Neural Computing & Applications, 3(4), 202-212,
Springer-Verlag, 1995.

Efficient Detection of Spurious Inputs for Improving the Robustness of MLP Networks in Practical Applications

G. C. Vasconcelos

M. C. Fairhurst

D. L. Bisset

Electronic Engineering Laboratories
University of Kent at Canterbury
Canterbury, Kent CT2 7NT
United Kingdom

e-mail : {gcv|mcf|dlb}@ukc.ac.uk

Efficient Detection of Spurious Inputs for Improving the Robustness of MLP Networks in Practical Applications

Abstract

The problem of the rejection of patterns not belonging to identified training classes is investigated with respect to multilayer perceptron networks (MLPs). The reason for the inherent unreliability of the standard MLP in this respect is explained and some mechanisms for the enhancement of its rejection performance are considered. Two network configurations are presented as candidates for a more reliable structure and are compared to the so-called "negative training" approach. The first configuration is an MLP which uses a Gaussian as its activation function and the second is an MLP with direct connections from the input to the output layer of the network. The networks are examined and evaluated both through the technique of network inversion and through practical experiments in a pattern classification application. Finally, the model of radial basis function networks (RBFs) is also considered in this respect and its performance is compared to that attained with the other networks described.

Keywords : Multilayer Perceptrons, Pattern Classification, Spurious Pattern Rejection.

1 Introduction

Pattern recognition has been one the most successful areas in which neural networks have found practical application. In this application area, the models developed are frequently evaluated with respect to their ability to correctly classify patterns belonging to the classes on which they have been trained. However, a further important characteristic in assessing the use of neural networks in reliable systems is the ability of the model employed to deal with the occurrence of patterns that do not share real membership with any of the training classes present in the application domain. Many networks have been shown to provide very good classification performance when they are tested with patterns of the same classes as those in the training set but, generally, little attention is paid to the question of how these models will behave if a pattern completely dissimilar to the training examples is presented to the network. The importance of developing appropriate methods for dealing with this kind of problem is particularly clear in applications where decision-making about the identity of the input patterns demands high reliability. In such situations, some form of rejection mechanism needs to be present in the environment either based on the built-in properties of the network itself, or through the introduction of an additional process integrated with the system. The work reported here concentrates specifically on investigating neural networks with respect to the former case, where the imposition of a threshold for measuring the confidence level of the network's outputs is the criterion used to accept or reject an input pattern.

It has been shown that one of the most effective networks used in pattern recognition, the multilayer perceptron architecture trained with backpropagation [1, 2, 3], can classify with a high degree of confidence patterns with random characteristics as if they were legitimate members of the trained classes [4, 5]. Few solutions have been proposed to overcome the problem, and these have usually been based on the "negative training approach", an approach which makes use of negative examples of random patterns presented to the network during its training phase. This approach was first introduced by Linden & Kindermann [4], subsequently investigated by Smieja & Muhlenbein [6] and also re-studied by Bromley & Denker [7]. More recently, more universal approaches have been presented which do not depend on *a priori* information about the classes of patterns that are to be rejected [8, 9].

In this paper, the reasons for the inherent unreliability of the standard MLP in this respect are explained and some mechanisms for enhancing its ability to reject spurious inputs are described. Two modified MLP configurations are investigated as candidates for a more reliable structure : the first is an MLP which uses a Gaussian as its activation function (GMLP) and the second consists of an MLP which uses additional direct connections from the input to the output layer of the network (DMLP). It is shown how these networks construct decision regions in the pattern space, and the implications of these constructions for pattern rejection are discussed. The negative training approach is also examined and its performance is compared to the GMLP and the DMLP networks. The technique of network inversion [4] is used to evaluate the various approaches, and practical experiments on the classification of handwritten characters are reported. Finally, the model of radial basis function (RBF) networks [10, 11] is also considered with respect to the rejection problem and a comparison is made with the standard MLP, the GMLP and the DMLP networks.

2 The Unreliability of the Standard MLP Network

In order to understand the factors which influence the rejection performance of a network in a classification problem, an examination of the network's operation at the level of the function computed by the processing units in the network is necessary. With respect to the standard MLP, one important feature is the fact that each unit in the network implements a global mapping through a non-linear discriminant function that divides the input space into two portions bounded only by the extreme limits of the input space. This is the result of the combination of the weighted sum of the inputs as the network's propagation rule ($net = \sum_j w_{ij} x_j$) and of the sigmoid as its activation function ($(1 + exp(-net))^{-1}$ or $tanh(net)$), which creates a global receptive field such as the one shown in Figure 1, for the 2-dimensional case. This property can bring advantages in terms of maximizing generalization capabilities, but together with the backpropagation learning algorithm, it is primarily responsible for the confident recognition of spurious inputs observed in MLP networks.

