Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

ORIGINAL ARTICLE

Using incremental general regression neural network for learning mixture models from incomplete data

Ahmed R. Abas

Department of Computer Science, University College in Lieth for Male Students, Umm Al-Qura University, Makka Al-Mukarrama, Lieth, Kingdom of Saudi Arabia

Abstract   Finite mixture models (FMM) is a well-known pattern recognition method, in which parameters are commonly determined from complete data using the Expectation Maximization (EM) algorithm. In this paper, a new algorithm is proposed to determine FMM parameters from incomplete data. Compared with a modified EM algorithm that is proposed earlier the proposed algorithm has better performance than the modified EM algorithm when the dimensions containing missing values are at least moderately correlated with some of the complete dimensions.

© 2011 Faculty of Computers and Information, Cairo University.
Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

FMM is a partitional probabilistic algorithm that is commonly used in cluster analysis [1–17]. Parameters of the FMM are usually determined via the EM algorithm [18] from complete data. It is shown that when the mechanism that describes the

E-mail address: armohamed@uqu.edu.sa

Production and hosting by Elsevier

occurrences of missing values in the data is missing at random (MAR), maximum likelihood inferences about parameters of the FMM that represents the data can be made from complete data [19]. In other words, the likelihood is simply the density of the complete data, which is a function of the FMM parameters [20]. Missing data are considered MAR if the probability of having missing values in a certain dimension of the data depends on values of the other complete dimensions but not on the true values of missing values [21]. However, determining FMM parameters using only complete data requires the data size to be large in order to obtain good fitting of the data distribution [22]. Due to practical problems in factorising the likelihood, it is commonly maximized iteratively via the EM algorithm for complete data. It is shown that the EM algorithm can be used in determining parameters of a multivariate normal distribution from incomplete data [21]. Missing values in the data are estimated using multivariate regression. The complete data with the residual covariances of the estimated values are then used in estimating the mean and the covariance matrix of the multivariate normal distribution. The regression

coefficients and the residual covariances are estimated from the current model parameters such that the likelihood is maximized.

The EM algorithm is modified such that it can determine parameters of a mixture of multivariate normal distributions from incomplete data [23]. Missing values and some other statistics are estimated in the E-step from each model component as they are come from that component. These values with the observed ones are then used in the M-step to determine the parameters of that model. However, accuracy of the estimated values is limited due to the small number of estimators (kernels) used in the estimation of missing values. The number of these estimators equals to the number of components in the FMM. This in turn affects the accuracy of the leaned FMM parameters and hence its clustering results. This problem can be overcome if missing values in the data are estimated with high accuracy first, and then FMM parameters are learned via the EM algorithm. The computational complexity of the modified EM algorithm [23] is reduced during the EM iterations by incorporating two types of auxiliary binary indicator matrices corresponding to the observed and unobserved components of each datum [24]. The resulting EM algorithm is compared with a new proposed Data Augmentation (DA) computational algorithm for learning normal mixture models when the data are missing at random [24]. Experimental results show that DA imputation has considerable promising accuracy in the prediction of missing values when compared to the EM imputation, especially when the missing rate increases. However, both algorithms impute missing values using mixture model parameters and hence their imputations are sensitive to the prior information about density functions of mixture components and the size of data that are fully observed. A supervised classification method, called robust mixture discriminant analysis (RMDA), is proposed to handle label noised data [25]. The RMDA algorithm uses only fully observed data to learn mixture model parameters, and then uses the resulting mixture model to estimate labels and detect noisy ones. However, imputations made by the RMDA algorithm are sensitive to prior information about density functions of mixture components, the size of data that are fully observed and assumptions such as all the uncertain labels are in one feature.

In this paper, a new algorithm for determining FMM parameters from incomplete data is proposed. The proposed algorithm is a combination of the Incremental General Regression Neural Network (IGRNN) [26] and the EM algorithm [18]. It estimates missing values in the data using the IGRNN and then uses the EM algorithm to determine FMM parameters. Performance of the proposed algorithm is investigated against the use of the modified EM algorithm [23] in learning FMM from incomplete data. The motivation of this investigation is to test the effect of the accuracy of the estimated values of missing data from both algorithms on the clustering behaviour of the resulting mixture models. This paper is organised as follows. Section 2 describes the modified EM algorithm [23]. Section 3 describes the proposed algorithm. Section 4 describes experiments that are carried out to compare both of the described algorithms. Section 5 discusses the results obtained from experiments. Section 6 concludes the paper and summarises its findings.

2. The modified EM algorithm [23]

The modified EM algorithm [23] is proposed for determining parameters of a mixture of multivariate normal distribution from incomplete data provided that missing values are Missing At Random (MAR) [21]. We refer to this algorithm as the EMH algorithm in the rest of this paper. The EMH algorithm is described as follows.

Suppose that $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a data set that is composed of n patterns and d dimensions such that each pattern is represented as $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T$. This data set is assumed to be generated randomly from a mixture of K multivariate distributions with unknown mixing coefficients $p(c)$, where $c = 1, 2, \ldots, K$. Let the probability density component of $\mathbf{x}_i$ from the kth multivariate distribution be $p(\mathbf{x}_i | k)$. The commonly used distribution is the Gaussian $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ are the mean and the covariance matrix, respectively [23]. This distribution is preferred to other distributions for the EM algorithm because it has a small number of parameters that need to be estimated and also computing its derivative is simple. The density of $\mathbf{x}_i$ can be written as $p(\mathbf{x}_i) = \sum_{c=1}^{K} p(c) p(\mathbf{x}_i | c)$, where $\sum_{c=1}^{K} p(c) = 1, 0 \leqslant p(c) \leqslant 1$, for $c = 1, 2, \ldots, K$.

When $\mathbf{X}$ contains MAR missing values, the pattern $\mathbf{x}_i$ can be denoted as $\mathbf{x}_i = (\mathbf{x}_{i,obs}, \mathbf{x}_{i,mis})$, where $\mathbf{x}_{i,obs}$ stands for the observed values, and $\mathbf{x}_{i,mis}$ stands for the missing values for pattern $\mathbf{x}_i$. In fitting the FMM, there are two types of missing values that have to be considered; one is the values of the cluster membership dimensions for each pattern $\mathbf{z}_i = [z_{i1}, z_{i2}, \ldots, z_{iK}]^T$, where $i = 1, 2, \ldots, n$ and the other is the missing values in the data matrix $\mathbf{X}$. Each value in the cluster membership vector z_{ij} represents the probability by which a certain pattern $\mathbf{x}_i$ in the data matrix $\mathbf{X}$ is generated from the jth component in the FMM. In the EM algorithm, $\mathbf{z}_i$ is approximated by the posterior probabilities when $\mathbf{x}_i$ is fed to the model. In the E-step, all $\hat{\mathbf{z}}_i$'s are determined besides some statistical moments necessary for the M-step using observed values for each pattern. While in the M-step, the new estimates of the FMM parameters are determined using the observed data and the statistical moments determined in the E step. Both the E and the M steps are alternated until convergence. For more details and description of this algorithm see [23].

