A COMPARISON OF TWO FUZZY CLUSTERING TECHNIQUES

Samarjit Das,[1] and Hemanta K. Baruah,[2]

[1]Department of Computer Science & IT, Cotton College, Assam, India

[2]Department of Statistics, Gauhati University, Assam, India

[1]ssaimm@rediffmail.com, [2]hemanta_bh@yahoo.com

Abstract- In fuzzy clustering, unlike hard clustering, depending on the membership value, a single object may belong exactly to one cluster or partially to more than one cluster. Out of a number of fuzzy clustering techniques Bezdek's Fuzzy C-Means and Gustafson-Kessel clustering techniques are well known where Euclidian distance and Mahalanobis distance are used respectively as a measure of similarity. We have applied these two fuzzy clustering techniques on a dataset of individual differences consisting of fifty feature vectors of dimension (feature) three. Based on some validity measures we have tried to see the performances of these two clustering techniques from three different aspects- first, by initializing the membership values of the feature vectors considering the values of the three features separately one at a time, secondly, by changing the number of the predefined clusters and thirdly, by changing the size of the dataset.

Keywords: Fuzzy c-means clustering, Gustafson-Kessel clustering, feature vectors, Euclidian distance, Mahalanobis distance, validity measures.

1. Introduction

Clustering is a process by means of which objects with similarities are placed to belong to same group called a cluster. In hard clustering an object is allowed to belong exactly to one cluster only. But with the advent of the concept of fuzzy set theory (FST) developed by Zadeh (1965) which particularly deals the situations pertaining to non-probabilistic uncertainty, the traditional hard clustering technique has unlocked a new way of clustering known as fuzzy clustering, where a single object may belong exactly to one cluster or partially to more than one cluster depending on the membership value of that object. A complete presentation of all aspects of FST is available in the work of Zimmermann (1991). The applications of FST in dealing with ambiguous problems where non-probabilistic uncertainty prevails have been reflected in the works of Dewit (1982) and Ostaszewski (1993). Baruah (2011a, 2011b) has shown that the membership value of a fuzzy number can be expressed as the difference between the membership function and a reference function, and therefore the fuzzy membership function and the fuzzy membership value for the complement of a fuzzy set are not the same. Based on this concept, Das (2012) tried to modify the design of Park and Park (2010) and was able to overcome the limitations of their work by visualizing the complement of a fuzzy set in a correct manner. Derring and Ostaszewski (1995) have explained in their research work a method of pattern recognition for risk and claim classification. Bezdek (1981) has discussed in his Fuzzy C-Means (FCM) technique that the data to be analyzed must be in the form of numerical vectors called feature vectors, and the number of clusters must be predefined for obtaining the membership values of the feature vectors. Das (2013) has tried the FCM algorithm of Bezdek with three different distances namely Euclidean distance, Canberra distance and Hamming distance which revealed that out of the three distances, the algorithm produces the result fastest as well as the most expected when Euclidean distance is considered and the slowest as well as the least expected when Canberra distance is considered. Das and Baruah (2013) have shown the application of the FCM algorithm of Bezdek on vehicular pollution. Wang, Huang, Yao, Qian and Jiang (2011) have applied the Gustafson-Kessel (GK) clustering algorithm of Gustafson and Kessel (1979) in the pattern recognition for gas insulated

switchgear (GIS). Although a number of fuzzy clustering techniques are available in the literature, FCM and GK clustering techniques are well known for analyzing data which are in the form of numerical vectors. In both FCM and GK clustering techniques the number of clusters must be predefined but Euclidian distance and Mahalanobis distances are used respectively in FCM and GK as a measure of similarity. We have applied these two fuzzy clustering techniques on a dataset of individual differences (see table1) consisting of fifty feature vectors of dimension (feature) three. In our present work, based on three validity measures namely Partition Coefficient (PC), Clustering Entropy (CE) and Partition Index (PI) (see section3) we have tried to see the performances of these two clustering techniques from three different aspects- first, by initializing the membership values of the feature vectors considering the values of the three features separately one at a time, secondly, by changing the number of the predefined clusters and thirdly, by changing the size of the dataset.

In Section 2, we shall provide the steps of FCM and GK algorithms. In Section 3, we explain our present work. The results and analysis of our work have been given in Section 4. Finally we put the conclusions in Section 5.

2. Mathematical calculations of FCM and GK algorithms

The basic task of a clustering technique is to divide n patterns, where n is a natural number , represented by vectors in a p-dimensional Euclidean space, into c, $2 \leq c < n$, categorically homogeneous subsets which are called clusters. Let the data set be

$X= \{x_1, x_2, \ldots\ldots, x_n \}$, where $x_k =\{ x_{k1}, x_{k2}, \ldots\ldots, x_{kp} \}$, k= 1,2,3,\ldots\ldots,n.

Each x_k is called a feature vector and x_{kj} where j=1,2,\ldots.p is the jth feature of the kth feature vector.

A partition of the dataset X into clusters is described by the membership functions of the elements of the cluster. Let S_1, S_2,\ldots\ldots,S_c denote the clusters with corresponding membership functions μ_{S_1}, μ_{S_2}, \ldots, μ_{S_c}. A c x n matrix containing the membership values of the objects in the clusters

$\tilde{U} = [\mu_{S_i}(x_k)]_{i=1,2,\ldots,c,k=1,2,\ldots,n}$ is a fuzzy c-partition if it satisfies the following conditions

$$\sum_{i=1}^{c} \mu_{s_i}(x_k) = 1 \qquad \text{for}$$
each k= 1,2,….,n. (1)

$$0 \leq \sum_{k=1}^{n} \mu_{s_i}(x_k) \leq n \qquad \text{for}$$
each i= 1,2,….,c. (2)

Condition (1) says that each feature vector x_k has its total membership value 1 divided among all clusters. Condition (2) states that the sum of membership degrees of feature vectors in a given cluster does not exceed the total number of feature vectors.

In sections 2.1 and 2.2 we provide respectively the steps of FCM and GK algorithms.

2.1 Bezdek's FCM algorithm

Step 1: Choose the number of clusters, c, $2 \leq c < n$, where n is the total number of feature vectors. Choose m, $1 \leq m < \alpha$. Define the vector norm $\| \ \|$ (generally defined by the Euclidean distance) i.e.

$$\| x_k - v_i \| = \sqrt{\sum_{j=1}^{p} (x_{kj} - v_{ij})^2} \qquad (3)$$

where x_{kj} is the jth feature of the kth feature vector, for k=1,2,\ldots\ldots,n; j=1,2,….,p and v_{ij}, j-dimensional centre of the ith cluster for i=1,2,\ldots\ldots,c; j=1,2,….,p; n, p and c denote the total number of feature vector , no. of features in each feature vector and total number of clusters respectively.

