Using RBF reduced by QLP decomposition for Probability Density Estimation

EDWIRDE LUIZ SILVA
Departamento de Matemática e Estatística
Universidade Estadual da Paraíba - UEPB
Rua Juvêncio Arruda, S/N Campus Universitário (Bodocongó).
Cep: 58.109 - 790 Campina Grande - Paraíba
BRASIL
http://www.uepb.edu.br

Abstract: - This paper is intender to be a simple example illustrating some of the capabilities of Radial basis function by pruning with QLP decomposition. The applicability of the radial basis function (RBF) type function of artificial neural networks (ANNS) approach for re-estimate the Box, Traingle, Epanechnikov and Normal densities. We propose an application of QLP decomposition model to reduce to the class of RBF neural models for improving performance in contexts of density estimate. Has been found in the QLP that such a coupling leads to more precise extraction of the relevant information, even when using it in a heuristic way. This paper is concerned with re-estimation these four densities estimated by pruning a Radial Basis Function network using pivoted QLP decomposition. For comparison all RBF type functions with the same Gaussian mixture model as the sample data is superimposed on the plot. This application tool can be used to identify the density estimate from empirical data where presents many type density estimative. The QLP methods proves efficient for reducing the network size by pruning hidden nodes, resulting is a parsimonious model which identify RBF type multiquadric to re-estimate kernel function Box and Normal distributions.

Key-Words: -Density estimate, RBF , Pivote QLP Decomposition, adjust function, Bayesian Information Criterion

1 Introduction

Scott [1] shows that as the number of histograms *m* approaches infinity, the averaged shifted histogram becomes a kernel estimate of the probability density function. In [2] introduced the basic algorithm of nonparametric density estimation. Estimating probability density functions is required in many areas of computational statistics. Another application where probability density estimation is used is in statistical pattern recognition. In other applications, we might need to determine the probability that a random variable will fall within some interval, so we would need to evaluate the cumulative distribution function The first published paper describing nonparametric probability density estimation was by Rosenblatt [3], where he described the general kernel estimator. Many papers that expanded the theory followed soon after. They addressed the problem of statistical discrimination when the parametric form of the sampling density was not known.

In this paper we show how RBFs with reduction neuron thought the network decomposition using QLP (a lower diagonal matrix L between orthogonal matrices Q and P [4]) using the different basis functions networks Cauchy and multiquadric, and Inverse multiquadric type function. This can resulting an approximation of the densities estimates Box and Triangle, and Epanechnikov. The performance of the RBF reduction with QLP is compared with model selection criteria as the Schwartz Bayesian Information Criterion (BIC) and mean squared error.

The rest the paper is organized as follows. Section 2 presents Kernel Density estimate, section 3 Gaussian Mixture Models, section 4 Design Cauchy RBF Neural, section 5 Detection of the Numerical rank of the QLP, 6 proposed reduction RBF to identification, section 7 Performance estimation, 8 Simulation results and 9 Conclusion.

2 Kernel Density estimate

The estimated distribution function is calculated for a number of equidistant points that cover the range of the sample data. For each point p, the estimated density depends on the closeness of the sample data values to the point, such that data values close to p have a larger effect than further away.

The basic kernel estimates may be written compactly by

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\theta\left(\frac{x-x_i}{h}\right) \qquad (1)$$

Where x_i represents each data point in the sample of size n, and the function is a standard normal distribution with mean 0 and variance 1.

$\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$, the smoothness of the estimate depends on the parameter h, kwon as the bandwidth. If h is small, only data values vary close to the point p have influence on the estimated density, and this tends to make the estimate rather jagged. As h increases, data values further away from p star to influence the distribution, and it tends to became smoother. Where $\theta(x) = \theta(x/h)/h$ [a notation introduced by [3]]. The kernel estimate can be motivated not only as the limiting case of the averaged shifted histogram (ASH).

3 Gaussian Mixture Models

Mixture Models are a type of density model which comprise a number of component functions, usually Gaussian. These component function are combined to provide a multimodal density.

Mixture models are a semi-parametric alternative to non-parametric histograms [4] (which can also be used as densities) and provide greater flexibility and precision in modeling the underlying statistic of sample data.

