Clustering Based on LVQ and a Split and Merge Procedure

Fujiki Morii

Dept. of Information and Computer Sciences,
Nara Women's University, Nara 630-8506, Japan
fmorii@ics.nara-wu.ac.jp

Abstract. Although learning vector quantization (LVQ) based on learning concept is a typical clustering method, we cannot necessarily obtain satisfactory classification results for linearly separable data. In this paper, a new clustering method based on LVQ and a split and merge procedure is proposed to realize reliable classification. Introducing a criterion of whether or not there is only one cluster in each class after clustering by LVQ, split subclasses in a class are merged into appropriate neighboring classes except one subclass. And the validity of the classification result is checked. Under several classification experiments, the performance of the proposed method is provided.

1 Introduction

An important and fundamental research issue for pattern recognition, image processing and data mining is clustering [1-9], whose aim is to classify unlabeled data forming clusters to classes correctly. In this paper, a new clustering method based on learning vector quantization (LVQ) [1-3] and a split and merge procedure is investigated.

Focusing on split and merge procedures for classification, Ueda et al. [10,11] provide remarkable results on parameter estimations of mixture models, where split and merge procedures are incorporated into the EM algorithm and the Variational Bayesian learning to avoid local minima of object functions.

The ideal of clustering is to classify data without any external restrictions. Although it is assumed that the number of clusters is provided, the K-Means algorithm (KMA) [8,9] and LVQ are typical algorithms for clustering, where KMA is derived from the viewpoint of minimizing the sum of squared-error distortion and LVQ is derived on the basis of learning concept. The value of the distortion function in KMA like the EM algorithm depends on initial cluster centers, and a local minimum may be captured. To avoid its defect and acquire the global minimum, a split and merge procedure is introduced into vector quantization (VQ) by Kaukoranta et al. [12], where VQ being almost equivalent to KMA is used for data compression. By using KMA with the split and merge procedure, i.e., Kaukoranta's method, we may obtain the minimum distortion or its approximations.

M. Ishikawa et al. (Eds.): ICONIP 2007, Part II, LNCS 4985, pp. 57–66, 2008.

However, when there are big differences among statistical distributions of class data, even KMA attaining the minimum distortion reveals bad classification results, which are well known and stated in the research book [9]. Although KMA provides good classification with the high possibility, clustering by the criterion of minimizing the distortion cannot necessarily find correct clusters, especially for those distributions. The motivation of this research is to recover those bad situations.

Fig.1 shows a typical bad classification result by KMA, where data are divided into three regions by three lines, and those cluster centers seem to attain the minimum squared-error distortion. Concerning the detail of the data, refer Section 3. Table 1 shows three of the sum of squared-distortion D_{KMA}, D_{CEN} and D_{OPT}, where D_{KMA} may be the minimum distortion by KMA, D_{CEN} is the distortion by correct centroids of clusters and the Voronoi partition, and D_{OPT} is the distortion by the optimum partition with no classification error. Note that $D_{OPT} > D_{KMA}$.

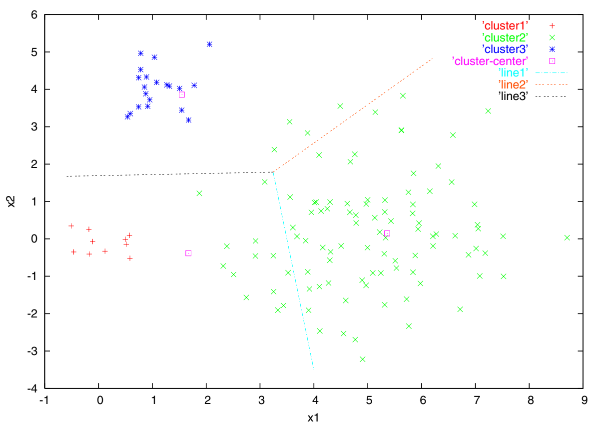

Fig. 1. A typical bad classification result by KMA with $K = 3$

Table 1. Comparison of the distortion D_{KMA}, D_{CEN} and D_{OPT}

	D_{KMA}	D_{CEN}	D_{OPT}
Distortion	373.0	397.8	412.9

After classifying data by an appropriate method such as Kaukoranta's method and obtaining good cluster centers, in order to deal with these bad situations too, LVQ started with those cluster centers is applied to the data, and a new split and merge procedure and another classification criterion being not the distortion criterion are introduced. KMA and LVQ are relations just like brothers. LVQ in comparison with KMA seems to have the high possibility of acquiring correct centroids of clusters. Hence we adopt LVQ in this paper.