Choose the initial fuzzy partition

$$U^{(0)} = [\mu_{s_i}^{(0)}(x_k)]_{1 \leq i \leq c, 1 \leq k \leq n}$$

2

Choose a parameter $\in > 0$ (this will tell us when to stop the iteration). Set the iteration counting parameter l equal to 0.

Step 2: Calculate the fuzzy cluster centers $\{v_i^{(l)}\}_{i=1,2,.....,c}$ given by the following formula

$$v_i^{(l)} = \frac{\sum_{k=1}^{n} (\mu_{s_i}^{(l)}(x_k))^m x_k}{\sum_{k=1}^{n} (\mu_{s_i}^{(l)}(x_k))^m} \qquad \text{for}$$

$i = 1, 2, \ldots c; \quad k = 1, 2, , \ldots n.$ \qquad (4)

Step 3: Calculate the new partition matrix (i.e. membership matrix)

$$U^{(l+1)} = [\mu_{s_i}^{(l+1)}(x_k)]_{1 \le i \le c, 1 \le k \le n} \quad , \qquad \text{where}$$

$$\mu_{s_i}^{(l+1)}(x_k) = \frac{1}{\sum_{j=1}^{c} \left(\frac{\| x_k - v_i^{(l)} \|}{\| x_k - v_j^{(l)} \|} \right)^{\frac{2}{m-1}}} \qquad (5)$$

for $i=1,2,\ldots,c$ and $k=1,2,\ldots,n$. If $x_k = v_i^{(l)}$, formula (5) cannot be used. In this case the membership function is

$$\mu_{s_i}^{(l+1)}(x_k) = \{_{0 \, if \, k \ne i, \, i=1,2,....,c}^{1 \, if \, k = i} \qquad (6)$$

Step 4: Calculate $\quad \Delta = \| U^{(l+1)} - U^{(l)} \|$

If $\Delta > \in$, repeat steps 2, 3 and 4. Otherwise, stop at some iteration count l^*. To make the result operational the fifth step had been introduced by Derring and Ostaszewski (1995).

Step 5: The final fuzzy matrix U^{l^*} is structured for operational use by means of the normalized α -cut, for some $0 < \alpha < 1$. All membership values less than α are replaced with zero and the function is renormalized (sums to one) to preserve partition condition (1).

2.2 The GK algorithm

Let n, p, c, m(>1),$\in$(>0) and A denote the total number of feature vectors , no. of features in each feature vector ,total number of clusters, the weighting exponent ,the termination

tolerance and the norm- inducing matrix respectively for a given data set.

Following are the steps of GK algorithm.

Repeat for $l = 1, 2, \ldots$

Step 1: Compute the cluster prototypes (centers) :

$$v_i^{(l)} = \frac{\sum_{k=1}^{n} (\mu_{ik}^{(l-1)^m} x_k}{\sum_{k=1}^{n} (\mu_{ik}^{(l-1)^m}}, 1 \le i \le c$$

Step 2: Compute the cluster covariance matrices:

$$F_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{(l-1)^m} (x_k - v_i^{(l)})(x_k - v_i^{(l)})^T}{\sum_{k=1}^{n} (\mu_{ik}^{(l-1)^m}},$$

$1 \le i \le c$

Step 3: Compute the distances (Known as Mahalanobis distance):

$$D_{ikA_i}^2 = (x_k - v_i^{(l)})^T A_i (x_k - v_i^{(l)})$$

where $A_i = \rho_i \det(F_i)^{\frac{1}{p}} F_i^{-1}$, (ρ_i =1 usually)

, $1 \le i \le c$, $1 \le k \le n$

Step 4: Update the partition matrix:
for $1 \le k \le n$

if $D_{ikA_i} > 0$ for $1 \le i \le c$,

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} (D_{ikA_i} / D_{jkA_j})^{2/(m-1)}},$$

Otherwise

$$\mu_{ik}^{(l)} = \{_{0 \, if \, k \ne i, \, i=1,2,....,c}^{1 \, if \, k = i}$$

until $\| U^{(l)} - U^{(l-1)} \| < \in$.

Step 5 has been introduced in the same way as in FCM algorithm for operational use.

3. Our present work

In our present work we have applied the FCM and GK algorithms on a dataset of individual differences(see table 1) which consists of fifty (50) feature vectors with three features namely Intelligent Quotient(IQ), Achievement Motivation(AM) and Social adjustment(SA) in each feature vector. We have tried to see the performances of these two clustering techniques on the same dataset using three validity measures PC, CE and PI from three different aspects- the initial membership values of the feature vectors, the no. of the predefined clusters and the size of the dataset. For this purpose first, we have executed both the algorithms thrice on the same dataset by initializing the membership values of the feature vectors considering the values of the features IQ, AM and SA separately one at a time in each round of execution. Secondly, both the algorithms have been executed thrice on the dataset by considering the no. of predefined clusters as four(04), three(03) and two(02) respectively for first , second and third execution. Thirdly, we have reduced the size of the dataset by ten(10) in each round and executed the FCM and GK algorithms for three times. To measure the performances of both the clustering techniques we have used three validity measures. The mathematical formulae of these three validity measures have been given in the following.

(a) Partition Coefficient (PC): measures the overlapping between clusters.

$$PC(c) = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^2$$

(b) Clustering Entropy (CE): measures the fuzziness of the cluster partition

$$CE(c) = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} \log(\mu_{ij})$$

(c) Partition Index (PI): is the ratio of the sum of the compactness and separation of the clusters.

$$PI(c) = \sum_{i=1}^{c} \frac{\sum_{j=1}^{n} (\mu_{ij})^m \| x_j - v_i \|^2}{n_i \sum_{k=1}^{c} \| v_k - v_i \|^2}$$

Table 1: Data set of individual differences of fifty(50) feature vectors with dimension(feature) three(03).

FV	IQ	AM	SA	FV	IQ	AM	SA
1	91	18	55	26	110	18	55
2	85	16	40	27	100	16	40
3	120	19	74	28	100	18	75
4	90	18	75	29	70	14	30
5	92	17	74	30	105	17	55
6	82	17	55	31	79	14	35
7	95	19	75	32	80	15	34
8	89	18	74	33	125	20	75
9	96	19	75	34	100	19	75
10	90	17	55	35	125	19	85
11	97	16	54	36	80	18	60
12	125	21	74	37	85	18	70
13	100	19	75	38	145	25	90
14	90	17	54	39	80	18	74
15	100	18	84	40	92	17	55
16	95	19	75	41	120	18	70
17	130	23	85	42	145	30	80
18	130	19	75	43	95	18	50
19	90	17	55	44	80	16	36
20	91	17	56	45	90	17	55
21	140	22	82	46	115	23	84
22	92	18	75	47	100	18	80
23	101	18	55	48	80	14	35
24	85	16	54	49	105	19	75
25	97	19	54	50	120	21	74

FV: feature Vector, IQ: Intelligent Quotient, AM:Achievement motivation, SA: Social adjustment

4. Results and analysis

In this section we place the results obtained in our work and an analysis of these results. The graphical representations of the membership values (full or partial) of the feature vectors in the final clusters obtained in different executions have been given in figures 1 to 20. In table2 we have recorded the values of different validity measures of FCM algorithm when the membership values of the feature vectors are initialized with different features at different no. of clusters. The same for GK algorithm have been recorded in table3. The values of different validity measures of FCM and GK algorithms at different sizes of dataset have been given in table4 and table5 respectively.

