Pattern Recognition Letters, 16(2), 207-212,

Elsevier-Science B.V., 1995.

Investigating Feedforward Neural Networks with Respect to the Rejection of Spurious Patterns

G. C. Vasconcelos M. C. Fairhurst D. L. Bisset

Electronic Engineering Laboratories

University of Kent at Canterbury

Canterbury, Kent CT2 7NT, United Kingdom

Abstract

Abstract : The reliability of feedforward neural networks with respect to the rejection of patterns not belonging to the defined training classes is investigated. It is shown how networks with different activation functions and propagation rules construct the decision regions in the pattern space and, therefore, affect the network's performance in dealing with spurious information. A modification to the standard MLP structure is described to enhance its reliability in this respect.

Key words : Feedforward neural networks, Pattern classification.

1 Introduction

In many pattern recognition applications such as optical character recognition, voice recognition and face recognition, a relevant aspect to be considered is the ability of the method employed to deal with patterns that do not share real membership with any of the training classes present in the domain of application. One of the most commonly used neural networks, the multilayer perceptron (MLP) trained with the backpropagation algorithm [Rumelhart (1986)], has been shown to perform poorly in this respect, classifying with a high degree of confidence patterns with random characteristics [Vasconcelos (1993)]. In this letter, three feedforward neural networks are compared with respect to the definition of the spatial classification regions generated. These are the standard MLP network, an MLP which employs a Gaussian as its activation function (GMLP) [Dawson & Schopflocher (1992)], and the radial basis function network (RBF) [Broomhead & Lowe (1988)]. It is shown how the standard MLP network can be modified to generate boundaries surrounding the training data to enhance its reliability.

2 MLP networks

2.1 Recognition of spurious patterns

Two of the main characteristics of the standard MLP network are, first, the use of the inner product as the network's propagation rule ($net = \sum_j w_{ij}x_j$) and, second, the use of the sigmoid as its activation function ($(1 + exp(-net))^{-1}$). Consequently, the network builds decision regions where each unit responds positively for an infinite portion of a half space. This property can be advantageous in some situations, mainly in terms of maximizing generalization capabilities, but it can be a major disadvantage when the primary issue to be considered is the network's reliability in dealing with spurious information. Indeed, it has been observed in experiments with handwritten characters [Vasconcelos (1993), Lee (1991)] that MLP networks can confidently accept patterns with random shapes as if they were legitimate representatives of the training classes. This problem can be investigated through the use of the technique of network inversion [Linden & Kindermann (1989)], which consists of clamping the network's weights after it has been trained and of modifying iteratively the initial pattern fed to the input matrix until it has been confidently classified as belonging to one of the training classes. By proceeding in this way, sets of patterns can be obtained which are considered by the network as valid members of the classes, such as those shown in Figure 1. In this example, random spurious patterns are classified, from right to left, top to bottom, as belonging to the classes of handwritten digits 0 to 9.

2.2 Explaining the network's unreliability

The difficulty in understanding the characteristics of a network that influence its reliability motivates the development of a very useful experimental procedure which allows the visualisation of the classification regions generated by the network. This consists of training a 2-hidden layer MLP network with a sample of training patterns in a 2-class classification problem in a 2-dimensional input space, and of testing its classification response for a large matrix of points covering the input space. The classification decision associated with each point in the matrix is determined in accordance with a confidence level imposed at the network's output. A pattern is classified as belonging to one of the 2 classes if the output (in the range [0,1]) for that class exceeds that of the other by a chosen confidence level (here taken as 0.65); otherwise, the pattern is rejected. The result of the experiment (Figure 2) demonstrates that the vast majority of the points, even those most dissimilar to the training patterns, are still confidently classified by the MLP network (points represented by the elevated surfaces in the diagram). As can be seen, only the points situated in between the training data are rejected. The reason for this is that the only constraint involved in the determination of the placement of the hyperplanes is the nature of the learning algorithm which places hyperplanes directly between classes rather than around them, and therefore open decision regions are created in those parts of the input space for which no information is available. This characteristic of standard MLPs trained with backpropagation is not always taken into account when practical applications are considered. It has been shown that MLP networks with different numbers of hidden layers are capable of generating any form of closed decision regions. An important point to make, however, is that closed regions are created only provided that there is sufficient information available to the network during the training phase. Not only is the necessary information extracted from the patterns in the training classes but also information from the other many parts of the input space is required, if boundaries are to be defined around the training points. A very useful additional experiment can be described to clarify this point. It consists of modifying the experiment above to include a group of chosen points surrounding the training classes in the original training set. These points are used to train the network negatively so that, every time one of them is presented, the network is taught to reject it. Figure 3 illustrates the result of this experiment which shows that, with the introduction of the additional points, the network is able to generate closed decision regions. This procedure has, however, very limited practical application, since in real world problems there is generally very little (or no) a priori information about the kinds of spurious patterns that should be rejected by the network and, therefore, there is no effective way of choosing the negative training examples.

