# A KFCM-based Fuzzy Classifier

Aimin Yang[1,2] Lingmin Jiang[1] Yongmei Zhou[1]
*[1]School of Informatics, Guangdong University of Foreign Studies, Guangzhou 510420, China*
*[2]School of Computer, National University of Defense Technology, Changsha 410073, China*
*E-mail:amyang18@163.com*

## Abstract

*A proposed KFCM-based fuzzy classifier was introduced. As for the process of constructing such classifier, firstly, the original sample space is mapped into a high dimensional feature space by selecting appropriate kernel function. Then in the feature space, training samples of each class are divided into some clusters by proposed KFCM algorithm. The optimal cluster number of every class is selected by our proposed method. For each created cluster, a fuzzy rule is defined with ellipsoidal regions. Finally, fuzzy classification rules are adjusted by GAs. The accuracy of the constructed classifier by our proposed methods are comparable to the maximum accuracy of the multilayer neural network classifier and the relational reference,and the training time is much shorter.*

## 1. Introduction

Mercer kernels have recently become very popular due in large part to many recent successes in applying kernel methods such as support vector machines(SVM) to many real world problems[1,5].Besides SVM , there are other kernel methods proposed such as kernel fisher discriminant and kernel principal component analysis[3,4].Many researchers think that kernel method is a powerful tool of generalizing a standard linear method to nonlinear method in the application of real world. Hence it becomes a research hotspot recently. The kernel-based concept has also been adopted for unsupervised learning. The initial work of using Mercer kernel in clustering has been introduced[2,6] .The method[2] is hard clustering and just suits for Hyper-spherical or hyper-ellipsoidal clusters. An another kernel-based hard clustering algorithm[6] was presented and only applied on datasets with different Gaussian density.

Fuzzy classification rules[7,8,9] are widely considered a well-suited representation of classification knowledge, with readability and interpretation. Fuzzy classification has been widely applied in many fields, such as, image processing, words recognition, text classify-cation, remote sensing etc. But there are some problems in generating and optimizing fuzzy classification rules, and in how to select membership function, that is, it is difficult to gain fuzzy rules with good performance. So, there are many researchers on studying these problems and developing variety of systems which were based on neural network [7], genetic algorithm[9], clustering[8] and so on. But fuzzy classification systems based on neural network and genetic algorithm have long training time and too many rules, when the dimensions of input samples or the number of class is many. In the system based on clustering, it is difficult to optimize the rules. Especially, when the distribution of samples is as Figure 1(a),the general clustering method has poor classification accuracy.

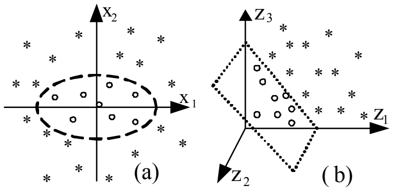

**Figure 1. A distribution of samples**

In order to train classifiers fast and obtain well generalization ability as neural network classifier, we propose an approach of generating and optimizing fuzzy rules and construct a new fuzzy classifier with these rules. For this approach, firstly, the initial sample space is mapped into a high dimensional space which is called feature space by selecting appropriate kernel function, as Figure 1,$(x_1,x_2)\to$ $(z_1,z_2,z_3)$,where, $z_1=x_1^2$, $z_2=2^{1/2}x_1x_2$, $z_3=x_2^2$, and the dimen-sion of sample space becomes from two to three.Then in the feature space,training samples of each class are separ-ated into some clusters respectively by Fuzzy C-means (FCM) clustering algorithm, and here such clustering method is called kernel FCM(KFCM). For each cluster, a fuzzy rule is defined with ellipsoidal region. Finally, the rules are adjusted by Genetic Algorithms (GAs). Such classifier is called a KFCM-based fuzzy classifier.

This paper organizes as follows. Section 2 presents KFCM Algorithm. In section 3, the expression of fuzzy classification rule is introduced. Section 4 describes the proposed adjusting approach for fuzzy classification rules by GAs. Section 5 is performance evaluation for proposed classifier. The last is conclusion.