Table 2: The values of different validity measures of FCM algorithm with different features for initializing the membership values at different no. of clusters

Feature	C	Itn	PC	CE	PI
IQ	4	6	0.90248	0.14535	0.03908
AM	4	8	0.93592	0.09099	0.06822
SA	4	6	0.92743	0.1033	0.06726
IQ	3	5	0.92794	0.10711	0.07536
AM	3	6	0.92721	0.10793	0.07995
SA	3	5	0.92147	0.11724	0.07775
IQ	2	4	0.86863	0.18854	0.11705
AM	2	2	0.85763	0.20356	0.13537
SA	2	3	0.85774	0.20341	0.15277

C: the no. of predefined clusters, Itn: the no. of iterations to reach the final clusters.

Table 3: The values of different validity measures of GK algorithm with different features for initializing the membership values at different no. of clusters

Feature	C	Itn	PC	CE	PI
IQ	4	11	0.85852	0.20914	0.05578
AM	4	5	0.83772	0.23248	0.11549
SA	4	5	0.77931	0.32711	0.07821
IQ	3	6	0.79372	0.30356	0.14256
AM	3	7	0.85369	0.21217	0.33906
SA	3	5	0.92444	0.10967	0.20339
IQ	2	6	0.82051	0.26268	0.5176
AM	2	4	0.73348	0.38106	0.27783
SA	2	4	0.80529	0.28523	0.28288

C: the no. of predefined clusters, Itn: the no. of iterations to reach the final clusters.

Table 4: The values of different validity measures of FCM at different sizes of dataset.

FCM ALGORITHM					
N	C	Itn	PC	CE	PI
50	4	6	0.90248	0.14535	0.03908
40	4	5	0.91737	0.12211	0.0387
30	4	5	0.92787	0.11276	0.03779

N: the size of the datasets, C: the no. of predefined clusters, Itn: the no. of iterations to reach the final clusters.

Table 5: The values of different validity measures of GK at different sizes of dataset

GK ALGORITHM					
N	C	Itn	PC	CE	PI
50	4	11	0.85852	0.20914	0.05578
40	4	No output as the norm inducing matrix becomes singular			
30	4	No output as the norm inducing matrix becomes singular			

N: the size of the datasets, C: the no. of predefined clusters, Itn: the no. of iterations to reach the final clusters.

Based on the values recorded in table 1, 2, 3, 4 and 5 we have tried to analyze the performances of each algorithm separately from all three aspects i.e. when the membership values of the feature vectors are initialized considering the values of the features one at a time, when the no. of the predefined clusters is changed and when the size of the dataset is reduced. Next we have tried to compare the performances of both the algorithms for each of these three aspects. In our analysis we have also considered the no. of iterations required by both the algorithms to reach the final clusters in different executions. The graphical representations of all these analysis have been given in figures 21 to 49.

A classification algorithm which exhibits greater value of Partition Coefficient (PC) and smaller values of Classification Entropy (CE) and Partition Index (PI) is considered to have better performance.

The no. of cluster wise performances of FCM when the membership values of the feature vectors are initialized with different features have been shown in figures 21, 22 and 23. It is seen that when the no. of clusters is 4 FCM performs better with feature AM, when the no. of clusters is 3 it performs better with feature IQ and when the no. of clusters is 2 it performs better with feature IQ. Again the feature wise performances of FCM at different no. of clusters have been shown in figures 24, 25 and 26. In these figures we see that with feature IQ FCM has better performance when the no. of clusters is 3, with feature AM it has better performance when the no. of clusters is 4, and with feature SA it has better performance when the no. of clusters is 4. Similar performances of GK algorithm based on these two aspects have been shown in figures 27 to 29 and figures 30 to 32 respectively. These reveal that when the no. of clusters is 4 GK performs better with feature IQ, when the no. of clusters is 3 it performs better with feature SA and when the no. of clusters is 2 it has better performance with feature IQ. Again with feature IQ GK performs better when the no. of clusters is 4, with feature AM it performs better when the

no. of clusters is 3 and with feature SA it has better performance when the no. of cluster is 3.

In figures 33 to 41 we provide the comparison of different validity measures of both the algorithms. It is seen that almost in all the cases FCM has better performance except in one where GK performs partially better (see figure 38).

We have provided the graphical representations of different validity measures and the no. of iterations of FCM at different sizes of datasets respectively in figures 48 and 49. It is seen that FCM performs better when the size of the dataset is 30. Also it performs faster when the dataset's size is 40 and 30. On the other hand the performance of GK algorithm deteriorates significantly when the size of the dataset becomes very small as the respective norm inducing matrix becomes singular (see table5).

From iterations point of view GK performs faster when initialized with feature AM and the no. of clusters is 4 (figures 43 and 45). It also performs faster when initialized with feature SA and the no. of clusters is 4 (figures 44 and 45). In rest of the cases FCM performs faster.
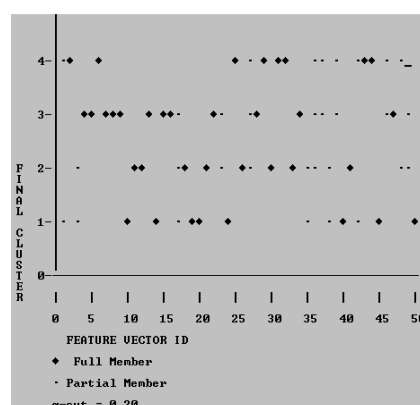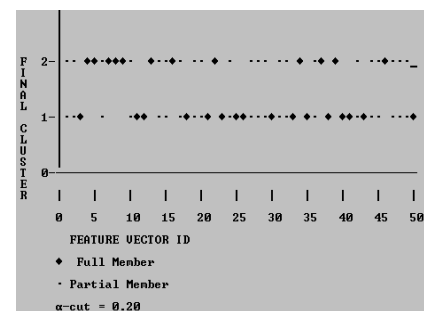

