A Modular Feedforward Network for Solving Classification Problems

Iulian B. Ciocoiu

Technical University of Iasi, Romania
Faculty of Electronics and Telecommunications, P.O. Box 877, Iasi, 6600, Romania
Email: iciocoiu@etc.tuiasi.ro

Abstract. We present a modular feedforward neural network for solving difficult classification problems. It is based on a hybrid architecture including RBF, PCA, and MLP modules, and is extremely efficient in terms of computational effort and number of parameters. Simulation results for the well-known 2-spirals benchmark problem are presented, and the possibility of using the proposed architecture as an autoassociative network is also outlined.

1 Introduction

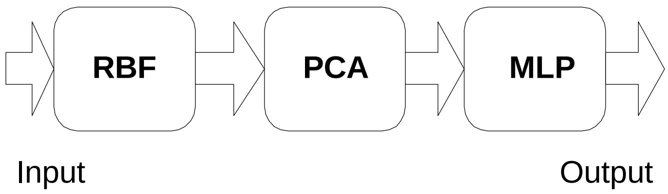
Pattern recognition is one of the most important applications of artificial neural networks. Such systems may be seen as semiparametric classifiers and their relation to the statistical approach has been extensively covered in the literature [1]. Feedforward networks have been mainly used for solving classification tasks, due to the their well-known universal approximation capabilities. In principle, any complicated decision surface may be closely modeled by such systems, given proper training data and learning algorithm [2]. Basically, two types of architectures have been considered: Radial Basis Functions (RBF) and Multilayer Perceptron (MLP). They both share the capacity of closely approximating any nonlinear multidimensional mapping, but use different "philosophies" in order to achieve discrimination: RBF's enlarge the dimensionality of the input data in order to increase the probability that originally nonlinearly separable classes become linearly separable (Cover's theorem [2]), whereas MLP's construct possibly non-convex and/or disjunct decision boundaries as a superposition of hyperplanes [2].

Another interesting approach to classification has been recently proposed in [3]. The basic idea is to train a neural network to learn an identity mapping: input and output layers have identical dimensions and input and desired data is the same. Typically, hidden layer(s) have lower dimension in order to force the extraction of significant (nonredundant) information from the input data. If trained properly, such networks would provide much lower *reconstruction errors* when fed with test data similar to the training database and much larger reconstruction errors when significantly different data is applied as input. This approach is especially useful for binary (2-class) classification problems, when input data is selected only from one of the classes. After training is completed data from both classes is applied at the input of the network, and reconstruction errors are computed for each

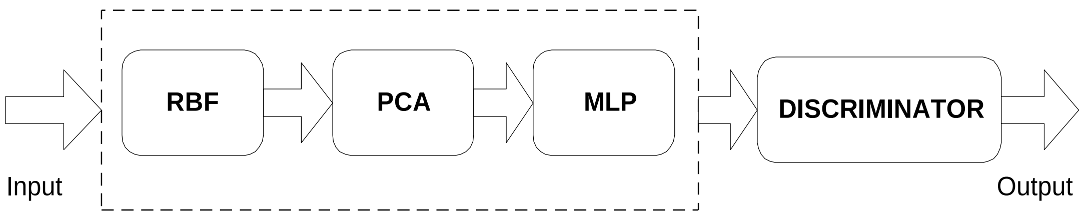
exemplar. If error values are clearly different (which should be the case if learning was accurate and the training database is noiseless) a discrimination threshold is chosen that may be used to yield correct class labels for subsequent test experiments.
We present two novel solutions to the well-known 2-spirals benchmark problem. The task is to correctly classify two sets of 194 training points that lie on two distinct spirals in the x-y plane. The spirals twist three times around the origin and around each other. This is a benchmark problem considered to be extremely difficult for standard MLP networks trained with classical back-propagation class algorithms, although successful results were reported using other architectures or learning strategies [4, 5]. The first solution is based on a modular system combining three distinct feedforward neural networks, namely a gaussian-type Radial Basis Function (RBF) network, a Principal Component Analysis (PCA) section, and a multilayer perceptron (MLP). The second one uses basically the same architecture but the dimension of the output layer is increased to match that of the input in order to enable autoassociation. Clearly distinct reconstruction errors are obtained for data belonging to each of the two classes, which makes (test data) discrimination very simple.