The following experiment, which visualizes the classification surfaces generated by a MLP network applied to a classification problem in the 2-dimensional space, provides a useful way to investigate network rejection performance. The experimental procedure consists of training a 2-hidden layer MLP with a sample of training patterns in a 2-class problem, 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; otherwise, the pattern is rejected. The result of the experiment using backpropagation with a high confidence level (0.65), shown in Figure 2, demonstrates that the vast majority of the points, even those most dissimilar to the training patterns (i.e at the extreme edges of the input space), are still confidently classified by the MLP network (points represented by the elevated surfaces in the diagram). As can be seen, only the points falling in between the training data points are rejected. The combined reasons for this are, first, that the discriminant function used by the processing units in the network corresponds to hyperplanes and, second, that the only constraint involved in the determination of the placement of the hyperplanes is the characteristic of the learning algorithm, which only places hyperplanes directly between classes rather than (as might be desired) around them. This means that 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. Although MLP networks are known to be capable of creating any form of complex closed decision regions [12], this experiment demonstrates that this does not occur with backpropagation when the training data only represents the classes to be learned, and hence does not provide information about the space *surrounding* the classes.

In many practical aplications it is clearly important that spurious unseen patterns are detected as such, and rejected. For example, in a factory process controlled by a neural network it is important to detect abnormal operating conditions so that the appropriate remedial action can be executed. It is therefore of considerable practical benefit to construct systems which are inherently able to decide when input patterns are genuine members of the known classes and when they simply correspond to spurious pattern classes, never seen before.

3 Improving the Rejection Capability of MLP networks

3.1 Training MLPs with Negative Examples

One of the possible methods for improving the rejection capability of MLP networks consists of presenting "negative" examples of random patterns to the network during its training phase and of teaching the network to reject them. This can be done in one of two ways : through the introduction of an additional unit in the output layer of the network representing the "rejection" class and mapping the random patterns to this unit, or through the minimization of the responses of all the network's output units when a random pattern is presented to the network. The purpose of this approach is to create attractors in the pattern space generated by the random examples so that when another random pattern is presented it is more likely to be classified as 'garbage' than as a valid member of the trained classes. The problem with this method is that there is no guarantee that other patterns both different from the training classes and also from the negative examples will not be accepted by the network. The decision regions generated to separate the other parts of the input space from the training classes are not uniform and are very much dependent on the expectation that the random patterns arbitrarily chosen to represent unseen patterns correspond to good representatives of the portions of the input space desirable to be considered as rejection areas. This approach was studied previously by Linden & Kindermann [4] and also by Smieja & Muhlenbein [6].

The effect of applying this method in the 2-class problem described above, in this case not with random patterns but with points carefully chosen to illustrate important characteristics, can be visualized in Figure 3. Here, a group of negative points surrounding the training classes is included in the original training set and the network is trained to reject these peripheral points. It is clear that the network is now able to generate closed decision regions because it has information about the whole input space.

3.2 The Gaussian MLP Network

An alternative approach to training a network to become a strong rejector is to consider a modification to the structure of the MLP network itself in order to transform it into an inherently reliable model. The unreliability of standard MLPs, caused by the creation of open decision regions in the input space, can be partially attributed to the type of activation function used in the network (i.e the use of an inherently "open" function). This implies that the use of a different activation function might profitably be considered. A multilayer network which employs a Gaussian activation function ($exp(-net^2/c^2)$) (GMLP) instead of the usual sigmoid has recently received some attention [13]. This modified architecture is claimed to provide faster training times and to reduce the number of units necessary for particular problems. This network is now examined with respect to spurious pattern rejection.

With the substitution of the sigmoid by a Gaussian function, the receptive field of each network unit is transformed into a hyper-hill in the pattern space which forces the unit to generate strong responses only if its input falls within a certain range. This is the result of the combination of the inner product as the propagation rule and the Gaussian as the activation function, which creates a semi-local receptive field such as that illustrated in Figure 4. As can be seen, points falling along the discriminant line perpendicular to the unit's weight vector produce the same unit output and as they move in parallel to the weight vector the unit response becomes weaker. When projected onto the plane, the decision region defined by the unit is bounded by two parallel hyperplanes.

Repeating the visualization experiment for the GMLP, this time for a problem with 9 classes, generates the characteristics shown in Figure 5. 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. This situation is more reliable for rejecting spurious patterns than that obtained with the standard MLP which provides, in the same situation, the solution illustrated in Figure 6. 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. This problem is minimized with the GMLP since the rejection areas are substantially increased.