## 2.KFCM Algorithm

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

Let $X=\{X^1,\ldots,X^k,\ldots,X^K\}(k=1,2,\ldots,K)$, present a given training sample set,where K is the number of classes. Let $X^k=\{x_1,x_2,\ldots,x_{Nk}\}$,where $x_i \in R^q$ (q is dimensions of original sample space) , present the k-th class sample subset, and Nk is the number of the k-th class training samples. For training samples of the k-th class, the objective of the FCM algorithm[10] is to minimize the Fuzzy C-means cost function formulate as Eq.(1).

$$J_m(X^k,U,V)=\sum_{j=1}^{Ck}\sum_{i=1}^{Nk}(\mu_{ji})^m\left\|x_i-v_j\right\|^2 \qquad (1)$$

$V=\{v_1,v_2,\ldots,v_{Ck}\}$,where $v_j \in R^q$, are the cluster centers. $U=(u_{ji})_{Nk\times Ck}$ is a fuzzy partition matrix, in which each member indicates the degree of membership of data vector $x_i$ in cluster j, where Ck denotes the cluster number. The elements of matrix U should satisfy the following conditions:

$$u_{ji}\in[0,1],i=1,\ldots,Nk,j=1,\ldots,Ck \qquad (2)$$

$$\sum_{j=1}^{Ck}(\mu_{ji})=1,i=1,\ldots,Nk \qquad (3)$$

The exponent $m\in[1,\infty]$ is the weighting exponent determining the fuzziness of the clusters.

Let $\Phi$: $R^q \rightarrow F$, $x \rightarrow \Phi(x)$, F is feature space. $\Phi$ is a map function. $\Phi(x)$ denotes a sample in feature space F, which is x in original space.Let K(,) denote kernel function[8]. Its definition is as Eq.(4).

$$K(x_i,x_j)=\Phi(x_i).\Phi(x_j) \qquad (4)$$

Where, "." is the inner product. In this paper, the kernel function of as Eq.(5) is used.

$$K(x_i,x_j)=\exp(-\gamma\|x_i-x_j\|^2)\ (\gamma\rangle 0) \qquad (5)$$

If Euclidian distance is applied in feature space, the objective function (as Eq.(6)) of fuzzy clustering could be derived from Eq. (1).

$$\begin{aligned}
J_{Km}(X^k,U,V)&=\sum_{j=1}^{Ck}\sum_{i=1}^{Nk}u_{ji}^m\left\|\Phi(x_i)-\Phi(v_j)\right\|^2\\
\left\|\Phi(x_i)-\Phi(v_j)\right\|^2&=(\Phi(x_i)-\Phi(v_j))^T(\Phi(x_i)-\Phi(v_j))\\
&=\Phi(x_i)^T\Phi(x_i)-\Phi(v_j)^T\Phi(x_i)-\Phi(x_i)^T\Phi(v_j)+\Phi(x_i)^T\Phi(v_j)\\
&=K(x_i,x_i)+K(v_j,v_j)-2K(x_i,v_j)
\end{aligned} \qquad (6)$$

Where, $2\le Ck\le Nk$. Here, the kernel function as Eq.(5) is selected, so $K(x,x)=1$, Eq.(6) can be rewritten as Eq.(7).

$$J_{Km}(X^k,U,V)=2\sum_{j=1}^{Ck}\sum_{i=1}^{Nk}u_{ji}^m(1-K(x_i,v_j)) \qquad (7)$$

So, optimal problem above is transformed to seek the minimal value for Eq. (8).

$$\min_{\mu_{ji},\{v_j\}_{j=1}^{Ck}} J_{km}(X^k,U,V) \qquad (8)$$

We can get the update expressions (seeing Eq.(9) and Eq.(10))of membership degree and center of cluster with Lagrange and the constraint condition.

$$u_{ji}^m=\frac{(1-K(x_i,v_j))^{-1/(m-1)}}{\sum_{j=1}^{Ck}(1-K(x_i,v_j))^{-1/(m-1)}} \qquad (9)$$

$$v_j=\frac{\sum_{i=1}^{Nk}u_{ji}^mK(x_i,v_j)x_i}{\sum_{i=1}^{Nk}u_{ji}^mK(x_i,v_j)} \qquad (10)$$

The KFCM algorithm for the k-th class is as follows.

**KFCM algorithm:**

*Step 1*:Choose the cluster number Ck for the training samples of the k-th class and the termination criterion $\varepsilon\in(0,1)$,maximal number of iteration $T_{max}$. Set the iteration count t=0.

*Step 2*:Choose kernel function K(.) and its parameters.

