M.Bhuvaneswari
Department of EEE, PPG Institute Institute of Technology, Coimbatore-641035, Tamil Nadu, India

Gaussian mixture model: An application to parameter estimation and medical image classification

M.Bhuvaneswari*

Abstract

Gaussian mixture model based parameter estimation and classification has recently received great attention in modelling and processing data. Gaussian Mixture Model (GMM) is the probabilistic model for representing the presence of subpopulations and it works well with the classification and parameter estimation strategy. Here in this work Maximum Likelihood Estimation (MLE) based on Expectation Maximization (EM) is being used for the parameter estimation approach and the estimated parameters are being used for the training and the testing of the images for their normality and the abnormality. With the mean and the covariance calculated as the parameters they are used in the Gaussian Mixture Model (GMM) based training of the classifier. Support Vector Machine a discriminative classifier and the Gaussian Mixture Model a generative model classifier are the two most popular techniques. The performance of the classification strategy of both the classifiers used has a better proficiency when compared to the other classifiers. By combining the SVM and GMM we could be able to classify at a better level since estimating the parameters through the GMM has a very few amount of features and hence it is not needed to use any of the feature reduction techniques. In this the GMM classifier and the SVM classifier are trained using the parameters and they are to be compared.

Keywords: Expectation Maximization (EM), Gaussian Mixture Model (GMM), Maximum Likelihood Estimation (MLE), Support Vector Machine (SVM).

INTRODUCTION

Classification, as a big part of supervised learning problem, has always attracted lots of attention for its various applications. Also, many methods are brought forward to tackle this problem. Statistical methods manipulate probabilities and try to model the entire hypothesis space and data distribution using probabilities distribution density, thus providing more complete description of the actual problems; however, it also asks for huge complexity to achieve the goal. Gaussian mixture modelling performs classification by extracting global statistics from Gaussian distributions of pixel intensity in an image data set. GMM is especially well suited for parameter estimation and classification because of its implementation facility and efficiency in the representation of data. The histogram gives the main idea about the probability density function (pdf) of the pixel values[1]. In the statistical approaches it has been assumed that pixel values follow a particular distribution and hence mixture model approach is applied on those values. The well known approach for estimating parameters of model is either by Maximum Likelihood Estimation (MLE) via Expectation Maximization algorithm (EM) or Maximum A Posteriori Estimation (MAP)[2].
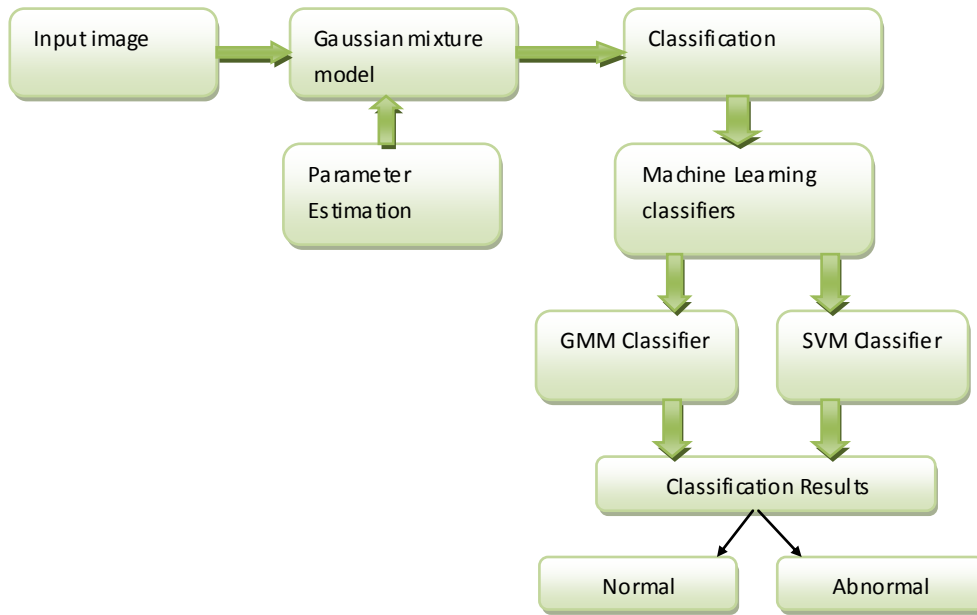
METHODOLOGY