3.3 MLP Networks with Direct Connections

The GMLP represents an improvement in rejection performance brought about by changing the activation function in the network. It is also possible to achieve a similar enhancement through structural modifications, introducing direct connections from the input to the output layer of the network. Apart from the theoretical work of Sontag [14] considering the recognition power of this variation of MLP networks in the particular case of threshold units, no work has been reported which investigates the implications, especially in practical applications, of using such direct connections. This section explores this novel network configuration (DMLP) with respect to its reliability as a pattern classifier. The motivation for this is that the representations stored in the direct connections tend to correspond more closely to template-like representations of the training classes, instead of the normal feature based representations obtained in the hidden layer(s) of the network. As a result, this introduces an additional constraint to be satisfied by the backpropagation algorithm when looking for a solution to a problem, since the response of the network's output units to the input patterns is more directly dependent on the similarity of the patterns to a template representation of the training examples. Figure 7 shows the result of the classification experiment with the DMLP network. Here, it can be seen that boundaries start to appear in the pattern space surrounding the training points. Similarly to the case of GMLP networks, it can be argued that this DMLP network configuration should provide a more reliable structure for rejecting invalid patterns.

4 Evaluating the Different Network Configurations

4.1 The Technique of Network Inversion

The previous sections have discussed the problem of spurious pattern rejection with respect to the simple, visualizable, 2-dimensional case. It is important to establish that these observations scale up to a higher dimensional class of problems, more appropriate to a practical application. This, however, leads to the problem of visualizing the results. One technique that can be used is to test the tolerance of the networks with respect to the visual appearance of the input patterns presented. Clearly, a network with limited tolerance to input pattern distortions will be good at rejecting spurious patterns. The method of network inversion for multilayer networks [4] provides a means of testing this tolerance. The technique consists of clamping the network weights after it has been trained and of modifying the pattern fed into the input matrix through gradient descent according to the least-mean-square error between the current network output and a given output target. In other words, the input pattern is successively modified until the output reaches the target. For example, if it is required to modify the pattern in the input matrix in order for to it be recognized as a '0', the network's target output for class '0' is clamped into the network's teaching vector and, with the weights of the network frozen, the input pattern is modified until it has been correctly classified as belonging to class '0'. The process stops when the global error between the current network output and the target output is less than a specified value. A convenient way to initialize the input matrix is to vary randomly the density of 'on' and 'off' pixels in order to transform the input pattern in an arbitrary fashion. Using this procedure, sets of patterns can be obtained which are considered by the network as valid members of the classes, such as those shown in Figure 8(a) for the case of the standard MLP. In this example, it is observed that patterns with a high degree of deformation are classified, from left to right, as belonging to the classes of machine printed digits 0 to 9, respectively.

With the application of the negative training approach, inverted patterns such as those shown in Figure 8(b) are obtained. In this case, there is some improvement in the overall appearance of the digits classified, although the patterns are still very indistinct. Experiments are also repeated for the DMLP architecture and the result of several runs shows a substantial improvement in the patterns accepted by the network. Figure 8(c) illustrates the result of one of the trials and it can be seen that the inverted patterns approximate much more the usual shapes of digits, indicating that the network is much less tolerant to deformed or unknown patterns. For the case of the GMLP network, the result of several trials also demonstrated an enhancement in the shapes of the inverted patterns (Figure 8(d)), when compared to the standard MLP, although less apparent than that attained with the DMLP. Such an analysis indicates that the observations made on the simple 2D-problem scale up to higher dimensions, that both DMLP and GMLP networks should provide more reliable structures for spurious pattern rejection, and that even the use of a simple technique such as negative training can produce an improvement over the standard MLP network.

4.2 Results in a Character Recognition Application

It is now important to establish the link between the performance observations discussed above and the rejection performance of networks in practical pattern recognition applications. Experiments were conducted on the application of the classification of handwritten characters with the MLP, DMLP, GMLP and the MLP with negative training (MLP^{neg}) networks.

The experiments consist of training the networks to recognize handwritten digits extracted from postcode information on envelopes in the UK mail, and of testing their performance in the rejection of spurious patterns using for this purpose alphabetic letters presented to the network during the recall phase. The database used to train the networks is composed of 1000 handwritten digits and the recognition performance of each network is tested on an independent set of 2000 digits. Each character is presegmented and size normalized as a binarised matrix of 16×24 pixels. A second database consisting of 7800 handwritten alphabetic letters is used to test the network's ability to reject patterns not belonging to the training classes. All the networks are single-hidden-layer networks tested with different numbers of units in the hidden layer. The results reported for the DMLP network correspond to a slightly modified version of the network which employs the normalization of the input pattern before feeding it to the hidden units. This configuration represents the best compromise between generalization error and the rejection of invalid patterns.