*Step 3*:Initialize centers $v^{(t)}_j$, and Calculate $u_{ji}^{(t)}$ according to Eq.(9) ,j=1,2,…,Ck, i=1,2,..,Nk.

*Step 4*:t=t+1,update $v_j^{(t)}$ according as the Eq.(10),and update $u_{ji}^{(t)}$ according as the Eq.(9).

*Step 5*: calculate $E^{(t)}=\max_{j,i}\left|u_{ji}^{(t)}-u_{ji}^{(t-1)}\right|$. If $E^{(t)}\le\varepsilon$, or $t=T_{max}$ then Stop, otherwise, go to Step 4.

Above KFCM algorithm can determine the center of each cluster under only knowing the number of cluster. How to determine parameter Ck is a key problem for constructing a classifier. Here, we adopt the following process to decide Ck. Firstly, make sure all the possible values of Ck. Then, calculate all the center and membership degree respectively for every Ck by KFCM algorithm. Finally, determine Ck by our proposed method which selects optimal Ck using average scatter degree(S) within clusters and average divided degree(D) among clusters. The definition of S and D is as Eq.(11).

$$\begin{aligned}
S&=\frac{1}{Ck}\sum_{j=1}^{Ck}\frac{1}{Nk_j}\sum_{i=1}^{Nk_j}\left\|\Phi(x_i)-\Phi(v_j)\right\|\\
D&=\frac{1}{Ck}\sum_{j=1}^{Ck}\frac{1}{Ck-1}\sum_{l',l'\neq l}^{Ck}\left\|\Phi(v_l)-\Phi(v_{l'})\right\|
\end{aligned} \qquad (11)$$

Where, $Nk_j$ is training sample number of the j-th cluster. S represents different degree among training samples, the smaller S, the more similar features of samples. D represents distance among cluster centers, the larger D, the larger different clusters[11]. So, clustering with small S and large D has optimal ck.

$$ck=\underset{ck\in\{2,3,\ldots,Ck\}}{\arg\min}\ r_{ck}=\underset{ck}{\arg\min}\,{}^{S_{ck}}\!/_{D_{ck}} \qquad (12)$$

## 3. Fuzzy Classification Rules

With the proposed approaches in section 2,training samples of every class can be divided into some clusters. Fuzzy classification rules are described as follows. Let x ($x\in R^q$ ) and $\Phi(x)$($\Phi(x)\in F$) be an input sample in the original space and in feature space respectively. The class number of the classified patterns is K.$C_k$(k=1,…,K) denotes the k-th class. $C_{kj}$(j=1,…) denotes the j-th cluster

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

of $C_k$. For each cluster, a fuzzy classification rule is generated and the expression is as (13).

$$R_{kj}: IF\ \Phi(x)\ is\ around\ C_{kj} \\ THEN\ x\in C_k\ with\ CF=\alpha_{kj} \tag{13}$$

Where, $R_{kj}$ is the label of rule and generated from $C_{kj}$ cluster, CF represents the degree of $\Phi(x)$ belonging to this rule, $\alpha_{kj}\in[0,1]$ is described as Eq.(18).

Let $\Phi(v_{kj})(v_{kj}\in R^q)$ denote the center of $C_{kj}$ cluster. $mf_{kj}(\Phi(x))$ denotes the membership function for the input sample x. $u_{kj}(\Phi(x))$ is the degree of membership function. The following equations are defined.

$$mf_{kj}(\Phi(x))=\exp(-h_{kj}^2(\Phi(x))) \tag{14}$$

$$h_{kj}^2(\Phi(x))=\frac{d_{kj}^2(\Phi(x))}{\delta_{kj}} \tag{15}$$

$$\begin{aligned} d_{kj}^2(\Phi(x)) &= \left\|\Phi(x)-\Phi(v_{kj})\right\|^2 \\ &= K(x,x)-2K(x,v_{kj})+K(v_{kj},v_{kj}) \\ &= 2(1-K(x,v_{kj}) \end{aligned} \tag{16}$$

$$u_{kj}(\Phi(x))=mf_{kj}(\Phi(x)) \tag{17}$$

$$\alpha_{kj}=u_{kj}(\Phi(x)) \tag{18}$$

Where, the membership function $mf_{kj}(\Phi(x))$ is a typical exponential function like Gauss function. $h_{kj}(\Phi(x)$ is called adjusting distance. $d_{kj}(\Phi(x))$ is the Euclidean distance between $\Phi(x)$ and the center of $C_{kj}$ cluster. $\Phi(v_{kj})$ is determined by our proposed KFCM algorithm in setion 2.$\delta_{kj}$ is a adjusting parameter. Figure 2 shows the space view of rule regions which are ellipsoidal.