By classifying the data by LVQ, suppose that we obtain samples classified into K classes. A method determining whether or not there is only a cluster in each class is introduced, where LVQ is applied to samples in each class again, and a measure of splitting a class into subclasses is introduced. When it is determined that the class must be split into subclasses, by comparing the dissimilarities between samples in the subclasses and samples in adjacent classes, the other subclasses except one subclass are merged into appropriate adjacent classes. This classification method by LVQ is a LVQ type version of the method by KMA [13].

This procedure is described in Section 2 in detail. In Section 3, several experimental results by this clustering method using LVQ and the split and merge procedure are shown for data composed from pseudo random numbers [14].

2 Clustering Based on LVQ and a Split and Merge Procedure

Let us consider classification of linearly separable data, where a set X of n samples $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iD}), i = 1, \ldots, n$ is partitioned into K disjoint subsets (classes) $X_k, k = 1, \ldots, K$.

Assume that we obtain appropriate cluster centers which satisfy the minimum squared-error distortion or its approximation by using random initialization or a classification method such as Kaukoranta's method. Subsequently, to realize more reliable classification, a clustering method based on LVQ and a split and merge procedure shown below is excuted.

(LVQ Algorithm)
(LVQ1) Set initial values of cluster centers $\{\boldsymbol{c}_k(1), k = 1, \ldots, K\}$. Repeat (LVQ2) and (LVQ3) for $t = 1, 2, \ldots$ untill convergent.
(LVQ2) Set

$$\boldsymbol{c}_l(t) = \arg \min_{1 \leq k \leq K} \| \boldsymbol{x}(t) - \boldsymbol{c}_k(t) \| . \tag{1}$$

(LVQ3) Compute

$$\boldsymbol{c}_l(t+1) \leftarrow \boldsymbol{c}_l(t) + \alpha(t)[\boldsymbol{x}(t) - \boldsymbol{c}_l(t)], \tag{2}$$

and determine $\boldsymbol{x}(t) \in$ class X_l.
(End of LVQ)

In the LVQ algorithm, we use $\boldsymbol{x}(1) = \boldsymbol{x}_1, \cdots, \boldsymbol{x}(n) = \boldsymbol{x}_n, \boldsymbol{x}(n+1) = \boldsymbol{x}_1, \cdots, \boldsymbol{x}(2n) = \boldsymbol{x}_n, \boldsymbol{x}(2n+1) = \boldsymbol{x}_1, \cdots$, and a learning rate $\alpha(t) = \text{constant}/t$.

Next, for classes $X_k, k = 1, \ldots, K$ classified by LVQ, let us consider a criterion determining whether or not there is only a cluster in each class by using LVQ again.

After classifying the samples in X by using LVQ, assume that we obtain classes $\{X_k\}$ and cluster centers $\{\boldsymbol{c}_k\}$. When it is determined that there is only a cluster in each X_k, the processing of clustering stops. However, if it is determined that there are two or more clusters in X_k, only a correct subcluster must survive in X_k by splitting X_k and merging other incorrect subclusters into adjacent classes. A method to dissolve this issue is proposed and investigated, where LVQ is used for the samples in each X_k again, and a split and merge procedure is applied to X_k and the subclasses of X_k.

After LVQ with $K = m$ for $2 \leq m \leq M$ is applied to the samples in X_k, X_k is split into m subclasses, whose subclasses and their cluster centers are denoted by $\{X_{k,p}^{(m)}, p = 1, \ldots, m\}$ and $\{\boldsymbol{c}_{k,p}^{(m)}, p = 1, \ldots, m\}$, respectively. In the ordinary situations of clustering, 2 or 3 as the value of M is used.