The methodology of this paper is well explained in the block diagram below. It shows the clear idea of the inputs given and the methods used for the estimation of the parameters and how these parameters are about to be used in the training and the testing of both the classifiers that are used here for the classification of the medical images. The database of medical images is used for both training and testing of which 25% of the images were abnormal and the classifier could able to correctly classify the images. The images were taken from different aspects both normal and abnormal combined and they are trained. Then the rest of the images were used for the testing phase for both the classifiers and they were able to correctly classify the abnormal from the normal images.

Correspondence:
M.Bhuvaneswari
Department of EEE, PPG Institute Institute of Technology, Coimbatore-641035, Tamil Nadu, India

Gaussian Mixture Model

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal -tract related spectral features in speaker recognition systems. GMM parameters are estimated from training data using the iterative Expectation-Maximizations (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well trained prior model.

Block Diagram

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation,

$$p(X|\lambda) = \sum_{i=1}^{M} w_i \quad g(X|\mu_i, \Sigma_i) \qquad (1)$$

where x is a D-dimensional continuous valued data vector (measurement of features), W_i, i=1......,M, are the mixture weights, and $g(x|\mu_i, i)$, i=1... ..,M, are the component Gaussian densities.

D-variate Gaussian function of the form in each component density is given by

$$g(x|\mu i, \Sigma i) = \frac{1}{(2\Pi)^{(D/2)} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1}(x-\mu_i)\right] \quad (2)$$

Where mean vector μ_i and the covariance matrix Σ_i.

The constraint of the mixture density is met and given by the form,

$$\sum_{i=1}^{M} w_i = 1$$

The mean vectors, covariance matrices and mixture weights from all component densities were used for the complete Gaussian distribution parameterisation. These parameters are collectively represented by the notation

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \quad i=1..............M \quad (3)$$

with this the Expectation Maximisation and the Maximum Likelihood are used for the estimation of the parameters as follows

Maximum Likelihood Parameter Estimation

Given training vectors and a GMM configuration, we wish to estimate the parameters of the GMM, λ, which in some sense best matches the distribution of the training feature vectors. Several techniques are available for estimating the parameters of a GMM and they have their own form of estimating which is clearly given in Dempster[3] *et al.* Among the various techniques available, the most popular and well established method is the Maximum Likelihood Estimation (MLE).

To maximize the likelihood of the GMM given the ML estimation of model parameters is used which will be better in the training data. For a sequence of T training vectors X={$x_1,...,x_T$}, the GMM likelihood assuming independence between the vectors 1 can be written as,

$$p(x|\lambda) = \prod_{t=1}^{T} P(X_t|\lambda) \qquad (4)$$

Unfortunately, this expression is a non-linear function of the parameters λ and direct maximisation is not possible. However, ML parameter estimates can be obtained iteratively using a special case of the expectation maximisation (EM) algorithm[4,5]. The new model that becomes the initial model for the next iteration and the process is repeated until some convergence threshold is reached. The initial model is typically derived by using some form of binary vector quantisation.

On EM iteration, re estimation formulas are used which guarantee a monotonic increase in the model's likelihood value and are as follows,

Mixture weights:

$$\overline{W_i} = \frac{1}{T} \sum_{t=1}^{T} P_r(i|X_t, \lambda) \qquad (5)$$

Mean:

$$\overline{\mu_i} = \frac{\sum_{t=1}^{T} P_r(i|X_t \lambda) X_t}{\sum_{t=1}^{T} P_r(i|X_t \lambda)} \qquad (6)$$

Variance:

$$\overline{\sigma_{i^2}} = \frac{\sum_{t=1}^{T} P_r(i|X_t \lambda) x^2_t}{\sum_{t=1}^{T} P_r(i|X_t \lambda)} \qquad (7)$$

where σ_{i^2}, x_t, and μ_i refer to arbitrary elements of the vectors σ_{i^2}, x_t, and μ_i, respectively [5].