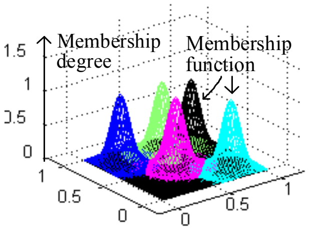

**Figure 2. Space view of rule regions**

After the creation of rules, rules should be adjusted because the overlapping of rule regions is unconsidered when rules were generated.This is discussed in next section. After the rules are adjusted, input samples may be classified.CF can be calculated corresponding to each input sample $\Phi(x)$.If the value $\alpha_{pj}$ for rule $R_{pj}$ is the maximum, x can be classified to the p-th class.

## 4. Fuzzy Classification Rule Adjusted By GAs

Because the overlapping of rule regions is unconsidered during the generation of rules, these rules should be adjusted. In this selection, we propose a method of adjusting rule by Genetic Algorithms(GAs). Here, we only adjust the value of $\delta_{kj}$ by GAs to get the maximal recognition rate[12].According to Eq.(15),when $\delta_{kj}$ is increased, the degree of membership function will be increased, vice versa.

### (1)Gene coding

Genetic information of fuzzy rule $R_{kj}$ is $\delta_{kj}$ and the firing information. A chromosome (seeing Figure 3(a)) consists of $\delta_{kj}$ and firing information. Each training sample is normalized to [0 1],and $0<\delta_{kj}\leq 5$.$\delta_{kj}$ is expressed by a 8-bit binary string and firing information is expressed by a bit. For firing information,“1” is fired for rule and “0” is not fired. Figure 3(b) shows the composition of an ind-ividual. Figure 4 shows the composition of a population. $I_p$(p= 1,2,…,P) denotes an individual,P denotes the size of this population.

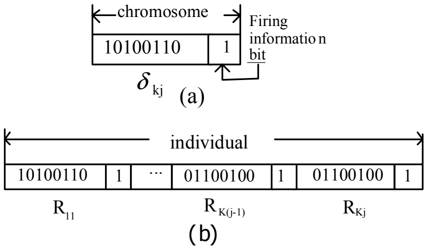

**Figure 3.Component of chromosome and its individual**

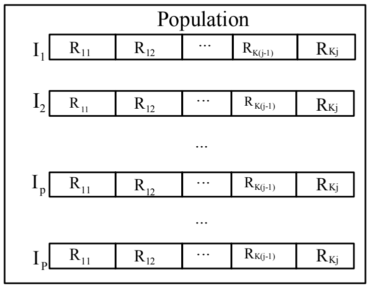

**Figure 4. Component of a population**

### (2) Fitness Function

For each individual, the fitness function is Eq.(21). The recognition rate of individual $I_p$ for training samples can be calculated according to Eq.(19). In Eq.(19),$Num_{all}$ is the total number of training samples and $Num_{cop}$ is the number of training samples which have been classified correctly. Eq(20) is the average classification accuracy of rule $R_{kj}$.In (20),$\Phi(x_L)$ is the training samples which have been classified correctly, $N_{corr}$ is the number of training samples which have been classified correctly by rule $R_{kj}$ and $N_{ncorr}$ is the number of training samples which have been classified incorrectly by rule $R_{kj}$. Eq.(21) is the fitness function of individual $I_p$. In Eq.(21), K is the number of class and $N_{R\_all}$ is total number of created rules.

$$\text{Accuracy} = \frac{\text{Num}_{\text{cop}}}{\text{Num}_{\text{all}}} \qquad (19)$$

$$\text{ACF}_{R_{ij}} = \frac{\sum_{L=1}^{N_{corr}} u_{kj}(\Phi(x_L))}{N_{corr} + N_{ncorr}} \qquad (20)$$

$$\text{Fit}_p = \text{Accuracy} \times \frac{\sum_{k=1}^{K}\sum_{j=1} \text{ACF}_{R_{kj}}}{N_{R\_all}} \qquad (21)$$