The squared-distortion for X_k is defined as

$$D_k^{(m=1)} = \sum_{\boldsymbol{x}_i \in X_k} \|\boldsymbol{x}_i - \boldsymbol{c}_k\|^2. \tag{3}$$

Under the definition of the distortion for $X_{k,p}^{(m)}$ by

$$D_{k,p}^{(m)} = \sum_{\boldsymbol{x}_i \in X_{k,p}^{(m)}} \|\boldsymbol{x}_i - \boldsymbol{c}_{k,p}^{(m)}\|^2, \; m = 2, \ldots, M, \tag{4}$$

the distortion for $X_k^{(m)}$, which means X_k with m subclasses, is provided as

$$D_k^{(m)} = \sum_{p=1}^{m} D_{k,p}^{(m)}. \tag{5}$$

Let us introduce a measure of splitting X_k given by

$$\rho_k(m) = D_k^{(m)} / D_k^{(m-1)}, \; m = 2, \ldots, M. \tag{6}$$

The abrupt decrease of $\rho_k(m)$ on m states that X_k should be split into m subclasses when X_k has m clusters. Consider the situation that for the partition of X_k into $m-1$ subclasses, each cluster center does not become a correct representative of the cluster in the subclass, but for the partition of X_k into m subclasses, each cluster center has the high possibility of becoming a correct representative. Then, the value of $D_k^{(m)}$ decreases abruptly in comparison with the value of $D_k^{(m-1)}$. This matter is demonstrated through classification experiments shown in Section 3.

Calculating

$$\rho_k(m^*) = \min_m \{\rho_k(m), \; m = 2, \ldots, M\}, \tag{7}$$

for a predetermined value ζ, if

$$\rho_k(m^*) < \zeta, \tag{8}$$

we split X_k into m^* subclasses. Otherwise, X_k is not split. The value of ζ is usually chosen on the basis of some experiential results, and the small value of ζ lowers the possibility of splitting.

When the class X_k must be split into the subclasses $\{X_{k,p}^{(m^*)}, p = 1, \ldots, m^*\}$ and the cluster centers $\{\boldsymbol{c}_{k,p}^{(m^*)}, p = 1, \ldots, m^*\}$ of $\{X_{k,p}^{(m^*)}\}$ are obtained by LVQ, only one subclass becomes the new class X_k renewing the old X_k and the other $m^* - 1$ subclasses must be merged into adjacent classes.

Let us define the dissimilarity between $X_{k,p}^{(m^*)}$ and the classes being adjacent to X_k as

$$d(X_{k,p}^{(m^*)}) = \min_{\boldsymbol{x}_i \in X_{k,p}^{(m^*)}, \boldsymbol{x}_j \in X_l, l \neq k} d(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{9}$$

where $d(\cdot, \cdot)$ expresses the Euclidean distance. Then, the subclass $X_{k,p^*}^{(m^*)}$ given by

$$\hat{d}(X_{k,p^*}^{(m^*)}) = \max_p \{d(X_{k,p}^{(m^*)}),\ p = 1, \ldots, m^*\} \tag{10}$$

becomes the new class X_k. The other subclass $X_{k,p}^{(m^*)}$ being $p \neq p^*$ are merged into the adjacent classes satisfying (9).

(Validity of classification)
After the classification by the split and merge procedure, the validity of the classification result must be checked. When there is only one cluster in each class X_k by using (8), the classification result is adopted. When otherwise, we do not adopt the classification result and the processing process is outputted.

When happening the inconsistency of processing, for example the exchange of subclasses in different classes, the classification result is not adopted.

3 Clustering Experiments

Let us consider the data shown by Fig.2, whose data are composed from three clusters. Cluster 1 is composed from 10 pseudo random numbers with mean $(x_1, x_2) = (0.086, -0.113)$, variance$(x_1, x_2) = (0.167, 0.076)$. Cluster 2 is composed from 100 pseudo random numbers with mean$(x_1, x_2) = (4.98, 0.163)$, variance$(x_1, x_2) = (1.78, 2.23)$. Cluster 3 is composed from 20 pseudo random numbers with mean$((x_1, x_2) = (1.10, 4.04)$, variance$(x_1, x_2) = (0.17, 0.306)$. Then the centroids for the clusters are provided by $(0.0861, -0.113)$, $(4.98, 0.163)$ and $(1.10, 4.04)$.