Expectation Maximisation Map Algorithm

In general, EM iterates through two steps to obtain estimates. The first step is an Expectation (E) step, in which missing values are filled in with a guess, that is, an estimate of the missing value, given the observed value in the data. The second step is a Maximization (M) step, in which the completed data from the E step are processed using ML estimation as though they were complete data, and mean and the covariance estimates are updated. Using the newly updated mean and variance matrix, the E step is repeated to find new estimates of the missing values.

The two steps E-step and the M-step are repeated until the maximum change in the estimates from one iteration to the next do not exceed a convergence criterion. The result of this process is a mean vector and covariance matrix that uses all available information. In the EM [6] estimates of the mean vector and covariance matrix can then be used in multivariate analyses to obtain estimates of model parameters and the standard errors, to test hypothesis, and to score or predict values for observations using the model selected.

There are popular algorithm for GMM[7] and this is modified here as the name of the EM-MAP algorithm. This algorithm is described as follows,

Input:

Observed Image in a vector X_j, j=1,2,....,n and i$\in$\{1,2,.....,k\} labels set.

Initialize:

$$\theta^{(0)}=(p_1^{(0)},.... p_k^{(0)}, \mu_1^{(0)},.... \mu_k^{(0)}, \sigma_1^{2(0)},.... \sigma_k^{2(0)}) \qquad (8)$$

E-Step:

$$p_{ij}^{(r+1)}=p^{(r+1)}(i|x_j)=\frac{p_i^{(r)} N(x_j|\mu_i^{(r)},\sigma_i^{2(r)})}{f(x_j)} \qquad (9)$$

M-Step:

$$\hat{p}_i^{(r+1)}=\frac{1}{n}\sum_{j=1}^{n} p_{ij}^{(r)} \qquad (10)$$

$$\mu_i^{(r+1)}=\frac{\sum_{j=1}^{n} p_{ij}^{(r+1)}}{n\hat{p}_i^{(r+1)}}x_j \qquad (11)$$

$$\sigma_i^{2(r+1)}=\frac{\sum_{j=1}^{n} p_{ij}^{(r+1)}}{n\hat{p}_i^{(r+1)}}(x_j-\hat{\mu}_i^{(r+1)}) \qquad (12)$$

Iterate steps 3 and 4 until an specify error (i.e)$\sum_i e_i^2 < \varepsilon$.

Compute:

$$P_{1j}= ArgMax_i P_{ij}^{(final)},j=1,2,...,n \qquad (13)$$

In each and every iteration the corresponding values of the number of classes, mean, probability and the variance are being calculated and are used in the iteration.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) was first proposed by Vapnik[8] and has since attracted a high degree of interest in the machine learning research community. In recent studies it has been reported that the SVM (support vector machines) generally are capable of delivering higher performance in terms of classification accuracy than other data classification algorithms. SVM is a binary classifier based on supervised learning which gives better performance than other classifiers. SVM classifies between two classes by constructing a hyper plane in high-dimensional feature space which can be used for classification. Hyper plane can be represented by equation-

$$w.x+b=0 \qquad (14)$$

w is weight vector and normal to hyperplane. b is bias or threshold.

Kernel functions are used with SVM classifier. Kernel function provides the bridge between from non-linear to linear[9,10] when compared. By using kernel function the mapping of the image in the low dimensional data into the high dimensional feature space where data points are linearly separable is possible. The radial bias based kernel function is being used in this SVM based classification.

Classifier Performance

Every classification results could have an error rate and on occasion will either fail to identify an abnormality, or identify an abnormality which is not present. This error rate is generally explained in the terms of true and false positive and true and false negative as follows:

True Positive (TP): the result of the classification is positive in the presence of the clinical abnormality.

True Negative (TN): the result of the classification is negative in the absence of the clinical abnormality.

False Positive (FP): the result of the classification is positive in the absence of the clinical abnormality.

False Negative (FN): the result of the classification is negative in the presence of the clinical abnormality.

