

Differentiating features for the F distributions with different degrees of freedom through RBF network pruning with QLP

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 proposes an artificial neural network RBF to classification using feature descriptors. The theoretical and practical aspects of theory F distributions with different degrees of freedom introduced. The distribution F densities are similar in shape, making it difficult to identify the differences between the two densities. This paper is concerned with separating these same probability densities with different degrees of freedom using feature descriptors, identified by pruning a Radial Basis Function (RBF) network using pivoted QLP decomposition generated for densities function, and its validity were evaluated by the rate of correct classification. The QLP method proves efficient for reducing the network size by pruning hidden nodes, resulting is a parsimonious model which identifies four main features (namely kurtosis and skewness and mean). The classification model induced by the methodology show, in general, good results.

Key-Words: - Radial Basis Function, Pivoted QLP Decomposition, Pruning, F distribution

1 Introduction

Recently, the Radial Basis Function network using pruning with pivoted QLP decomposition have been studied by researchers [1,2] that identified three densities function, very simile in shape, although RBF reduction by QLP decomposition. This paper applies this model to the discrimination, from data, between two F distribution with numerator degrees of freedom V_1 and denominator degrees of freedom V_2 . Several results show that RBF reduced by QLP decomposition are very powerful in classification and adjust [2].

The characteristic features and number of neurons RBF resulting from the network decomposition using QLP (a lower diagonal matrix L between orthogonal matrices Q and P) [3] for this particular choice of density functions has a very interesting interpretation. The distributions of interest are F(10,9) and F(9,8) density functions within these particular degrees of freedom designed. The analysis presented in this paper identifies four key discriminate features (kurtosis and skewness and mean).

This paper is organized as follows. Section 2 describes the F density probability, section 3 formulates the detection of the Numerical rank of the QLP, section 4 presents the simulated data to test the proposed approach to discrimination, section 5 reports on the quantitative performance comparisons for the

pruned RBF networks, section 6 experiment results. Finally, section 7 offers conclusions from this study.

2 Probability density function F

Named for the great statistician R.A.Fisher, the statistic F distribution has a natural relationship with the chi-square distribution. If χ_1 and χ_2 are both chi-square with 1 and 2 degrees of freedom respectively, then the statistic F below is F distributed. The degrees of freedom is a set of observations in a given context, the number of values that can be assigned freely, without restriction.

$$F(v_1, v_2) = \frac{\chi_1 / v_1}{\chi_2 / v_2}$$

The two parameters, χ_1 and χ_2 , are the numerator and denominator degrees of freedom. That is, 1 and 2 are the number of independent pieces of information used to calculate 1 and 2, respectively [6]

The pdf for the F distribution is

$$y = f(x | v_1, v_2) = \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} \cdot \frac{x^{\frac{v_1-2}{2}}}{\left[1 + \left(\frac{v_1}{v_2}\right)x\right]^{\frac{v_1+v_2}{2}}} \quad (1)$$

Where $\Gamma(\cdot)$ is the Gamma function.

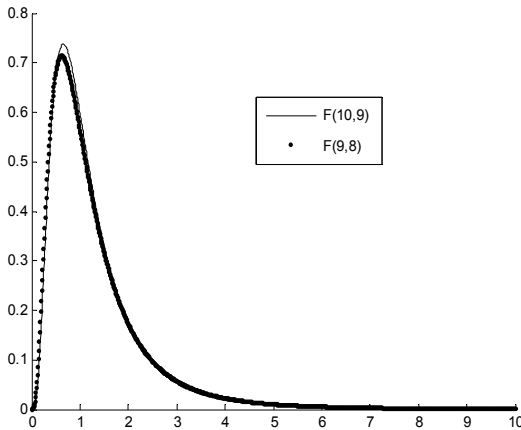


Fig. 1 The plot shows that the F distribution exists on the positive real numbers and is skewed to the right.

3 Detection of the Numerical rank of the QLP

The algorithm to compute the QLP decomposition can be used as an alternative to SVD and QR [3]. Since the gaps and singular subspace are defined in terms of the singular value decomposition, the natural way to compute them is to compute the singular value decomposition. Unfortunately, the computation of this decomposition is expensive. For this reason, researchers have proposed many alternatives. Of these the pivoted QR decomposition is widely recommended because of its simplicity. The pivoted QR decomposition has certain drawbacks: it gives only fuzzy approximations to the singular values, and fails to provide orthonormal bases for some of the fundamental subspaces [4].