Hopefully, the reader can see the connection between finite mixtures and kernel density estimation. Recall that in the case of univariate kernel density estimators, we obtain these by evaluating a weighted kernel centered at each sample point, and adding these n terms. So, a kernel estimate can be considered a special case of a finite mixture where $c = n$.

Therefore, the estimate of a finite mixture (2) would be written as

$$\hat{f}_{FM}(x) = \sum_{i=1}^{c} \hat{p}_i \theta(x; \hat{\mu}_i, \hat{\sigma}_i^2) \qquad (2)$$

Where $\theta(x; \hat{\mu}_i, \hat{\sigma}_i^2)$ denotes the normal probability density function with mean μ_i, variance σ_i^2 and FM is finite mixture.

4 Design Cauchy RBF Neural

The network consists of n input features x, M hidden units with center C_j and y output. The θ_j are the basis functions, and w_{kj} are the output layer weights. The basis function activations are then calculated using a method which depends on the nature of the function. Suppose at a set of fixed point $x_1, \ldots, x_j$, $\theta_j = \theta_j(x)$ can be

Table 1. BRF based interpolation

Gaussian	$\theta_j(x) = \exp(-\|x - c\|^2 / 2r^2)$
Cauchy	$\theta_j(x) = \frac{1}{r + c}$
Multiquadric	$\theta_j(x) = \sqrt{r^2 + c^2}$
Inverse Multiquadric	$\theta_j(x) = 1/\sqrt{r^2 + c^2}$

We shall write the RBF network mapping in the following form [8]

$$y_n(x) = \sum_{j=1}^{M} w_{nj}\theta_j(x) + w_{no}, \quad n = 1,2,\ldots,L \quad (11)$$

Finally the network outputs are calculated by

$$y_n(x) = \sigma\left[\sum_{j=1}^{M} w_{nj}\theta_j(x) + w_{no}\right], \quad n = 1,2,\ldots,L$$

where σ_n is a linear transfer function.

5 Detection of the Numerical rank of the QLP

The QLP (a lower diagonal matrix L between orthogonal matrices Q and P) decomposition [4] is computed by applying pivoted orthogonal triangularization to the columns of the matrix design Θ in question to get an upper triangular factor R and then applying the same procedure to the rows of R to get a lower triangular matrix L. The diagonal elements of R are called the R-values of Θ; those of L are called the L-values [4]

Numerical examples show that the L-values track the singular values of Θ with considerable fidelity-far better than the R-values. At a gap in the L-values the decomposition provides orthonormal bases of analogues of row, column, and null spaces provided of Θ.The decomposition requires no more than twice the work required for a pivoted QR decomposition. The computation of R and L can be interleaved, so that the computation can be terminated at any suitable point, which makes the decomposition especially suitable for low-rank determination problems.We will call the diagonals of R the R-values of Θ.The folklore has it that the Rvalues track the singular values well enough to to expose gaps in the latter.

The implementation of the Matlab package permits the QLP: [P,Q,L,pr,pl]=qlp(Θ) [6,7] to determine the numerical rank of matrix design Θ applied on the Gaussian and Cauchy and Multiquadric.

Thus, we have a simple QLP algorithm as follows:

1. **define** matrix design Θ, which consists of Θ (9).
2. **calculate** the orthonormal matrices Q and P which reduce the matrix design Θ to lower diagonal form.
3. **identify the** diagonal of lower-triangular matrix L
4. **sort** the diagonal elements by size.

In the Table 1 shows the computational speed advantage by over an order of magnitude from using QLP decomposition compared with the SVD algorithm.

The speed to resulting the QLP [7,8] of matrix design on the training RBF network also makes them attractive for use as a component in more complex model.

Table 2. Comparison between the times required to calculate matrix design Θ using SVD and QR and QLP decompositions

Decompositions	Gau	Cauchy	Mul	InvMul
QLP and QR	0.01	0.01	0.01	0.02
SVD	0.05	0.02	0.03	0.03

Moreover, it is likely that the pivoted QLP decomposition may also provide better approximations to the singular values of the original matrix design.