Table below is the contingency table which defines various terms used to describe the clinical efficiency of a classification based on the terms above and,

Sensitivity = TP/ (TP+FN) *100%

Specificity = TN/ (TN+FP) *100%

Accuracy= (TP+TN)/ (TP+TN+FP+FN)*100 % are used to measure the performance of a classifier.
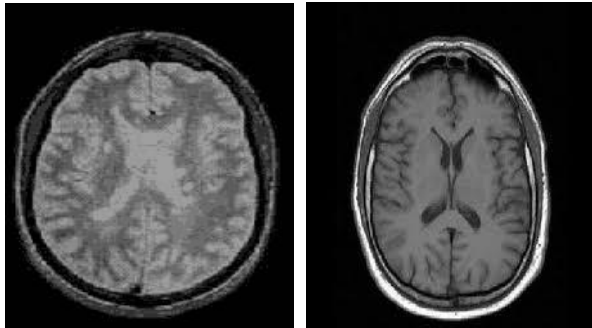
Contingency Table:

Table 1: Contingency table for the classifier performance

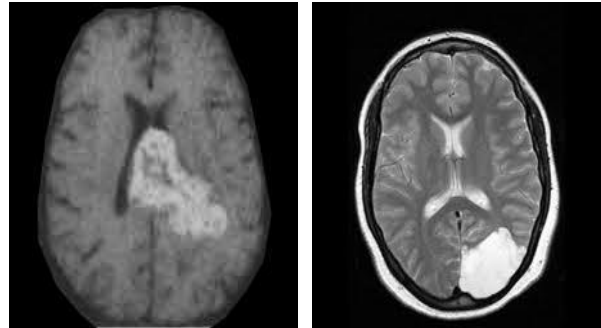| Actual Group	Predicted Group	
	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

RESULTS

In this the features or the parameters are being extracted by using the GMM based technique. And here they are used as the inputs for both the GMM based classifier and the SVM classifier and their results varies with each classifiers however there is no much variation in the for the image given as the input. The images given as the inputs that are both normal and abnormal are given as follows;

classification performance. Here in the RBF kernel is used in the SVM based classifier and in the GMM the K-nearest neighbour method of classification is used. In the SVM classifier the other kernel based functions are also compared which has been taken from the other paper results. With this the Gaussian plot of the images can also be obtained
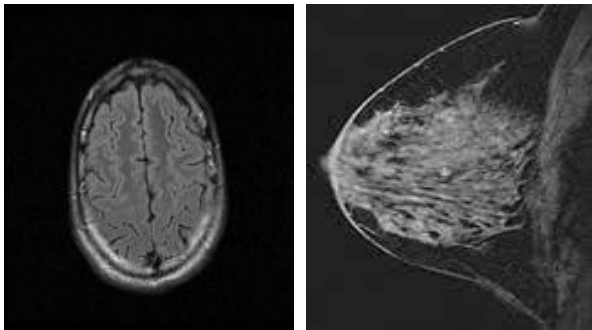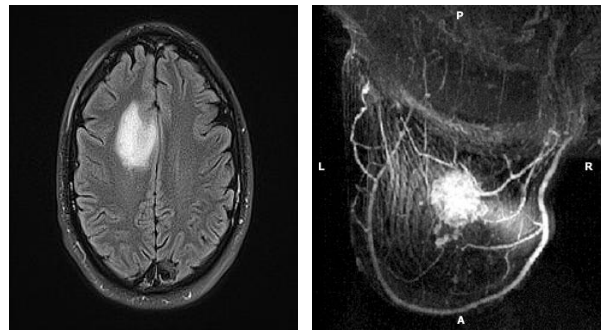