It is clear that Singular value decomposition may be easily avoided by computation the rank-r principal subspace of

$$A_{xx}^{-1} = R_{yx} R_{xx}^{-1} R_{yx}$$

But the product these three matrices, involves additional computational cost and this is made worse by the calculation of the inverse matrix R_{xx}^{-1} . The QR decomposition produces an upper triangular matrix R

of the same dimension as A and a unitary matrix Q so that $A_{xx} = QR$, where $R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$, R_1 is an upper triangular matrix. The diagonal elements are called the R-values of A_{xx} . The column permutation E is chosen so that is monotonically decreasing, $A_{xx}(:, E) = QR$. To motivate the decomposition QLP consider the partitioned R-factor.

$$R = \begin{pmatrix} \kappa_{11} & \kappa_{12}^T \\ 0 & R_{22} \end{pmatrix}$$

A better estimate is the norm $l_{11} = \sqrt{\kappa_{11}^2 + \kappa_{12}^T \kappa_{12}}$ of the first row R.

$$RH_1 = \begin{pmatrix} l_{11} & 0 \\ l_{12} & \hat{R}_{22} \end{pmatrix}$$

We can obtain an even better value if we interchange (Π) the largest row with the first

$$\Pi_1 RH_1 = \begin{pmatrix} l_{11} & 0 \\ l_{12} & \hat{R}_{22} \end{pmatrix}$$

Now if we transpose, we see that it is the first step of pivoted Householder triangularization applied to RT. If we continue his reduction and transpose the result, we obtain a triangular decomposition of the form [3]

$$\Pi_L^T Q^T A_{xx} \Pi_R P = \begin{pmatrix} L \\ 0 \end{pmatrix}$$

We will call this the pivoted QLP decomposition of A_{xx} and will call the diagonal elements of the L-values of A_{xx}

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 A_{xx} . The folklore has it that the R-values track the singular values well enough to expose gaps in the latter. For example, a matrix A of order 100 was generated in the form [3]

$$A_{xx} = U \Sigma V^T + 0.1 \sigma_{50} E \quad (2)$$

where Σ is formed by creating a diagonal matrix (of size 1x100) decreasing geometrically from one to 10^{-3} and setting the last fifty diagonal elements to zero. U and V are random orthogonal matrices of size 100x100. E is a matrix of standard normal deviates.

Thus A represents a matrix of rank 50 perturbed by an error whose elements are one-tenth the size of the last nonzero singular value. In Figure 3, the values of $r_{50,50}$ and $r_{51,51}$ shows that there is a well-marked gap in the R-values, though not as marked as the gap in the singular values [4].

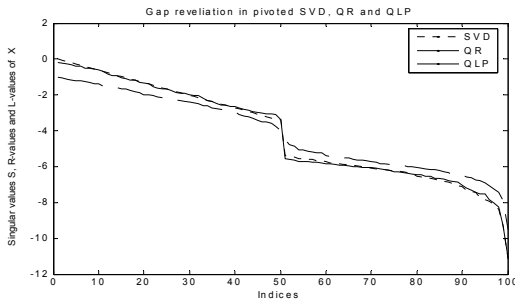


Fig. 2 Gap identification in pivoted SVD, QR and QLP

The gap in the L-value of the decomposition provides orthogonal bases of analogues of row, column, and null space provided by A_{xx}

The implementation of the Matlab package permits the QLP: $[P,Q,L,pr,pl]=qlp(A_{xx})$ to determine the numerical rank of matrix A_{xx}

Thus, we have a simple QLP algorithm as follows:

1. **define** matrix A, which consists of the hidden node activations in the RBF network.
2. **calculate** the orthogonal matrices Q and P which reduce the matrix A to lower diagonal form.
3. **identify** the diagonal of lower-triangular matrix L
4. **sort** the diagonal elements by size.

The set of input features used in this analysis are shown in table 1 [5]

Table 1 Resulting of the characteristics descriptive for comparison of F densities.

Descriptive features	order
Mean	1
Median	2
Q1 – first quartile	3
Q2 – second quartile	4
interquartile range (iqr)	5
sqrt	6
kurtosis	7
skewness	8

4 Design RBF Neural Classifiers

Figure 3 depicts the architecture for a fully connected RBF network. The network consists of n input features x, M hidden units with center C_j and y output. The ϕ_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. We shall write the RBF network mapping in the following form:

$$y_n(x) = \sum_{j=1}^M w_{nj} \phi_j(x) + w_{no}, \quad n = 1, 2, \dots, L \quad (3)$$

Finally the network outputs are calculated by

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

where σ_n is a linear transfer function.