3. The proposed algorithm for determining FMM parameters from incomplete data

The Incremental General Regression Neural Network (IGRNN) [26] is proposed for estimating missing values in numeric data sets. It is shown that the IGRNN produces highly accurate estimations for missing values in the case of a data set that has strong correlations among its dimensions [26]. In this Section, it is proposed to combine this algorithm with the EM algorithm [18] for determining FMM parameters from incomplete data. First, the proposed algorithm estimates missing values in the data set using the IGRNN. Second, it estimates parameters of the FMM that can be used in clustering the data using the resulting complete data and the EM algorithm. The proposed algorithm is referred to as the IGRNNEM algorithm in the rest of this paper. The IGRNNEM algorithm is described as follows:
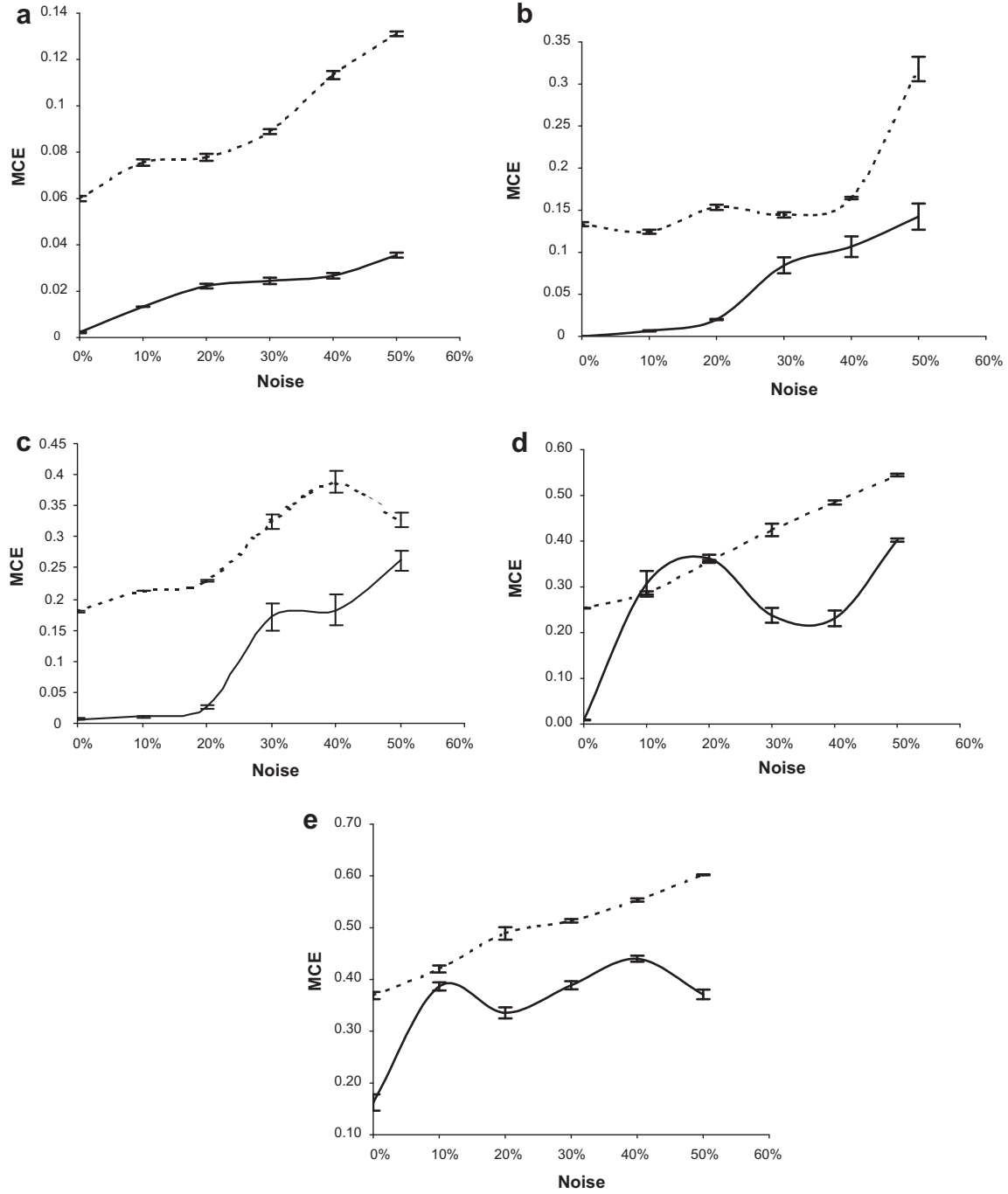
Figure 1 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the first data set. Different amounts of missing values that are MCAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

3.1. The IGRNNEM algorithm

Step 1. The IGRNN algorithm [26] is proposed to estimate missing values when correlations among dimensions of the data are large. Suppose a data set of size n patterns that contains a group of fully observed dimensions X and a dimension Y that contains missing values. In this algorithm, estimating missing values in Y requires all data patterns that have missing values on that dimension to be sorted according to how close they are to complete data patterns in the sub-space that is composed of X dimensions. Patterns that contain missing values are sorted in a descending order according to their closeness to complete patterns.