Figure 1: Normal images (a), (b), (c) brain images, (d) mammogram image

Figure 2: The abnormal images (a), (b), (c) brain images, (d) breast image used in the training and testing
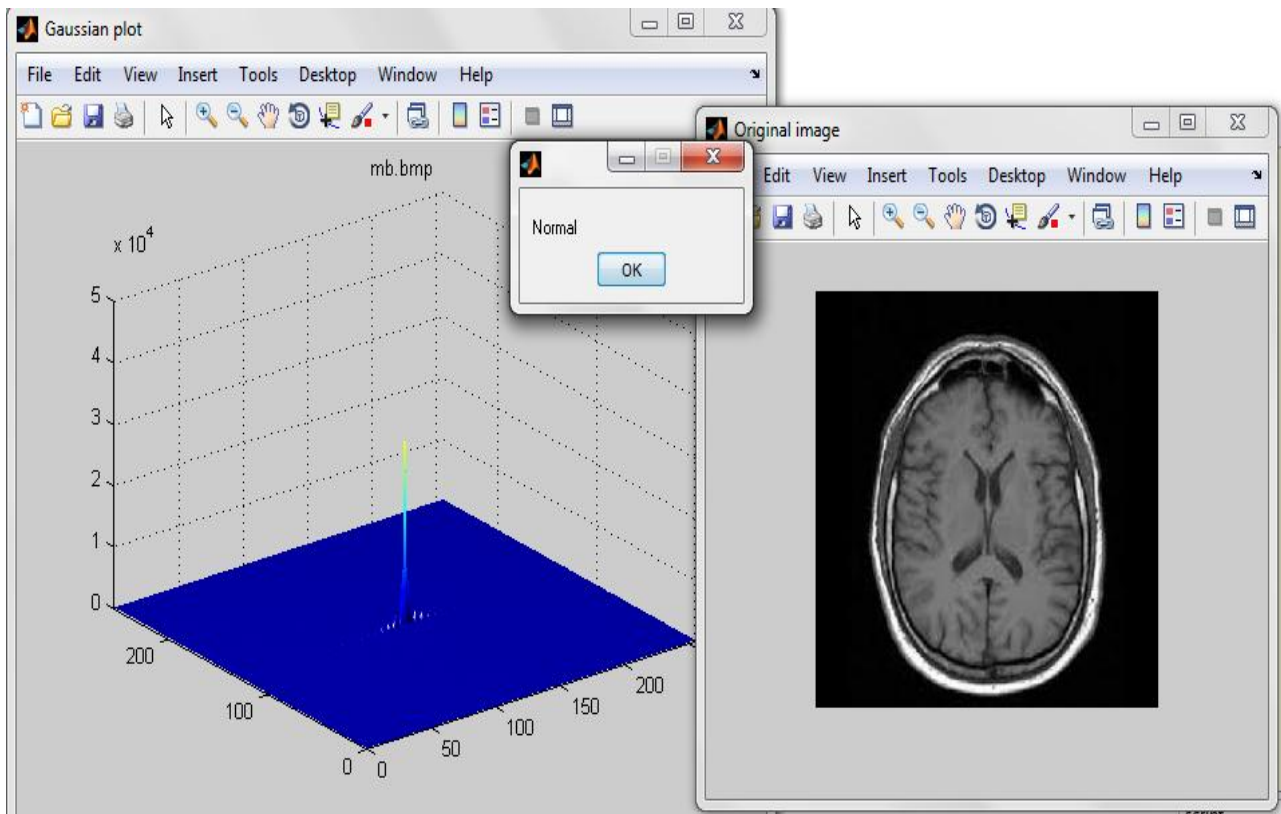