Tables 1 and 2 present the classification results obtained for networks with 80 hidden units. The experiments are repeated with different values for the classification confidence level imposed at the network's output. A pattern is classified as belonging to one of the classes if the output unit representing that class exceeds that of the others by the given confidence level. Otherwise, the pattern is rejected. Table 1 shows the correct recognition rates for the digits and also the proportion of letters rejected by the networks. Finally, Table 2 illustrates the percentage of digits erroneously classified together with the percentage of digits rejected by the network.

It is important to note that this test of rejection performance is a strict one since letters are similar in many respects to digits. In all the experiments carried out it has been observed that certain pairs of numerals and letters such as 0/O, 1/I, 8/B, 5/S, 2/Z, etc. are very difficult to differentiate. For example, in a practical situation it is impossible to differentiate between '0's (the numeral) and 'O's (the letter) unless some contextual information is provided. The results presented therefore reflect a particularly difficult task environment, and far higher rejection rates would be experienced if random or other non-character based patterns were used.

The results observed with the MLP^{neg} network show a relatively inconsistent pattern of behaviour. In some cases, a substantial improvement in the rejection performance is noticed when compared to the standard MLP, while in other cases only a small increase in the rejection rate is observed. In other instances no significant difference is apparent. This suggests that this approach is very dependent on the extent to which the negative random patterns chosen during the training of the network are representative of the parts of the pattern space considered as rejection areas. The experiments reported here correspond to the best case achieved with this approach. For example, for a confidence level of 0.15 the rejection rate is improved by 11 %, for a confidence level of 0.55 the improvement is by 5 %, whereas for a confidence level of 0.65 there is no improvement at all (indeed, in this case, there is actually a decrease in the rejection rate). With respect to the digit (valid class) recognition rates, some degradation in performance is also introduced, as can be seen in Table 1.

In the case of both the GMLP and the DMLP, a consistent increase in the letter rejection rate is observed when compared to the standard MLP as well as in comparison to the MLP^{neg} network. For the case of the GMLP, the rejection rates are increased by 20 % on average for small values of the confidence level when compared to the standard MLP, whereas for larger confidence level values the rejection ability is enhanced by 21 % on average. As might be expected, there is also some degradation in the digit recognition performance as a result of the more strict definition of the classification surfaces. The reduction in the digit recognition rate for small values of the confidence level is smaller, around 1.4 %, whereas for larger values the degradation in digit recognition is greater (around 6 %). For the case of the DMLP network, the increase in letter rejection for small values of the confidence level is by 17 % with a decrease in the digit recognition by 1.3 %. For larger values of the confidence level, the rejection rate is enhanced by q18 % whereas the digit recognition drops by 5.5 %.

One very relevant aspect to consider regarding the GMLP and DMLP, shown in Table 2, is that there is a reduction in the rate of misclassified patterns for all values of confidence level introduced, especially in the case of the GMLP (except where there is no confidence level imposed on the network, value = 0). For instance, while the digit error rate is 5.3 % for the standard MLP, with the value of 0.15 for the confidence level, it is 3.8 % for the GMLP. This is a result of the fact that some patterns, for which the network has no confidence about their identity, are being rejected by the GMLP instead of being erroneously classified.

It becomes readily apparent that GMLPs and DMLPs provide a more reliable structure for rejecting invalid patterns than the standard architecture. However, there is a small trade-off between enhancing the rejection capabilities and losing some performance in the recognition of valid characters. The choice of which network to use should be made depending on the parameters imposed by a specific application. Standard MLPs tend to maximize generalization capabilities but can be very unreliable in use, making them unsuitable for problems which demand a high degree of reliability.

The pre-negative training of the network with random patterns has also been shown to augment the rejection capabilities of the standard MLP network, although the method has not proved to be very reliable. A massive presentation of random patterns to the network would be necessary to increase the probability of creating good representations of rejection areas.

5 RBF Networks

In this section, the MLP, the GMLP and the DMLP networks are compared to the class of models known as radial basis function (RBF) classifiers [10, 11]. The motivation for this comparison is the fact that RBF networks appear naturally as very strong candidate structures for the reliable detection of spurious inputs due to the characteristics of the processing units in the network which compute localized functions of the pattern space. The classification problem in the simple 2-dimensional space is used first to visualize the shape of the decision regions generated by RBF networks and, subsequently, some results are presented with RBFs, in the same context of the classification of handwritten alphanumeric characters described in the previous section.