6 Proposed reduction RBF to identification

Consider using a radial basis function (RBF) network to approximate a known density estimates. One hundred training data were generated from mixture Gaussian,

$$z(x) = 1.\theta(x,0,1) + 0.2\theta(x;0,1)$$

where the input x was normally distributed in $[0,1]$. There are 1000 testing data (x_i, z_i) with randomly distributed in the range $(0,1)$. Here we generate 1000 data sets, independently from z. The data set are indexed by $l = 1,..., L$, where L = 100, and for each data set we fit a model with 100 Gaussian or Cauchy or Multiquadric and regularized by $\lambda = 10^{-1}$ to give a prediction function.

The Gaussian basis function was used with a kernel $r = 0.1$. All the 100 training data points were used as the candidate RBF center set for c .

The design matrix from the input data, centre positions and radial factors has size of 100x100. We assume that $w = inv(\theta'\theta)\theta' y$ and $y_n(x) = \theta w$ with 100 neurons has been obtained. The network output for an input x_i is given by

$$y_1(x) = \sum_{j=1}^{N} w_j \theta_j(x) .$$

The target function to be approximately is the following one density estimate function. Training samples (x, y) density kernel estimate. The numbers of training samples kernel box, Epanechnikov, and triangle are 100. The approximation accuracy is estimated for test samples after incremental learning is completed. The test samples are also randomly drawn from the same regions, and the numbers of them are 1000 to box, Epanechnikov, and triangle, respectively. The estimate is based on a RBF reduced by QLP decomposition. The density is evaluated at 100 equally-spaced points covering the range of the data in x.

Each iteration with a RBF network requires a single matrix inversion. The first process is the creation of the matrix design Φ composed with inputs and centre. In here the matrix design has same radios and centre equal the data input.

7 Performance estimation

Before we can describe the various density estimation methods, we need to provide a little background on measuring the error in functions.

7.1 BIC Shwartz Bayesian Information Criterion.

This generic function calculates the Bayesian information criterion, also known as Schwarz's Bayesian criterion (SBC), for one or several fitted model objects for which a log-likelihood value can be obtained, according to the formula

$-2\log likehood + npar.\log(nobs)$, where $npar$ represents the number of parameters and $nobs$ the number of observations in the fitted model [11].

The number of parameters here considered is $\tau = p - trace(P)$, where $p = 100$ number row of matrix design Φ, and $P = I_p - \Phi A^{-1}\Phi^T$ is projection matrix, where $A = (\Phi^T\Phi + IU^TU)$, U is the upper triangular transform of dimension 100x100 and $I = \lambda = 10^{-3}$ parameters of regularization.

On the one hand, large number of parameters (and of corresponding basis function) ensures good quality of data description but, on the other hand, it complicates the model excessively.

The BIC criterion is defined as [5]:

$$BIC = N_{bic}(y^T P^2 y)/p$$

Where $N_{bic} = \dfrac{(p+(\log(p)-1).\tau)}{p-\tau}$ is the effective number of parameters and y is input training data.

In this paper the model is the best due to the fact that it shows the least BIC value. The value of the BIC statistic suggests also that the errors are not correlated.

7.2 Final Prediction Error (FPE)

The FPE is a network performance function is defined as [9]

$$FPE = \frac{p+g}{p-g}(y^T P^2 y)/p$$

8 Simulation results

8.1 RBF type Inverse Multiquadric to Kernel Density Estimation

The QLP decomposition in the Fig. 1 for the RBF reduced by QLP, the pruning threshold is chosen as 10 neurons.

Initial the 100 training data points were used to model as the candidate RBF centre set and the regularization parameter was fixed to $\lambda = 10^{-1}$. One method to choose this number of hidden is to use the minimum value of the BIC criterion (table 4) shows us that 10 hidden units is better.

In the Table 3, shows us the error squared mean with the different kernel density estimate about the training set and testing set. A good resulted is obtained with Triangle density estimate. If consider 10 neurons the mse would be $9.64\text{x}10^{-6}$ for training and 0.000150 for testing.

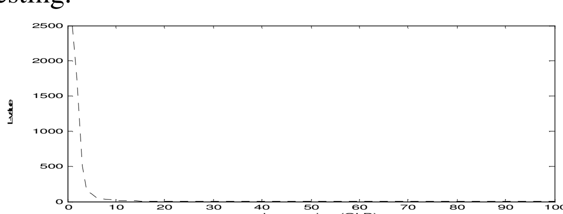

Fig 1. Decomposition QLP in the case Inverse Multipliquadric

Comparison of BIC to the case density estimate normal the values of 10 neurons against 7, 9 and 12 neurons were made. In the Table 4, shows us the least BIC value of $1.04\text{x}10^{-6}$ without pruning and a value 0.00010 with pruning by QLP decomposition in the matrix design (table 4).