Figure 1: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature IQ and no. of clusters is 4.
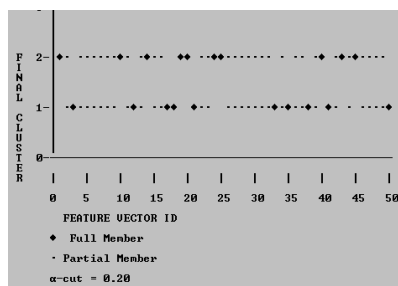
6
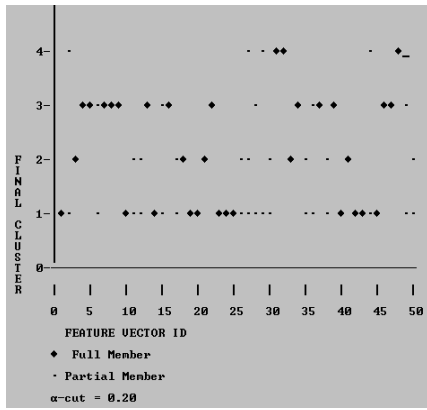
Figure 2: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature AM and no. of clusters is 4.
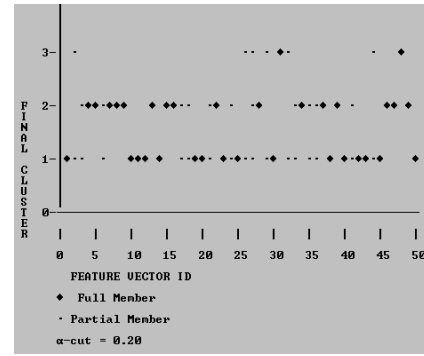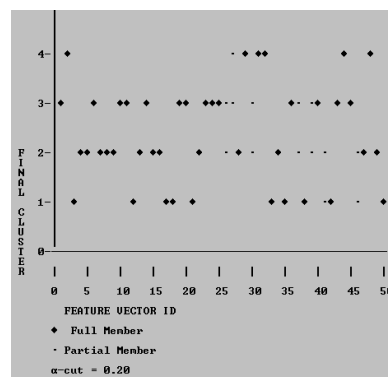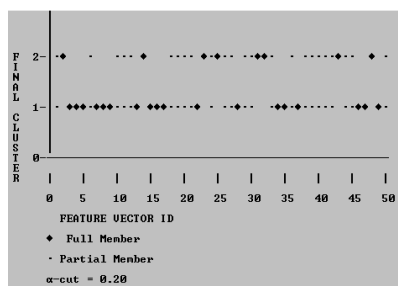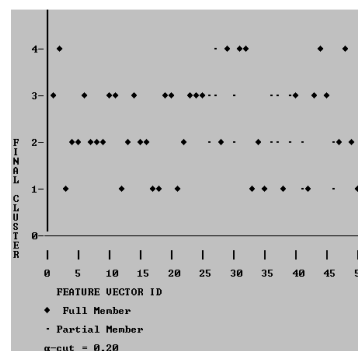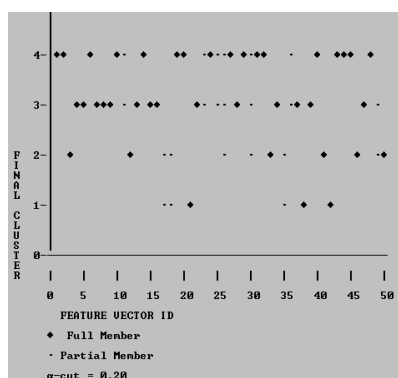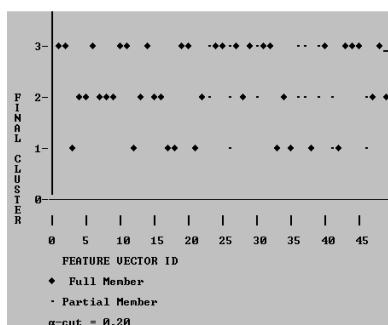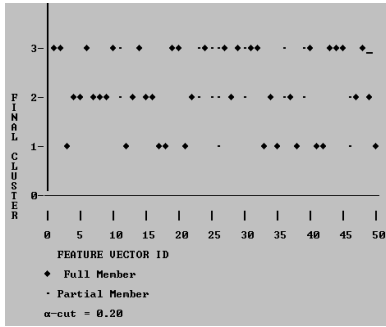

Figure 5: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature AM and no. of clusters is 3.
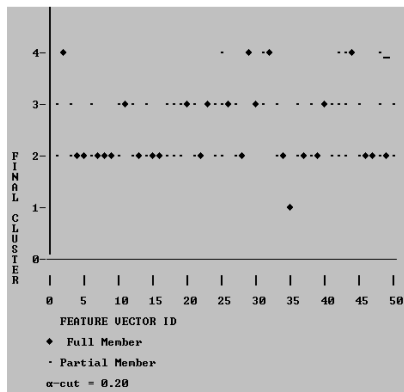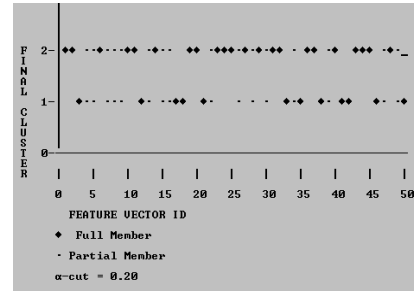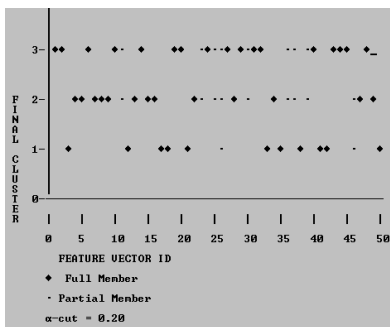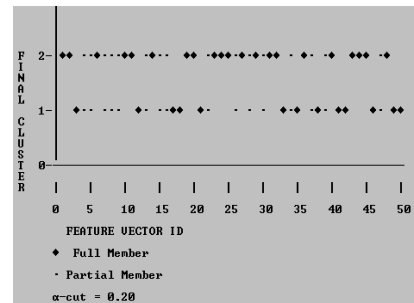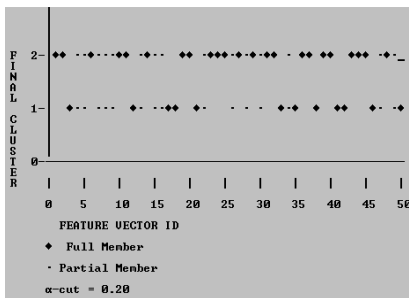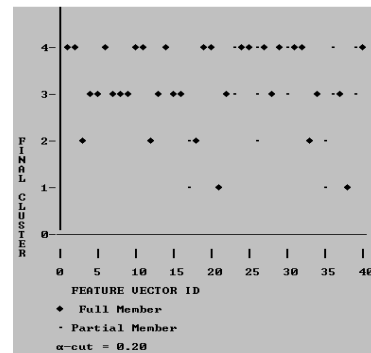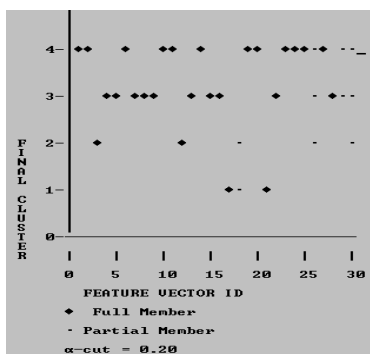