2 The proposed solutions

Typically, RBF and MLP's have been separately used in classification applications, but several benchmark tasks are known to be extremely hard to solve by standard solutions. We propose a novel approach based on a combination of the two, yielding a modular system with improved performances in terms of model parameters and computing time. The block diagram is presented in Figure 1a and includes a gaussian-type RBF network, a Principal Component Analysis (PCA) section, and a MLP.



a)



b)

Figure 1. a) Block diagram of the basic modular approach

b) Autoassociative network

Standard RBF networks implement general multidimensional mappings $f : R^m \rightarrow R$ according to:

$$f(\mathbf{X}) = w_0 + \sum_{i=1}^{M} w_i f(\|\mathbf{X} - \mathbf{C}_i\|) \qquad (1)$$

where ϕ is a nonlinear function selected from a set of typical ones, $\| \, . \, \|$ denotes the Euclidean norm, w_i are the tap weights and $C_i \in R^m$ are called RBF centers. It is easy to see that the formula above is equivalent to a special form of a 2-layer perceptron, which is *linear in the parameters* by fixing all the centers and nonlinearities in the hidden layer. The output layer simply performs a linear combination of the (nonlinearly) transformed inputs and thus the tap weights w_i can be obtained by using the standard LMS algorithm or its momentum version [2].

One of the recognized difficulties associated with such systems derives from the localized nature of the representation by the nonlinear function ϕ and is called 'curse of dimensionality': the quantity of training data needed to specify the mapping grows exponentially as the dimensionality of the input space increases [1]. This justifies the introduction of the second module, performing PCA. This represents a common preprocessing tool in pattern recognition applications, and defines the linear transformation that maximizes the variance (power) of the projection to the subspace spanned by the principal eigenvectors of the input covariance matrix. Examination of the corresponding eigenvalues would reveal the meaningful principal components (that is, those containing most of the original information) [2]. The reason for introducing the MLP module is based on the observation that in some cases the discriminant capacity of the output layer in the standard RBF architecture is insufficient: the decision boundaries can be so complex that a simple linear combiner could not yield satisfactory results.

The second approach uses the same basic architecture to implement an autoassociative network (Figure 1b). It may be included in the category of systems performing nonlinear PCA [7, 8], which have proven improved compression performances and discrimination power compared to the standard linear PCA .

3 Experimental results

Intensive computer simulations were performed to test the efficiency of the proposed solutions. We successfully classified correctly all the points in the training database using as few as 10 centers in a gaussian-type RBF module. We used random initialization and the simulations were repeated 10 times. From various existing neural approaches for implementing PCA we have selected Sanger's rule [2] due to its simplicity. The average values and the standard deviations of the eigenvalues obtained from the PCA analysis are presented in Table 1 and show that more than 90% of the information output by this module is contained only in the first 3 components.

Table 1: Eigenvalues obtained from PCA analysis for a RBF module with 10 centers (average values ± standard deviation)

Index	Value
λ_1	55.2±1.47
λ_2	43.6±1.53
λ_3	15.2±0.91
λ_4	4.9±0.43
λ_5	3.9±0.43
λ_6	1.9±0.34
λ_7	1.1±024
λ_8	0.0±0.00
λ_9	0.3±0.13
λ_{10}	0.3±0.09

The activation surfaces formed by each of the 3 output neurons of the PCA module are presented in Figure 2, showing distinct responses to the input data.