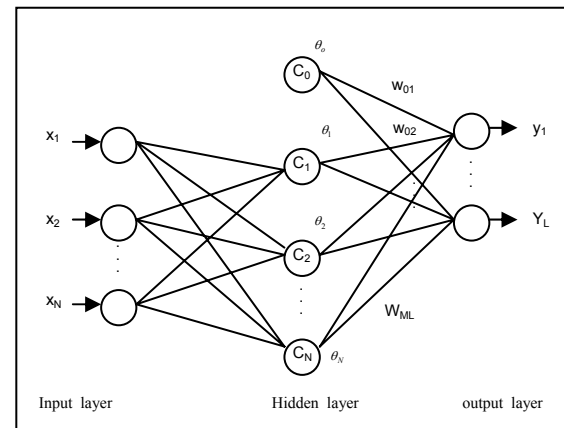


Fig. 3 A typical RBF network architecture

4.1 Proposed reduction RBF for the identification of important covariates

The RBF neural network is applied here to distinguish between two density functions F in pairs, for example discriminating between the F(10,9) versus, the F(9,8).

The first process is the creation of the matrix P and vector T (target matrix) of training for the problem. Each distribution has a matrix of size 200x500 comprising 500 samples of the distribution, each containing 200 observations, where x_i 's are uniformly randomly distributed in the (0,10). 8 features are extracted from each sample of 200 points, the features being described in table 1 [1]. The sample of 200x1 reduces the data matrix to 8x1 for each distribution, resulting in a data matrix of size 8x1000 of the two distributions. The training data matrix uses 8 features calculated from each column of the matrix, resulting in a data matrix of size 8x1000 when the data

from the two distributions is concatenated. RBF functions are trained to distinguish between these two distributions. Starting with the number of hidden nodes (ie radial basis functions) equal to the number of data points, the hidden nodes are reduced using QLP decomposition. The algorithm is based in the reduction of number hidden nodes with QLP.

For each trial, a training set (x_i, p_i) was established, where p is an indicator function representing the density function which generated the sample vector x. An out of sample test set was generated for performance estimation. The data were generated in a similar way to the training data, but with 300 instead of 500 samples [1,2].

The QLP algorithm is applied twice: first, to the training data matrix, of size 8x1000, to identify the most important inputs, which are features of the distribution types; then, again, to the matrix of output node weights, of size number of hidden nodes x number of output nodes, this time to identify the most important hidden nodes, i.e. radial basis functions. The key variables identified by the QLP method are those with the highest ranking values of the diagonal matrix L.

The reduction in the number of hidden nodes is carried out by sorting the diagonal elements generated by the QLP decomposition and selecting the corresponding nodes so as to give good classification accuracy with a small number of hidden nodes.

The results of network training are analyzed for various conditions used in the learning process. Thus, for RBF models, the number of hidden neurons is regarded as the result of training (which depends on two key parameters, namely mean squared error goal and spread). Note that of the results reported here, the reported computational effort does not include the centre selection phase and is solely for the training of the RBF with reduction neurons. The algorithm was the same the whole analysis and the stopping criterion for training was that the error function should be less than 1×10^{-2} and a maximum value of 0.5 for the radial function widths.

4.2 Apparent error rate (APER) and Correct classification rate (CCR).

We find that the overall performance of the RBF reductions classification approach is fairly good. A good classification procedure should result in few misclassifications. The apparent error rate (Aper) is the fraction of observations in the test set that are misclassified by RBF reductions.

$$Aper = \sum_{i=1}^k n_i / \sum_{i=1}^k N_i$$

where N_i is the number of

observations into two cases particulars, and n_i is the number of items misclassified in another probabilities other than i . The Correct classification rate (CCR) is given by $CCR=1-Aper$.

6 Experimental Results

The QLP shown in Fig. 4 has 4 or 5 significant inputs, which are skewness, kurtosis, mean, deviation standard were the principal features between F(10,9) and F(9,8) densities.

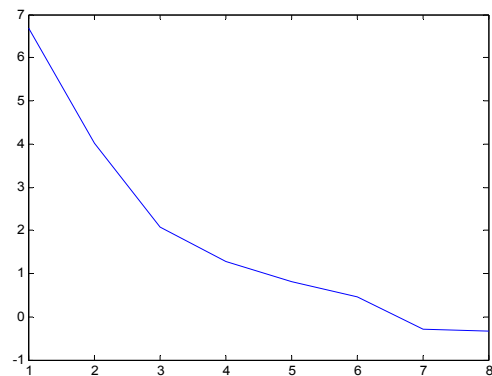


Fig. 4 Decomposition QLP between F(10,9) and F(9,8)

A preliminary calculation of the mean values of each feature, for each distribution, is shown in Table 2. In this Table note that the kurtosis and skewness presents best different between these two densities F. Thus justifying that the differences between the F(10,9) and F(9,8) become most significant in the kurtosis behavior. Also shows that the sensible differences between the F(10,9) and F(9,8) aren't in the tail (Fig 1).