Fig.3 shows the classification result by LVQ. We obtain $\boldsymbol{c}_1 = (0.695, -0.0738)$, $\boldsymbol{c}_2 = (5.24, 0.132)$ and $\boldsymbol{c}_3 = (1.28, 3.98)$ as the cluster centers. Then, 11 samples among 130 samples are misclassified. The classification result is summarized by Table 2.

Selecting $M = 3$ as the maximum number of subclasses, let us split each class into m subclasses by using LVQ with $K = m$. Partition results of the classes $\{X_k\}$ by LVQ with $m = 2, 3$ are shown by Table 3 and Table 4.

The distortions $\{D_k^{(m)}\}$ for $\{X_k\}$ and $\{X_k^{(m)}\}$ calculated from (3) and (5) are provided in Table 5. Table 6 shows the splitting measures $\{\rho_k^{(m)}\}$ by (6).

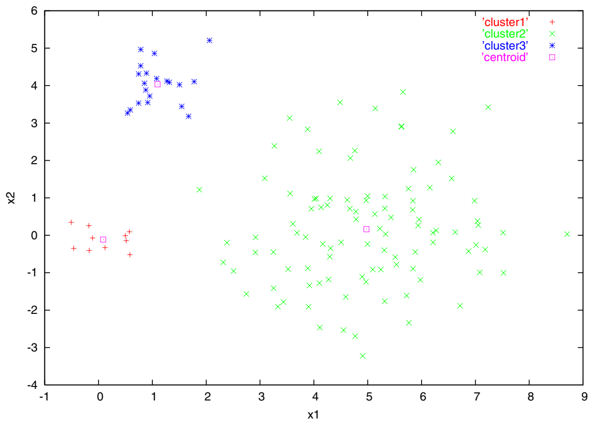

Fig. 2. Data composed from 3 clusters

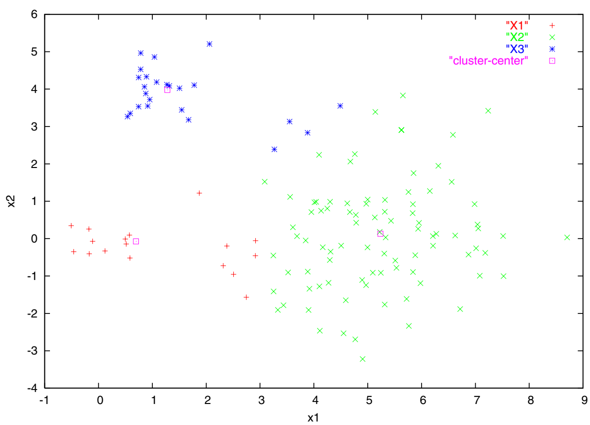

Fig. 3. Classification by LVQ with $K = 3$

Focusing on the values of $\{\rho_k^{(m)}\}$ in Table 6, we recognize that the values of $\rho_{k=1}^{(m=2)}$ and $\rho_{k=3}^{(m=2)}$ are very small. When $\zeta \approx 0.35$ in (8) is settled, it is determined that $m^* = 2$ and the classes X_1 and X_3 must be split into two subclasses, respectively.

Next, based on (10), appropriate subclasses among $\{X_{k,p}^{(m^*=2)}, k = 1, p = 1, 2\}$ and $\{X_{k,p}^{(m^*=2)}, k = 3, p = 1, 2\}$ must be merged into adjacent classes. The dissimilarities between the subclasses and the adjacent classes, and $\{d(X_{k,p}^{(m^*=2)})\}$ of (9) are provided by Table 7. From Table 7, $X_{1,2}^{(m^*=2)}$ is merged into X_2, and $X_{3,2}^{(m^*=2)}$ is merged into X_2.