3 GMLP networks

Since the inherent unreliability of standard MLPs, caused by the creation of open decision regions in the input space, can be partially atributed to the type of activation function used in the network, one possible way of developing an enhanced structure is to replace the sigmoid with a Gaussian ($exp(-net^2/c^2)$) [Dawson & Schopflocher (1992)]. The motivation for comparing a Gaussian MLP (GMLP) with the MLP network is due to the fact that, with the use of the Gaussian function, the receptive field of each network's unit corresponds to a hyper-hill in the pattern space which "prunes" the unit to respond only to part of the half space and which, therefore, tends to cause more confined regions surrounding the training data. The same experiment as described above, when repeated for the GMLP, provides the characteristics shown in Figure 4. In this case, the training patterns from one of the classes are confined to the region in the middle of the diagram and the other points are considered to belong to the second class. This situation can be considered more reliable for rejecting spurious patterns than that obtained with the standard MLP, especially when there is an increasing number of training classes present in the problem. A further experiment with 9 classes, with results shown in Figure 5, provides evidence to support this observation. In this case, the decision regions created by the network tend to correspond to closed regions with a much more confined area of the input space associated with each class. In the same situation, the standard MLP provides the solution illustrated in the diagram of Figure 6. As can be seen, the rejection areas in the pattern space for the case of the standard MLP correspond only to the regions in between the training classes and patterns far away from the training points are still very likely to be classified as valid patterns. The problem is minimized with the GMLP since the rejection areas are substantially increased. This difference in the definition of the classification regions is expected to have a consequent impact on the network's reliability and generalization performance.

4 RBF networks

There has been an increasing interest in the use of RBF networks in practical applications as an alternative to MLP networks trained with backpropagation. RBF networks have been compared with MLPs with respect to many characteristics such as speed of learning, memory requirements and generalization performance [Lippmann (1989)]. However, an important practical issue which has in the past received relatively little attention is the question of its reliability. In contrast with both MLP and GMLP networks, each hidden unit in an RBF network [Broomhead & Lowe (1988)] responds to a localized receptive field in the input space formed by the combination of a Euclidean distance measure (usually) as the propagation rule ($\sum_j (x_j - w_{ij})^2$) and a Gaussian as the network's activation function. As a result, the output of the units reaches its maximum when the input pattern is near a centroid and decreases monotonically when it becomes more distant from the centroid. Since the centroids are typically chosen randomly from the training patterns and units respond positively only to a certain local area around the centroids, inputs very dissimilar from the training patterns tend to receive low network output. It is because of this characteristic that RBF networks do not suffer from the problem of confidently classifying random patterns [Lee (1991)] and are expected to provide more reliable structures when compared to MLP and GMLP networks for rejecting spurious information in general practical problems. Figure 7 depicts the decision surfaces generated by an RBF network in the 9 class problem, and illustrates a situation where all the training points are bounded by nine circular regions.

5 A modification to the MLP structure

Another possible way of modifying the MLP network towards a more reliable structure comes from computational geometry theory, and is based on the observation that if a linear threshold unit receives input from an extended input space where the $(n+1)$ coordinate is the sum of the squares of the other n coordinates, its receptive field becomes a sphere [Omohundro (1990)]. The original input space is mapped into a paraboloid of revolution in a space of higher dimension and the hyperplanes in this space intersect the paraboloid creating ellipsoids that form localized receptive fields, when projected onto the original input space. Figure 8 illustrates the construction of the paraboloid for the 2-

dimensional case. The idea is based on the assumption that patterns belonging to the same class are expected to have approximate values for the $(n+1)$ coordinate and that the opposite occurs the more distant the patterns become from the training classes. Hence, the presentation of this additional information to the network enables it to define hyperplanes in the extended space which tend to separate the training points from the other parts of the input space. Here, it is shown how the implementation of this idea in the standard MLP modifies the decision regions generated by the network and, therefore, directly contributes to the enhancement of its rejection capabilities. The experiment described above with the 9 classes is carried out with this modified MLP structure and the result is shown in Figure 9. In contrast with the standard MLP and in common with the case of RBF networks, the decision regions created in this case tend to correspond to closed regions with a much more confined area of the input space associated with each class. The hyperplanes are placed in this situation not only to separate the training classes but also to create flexible boundaries surrounding the training points where patterns distant from them are rejected by the network. The effectiveness of this approach in different situations should be expected to vary according to the particular problem. Its application in spaces of lower dimensions is likely to be more successful since the information carried in the $(n+1)$ coordinate is, in this case, much more representative of the training classes. In problems involving higher dimensions a variation in the approach can be considered which corresponds to extending the original input space in m-dimensions instead of only one, and of associating the calculation of each one of the m coordinate values with subdivisions of the input space instead of the whole space.