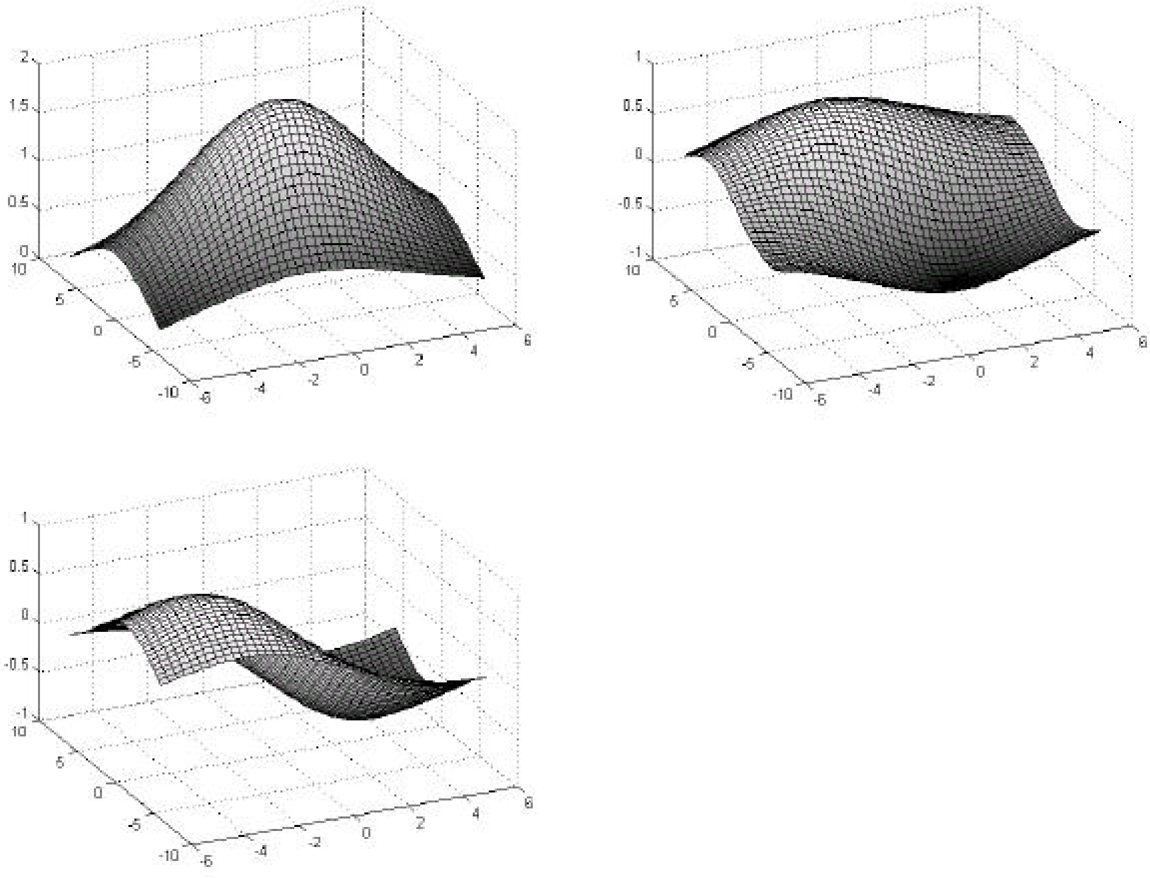
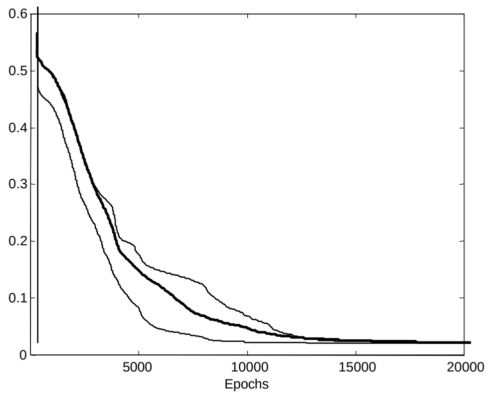
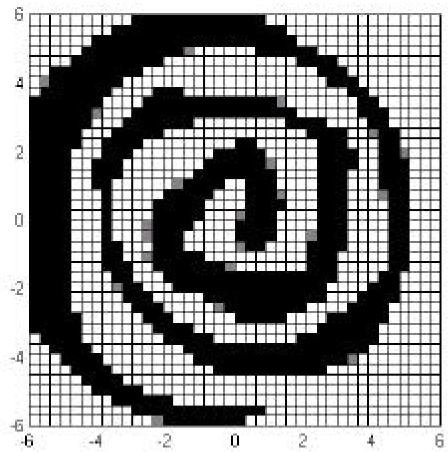