Fig.4 shows the situation of classification by applying LVQ with $m^* = 2$ to each X_k. The class X_1 is split into two subclasses $X_{1,1}, X_{1,2}$. The right

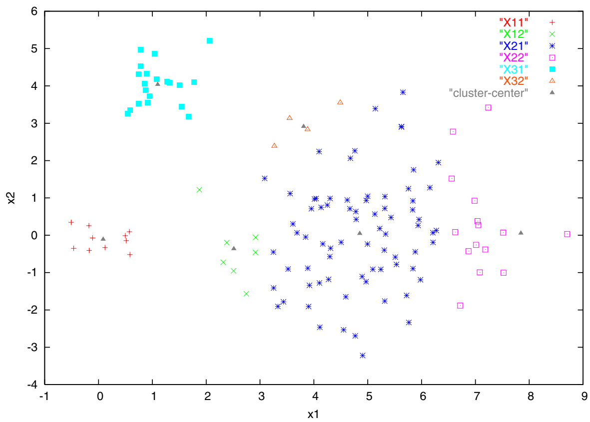

Fig. 4. Classification by LVQ for $K = 3$ and $m^* = 2$

Table 2. Classification result by LVQ

Class	Number	Cluster center
X_1	17	$\boldsymbol{c}_1 = (0.695, -0.0738)$
X_2	89	$\boldsymbol{c}_2 = (5.24, 0.132)$
X_3	24	$\boldsymbol{c}_3 = (1.28, 3.98)$

Table 3. Partition results by LVQ with $m = 2$

Subclass	Number	Cluster center
$X_{1,1}^{(m=2)}$	10	$\boldsymbol{c}_{1,1}^{(m=2)} = (0.0866, -0.110)$
$X_{1,2}^{(m=2)}$	7	$\boldsymbol{c}_{1,2}^{(m=2)} = (2.51, -0.363)$
$X_{2,1}^{(m=2)}$	74	$\boldsymbol{c}_{2,1}^{(m=2)} = (4.85, -0.0461)$
$X_{2,2}^{(m=2)}$	15	$\boldsymbol{c}_{2,2}^{(m=2)} = (7.85, 0.0548)$
$X_{3,1}^{(m=2)}$	20	$\boldsymbol{c}_{3,1}^{(m=2)} = (1.10, 4.04)$
$X_{3,2}^{(m=2)}$	4	$\boldsymbol{c}_{3,2}^{(m=2)} = (3.81, 2.91)$

subclass $X_{1,2}$ is merged into the class X_2. The class X_3 is split into two subclasses $X_{3,1}, X_{3,2}$. The right subclass $X_{3,2}$ is merged into the class X_2.

The final classification result is provided by the same figure as Fig.2. We recognize that the perfect classification for the data is realized with no error.

Lastly, the validity of the final classification result must be checked. According to (5)-(10), we apply LVQ with $K = m (2 \leq m \leq M)$ to the samples classified as Fig.2. We obtain Table 8 and Table 9. From Table 9 and (8), it is concluded that there is one cluster in each classified class.

Table 4. Partition results by LVQ with $m = 3$

Subclass	Number	Cluster center
$X_{1,1}^{(m=3)}$	6	$\boldsymbol{c}_{1,1}^{(m=3)} = (-0.222, -0.0676)$
$X_{1,2}^{(m=3)}$	4	$\boldsymbol{c}_{1,2}^{(m=3)} = (0.652, -0.143)$
$X_{1,3}^{(m=3)}$	7	$\boldsymbol{c}_{1,3}^{(m=3)} = (2.52, -0.355)$
$X_{2,1}^{(m=3)}$	37	$\boldsymbol{c}_{2,1}^{(m=3)} = (4.55, -0.967)$
$X_{2,2}^{(m=3)}$	40	$\boldsymbol{c}_{2,2}^{(m=3)} = (5.37, 1.31)$
$X_{2,3}^{(m=3)}$	12	$\boldsymbol{c}_{2,3}^{(m=3)} = (7.99, -0.171)$
$X_{3,1}^{(m=3)}$	19	$\boldsymbol{c}_{3,1}^{(m=3)} = (1.07, 4.08)$
$X_{3,2}^{(m=3)}$	1	$\boldsymbol{c}_{3,2}^{(m=3)} = (2.34, 3.67)$
$X_{3,3}^{(m=3)}$	4	$\boldsymbol{c}_{3,3}^{(m=3)} = (3.81, 2.91)$

Table 5. Distortions $\{D_k^{(m)}\}$ for $\{X_k\}$ and $\{X_k^{(m)}\}$

	$D_k^{(m=1)}$	$D_k^{(m=2)}$	$D_k^{(m=3)}$
$k = 1$	35.6	7.83	6.49
$k = 2$	306.8	250.0	142.6
$k = 3$	41.2	11.1	10.7

Table 6. Splitting measures $\{\rho_k^{(m)}\}$.