### (3) Determination of firing information

In formula (22), ACC($R_{kj}$) is the classification correction rate of rule $R_{kj}$ for training samples, $\text{Trnum}_{\text{corr}}(R_{kj})$ is the number of training samples which are classified correctly by rule $R_{kj}$, $\text{Trnum}_{\text{all}}(R_{kj})$ is the total number of the training samples classified by rule $R_{kj}$, and w is general in the range from 0.6 to 0.9.In each generation, if formula (22) is satisfied, rule $R_{kj}$ will be fired.

$$\text{ACC}(R_{kj}) = \frac{\text{Tr}num_{corr}(R_{kj})}{\text{Tr}num_{all}(R_{kj})} > w \qquad (22)$$

Each rule maintains a count for unfired information. nfc($R_{kj}$) is the count for $R_{kj}$. NFC is a threshold of this count. During the evolving procedure of the rules, if the rule is not fired, nfc($R_{kj}$) will be increased by 1. As soon as nfc($R_{kj}$)=NFC, rule $R_{kj}$ will be deleted, and these training samples for $R_{kj}$ are united with the rule which is the nearest to deleted rule $R_{kj}$. This strategy can delete some rules and make the number of rules more reasonable.

### (4)Genetic Operation

The procedure of genetic operation is as follows.

*Step 1*: Initial population is generated. Generally, this process is carried out randomly and all firing bits of rules are 1.

*Step 2*: Evaluate the fitness of each individual $I_p$ in the population as Eq.(21).

*Step 3*:Create **a** new population by repeating following steps until a population size is reached P.

*Selection Operation:* Select two parent chromosomes from a population according to their fitness by using the roulette wheel method.

*Crossover Operation:* A single point crossover(firing bit is excluded) method is conducted to form a new offspring.

*Mutation Operation*: With a mutation rate, mutate new offer spring at each locus(position in chromosome, firing bit excluded).Place new offspring in a new population.

*Step 4*：Genetic algorithm is done with newly generated populations. If the maximum number of genera-tions has been reached, Genetic algorithm is stopped. Otherwise, return to to Step 2.

## 5. Experiments

Two typical data sets are used to valuate the proposed methods. The results are compared with methods of neural network and reference [9].

### (1)Experiment with Wine data set

Wine data[13] include 178 datum with 13 input attributes and 3 classes.90 datum are used as training samples while 88 datum are used as testing samples. Parameters of experiment are as follows. $\gamma$=0.2(kernel function parameter).Initial value of $\delta_{kj}$ is 1,P=40,Generation=2500. Mutation probability is 0.002.Others,w=0.7, NFC=150. Table 1 is experimental result of the proposed classifier.

**Table 1 Experimental result of the proposed classifier**

| Generations | Time(sec.) | Accuracy(%) |
|---|---|---|
| 10 | 6 | 77.21 |
| 30 | 20 | 81.90 |
| 150 | 70 | 83.21 |
| 480 | 138 | 85.10 |
| 580 | 170 | 86.23 |
| 1000 | 300 | 87.57 |
| 1500 | 550 | 89.52 |
| 1800 | 603 | 92.16 |
| 2500 | 812 | 95.21 |

**Table 2 Comparison of accuracy for wine data**

| | | PM | RM |
|---|---|---|---|
| Training samples(90) | AA | 94.3% | 88.7% |
| | MA | 95.3% | 90% |
| Test samples(88) | AA | 92.6% | 81.8% |
| | MA | 93.8% | 86% |
| Average number of rules | | 16 | 3.5 |
| Number of evolution generations | | 3000 | 5000 |

* *PM: Proposed method,RM:Reference[12] method,* AA: Average accuracy,MA: Maximum accuracy

Table 2 is comparison of accuracy for proposed method and reference[9] method.

### (2)Experiment of Pen-Based Recognition of Handwritten Digits data[13]

This data set included 7494 training samples and 3498 test training samples with 16 input attributes and 10 classes. Table 3 is comparison of accuracy and training time for proposed method and neural network method.

Parameters of the experiment are as follows. Initial value of $\delta_{kj}$ is 1,P=70,Generation=3000, $\gamma$=0.15.Mutation rate is 0.002,w=0.75,NFC=300.The neural network has a 3-layer structure and 10 hidden layer units.Other parameters are as follows:initial weight is distributed in [-0.1, 0.1], Different initial values are used to train for 15 times.epoch is 2000.The learning speed is 0.8.The momentum is 0.