In the Figure 2 has demonstrated that RBF reduced by QLP is better in Triangle, Epanechnikov and normal in comparison with Box desnity. In the table 3, in the case inverse multiquadratic RBF with kernel triangle presents a minor error squared mean (MSE) to training and test in comparison other case.

The value final prediction error with QLP was $3.59\text{x}10^{-5}$ and final prediction error was of $2.19\text{x}10^{-5}$.

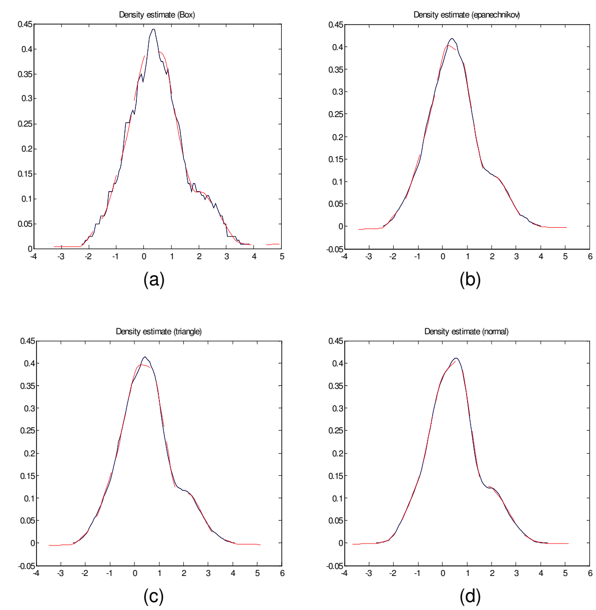

Fig. 2 Kernel Density Estimation (black solid line) and RBF reduced by QLP (red solid line)

Table 3 The error squared mean of method using RBF reduced by QLP decomposition

Density	Error train	Error test
Box	$4.85\text{x}10^{-5}$	0.000165
Triangle	$9.64\text{x}10^{-6}$	0.000150
Epanechnikov	$2.806\text{x}10^{-5}$	0.000159
normal	$3.399\text{x}10^{-5}$	0.000512

Table 4 Numerical results for BIC and EPF after reduction of matrix design.

Density	BIC		EPF	
	Φ	Φ_{QLP}	Φ	Φ_{QLP}
Box	$7.85\text{x}10^{-5}$	$18.1\text{x}10^{-5}$	$4.41\text{x}10^{-5}$	$11.1\text{x}10^{-5}$
Triangle	$1.73\text{x}10^{-6}$	$3.59\text{x}10^{-5}$	$9.63\text{x}10^{-7}$	$2.19\text{x}10^{-5}$
Epanechnikov	$1.04\text{x}10^{-6}$	0.00010	$1.04\text{x}10^{-6}$	$6.30\text{x}10^{-5}$
normal	$1.05\text{x}10^{-6}$	0.00012	$5.63\text{x}10^{-7}$	$7.37\text{x}10^{-5}$

*EPF=error prediction final and BIC=Schwartz Bayesian Information Criterion

8.2 RBF type Cauchy to Kernel Density Estimation

The QLP decomposition in the Fig. 3 for the RBF reduced by QLP, the pruning threshold is chosen as 40 neurons.