Figure 2: Activation surfaces of the output neurons in the PCA module

We used a 3-10-1 MLP network with tanh activation function for both hidden and output layers. The training procedure worked sequentially: first the positions of the RBF centers were obtained by using a competitive unsupervised learning algorithm, then PCA was performed, and finally the MLP network was trained by a fast variant of the backpropagation algorithm called delta-bar-delta [2]. In Figure 3a we present the evolution of the Mean-Square Error (MSE) during training. Classification accuracy was tested on a 41x41 test grid, and results are given in Figure 3b.
The same RBF-PCA-MLP architecture was used to implement the second approach, except for the output layer of the MLP network that was set to 2 neurons, similar to the input of the system. In the first stage of the training procedure, data vectors belonging to only one of the two classes were used both as input and desired data, and the same unsupervised-supervised sequential training described above was performed. In the second learning stage the network was fed with input vectors belonging to both classes and the (Euclidean) reconstruction error was computed for each exemplar. The error values are presented in Figure 4a. It is obvious that these values are clearly apart, which enables to set a discrimination threshold about 0.001. Finally, test vectors from a 41x41 grid were applied as input and class labels were given after comparing the corresponding reconstruction error to the above threshold value. Results are given in Figure 4b.

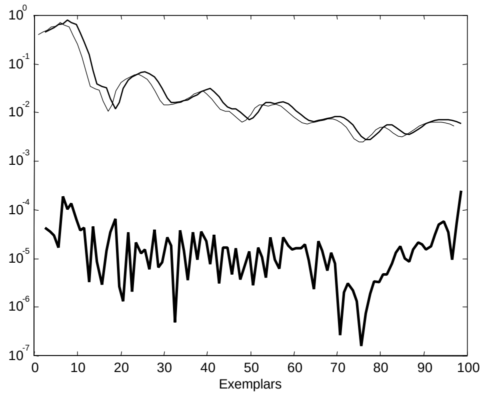


a)

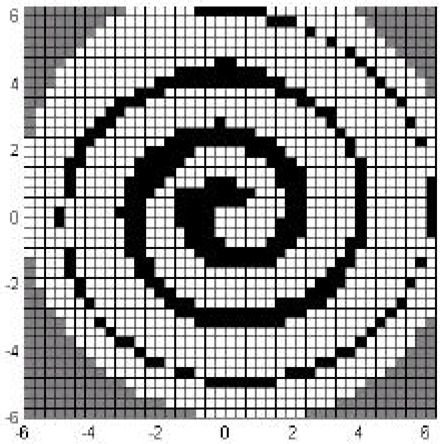


b)

Figure 3: a) Mean-Square Error (MSE) as a function of training epochs. The solid thick line represents the average training curve, and the dotted lines the minimum and maximum curves, respectively.
b) Classification performance on the test set for the modular network



a)



b)

Figure 4: a) Euclidean reconstruction errors for the autoassociative network: thick line – for exemplars in the training set (from only one class); thin line – for exemplars in the test set (the other class)
b) Classification performance on the test set for the autoassociative network

4 Conclusions

The first approach combines the capacity of RBF and MLP networks to form local, respectively global features. It defines a trade-off between the space selectivity provided by local kernel activation functions and augmented discriminant capacity of nonlinear networks. In terms of model parameters, our system configuration requires only 71 parameters, which compares favorably to the cascade-correlation solution in [4] (roughly a MLP network with 40 hidden units, that is 161 parameters), and the centroid based MLP in [5], with 77 parameters.
The second approach yields slightly worse classification accuracy. Its performances could improve if two autoassociative networks "tuned" to each of the classes are used. Anyway, as pointed out in [3], this is an interesting solution if training data for one of the classes is difficult to obtain. Other feedforward autoassociative architectures could also be used.

Acknowledgement

This work was supported by SCOPES 2000-2003 under grant 7RUPJ062381.

References

[1] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, New York, 1995.

[2] S. Haykin, Neural Networks - A Comprehensive Foundation, IEEE Press, New-York, 1994.

[3] N. Japkowicz, C. Myers, and M. Gluck, "A novelty detection approach to classification", Proc. 14th Joint Conf. on Artif. Int., 1995, 518-523.

[4] S.E. Fahlman, and C. Lebiere, "The cascade-correlation learning architecture", Proc. NISP, vol. 2, 1990, 524-532.

[5] M. Lehtokangas, "Determining the number of centroids for CMLP network", Neural Networks, vol. 13, 2000, 525-531.

[6] D.S. Broomhead, and D. Lowe, "Multivariable functional interpolation and adaptive networks", Complex Syst, vol. 2**,** 1988, 321-355.

[7] H. Bourlard, and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition", Biological Cybernetics, vol. 59, 1988, 291-294.

[8] G.W. Cottrell, and P. Munro, "Principal component analysis of images via back propagation", Proc. SPIE, Cambridge, MA, 1988.