6 Conclusion

This paper has presented a study of some feedforward neural networks with respect to the decision regions created when various network configurations are employed. It is possible to identify the reasons for the unreliability of standard MLP networks in rejecting unknown patterns and it is observed that alternative configurations such as the GMLP or the modified MLP network are candidates for a more reliable structure. MLPs, GMLPs and RBF networks can be compared with respect to various characteristics and a decision about which network to use should be made depending on the nature of a particular application. However, if the principal issue to consider is the reliability of the network in dealing with spurious information then the standard MLP is potentially the least attractive, unless an additional mechanism is provided to enhance its reliability. While RBF networks can be considered the most reliable of the three, GMLPs can be said to lie in between the other two models. One important aspect to emphasize is that, although the illustrations provided here are given in the context of a 2-dimensional problem, more extensive experiments carried out with handwritten characters show that the ideas presented scale up to the more general higher dimensional case [Vasconcelos (1994)].

Acknowledgment: G.C. Vasconcelos would like to acknowledge the support of the Brazilian Federal Agency for Postgraduate Studies (CAPES). Grant no. 2217/91-8.

References

[Rumelhart (1986)] Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). 'Learning internal representations by error propagation'. In Rumelhart, D.E., and McClelland, J.L.(Eds):'Parallel distributed processing, Vol. 1',, pp. 318-362. (MIT Press, Cambridge, MA), pp. 318-362.

[Vasconcelos (1993)] Vasconcelos, G.C., Fairhurst, M.C., and Bisset, D.L. (1993). 'Enhanced reliability of multilayer perceptron networks through controlled pattern rejection'. Electronics Letters, Vol. 29, No. 3, 261-263.

[Vasconcelos (1994)] Vasconcelos, G.C., Fairhurst, M.C., and Bisset, D.L. (1994). 'A study of mechanisms to enhance the reliability of multilayer perceptron networks with respect to the rejection of spurious patterns'. Internal Report, Electronic Engineering Labs, University of Kent.

[Dawson & Schopflocher (1992)] Dawson, M.R.W., Schopflocher, D.P. (1992). 'Modifying the generalized delta rule to train networks of non-monotonic processors for pattern classification'. Connection Science, Vol. 4, No. 1, 19-31.

[Broomhead & Lowe (1988)] Broomhead, D.S., Lowe, D. (1988). 'Multivariable functional interpolation and adaptive networks'. Complex Systems 2, 321-355.

[Linden & Kindermann (1989)] Linden, A., Kindermann, J. (1989). *Inversion of multilayer nets*. IJCNN Washington.

[Lee (1991)] Lee, Y. (1991). 'Handwritten digit recognition using K nearest neighbour, radial basis function, and backpropagation neural networks'. Neural Computation 3, 440-449.

[Lippmann (1989)] Lippmann, R. (1989). 'Pattern classification using neural networks'. IEEE Communications Magazine, November.

[Omohundro (1990)] Omohundro, S.M. (1990). 'Geometric learning algorithms'. Physica D 42, North-Holland, 307-321.

