ., o p, . .: o p rp p A for b s for or p o p. r or s 3 (1990) 277-291
6. , ., r , A., .: p d o p rp p for s r p A p s s, , p d r o p. r p s. r or s 4 (1993) 636-649
7. , .: p d o p rp p, p r , p d s s s- r p. p: ro . of p rp o p o o f r p o r or s ('98). o .3. r ss, A s (1998) 2525-2530
8. , ., p, .: A o s p o p A p p ro b o p p d o p rp p ( ). r p s. s s, p, p d b rp s p r ss.