Figure 3: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature SA and no. of clusters is 4.
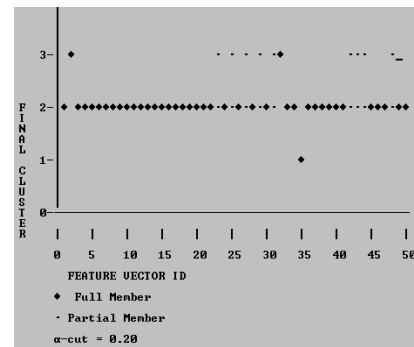
Figure 6: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature SA and no. of clusters is 3.
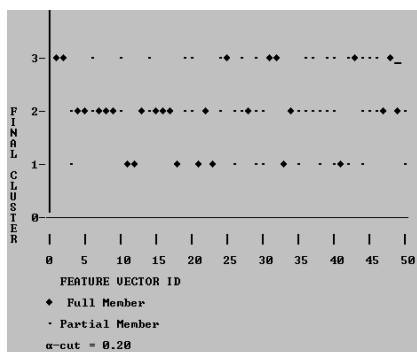
Figure 4: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature IQ and no. of clusters is 3.

Figure 7: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature IQ and no. of clusters is 2.

7

Figure 8: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature AM and no. of clusters is 2

Figure 11: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature AM and no. of clusters is 4

Figure 9: Graphical representation of the membership values of feature vectors produced by GK algorithm when initialized with feature SA and no. of clusters is 2

Figure 12: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature SA and no. of clusters is 4

Figure 10: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature IQ and no. of clusters is 4.

Figure 13: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature IQ and no. of clusters is 3

Figure 14: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature AM and no. of clusters is 3

Figure 17: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature AM and no. of clusters is 2

Figure 15: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature SA and no. of clusters is 3

Figure 18: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature SA and no. of clusters is 2

Figure 16: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature IQ and no. of clusters is 2

Figure 19: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature IQ, no. of clusters is 4 and no. of feature vectors is 40

9

Figure 20: Graphical representation of the membership values of feature vectors produced by FCM algorithm when initialized with feature IQ, no. of clusters is 4 and no. of feature vectors is 30
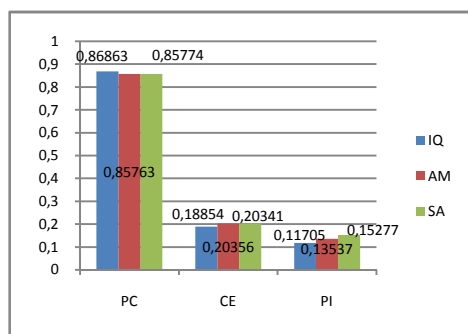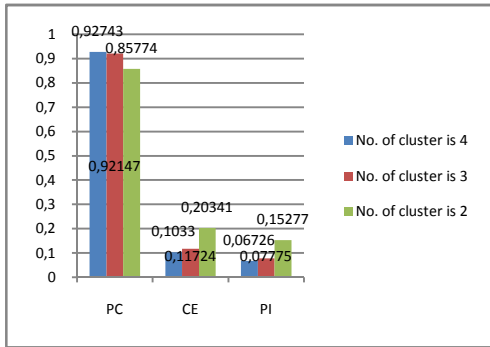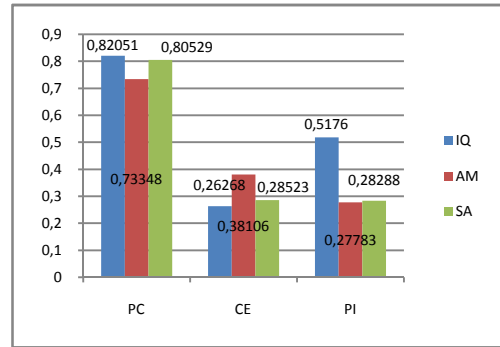

Figure 23: Comparison of the values of different validity measures of FCM algorithm with different features for initializing the membership values when the no. of clusters is 2.
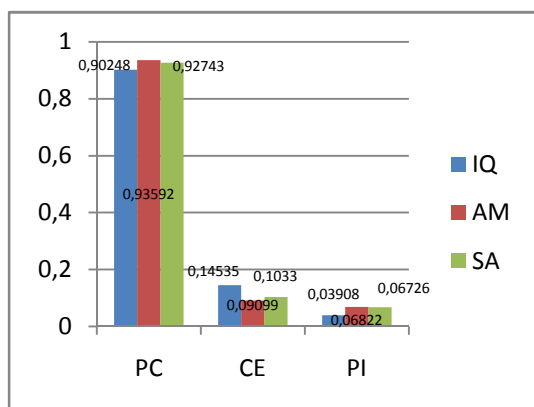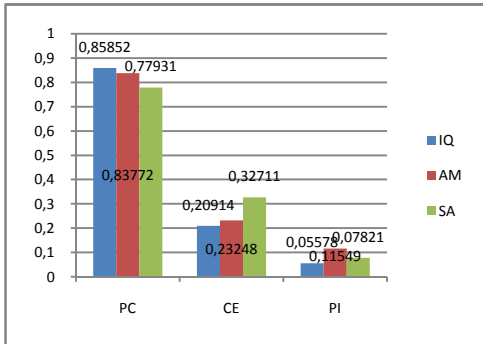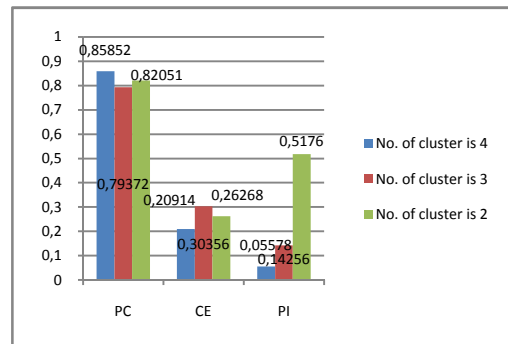