	$\rho_k^{(m=2)}$	$\rho_k^{(m=3)}$
$k = 1$	0.22	0.83
$k = 2$	0.82	0.57
$k = 3$	0.27	0.96

Table 7. Dissimilarity table

Subclass	X_1	X_2	X_3	$d(X_{k,p}^{(m^*)})$
$X_{1,1}^{(m^*=2)}$		2.66	3.09	$d(X_{1,1}^{(m^*)=2}) = 2.66$
$X_{1,2}^{(m^*=2)}$		0.333	1.82	$d(X_{1,2}^{(m^*)=2}) = 0.333$
$X_{3,1}^{(m^*=2)}$	1.97	2.18		$d(X_{3,1}^{(m^*)=2}) = 1.97$
$X_{3,2}^{(m^*=2)}$	1.82	0.628		$d(X_{3,2}^{(m^*)=2}) = 0.628$

Table 8. Validity: Distortions $\{D_k^{(m)}\}$ for $\{X_k\}$ and $\{X_k^{(m)}\}$

	$D_k^{(m=1)}$	$D_k^{(m=2)}$	$D_k^{(m=3)}$
$k = 1$	2.43	0.995	0.693
$k = 2$	401.0	260.0	195.9
$k = 3$	9.52	5.44	3.94

Table 9. Validity: Splitting measures $\{\rho_k^{(m)}\}$

	$\rho_k^{(m=2)}$	$\rho_k^{(m=3)}$
$k = 1$	0.410	0.697
$k = 2$	0.648	0.753
$k = 3$	0.571	0.725

4 Conclusion

We proposed a new clustering method based on LVQ and the split and merge procedure to improve the classification performance of the ordinary LVQ algorithm. After introducing the splitting measure and the dissimilarity measure for merging, the classification method proposed in this paper was applied to the data that reveal the typical bad performance by the ordinary LVQ algorithm. Under some classification experiments, the performance of this method was investigated. As a future issue, we would like to develop this method to a general and robust method under the consideration of research results by the papers [10,11,12,13]. And it is also an important issue to estimate the number of clusters correctly.

References

1. Kohonen, T.: Self-Organizing Maps, 2nd edn. Springer, Berlin (1997)
2. Pal, N.R., Bezdek, J.C., Tsao, C.-K.: Generalized Clustering Networks and Kohonen's Self-Organizing Scheme. IEEE Trans. Neural Network 4(4), 549–557 (1993)
3. Miyamoto, S.: Intoduction of Cluster Analysis: Theory and Applications of Fuzzy Clustering. Morikita-Syuppan (1999) (in Japanese)
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. John Wiley & Sons, INC., Chichester (2001)
5. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)
6. Gordon, A.D.: Classification, 2nd edn. Chapman & Hall/CRC, Boca Raton (1999)
7. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, NY (1981)
8. MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations. In: Proc. 5th Berkeley Symp. on Math. Stat. and Prob. 1, pp. 281–297. Univ. of California Press, Berkeley and Los Angeles (1967)
9. Linde, Y., Buzo, A., Gray, R.M.: An Algorithm for Vector Quantizer Design. IEEE Trans. Commun. 28, 84–95 (1980)
10. Ueda, N., Nakano, R.: EM Algorithm with Split and Merge Operations for Mixture Models. Systems and Computers in Japan 31(5), 930–940 (2000)
11. Ueda, N., Ghahramani, Z.: Bayesian model search for mixture models based on optimizing variational bounds. Neural Networks 15, 1223–1241 (2002)

12. Kaukoranta, T., Franti, P., Nevalainen, O.: Iterative split-and-merge algorithm for vector quantization codebook generation. Optical Engineering 37(10), 2726–2732 (1998)
13. Morii, F., Kurahashi, K.: Clustering by the K-means algorithm using a split and merge procedure. In: Proc. of SCIS and ISIS 2006, pp. 1767–1770 (2006)
14. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C. Cambridge University Press, Cambridge (1988)