In contrast with the global function used in the units of a standard MLP, each node in the hidden layer of a RBF network computes a localized mapping of the input space defined by the combination of a distance metric (usually the Euclidean distance) as its propagation rule and a Gaussian as its activation function. The effect of this combination, as illustrated in Figure 9, is the creation of a receptive field which forces the unit to respond positively to only a restricted region of the input space close to a centroid stored in its weights. The output of the unit is maximum when the input pattern is near the centroid and decreases rapidly as the input becomes more distant from it. Since the weights for the hidden layer are defined to be template-like representations of the training patterns (usually, for example, by choosing randomly a sample of patterns from each training class and assigning them to the weights) any input which is significantly different from the training patterns receives a low output from the units in the network and is, therefore, rejected by the criterion of the confidence level imposed on the network. Figure 10 illustrates the decision surfaces defined by an RBF network in the 9-class problem investigated previously. As can be clearly seen, closed boundaries are naturally created surrounding the patterns from the training classes.

Practical experiments in the classification of alphanumeric characters have been carried out with RBF networks. The RBF used in the experiments corresponds to a network with 1000 hidden units (defined with the same number of 100 training patterns per class used previously). Table 3 presents the digit recognition and letter rejection results achieved by the network together with the previous results obtained with the MLP, GMLP and DMLP networks. It is observed that even for a small value of the confidence level (0.15) the rejection rate obtained with the RBF (44 %) is almost three times better than the equivalent with the standard MLP (17 %). For bigger values of the confidence level, the improvement in rejection performance is also very evident when compared to the MLP.

There is, however, a degradation in the recognition of valid digits critically proportional to the increase in the required confidence level. For example, it can be seen that with the use of the value of 0.55 for the confidence level, the digit recognition rate drops by up to 25 %. This a result of the fact that many input patterns that receive low confidence outputs from the hidden units in the network are being rejected, in order to reduce the risk of a misclassification. The evidence for this argument can be illustrated by the results outlined in Table 4 which presents a comparison between the proportion of digits rejected by the network and the percentage of digits erroneously classified. This shows a decreasingly small error rate with the direct variation of the confidence level and, conversely, an increase in the rate of valid patterns rejected by the network. For example, for a confidence level of 0.55 the error rate generated by the RBF network is around 0.2 % whereas for the standard MLP it is 3 %. On the other hand, the rate of digit rejection is around 44 % for the RBF and 15 % for the MLP network. However, it is important to note that if the detection of spurious inputs is one of the most important aspects to be considered in a given problem then the results obtained with the RBF network offer a better solution, even in terms of the trade-off between valid pattern recognition and spurious pattern rejection. For example, if it is required to establish a maximum error of around 3 % for the valid patterns, the RBF is able to recognize correctly 85.7 of the digits with the small confidence level of 0.15 imposed on network, as opposed to 81.3 % achieved with the MLP with the high value of 0.55 for the confidence level (see Table 3). In this case, the rejection of letters achieved with the MLP network is around 48 % and with the RBF network it is around 44 %. Another example can be given when a minimum recognition rate of 80 % is established. In this example, the rejection of letters achieved by the RBF is 63 % for a digit error rate of 1.4 % and the rate obtained with the MLP is 46 % for a digit error rate of 2.6 %. It is clear that a better balance between valid pattern acceptance and spurious inputs detection is obtained with the RBF network.

Comparing the results obtained with the the DMLP and, especially, the GMLP network presented in this paper, the improvement shown with RBF networks in terms of letter rejection is not as marked as that observed when compared to the standard MLP. The increase in the letter rejection rate is by about 15 % on average with respect to the GMLP and by 19 % when compared to the DMLP network. However, the same comparisons made above when applied to these networks show that, in various cases, the compromise between valid pattern recognition and invalid patterns rejection offered by them can be similar or, in some cases even superior, to that attained with the RBF network. Consider, for example, the following correlation between the results obtained with RBFs and the results obtained with GMLPs. For an error rate of 3.4 % observed with the RBF network, 85.7 % of the digits are correctly classified with 44 % of letters rejected by the network. An even smaller error rate of 2.9 %, obtained with the GMLP, provides 86.7 % of digit recognition and the same 44 % of letter rejection rate. In another case, for a digit recognition rate of about 80 % with the RBF, the digit error rate is 2.6 % and the letter rejection rate is about 63 %. For a recognition rate of 83.5 % obtained with the GMLP and a digit error rate of 2.0 %, the rejection of letters is about 55 %. A similar level of evaluation can be made with the DMLP network although, in this case, its performance is slightly inferior to that achieved by the GMLP network.