**Table 3 Comparison for Handwritten Digits data[13]**

| CS | CT | RM | ATR(%) | ATE(%) |
|---|---|---|---|---|
| PM | 811.6 | 82 | 96.3 | 93.9 |
| NNM | 2350.6 | 10(units) | 93.9 | 91.2 |

**PM: Proposed method,NNM: Neural Network method,CS: Classifier System,CT: CPU time (s),RM: Rule numbers,ATR:Accuracy of training sample,ATE:Accuracy of testing sample*

These results and comparison show that the accuracy of our proposed classifier are comparable to the maximum accuracy of the multilayered neural network classifier and reference[12],the training time is much shorter.

## 6. Conclusion

In this paper, we introduce a KFCM-based fuzzy classifier. During constructing such classifier, firstly, the original sample space is mapped into a high dimensional feature space by selecting appropriate kernel function. Then in the feature space, training samples of each class are divided into some clusters by the proposed KFCM algorithm. The optimal cluster number of every class is selected by our proposed method.Finally,fuzzy classification rules are adjusted with parameter $\delta_{kj}$ by GAs. The accuracy of the constructed classifier by our proposed method are comparable to the maximum accuracy of the multilayered neural network classifier and reference [9],and the training time is much shorter.

## Acknowledgements

This work was Supported by the National Natural Science Foundation of China (60373062) and the Natural Science Foundation of Hunan Province of China (05JJ40101)

## References

[1] Burges CJC.Geometry and invariance in kernel based methods.In Scholkopf B,Burges C, Smola A, Eds. Advance in Kernel Methods Support Vector Learning. Cambridge, MA:MIT Press, Cambridge, 1999,pp:89-116.

[2] Girolami M. Mercer kernel based clustering in feature space. IEEE Trans. on Neural Networks, 2002,13(3):780-784.

[3] Krishnapuram R, Frigui H,Nasraui O.The fuzzy c qadric shell clustering algorithm and the detection of second degree. Pattern Recognition Letters ,1993,14(7) :545-552.

[4] Muller K, Mika S , Ratsch G, et al. An introduction to kernel-based learning algorithms.IEEE Tran. Neural Networks , 2001,12(2): 181-201.

[5] Scholkopf B ,Mika S, Burges C, et al. Input space versus feature space in kernel-based methods.IEEE Tran. Neural Networks , 1999 , 10(5):1000-1017.

[6] Zhang L , Zhou W D, Jiao LC. Kernel clustering algorithm. Chinese Journal of Computer, 2002, 25(6):587-590.

[7] F. Uebele, S. Abe, and M. -S. Lan.A Neural Network-Based Fuzzy Classifier. IEEE Trans. Systems,Man, and Cybernetics, Vol. 25, No. 2, pp. 353–361, 1999.

[8]S.Abe.Dynamic Cluster Generation for a Fuzzy Classifier with Ellipsoidal Regions.IEEE Trans. Systems,Man, and Cybernetics-Part B, Vol. 28, No. 6, pp.869–876, 1998.

[9]Hng, L.,Inoue, H.Automatic generation of fuzzy classification systems using hyper-cone membership function. Computational Intelligence in Robotics and Automation,2003. Proceedings. 2003 IEEE International Symposium on, Volume: 2 ,July 16-20, 2003 Pages:658-663.

[10] Wu Zhongdong,Gao Xinbo,Xie Weixin.Kernel method-based fuzzy clustering algorithm.Journal of Systems Engineering and Electronics,2005,16(1):160-166.

[11] DaeWon Kim,Kwang H. Lee,Doheon Lee(2004) A novel initialization scheme for the fuzzy c-means algorithm for color clustering.Pattern Recognition Letters, 25: 227-237.

[12]Kaieda,K,Abe,S.A kernel fuzzy classifier with ellipsoiddal regions.Neural Networks, 2003. Proceedings of the International Joint Conference on,Volume: 3,July,2003 Pages:2043–2048.

[13] [UCI Repository of Machine Learning Databases Network ocument].ftp://ftp.ics.uci.edu/pub/machinelearning-databases/

0-7695-2874-0/07 $25.00 © 2007 IEEE