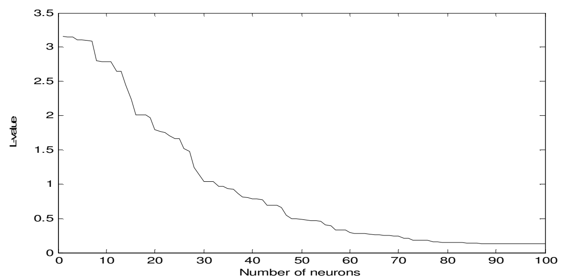

Fig 3. Decomposition QLP in the case Cauchy

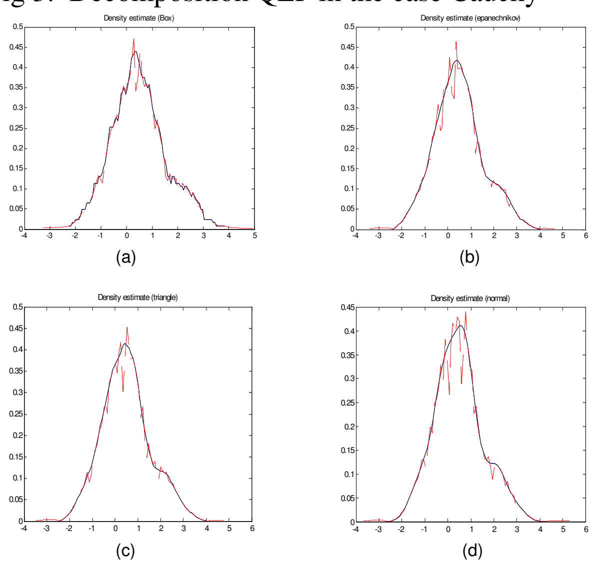

Fig. 4 Kernel Density Estimation (black solid line) and RBF reduced by QLP (red solid line)

Comparison of BIC to the case normal also the values of 40 against 30, 35 and 50 neurons were made. In this case (Cauchy) shows that the re-estimative to all kernel density by RBF reduced by QLP decomposition was less successful. The density estimates are roughly comparable, but the normal kernel produces a density that is rougher than the others.

Table 5 The error squared mean of method using RBF reduced by QLP decomposition (Cauchy)

Density	Error train	Error test
Box	0.00029	0.00016
Triangle	0.00033	0.00015
Epanechnikov	0.00038	0.0001
normal	0.00057	0.0001

Table 6 Numerical results for BIC and EPF after reduction of matrix design.

Density	BIC		EPF	
	Φ	Φ_{QLP}	Φ	Φ_{QLP}
Box	$7.37\text{x}10^{-5}$	0.00110	$3.82\text{x}10^{-5}$	0.00065
Triangle	$8.40\text{x}10^{-6}$	0.00073	$4.28\text{x}10^{-6}$	0.00073
Epanechnikov	$7.63\text{x}10^{-6}$	0.00144	$3.91\text{x}10^{-6}$	0.00085
normal	$1.03\text{x}10^{-5}$	0.00216	$5.20\text{x}10^{-6}$	0.00127

*EPF=error prediction final and BIC=Schwartz Bayesian Information Criterion

8.3 RBF type Multiquadric to Kernel Density Estimation

The QLP decomposition in the Fig. 5 for the RBF reduced by QLP, the pruning threshold is chosen as 12 neurons also.

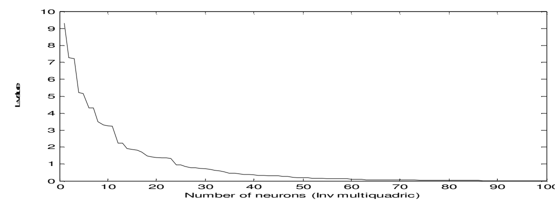

Fig 5. Decomposition QLP in the case Cauchy

Table 7 The error squared mean of method using RBF reduced by QLP decomposition

Density	Error train	Error test
Box	0.00031	$1.40\text{x}10^{-4}$
Triangle	$4.06\text{x}10^{-5}$	0.00014
Epanechnikov	$3.67\text{x}10^{-5}$	$1.39\text{x}10^{-4}$
normal	$9.09\text{x}10^{-6}$	0.00014

The Figure 8 consists of a linear combination of 12 radial functions to re-estimate all kernel density.

In this case kernel density by RBF reduced by QLP decomposition was less successful in their tail behavior (Figure 6). In this case the area of probability density estimate is not 1. Comparison of BIC values of 12 against 10, 15 and 20 neurons were made, the value was a value of $8.59\text{x}10^{-5}$ to BIC without reduction QLP and $5.96\text{x}10^{-5}$ to matrix design with reduction QLP.

Table 8 Numerical results for BIC and EPF after reduction of matrix design.