Figure 3: The Gaussian classifier result for one of the normal image used in the data set for the Testing
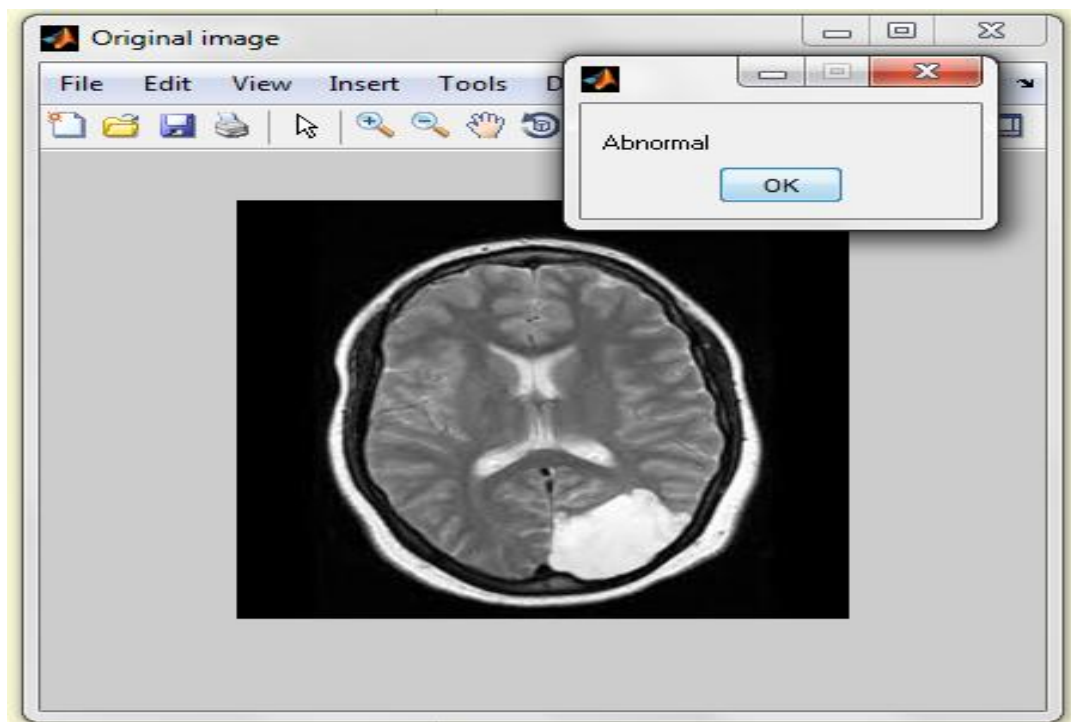
Support Vector Machine Classifier Result

Figure 4: Result of the brain image used in the dataset for testing by the support vector machine classifier

Classifier Performance:

Table 2: Classifier performance table

Feature Extraction Technique	Kernel function	Specificity	Sensitivity	Accuracy
GMM	RBF	90	92.56	**96.78**
	LINEAR	89	90.54	95.45
	QUADRATIC	91	93.67	92.34
GMM	K-means	91	93.23	**94.67**

Gaussian Mixture Model Classifier Result

In this the normal and the abnormal images from the datasets were used and they are correctly classified.

In this gaussian plot of the particular image is shown and the image identified as normal is given. The likelihood ratios of the classifier has also been estimated and they have range of values from -7.1202 to -7.9234.

The same method has also been used for the SVM based classifiers also where the parameters are estimated from the Gaussian mixture model itself that are used for the training of the both classifiers.

CONCLUSION

The parameters have been estimated using the GMM model and the same model is being used as one of the classifier along with the SVM classifier. In this, their performance were equally good but the GMM based model has the less accuracy of 94.67% when compared with the SVM which has about 96.78% in the case of using large dataset values, means that while having more number of features. When it is being trained with large dataset in SVM classifier it works well in both the cases.

REFERENCES

1. T. Elguebaly, N. Bouguila, Bayesian learning of finite generalized Gaussian mixture models on images, Signal Processing 91 (2011) 801–820.
2. J. Liu, Y.B. Ku, S. Leung, Expectation-maximization algorithm with total variation regularization for vector-valued image segmentation, Journal of Visual Communication and Image Representation 23 (2012) 1234–1244.
3. McLachlan, G., ed.: Mixture Models. Marcel Dekker, New York, NY (1988)
4. Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood from Incomplete Data via the EMAlgorithm. Journal of the Royal Statistical Society 39(1) (1977) 1–38
5. H. Permuter, J. Francos, I.H. Jermyn, Gaussian mixture models of texture and colour for image database retrieval, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Hong Kong, China, 2003

6. Buck, S.F. (1960). "A method of estimation of missing values in multivariate data suitable for use with an electronic computer". Journal of the Royal Statistical Society, B22, 302-306.

7. M.A.T.Figueiredo, Bayesian Image Segmentation Using Gaussian Field Prior, EMMCVPR 2005, LNCS 3757, pp. 74-89.

8. J. Tohka, I. D. Dinov, D. W. Shattuck, and A. W. Toga,"Brain MRI tissue classification based on local Markov random fields", Magn. Reson. Imag., vol. 28, no. 4, pp. 557-573, 2010.

9. C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation", IEEE Trans. Image Process., vol. 16, no. 4, pp. 1121-1130, Apr. 2007 Wiley, 1985

10. V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method", Automation and Remote Control 24, 1963.