Step 2. The estimation of the missing value in the first pattern is $\hat{y}(\mathbf{x})$ is computed as a weighted average of all observed values y_i along the Y dimension, where each
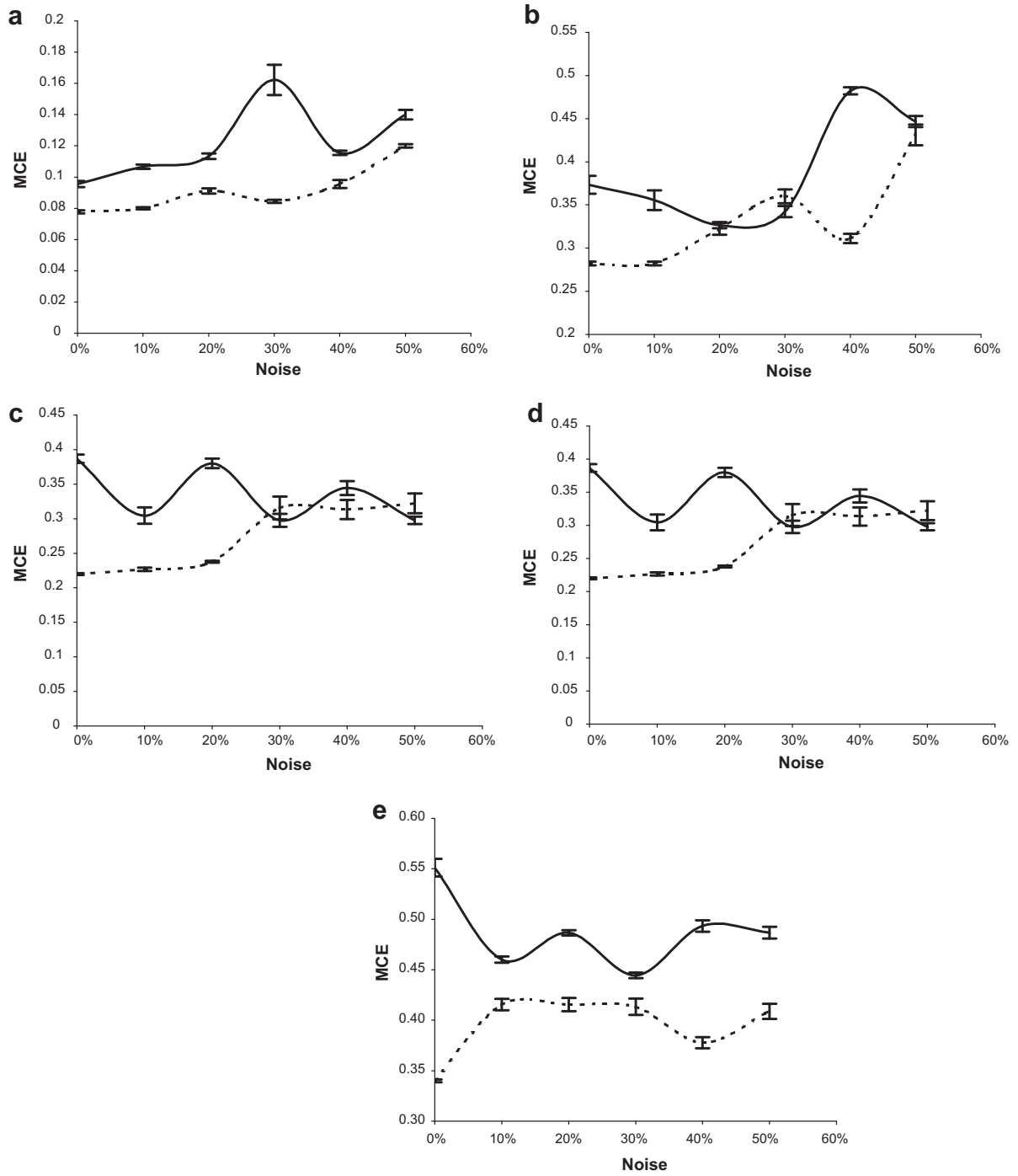
Figure 2 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the second data set. Different amounts of missing values that are MCAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

observed value is weighted exponentially according to its Euclidean distance from y along X dimensions as shown in Eq. (1).

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \tag{1}$$

where D_i^2: is the Euclidean distance between pattern $(\hat{y}, x)$ and pattern (y_i, x_i) and it is computed as shown by Eq. (2), σ: is a smoothing parameter and it is determined empirically as shown in [26].

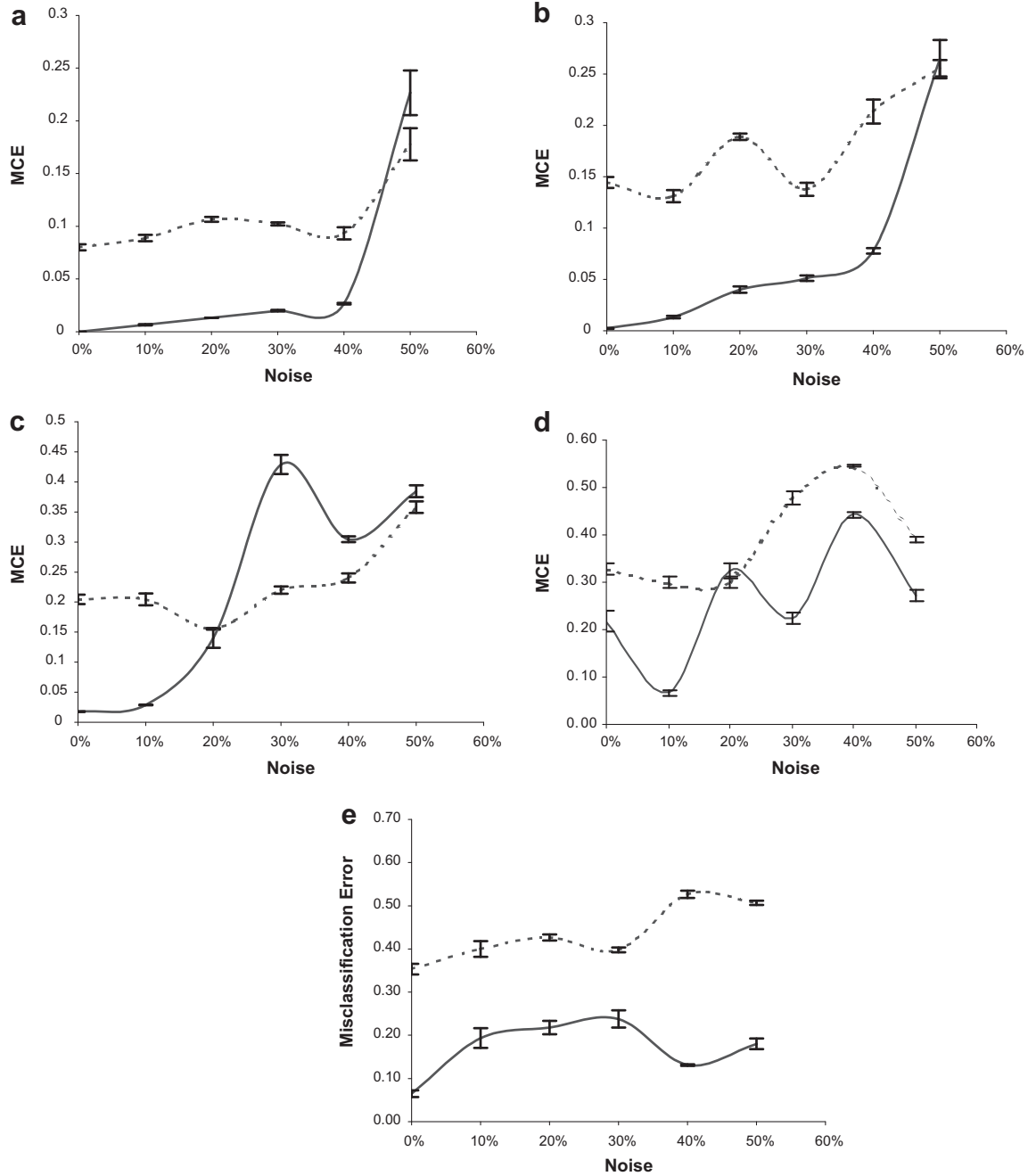$$D_i^2 = (x - x_i)^T(x - x_i) \tag{2}$$

Figure 3 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the first data set. Different amounts of missing values that are MAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

Step 3. After estimating its missing value, the pattern is added to the complete ones and then used in estimating missing values in the next incomplete patterns in the sorted group.