The results illustrated here indeed indicate that RBF networks are strong candidates for the reliable rejection of spurious patterns. However, the results also indicate that GMLPs and DMLPs can, in some situations, represent a better alternative in terms of the trade-off between valid pattern recognition and spurious pattern rejection.

6 Conclusion

The work reported in this paper has investigated MLP networks with respect to the problem of the detection of spurious inputs. It has been observed that when a confidence level imposed at the network's output is the criterion employed for accepting or rejecting a classification decision, the type of similarity measure computed by the processing units in the network plays an important role in its ability to reject spurious patterns. Units that compute global mappings of the input space, as in the case of standard MLPs, tend to classify into one single class extensive areas of the input space. Here, the simple application of a high confidence level for acceptance of a decision response is not sufficient, since in these conditions even patterns with very random characteristics are still confidently classified by the network. The GMLP network investigated in this paper represents a possible way of overcoming or, at least, minimizing this problem, where the combination of the inner product as the network's propagation rule and the Gaussian activation function creates a semi-localized receptive field in the pattern space. This network can offer the benefits of a much more enhanced structure for rejecting spurious data when compared to the standard MLP with only a small degradation in the recognition of valid patterns. The other alternative network configuration, the DMLP network, enhances the network's rejection ability through the imposition of additional constraints on the learning algorithm for the definition of the decision regions during the training process. The introduction of the direct connections allows for the generation of "template-like" representations in these connections which tend to force the learning algorithm to create boundaries surrounding the training examples.

In general, it can be said that networks in which the processing units compute a localized function of the input space should naturally provide more reliable structures for rejecting spurious inputs. One such class of networks correspond to the radial basis function architecture. In this case, because each processing unit in the network creates a localized receptive field in the input space, there is a very low probability of spurious patterns being accepted by the network. Indeed, the results shown here confirm that RBF networks are powerful candidates for applications that demand high reliability. However, an example illustrating the practical application of the classification of handwritten characters has shown that some "intermediate" network configurations between MLPs and RBF models such as the DMLP, and especially the GMLP, may offer a better compromise between the recognition of valid patterns and rejection of spurious patterns. The other advantageous feature of the GMLP and the DMLP when compared to RBFs is that they can improve the rejection abilities over the MLP network with a much smaller number of free parameters required in the network. Whereas the RBF network used 1000 hidden units for achieving the results reported in this paper, the GMLP and the DMLP each employed only 80 units as used in the standard MLP network. On the other hand, RBF networks offer the benefits of a very fast training process. It is clear, therefore, that the choice of which network to use in a practical application has to take into account the particular parameters demanded in the specific problem. This paper has explored and identified some of these parameters of choice.

Acknowledgements

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

References

[1] 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.

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

[3] Martin, G.L., Pittman, J.A. (1991). 'Recognizing Hand-Printed Letters and Digits Using Backpropagation Learning'. Neural Computation 3, 258-267.

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

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

[6] Smieja, F.J., Muhlenbein, H. (1992). 'Reflective Modular Neural Network Systems'. Technical Report, German National Research Centre for Computer Science, Germany.

[7] Bromley, J. , Denker, J. (1993). 'Improving rejection performance on handwritten digits by training with rubbish'. Neural Computation 5(3), 367-370.

[8] 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.

[9] Bishop, C. (1994). 'Novelty detection and neural network validation'. IEE Proceedings Vision, Image and Signal Processing, Vol. 141, No. 4, 217-222

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

[11] Moody, J., Darken, C. (1989). 'Fast learning in networks of locally-tuned processing units'. Neural Computation 1, 281-294.

[12] Lippmann, R. (1989a). 'Review of Neural Networks for Speech Recognition'. Neural Computation vol. 1(1), 1-38.

[13] 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.

[14] Sontag, E. (1990) On the recognition capabilities of feedforward nets. Technical Report SYCON 90-03, Department of Mathematics, Rutgers University.