Table 2 Descriptive characteristics of the probability density functions for each F density

Descriptive features	Density F(10,9)	Density F(9,8)
Mean	1.29	1.34
Median	1.07	1.02
Q1	0.25	0.13
Q2	0.25	0.14
iqr	0.86	1.05
sqrt	0.84	1.26
kurtosis	8.08	13.93
skewness	1.96	3.54

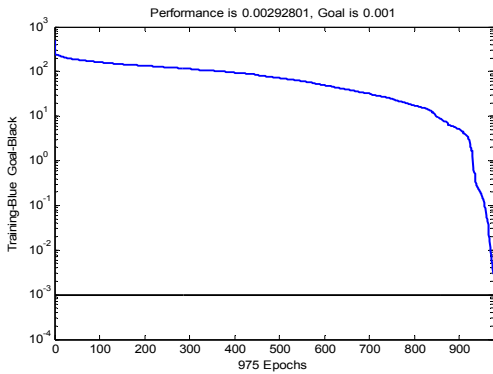


Fig. 5 Training of radial basis function

The training stopped after 975 iterations because the validation error increased. It is a useful diagnostic tool to plot the training. In the Fig.5 shows us that the RBF had learned these two F densities with a sum squared error value of 0.0029 and total neurons of 978.

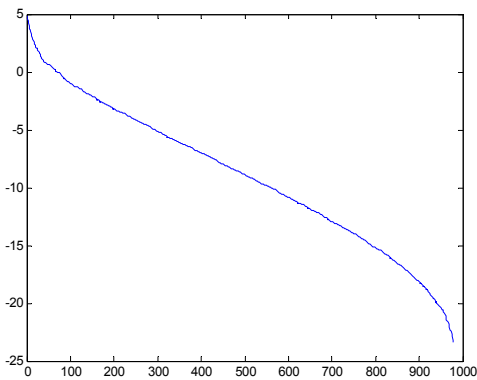


Fig.6 Decomposition QLP for identification number neurons RBF

In the Table 3 shows us that of 300 values F(10,9) and 300 values of F(9,8), the RBF reduction with 200 neurons has detected a total of 174 value for the F(10,9) and 187 for the F(9,8). In this case the reduced RBF (200 neurons) has better identification between these two densities highlighted by the characteristic statistics. The apparent error rates were 0.58 for F(10,9) and 0.62 for F(9,8). In the case 8 inputs the 200 neurons RBF has identified a value of 52% for F(10,9) and 79% for F(9,8). The apparent error rate (APER) is represented between brackets in Table 3.

Table 3 Resulting of RBF reduced of relationship for F(10,9) and F(9,8)

Case	Number inputs	Number neural	300 F(10,9)	300 F(9,8)
Network original Reduction	8	200	158(0.52)	238(0.79)
	6	100	168(.56)	178(.59)
	6	130	172(.57)	179(.59)

inputs and neural	6	150	173(.58)	181(.60)
	6	180	173(.58)	187(.62)
	6	200	174(.58)	187(.62)
	5	100	167(.55)	180(.60)
	5	150	170(.56)	187(.62)
5	180	171(.57)	188(.62)	

The reduced RBF by QLP decomposition identified these differences with approximately 200 neurons and 6 inputs because the curve of the frequency F(10,9) is bell-shaped and broadly similar to the F(9,8) distribution.

7 Conclusion

In this paper, a simple idea of using RBF reduction to identify characteristics among two densities has been developed. The resulting reduced networks show us that the kurtosis, skewness and mean were the principal characteristics that accurately separated the two densities. The RBF reduction by QLP decomposition can be effectively used for identification purposes. The F(10,9) and F(9,8) show us that 200 neurons with 6 inputs the RBF reduction identified a value of CCR=0.48 to F(10,9) and CCR=0.38 to F(9,8). The differentiation between these two distributions was less successful to below of 100 neurons. We conclude that specific class assignments to the F probability distributions with similar density functions can be accurately carried using an RBF neural network with QLP decomposition.

References:

- [1] Edwirde L.S and Paulo Lisboa. Comparison of Wiener Filter solution by SVD with descompositions QR and QLP. Conference WEAS Corfu Greece, 2007.
- [2] 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, 2007.
- [3] Stewart, G.W, *Matrix Algorithm: Basic Decompositions*. SIAM Publications. Philadelphia, U.S.A., 1998
- [4] Sterwart, G.W, On an Inexpensive Tringular Approximation to the Singular Value Decomposition. Department of Computer Science and Institute for USA, 1998, pp.1-16.
- [5] López P.C, *Matlab y sus Aplicaciones en las Ciencias y la Ingeniería*. Prentice Hall. Universidad Complutense de Madrid, Madrid, España, 2002
- [6] Matlab Reference Manual, Version 5.3, Mathworks, U.S.A., 1997.