Step 4. Steps 2 and 3 are repeated until all missing values in the data set $(\widehat{Y}, X)$ are estimated.

Step 5. The resulting complete data set is fed to the EM algorithm [18] to estimate parameters of the FMM as follows. In the E-step, compute the following quantities for each model component c in the FMM.

- The posterior probabilities vector $\mathbf{z}_i$ for all feature vectors in the data set.

$$\hat{z}_{ic} = \frac{\widehat{P}(c)p(\hat{y}_i, \mathbf{x}_i|\boldsymbol{\theta}_c)}{\sum_{j=1}^{K}\widehat{P}(j)p(\hat{y}_i, \mathbf{x}_j|\boldsymbol{\theta}_j)} \tag{3}$$

where $\sum_{c=1}^{K}P(c) = 1$, and $0 \leqslant P(c) \leqslant 1$, $p(\hat{y}_i, \mathbf{x}_i|\boldsymbol{\theta}_k) = N(\hat{y}_i, \mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean and the covariance matrix of the kth component in the FMM. The total density
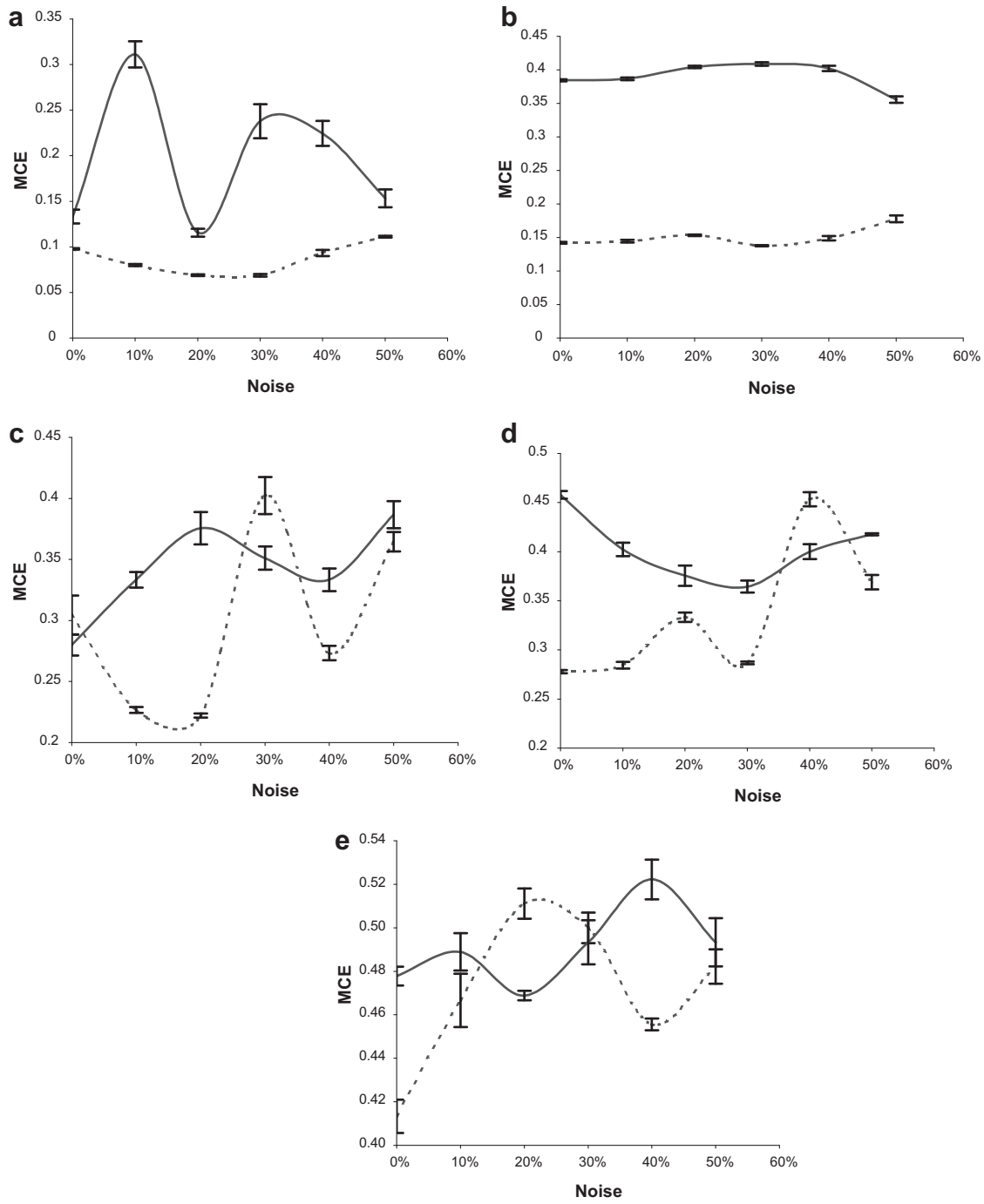
Figure 4 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the second data set. Different amounts of missing values that are MAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

of $(\hat{y}_i, \mathbf{x}_i)$ from the FMM is then computed as $p(\hat{y}_i, \mathbf{x}_i) = \sum_{c=1}^{K} P(c) p(\hat{y}_i, \mathbf{x}_i | \boldsymbol{\theta}_c)$.

Step 6. In the M-step, compute parameters of each component c in the FMM.

$$\widehat{P}(c) = \frac{1}{N} \sum_{j=1}^{N} \hat{z}_{jc} \tag{4}$$

$$\widehat{\boldsymbol{\mu}}_c = \frac{1}{N\,\widehat{P}(c)} \sum_{j=1}^{N} z_{jc} (\hat{y}_j, \mathbf{x}_j) \tag{5}$$

$$\widehat{\boldsymbol{\Sigma}}_c = \frac{1}{N\,\widehat{P}(c)} \sum_{j=1}^{N} z_{jc} (\hat{y}_j, \mathbf{x}_j)(\hat{y}_j, \mathbf{x}_j)^T - \widehat{\boldsymbol{\mu}}_c \widehat{\boldsymbol{\mu}}_c^T \tag{6}$$