Figure 21: Comparison of the values of different validity measures of FCM algorithm with different features for initializing the membership values when the no. of clusters is 4.
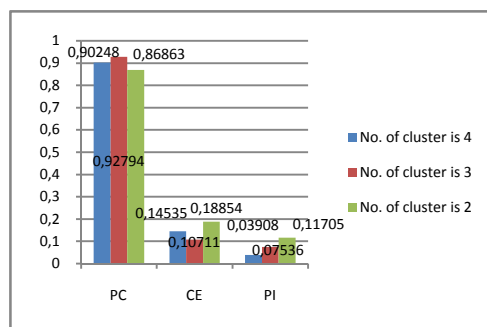
Figure 24: Comparison of the values of different validity measures of FCM algorithm at different no. of clusters when the feature for initializing the membership values is IQ.
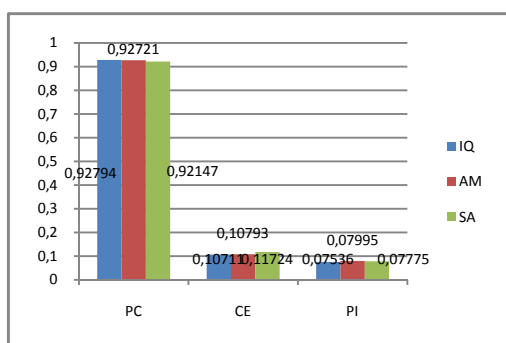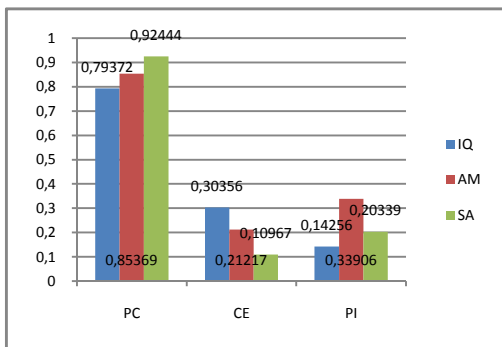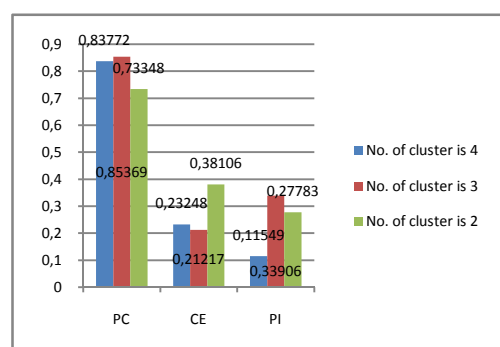

Figure 22: Comparison of the values of different validity measures of FCM algorithm with different features for initializing the membership values when the no. of clusters is 3.
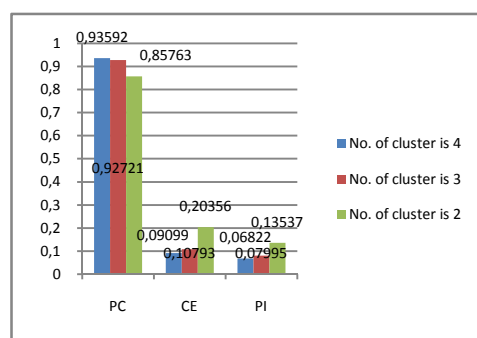
Figure 25: Comparison of the values of different validity measures of FCM algorithm at different no. of clusters when the feature for initializing the membership values is AM.
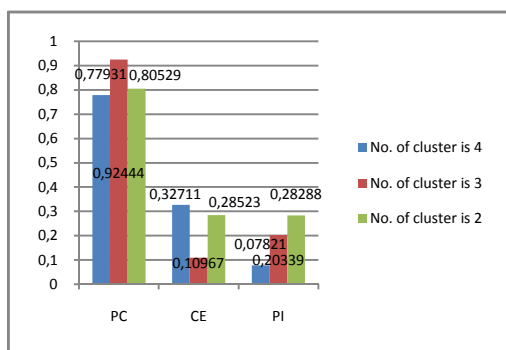
10

Figure 26: Comparison of the values of different validity measures of FCM algorithm at different no. of clusters when the feature for initializing the membership values is SA.
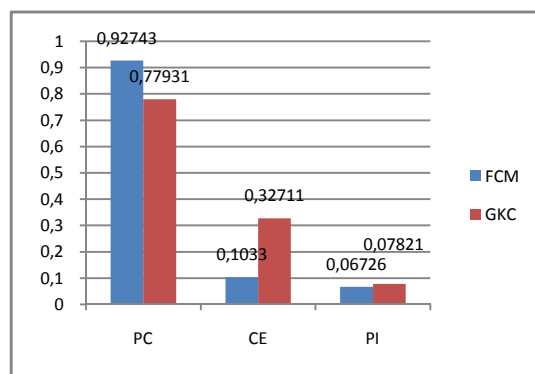

Figure 29: Comparison of the values of different validity measures of GK algorithm with different features for initializing the membership values when the no. of clusters is 2.

Figure 27: Comparison of the values of different validity measures of GK algorithm with different features for initializing the membership values when the no. of clusters is 4.

Figure 30: Comparison of the values of different validity measures of GK algorithm at different no. of clusters when the feature for initializing the membership values is IQ

Figure 28: Comparison of the values of different validity measures of GK algorithm with different features for initializing the membership values when the no. of clusters is 3.

Figure 31: Comparison of the values of different validity measures of GK algorithm at different no. of clusters when the feature for initializing the membership values is AM

11

Figure 32: Comparison of the values of different validity measures of GK algorithm at different no. of clusters when the feature for initializing the membership values is SA