List of Figures

List of Tables

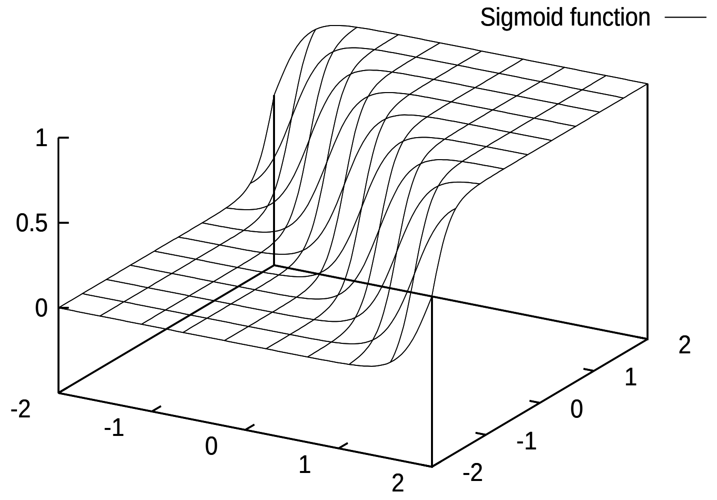

Figure 1: Combination of the inner product with the sigmoid activation function

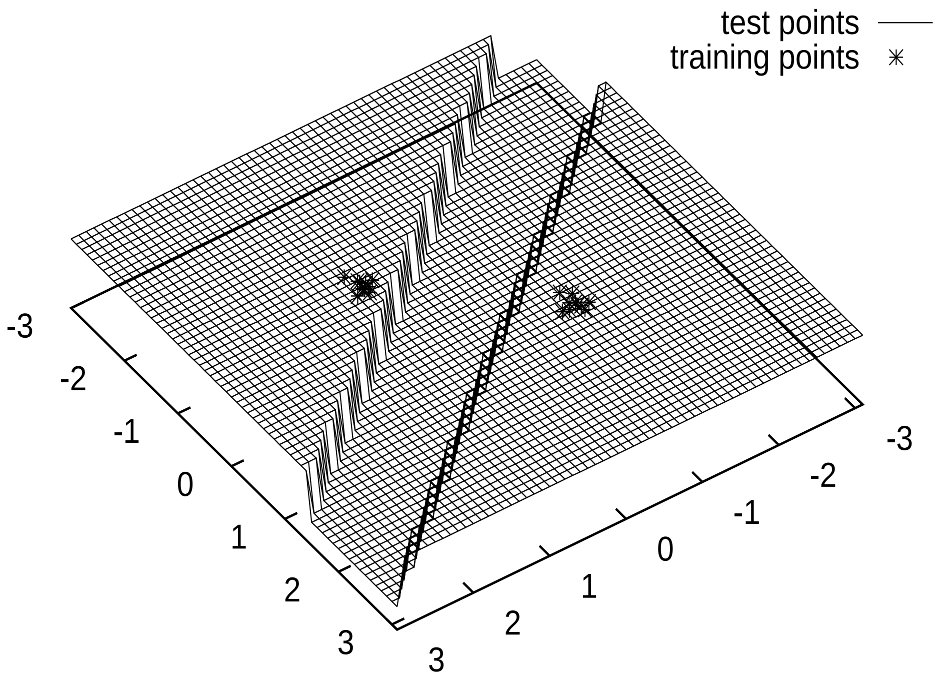

Figure 2: Decision regions for the standard MLP network

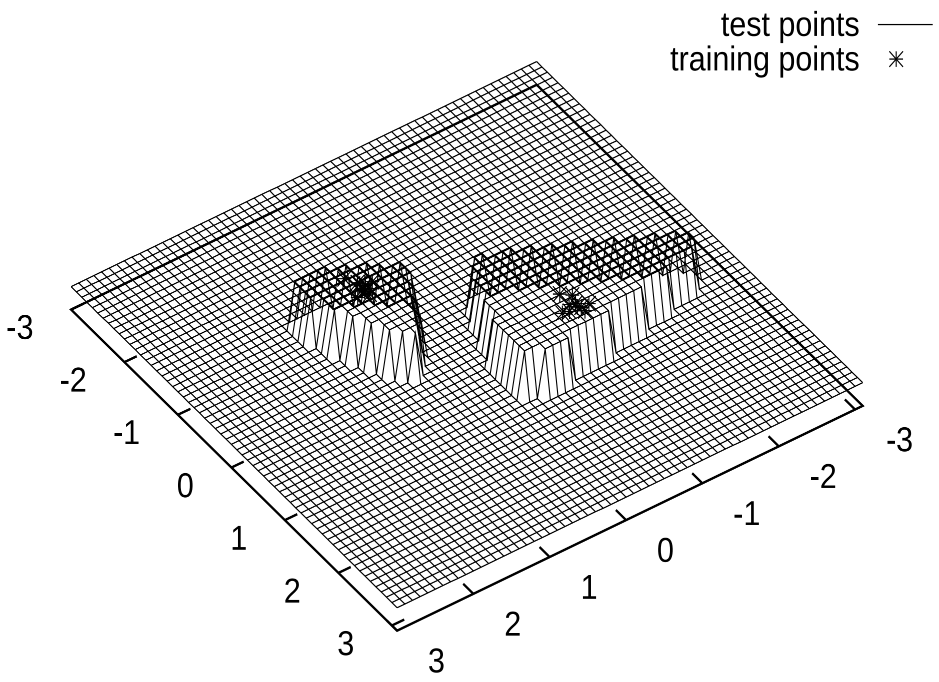

Figure 3: Decision regions for the MLP with negative training

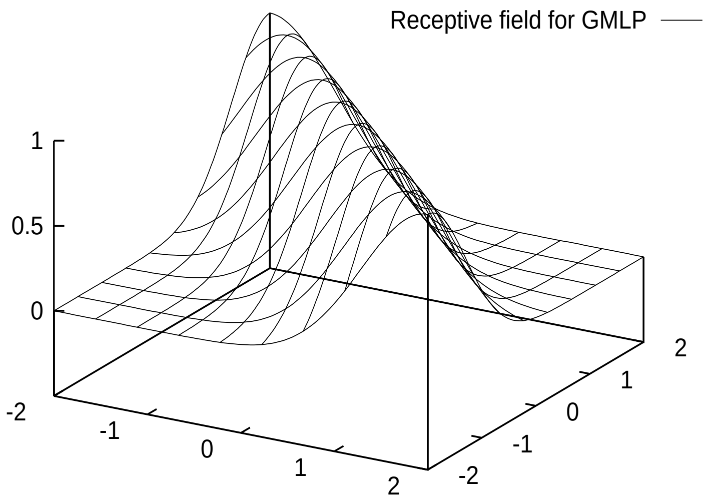

Figure 4: Combination of the inner product with the Gaussian activation function