Step 7. After convergence, save the resulting FMM. This FMM could be used in clustering the data set $(\widehat{Y}, X)$.
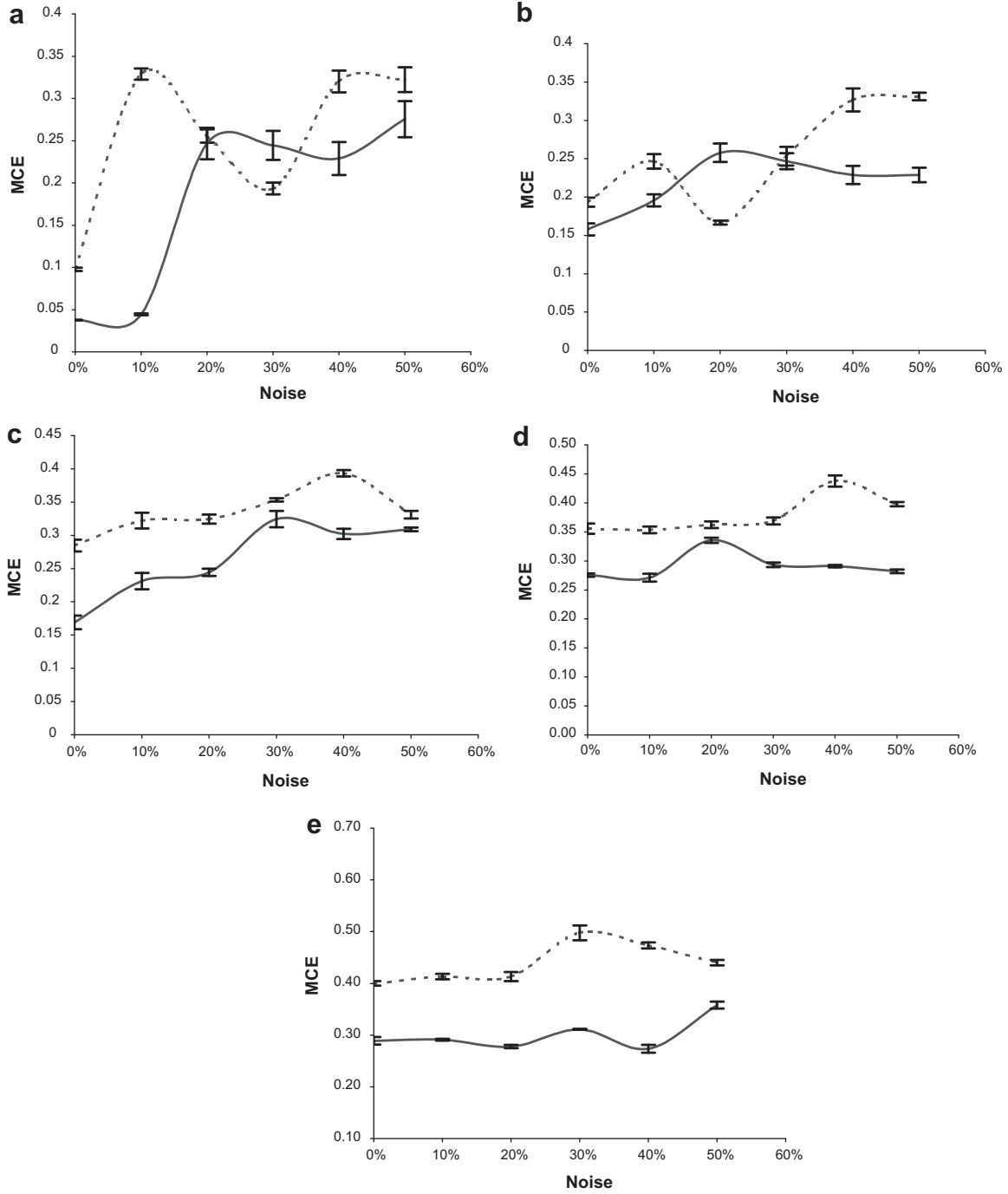
Figure 5 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the Iris data set. Different amounts of missing values that are MCAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

4. Experimental results and discussion

The IGRNNEM algorithm is compared with the EMH algorithm in estimating parameters of the FMM for clustering incomplete data. Different mechanisms of missing values are used in the comparison such as Missing Completely At Random (MCAR) and Missing At Random (MAR) [21]. Also, the comparison is made using different percentages of missing values in multiple dimensions and different percentages of noisy patterns in the data set. The added noise comes from a multivariate normal distribution whose parameters are $N(0, 0.1^*\Sigma)$, where the mean is at the origin of the space and the covariance matrix is 10% of the data covariance matrix. The noise is added randomly to the patterns. Four data sets are used in the comparison. The first and the second data sets
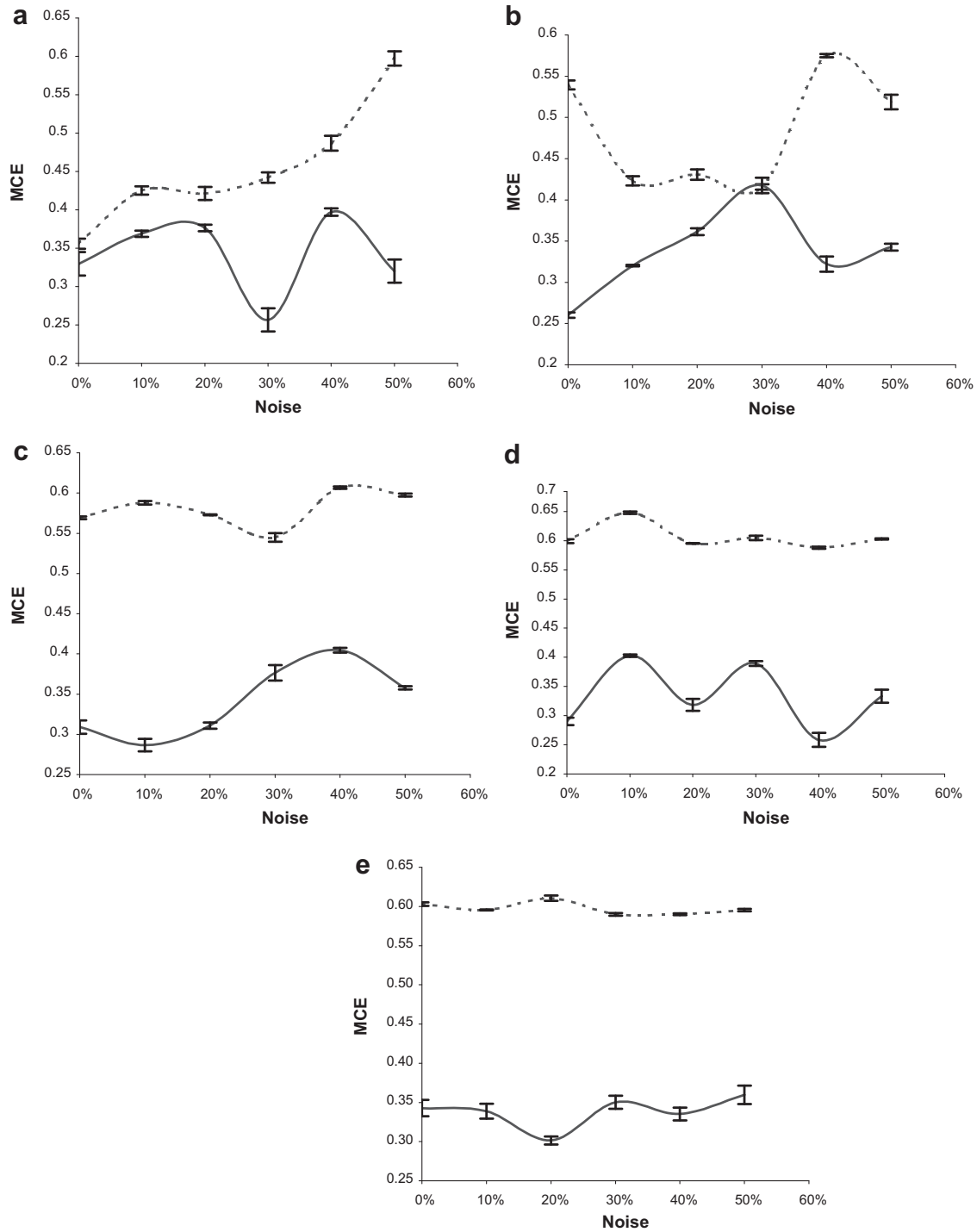
Figure 6 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the Wine data set. Different amounts of missing values that are MCAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