Figure 35: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is SA and the no. of cluster is 4
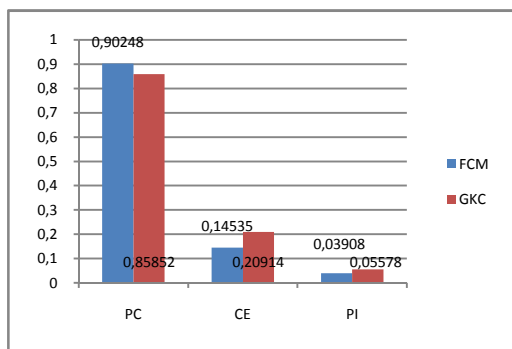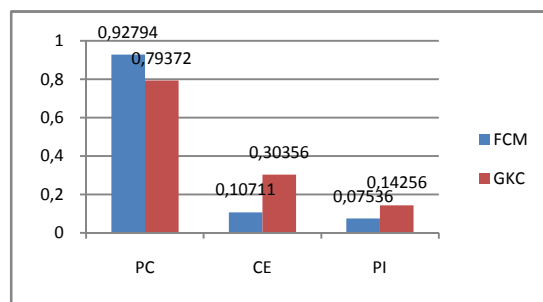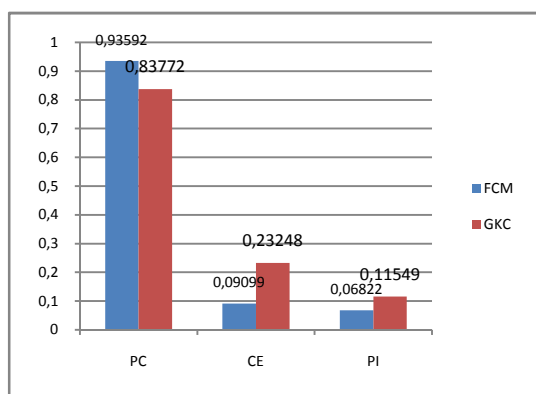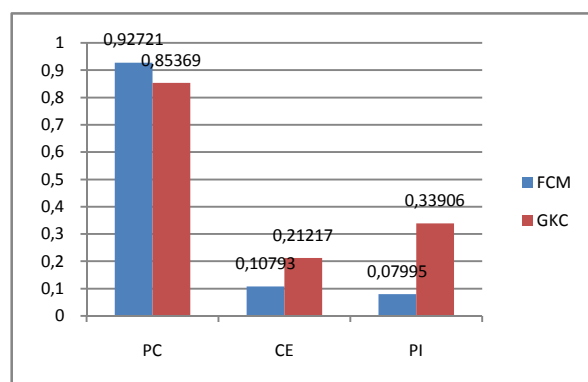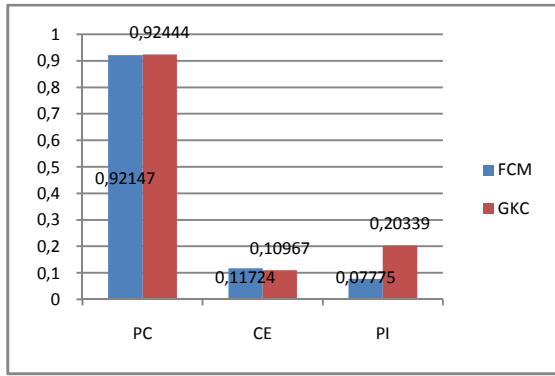

Figure 33: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is IQ and the no. of cluster is 4

Figure 36: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is IQ and the no. of cluster is 3
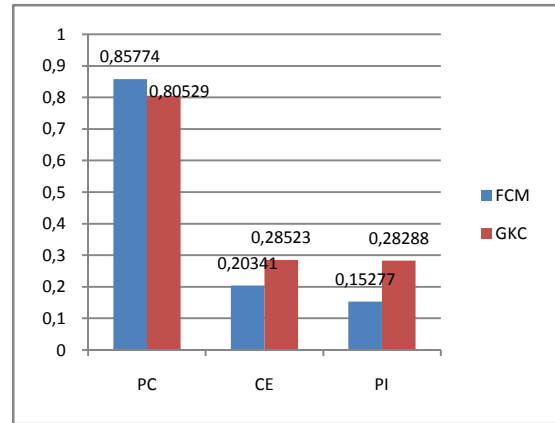

Figure 34: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is AM and the no. of cluster is 4

Figure 37: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is AM and the no. of cluster is 3

Figure 38: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is SA and the no. of cluster is 3
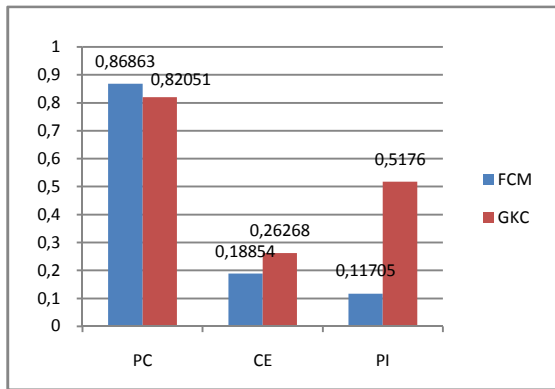

Figure 39: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is IQ and the no. of cluster is 2

Figure 40: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is AM and the no. of cluster is 2

Figure 41: Comparison of the values of different validity measures of FCM and GK algorithm when the feature for initializing the membership values is SA and the no. of cluster is 2
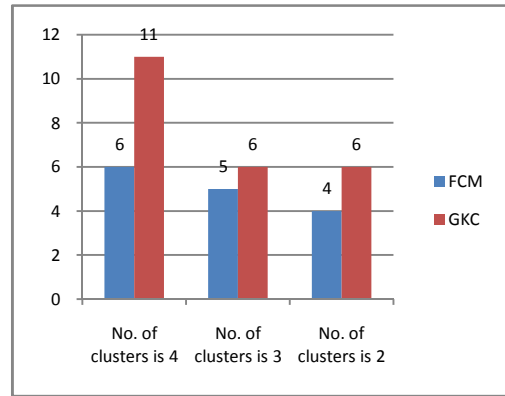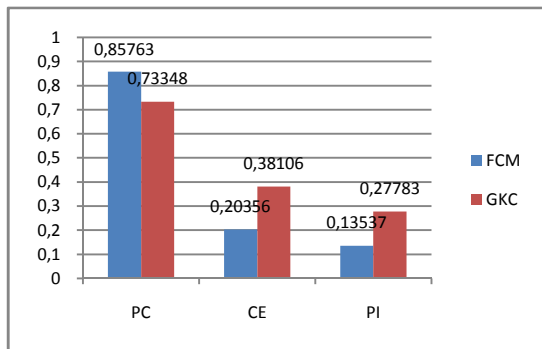

Figure 42: Comparisons of the no. of iterations of FCM and GK algorithm at different no. of clusters when the feature for initializing the membership values is IQ
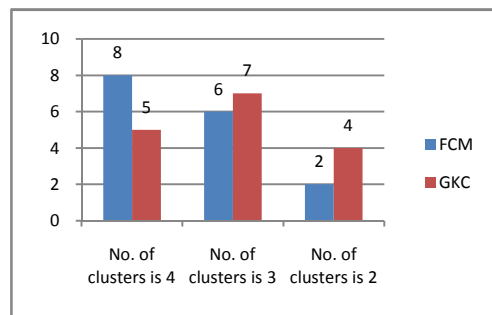