Density	BIC		EPF	
	Φ	Φ_{QLP}	Φ	Φ_{QLP}
Box	$8.44\text{x}10^{-5}$	0.00024	$4.96\text{x}10^{-5}$	0.00024
Triangle	$7.69\text{x}10^{-7}$	$6.60\text{x}10^{-5}$	$4.44\text{x}10^{-7}$	$5.16\text{x}10^{-5}$
Epanechnikov	$8.59\text{x}10^{-5}$	$5.96\text{x}10^{-5}$	$5.96\text{x}10^{-5}$	$4.66\text{x}10^{-5}$
normal	$7.00\text{x}10^{-10}$	$1.47\text{x}10^{-5}$	$3.99\text{x}10^{-10}$	$1.15\text{x}10^{-5}$

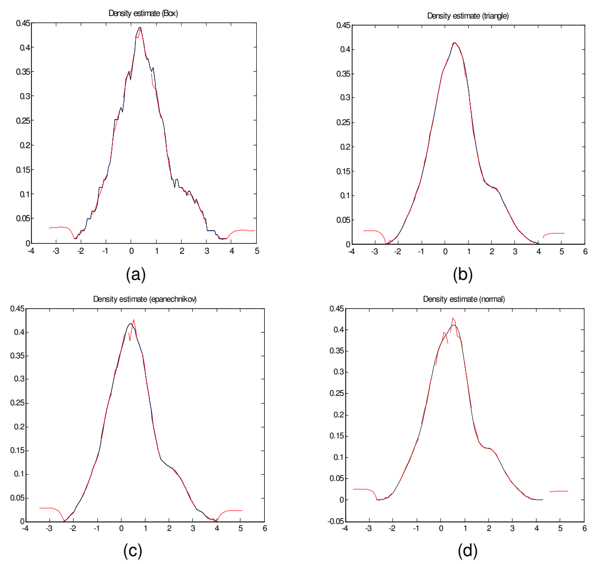

Fig. 6 Kernel Density Estimation (black solid line) and RBF reduced by QLP (red solid line)

9 Conclusion

In this paper, a simple idea of using RBF reduction by QLP decomposition to approximate density estimate has been develops. The experimental results demonstrate the potential of our proposed techniques, indicating that QLP is effective when the RBF centre aren't adjusted and the regularization parameters are kept fixed.

The value BIC to minor number of neurons confirm the QLP decomposition. In the Figures 4 and 8 shows the RBF network estimate after the reduction in the number of hidden units, a good resulted is obtained with 10 neurons in the case kernel density estimate triangle with function RBF type inverse multiquadric. We also showed that mean square error of selection RBF for training and testing shows us a better value in the case Inverse multiquadric. The re-estimative of the Cauchy and multiquadric RBF type function were less successful for all kernel estimates.

We conclude that specific density estimate can be accurately carried using a multiquadric RBF neural network pruning with QLP decomposition.

References:

[1] Scoott, David W. "On optimal and data-based histograms", Biometrika, 66, 1985, pp. 605-610

[2] Fix, E. and Hodges, J.L, Jr. "Nonparametric Discrimination: Consistency properties". Report Number 4, USAF Scholl of aviation Medicine, Randolph Field, Texas. 1951.

[3] Rosenblatt, M. "Remarks on some Nonparametric Estimates of a density function". Ann. Math. Statistic, 27, 1956, pp 823-837.

[4] Sterwart, G.W. *On an Inexpensive Tringular Approximation to the Singular Value Decomposition.* Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, U.S.A.

[3] Stewart, G.W. *Matrix Algoritms: Basic Decompositions*. SIAM Publications. Philadelphia, U.S.A., 1998.

[4] Christopher M. Bishop. Pattern Recognition and Machine Learning. Computer Science. Ed. Springer, pp 144-145. U.K.

[5] Mark J.L. Orr Matlab Routines for subset Selection and Ridge Regression in Linear Neural networks, Centre for Cognitive Science, University of Edinburgh, Edinburg Scotland, U.K., 1997.

[6] Edwirde L.S and Paulo Lisboa. Comparison of Wiener Filter solution by SVD with descompositions QR and QLP. Conference WEAS Corfu Greece, 2006.

[7] Edwirde L.S and Paulo Lisboa. Differentiating features for the Weibull, Gamma, Log-normal and normal distributions through RBF network pruning with QLP. Conference WEAS Corfu Greece, 2006.

[8] Ian T. Nabney. NETLAB. Algorithms for Pattern Recognition. Birmingham, U.K., 2002.