are artificial data sets while the third and the fourth data sets are true data sets.

4.1. The first and the second data sets

Both data sets are similar to the first and the second data sets that are used in [26]. Both data sets have 150 patterns and four dimensions. Each data set is generated from three clearly sep-

arated Gaussian shaped clusters such that 50 patterns are generated from each cluster. Cluster centres of the first data set are represented by the following vectors: $\boldsymbol{\mu}_1 = [2\ 2\ 2\ 2]^T$, $\boldsymbol{\mu}_2 = [4\ 4\ 4\ 4]^T$, and $\boldsymbol{\mu}_3 = [6\ 6\ 6\ 6]^T$, while cluster centres of the second data set are represented by the following vectors: $\boldsymbol{\mu}_1 = [2\ 2\ 2\ 2]^T$, $\boldsymbol{\mu}_2 = [2\ 2\ 6\ 2]^T$, and $\boldsymbol{\mu}_3 = [2\ 2\ 2\ 6]^T$. Covariance matrices of all clusters are similar ($\boldsymbol{\Sigma}_i = 0.5\mathbf{I}_4$, $i = 1{:}3$), where $\mathbf{I}_4$ is the identity matrix of order four. Missing values
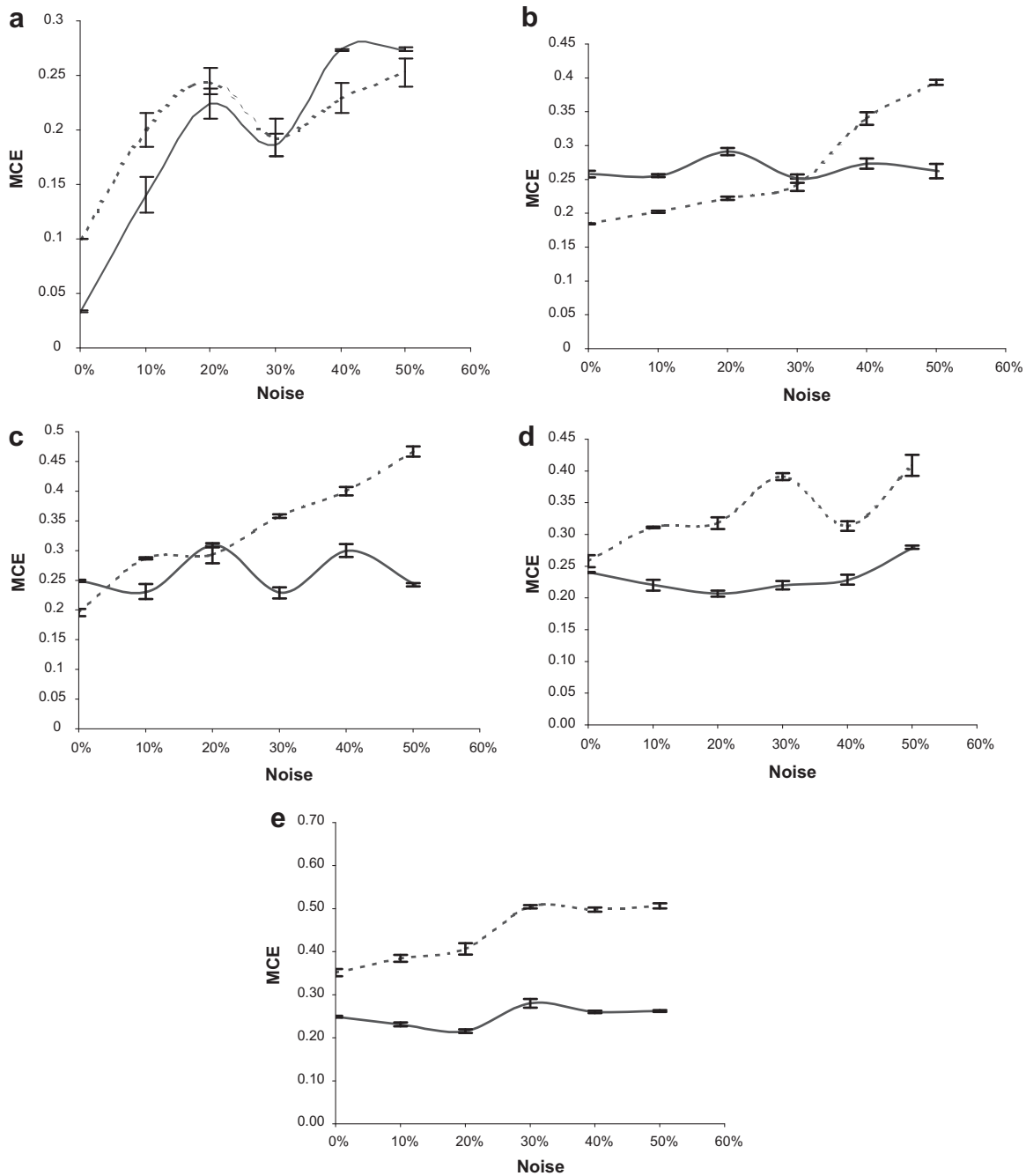
Figure 7 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the Iris data set. Different amounts of missing values that are MAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

are put in the third and in the fourth dimensions of each data set. These data sets compare the clustering performances of FMM's resulting from both algorithms in the extreme cases of correlations among data dimensions [26].