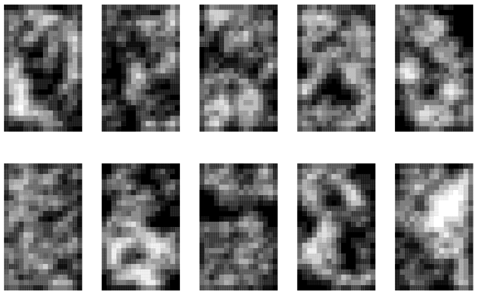

Figure 1: Random patterns classified by a MLP network

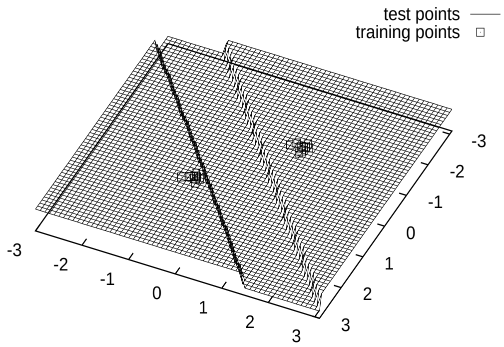

Figure 2: Decision regions for the standard MLP network

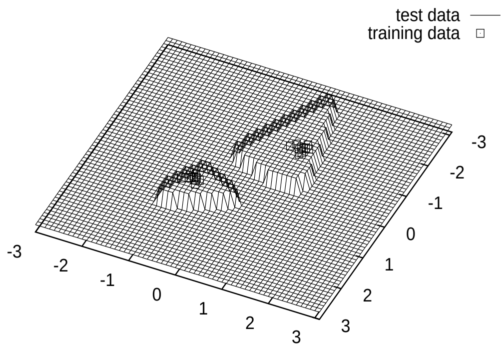

Figure 3: Decision regions for the MLP with negative training

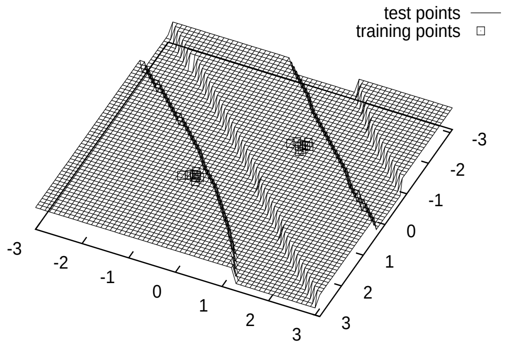

Figure 4: Decision regions for the GMLP network

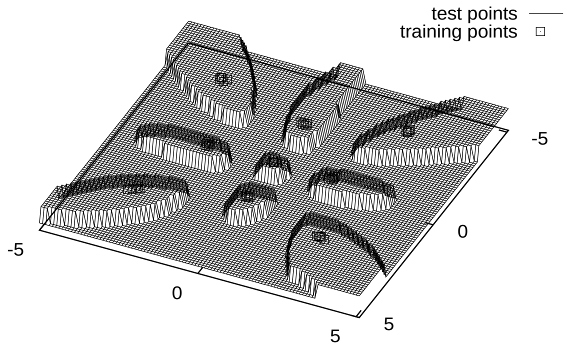

Figure 5: Decision regions for the GMLP network (9 classes)

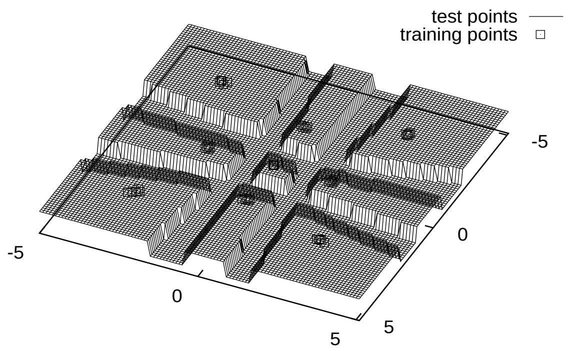

Figure 6: Decision regions for the MLP network (9 classes)

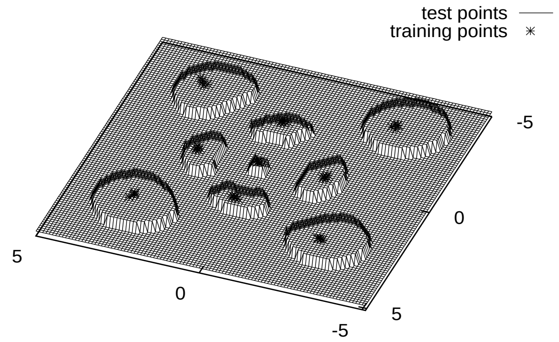

Figure 7: Decision regions for the RBF network (9 classes)

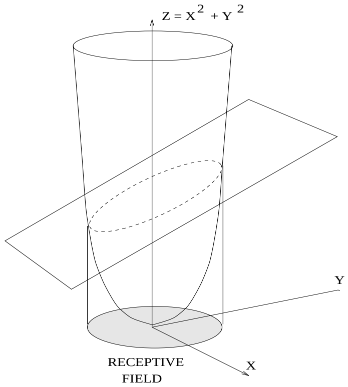

Figure 8: Construction of the paraboloid for the 2-dimensional case, redrawn from [Omohundro(1990)]

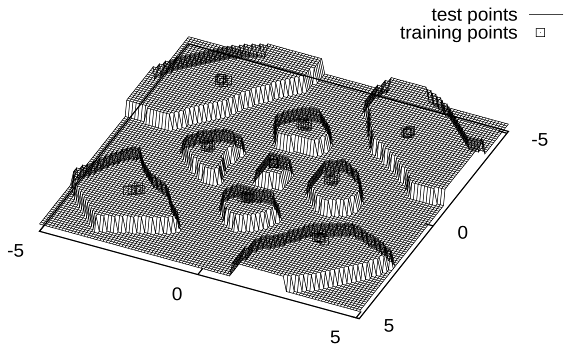

Figure 9: Decision regions for the modified MLP structure (9 classes)