Figure 43: Comparisons of the no. of iterations of FCM and GK algorithm at different no. of clusters when the feature for initializing the membership values is AM
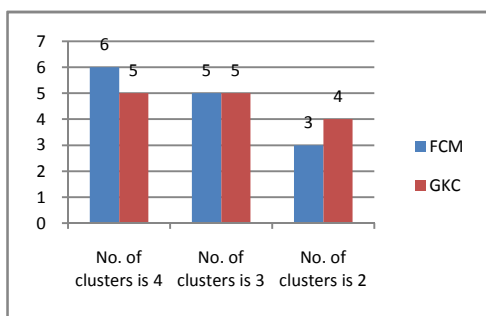
Figure 44: Comparisons of the no. of iterations of FCM and GK algorithm at different no. of clusters when the feature for initializing the membership values is SA
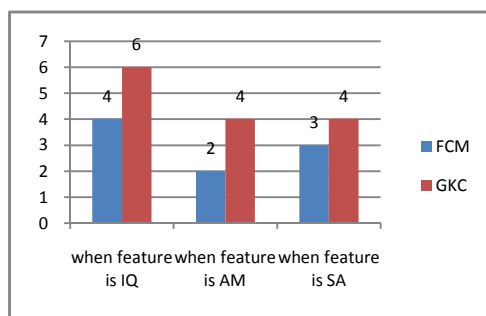
Figure 47: Comparisons of the no. of iterations of FCM and GK algorithm with different features for initializing the membership values and when the no. of clusters is 2
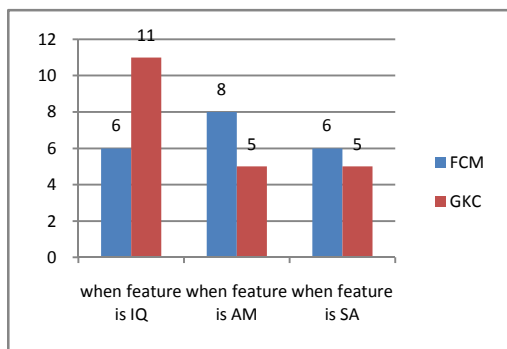

Figure 45: Comparisons of the no. of iterations of FCM and GK algorithm with different features for initializing the membership values and when the no. of clusters is 4
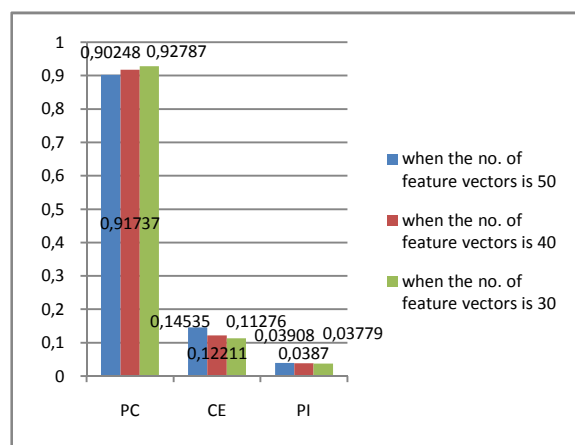
Figure 48: Comparisons of the values of different validity measures of FCM algorithm at different sizes of dataset.
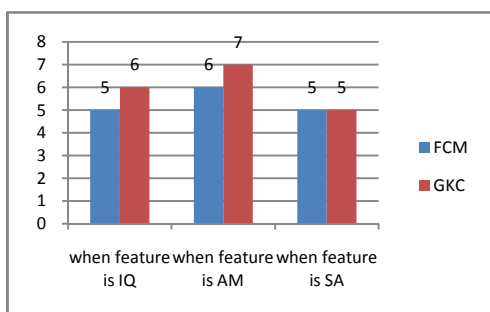

Figure 46: Comparisons of the no. of iterations of FCM and GK algorithm with different features for initializing the membership values and when the no. of clusters is 3
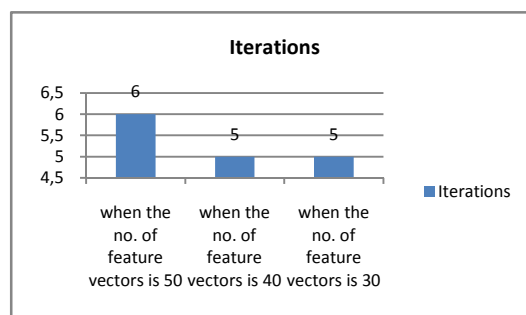
Figure 49: Comparisons of the no. of iterations of FCM algorithm at different sizes of dataset.

14

5. Conclusions

The results of our present work reveal that the performances of both FCM and GK algorithms depend on the initial membership values of the feature vectors, the no. of the predefined clusters and the size of the dataset. These performances do not show any specific rule for the first two aspects but for the third one. The performance of FCM increases as the size of the dataset reduces. On the other hand GK algorithm does not perform well when the size of the dataset is very small as the respective norm inducing matrix becomes singular. However the overall performance of FCM is better as compared to that of GK on the dataset we have used.

References

1. Zadeh, L. A. (1965), Fuzzy Sets, *Information and Control*, Vol.8, Issue 3, 338-353.
2. Zimmermann, H.J. (1991), Fuzzy Set Theory and its Applications, Second Edition, Kluwer Academic Publishers, Boston Massachusetts.
3. Dewit, G. W. (1982), Underwriting and Uncertainty, *Insurance: Mathematics and Economics*, Vol.1, Issue 4, 277-285.
4. Ostaszewski, K. (1993), An Investigation into Possible Applications of Fuzzy Sets Methods in Actuarial Science, Society of Actuaries, Schaumburg, Illinois.
5. Baruah, H. K. (2011a), Towards Forming a Field of Fuzzy Sets, *International Journal of Energy, Information and Communications*, vol. 2, Issue 1, 16-20.
6. Baruah, H. K. (2011b), The Theory of Fuzzy Sets: Beliefs and Realities, *International Journal of Energy, Information and Communications*, vol. 2, Issue 2, 1-22.
7. Das, S. (2012), A New Approach to Visualization of Fuzzy Set Operations, *International Journal of Energy, Information and Communications*, Vol. 3, Issue 3, 29-40.
8. Park, Y. and J. Park (2010), Interactive Visualization of Fuzzy Set Operations, *SPIE-IS&T*, Vol. 7530 753002, 1-12.
9. Derrig, R. A. and K. M. Ostaszewski (1995), Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification, *Journal of Risk and Insurance*, Vol. 62, Issue 3, 447-482.
10. Bezdek, J.C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
11. Das, S. (2013), Pattern Recognition using the Fuzzy c-means Technique, *International Journal of Energy, Information and Communications*, vol. 4, Issue 1, 1-14.
12. Das, S. and H. K. Baruah (2013), Application of Fuzzy C-Means Clustering Technique in Vehicular Pollution, *Journal of Process Management – New Technologies,* Vol. 1, Issue 3, 96-107.
13. Gustafson, D.E. and W.C. Kessel (1979), Fuzzy clustering with a fuzzy covariance matrix, *Proc. IEEE CDC*, San Diego, CA, USA, 761–766.
14. Wang, H. , C. Huang, L. Yao, Y. Qian and X. Jiang(2011), Application of Gustafson-Kessel clustering algorithm in the pattern recognition for GIS, *Przegland Elektrotechniczny (Electrical Review),* ISSN 0033-2097, R. 87 NR, 215-219.