4.2. The third data set

This data set is the Iris data set that is commonly used in statistical experiments since it was used in [27]. The data contains 150 patterns. Each pattern is a vector in four dimensions. The patterns are representing three clusters such that every cluster

contains 50 patterns. The missing values are put in the third and in the fourth dimensions.

4.3. The fourth data set

This data set is the Wine data set that is a well-known data set for statistical analysis. Both of this data set and the Iris data set can be obtained from the UCI machine learning repository.[1] The data contains 178 patterns. Each pattern is a vector

---

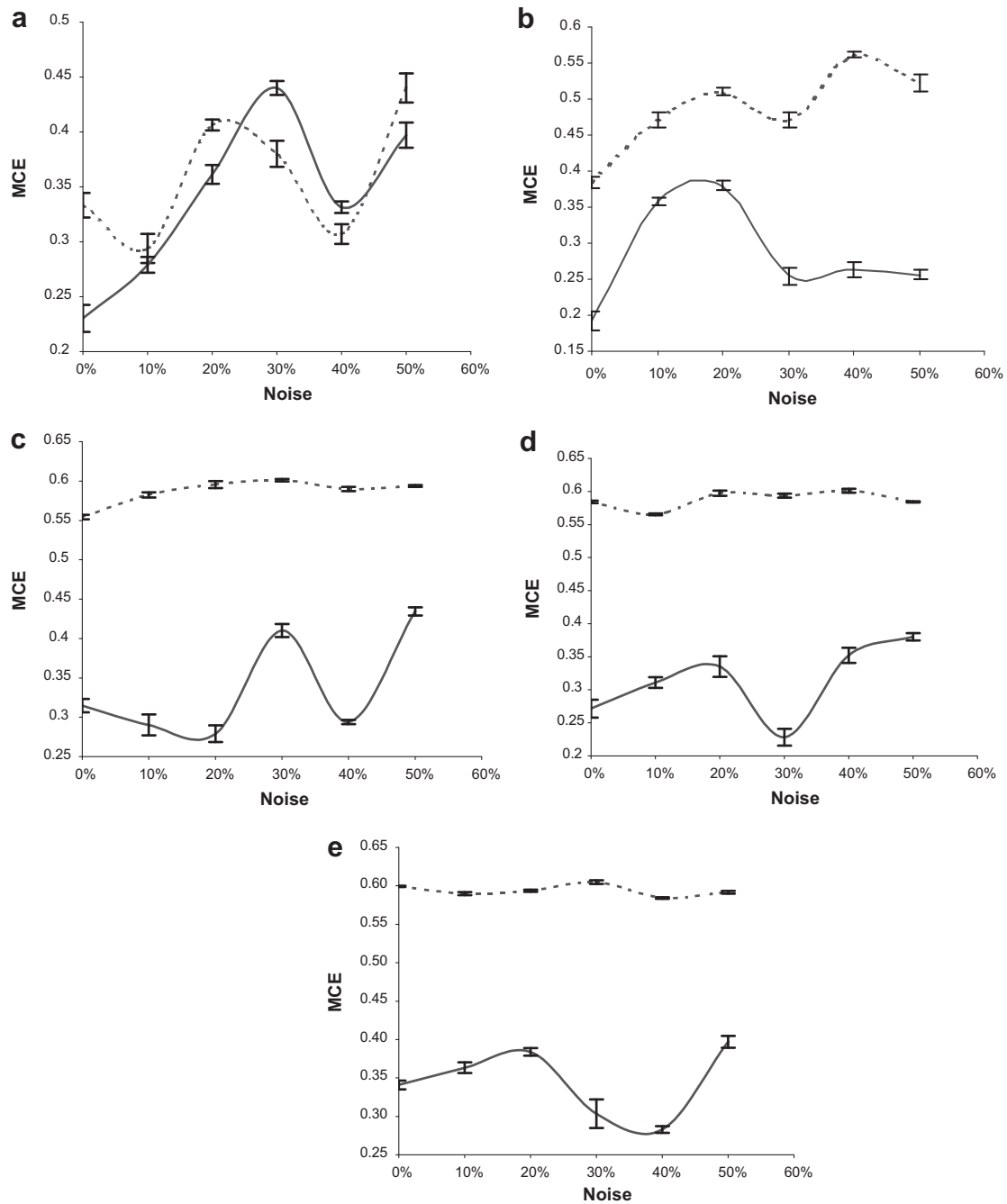[1] http://www.ics.uci.edu/~mlearn/MLRepository.html.

Figure 8 Comparing performances of the IGRNNEM (solid line) and the EMH (dashed line) in clustering incomplete data that contains different amounts of noisy patterns using the Wine data set. Different amounts of missing values that are MAR are used: (a) 10%, (b) 20%, (c) 30%, (d) 40%, and (e) 50%. Vertical bars represent the standard deviations of the errors from three different experiments.

in a 13-dimension space. The patterns represent three clusters, which contain 59, 71, and 48 patterns, respectively. The missing values are put in the sixth and in the seventh dimensions. When missing values are MCAR, they are randomly scattered in the specified dimensions for each data set. Meanwhile, when missing values are MAR, they are randomly scattered in the specified dimensions for each data set restricted by the value of the second dimension of each pattern. The value of the second dimension of any pattern that has missing values should be greater than either 3 in the first data set, 1.5 in the second data set, 2.99 in the Iris data set, or 1.85 in the Wine data set.

Since the EM algorithm converges to a local maximum for the likelihood function, the algorithm in both of the algorithms compared is initialised using 20 different random initialisations. After convergence, the results of the run that produces the maximum likelihood are selected as the parameters of the mixture model that represents the data. The convergence is achieved when the percentage difference in the likelihood between iterations $(t, t - 10)$ is less than or equal 0.01. This condition overcomes the EM problem of slow convergence near the local maximum of the likelihood function. The FMM resulting from the EM algorithm in both of the

algorithms compared is evaluated by computing the Mis-Clustering Error (MCE). Each pattern is allocated to the cluster that has the maximum posterior probability given that pattern, and the number of incorrectly clustered patterns (N_r) is computed. The MCE is then computed such that MCE = N_r/N, where N is the total number of patterns.

Figs. 1–8 (see, the Appendix) show the results of the experiments. Each figure shows the mean of the results obtained from three different experiments, in which different groups of patterns are selected to have missing values. The error bars represent the standard deviation of the results multiplied by ± 0.1. Figs. 1 and 3 show that the IGRNNEM algorithm produces smaller MCE and hence more accurate results than the EMH algorithm when correlations among all dimensions of the data set are large. Meanwhile, Figs. 2 and 4 show that both algorithms produce large MCE and hence inaccurate results but the IGRNNEM is the worst when correlations among all dimensions of the data set are too small. Figs. 5 and 7 show that the IGRNNEM algorithm generally produces smaller MCE and hence more accurate results than the EMH algorithm in the case of the Iris data set. This data set has moderate correlations among its dimensions. Finally, Figs. 6 and 8 show that the IGRNNEM algorithm produces smaller MCE and hence more accurate results than the EMH algorithm in the case of the Wine data set although the error is slightly high. In this data set, most of the correlations among dimensions are small but dimensions 6 and 7, which contain missing values, have moderate correlations with some of the other dimensions.

The above results show that the IGRNNEM algorithm produces smaller MCE and hence more accurate estimations for the missing values than the EMH algorithm when correlations among all dimensions or at least among dimensions that contain missing values and some of the other dimensions of the data are at least moderate. This is because the IGRNNEM algorithm uses more estimators (kernels) than the EMH algorithm. Therefore, the mixture model resulting from the IGRN-NEM algorithm fits the data better than that resulting from the EMH algorithm as more accurate data are contributing in fitting the model. On the other hand, both algorithms produce large errors in estimating missing values and therefore fitting of the mixture models are bad when correlations among dimensions of the data are small. In this case, the EMH algorithm produces better results than the IGRNNEM algorithm although its error level is high. This is because of that the EMH algorithm depends on local correlations among data dimensions given each cluster in estimating missing values in the data and hence in estimating parameters of the model component that represent that cluster. Local correlations can be better than the overall correlations among dimensions of the data provided that the shape of each cluster is not a circle, a sphere, or a hyper-sphere. These shapes results in too small correlations among dimensions given each cluster. This result agrees with conclusions of the recently published work that is based on local tuning of the General Regression Neural Networks [28].

5. Conclusions

Finite mixture model method is a well-known method in cluster analysis. Parameters of the finite mixture are commonly estimated using the expectation maximization algorithm. As this algorithm requires complete data, a modified version of it is proposed earlier [23] to deal with incomplete data. In this paper, a new algorithm is proposed to estimate parameters of the finite mixture model from incomplete data. A comparison study of the proposed algorithm and the algorithm proposed in [23] in clustering incomplete data is given. It is shown that the proposed algorithm produces more accurate results than the other algorithm when correlations among all dimensions or at least among dimensions that contain missing values and some of the other dimensions of the data are at least moderate.

On the other hand, both algorithms produce bad results when correlations among dimensions of the data are small. The proposed algorithm is the worst if the cluster shapes are not circles, spheres, or hyper-spheres.

Appendix A. Results of experiments

This appendix contains Figs. 1–8 that show the results of experiments carried out in this paper.

References

[1] Banfield J, Raftery A. Model-based Gaussian and non-Gaussian clustering. J Biometrics 1993;49:803–21.

[2] Bishop C. Neural networks for pattern recognition. Oxford University Press; 1995.

[3] Meila M, Heckerman D. An experimental comparison of several clustering and initialization methods. Technical report MSR-TR-98-06. Redmond, WA, USA: Microsoft Research, Microsoft; 1998.

[4] Fraley C, Raftery A. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput J 1998;41:578–88.

[5] Fraley C, Raftery A. MCLUST: Software for model-based cluster and discriminant analysis. Technical report 342. Seattle, WA, USA: Department of Statistics, University of Washington; 1998b.

[6] Fraley C. Algorithms for model-based Gaussian hierarchical clustering. Soc Ind Appl Math (SIAM) J Sci Comput 1998;20:270–81.

[7] Cadez I, Smyth P. Probabilistic clustering using hierarchical models. Technical report 99-16. Irvine, USA: Department of Information and Computer Science, University of California; 1999.

[8] Posse C. Hierarchical model-based clustering for large datasets. J Comput Graph Stat 2001;10(3):464–86.

[9] Roberts S, Everson R, Rezek I. Maximum certainty data partitioning. J Pattern Recognit 1999;33:833–9.

[10] Webb A. Statistical pattern recognition. Arnold; 1999.

[11] Figueiredo M, Jain A. Unsupervised learning of finite mixture models. IEEE Trans Pattern Anal Machine Intell 2002;24(3):381–96.

[12] Fraley C, Raftery A. Model-based clustering, discriminant analysis, and density estimation. Technical report no. 380. Seattle, WA, USA: Department of Statistics, University of Washington; 2000.

[13] Williams C. A MCMC approach to hierarchical mixture modelling. Proc Adv NIPS 2000;12:680–6.

[14] Roberts S, Holmes C, Denison D. Minimum-entropy data partitioning using reversible jump Markov chain Monte Carlo. IEEE Trans Pattern Anal Machine Intell 2001;23:909–14.

[15] Yeung K, Fraley C, Murua A, Raftery A, Ruzz W. Model-based clustering and data transformations for gene expression data. Technical report UW-CSE-2001-04-02. Seattle, WA, USA:

Department of Computer Science and Engineering, University of Washington; 2001a.

[16] Szymkowiak A, Larsen J, Hansen L. Hierarchical clustering for datamining. In: Proceeding of the fifth international conference on knowledge-based intelligent information engineering systems and allied technologies KES-2001; 2001.

[17] Larsel J, Szymkowiak A, Hansen L. Probabilistic hierarchical clustering with labeled and unlabeled data. Int J Knowledge-Based Intell Eng Syst 2002;6(1):56–62.

[18] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J Royal Stat Soc 1977;B(39):1–38.

[19] Rubin DB. Inference and missing data. J Biometrica 1976;63:581–93.

[20] Hunt L, Jorgensen M. Mixture model clustering for mixed data with missing information. J Comput Stat Data Anal 2003;41:429–40.

[21] Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley and Sons; 1987.

[22] Cang S, Partridge D. Analysis of TRAUMA missing data. In: Proceeding of the annual conference of chinese automation and computer society in the UK; 2002. p. 445–51.

[23] Hunt LA. Clustering using finite mixture models. Ph.D. thesis. New Zeland: Department of Statistics, University of Waikato; 1996.

[24] Lin TI, Lee JC, Ho HJ. On fast supervised learning for normal mixture models with missing information. J Pattern Recognit 2006;39:1177–87.

[25] Bouveyron C, Girard S. Robust supervised classification with mixture models: learning from data with uncertain labels. J Pattern Recognit 2009;42:2649–58.

[26] Abas AR. An incremental general regression neural network for missing value estimation. Egypt Comput J 2010;37(2):1–23.

[27] Fisher RA. The use of multiple measurements in taxonomic problems. J Ann Eugenics 1936;7:179–88.

[28] Abas AR. Using general regression with local tuning for learning mixture models from incomplete data sets. Egypt Inform J 2010;11(2):49–57.