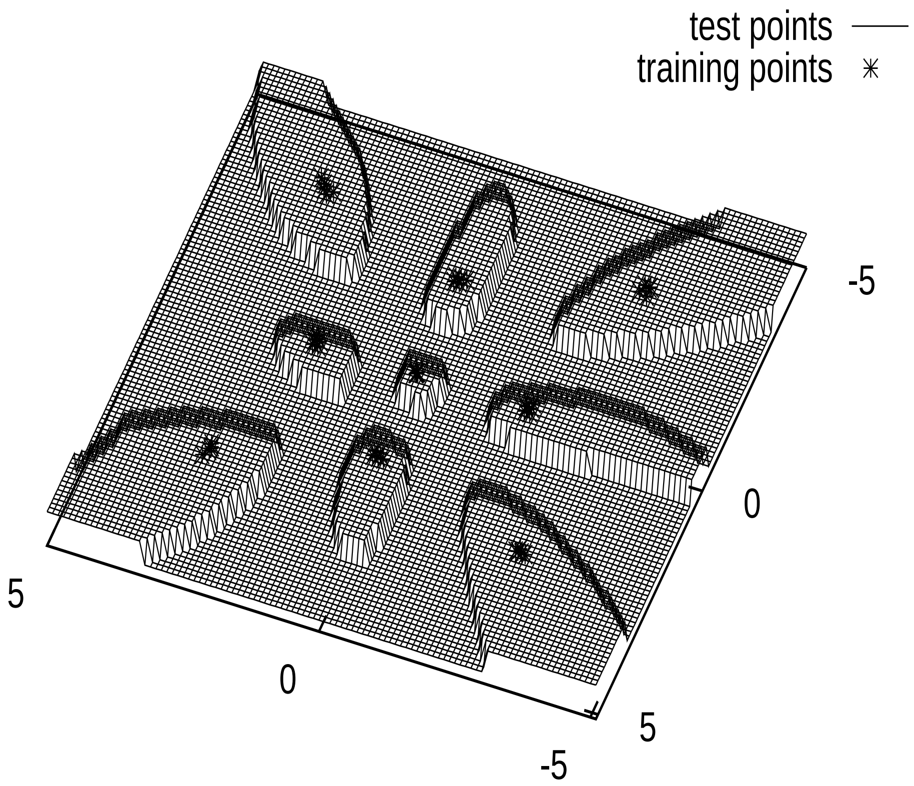

Figure 5: Decision regions for the GMLP network (9 classes)

	Digit Recognition Rate				Letter Rejection Rate			
confidence	MLP	MLP^{neg}	GMLP	DMLP	MLP	MLP^{neg}	GMLP	DMLP
0	92.0	91.5	91.0	92.2	0	0	0	0
0.15	89.4	87.2	87.5	88.3	17	28	36	29
0.25	88.2	86.2	86.7	85.7	25	36	44	42
0.35	84.4	80.3	83.5	82.9	32	48	55	53
0.45	80.8	79.4	77.6	78.4	46	56	65	62
0.55	81.3	80.2	71.4	73.6	48	53	76	71
0.65	72.9	78.4	67.5	66.5	64	60	80	79

Table 1: Digit recognition rate vs letter rejection rate

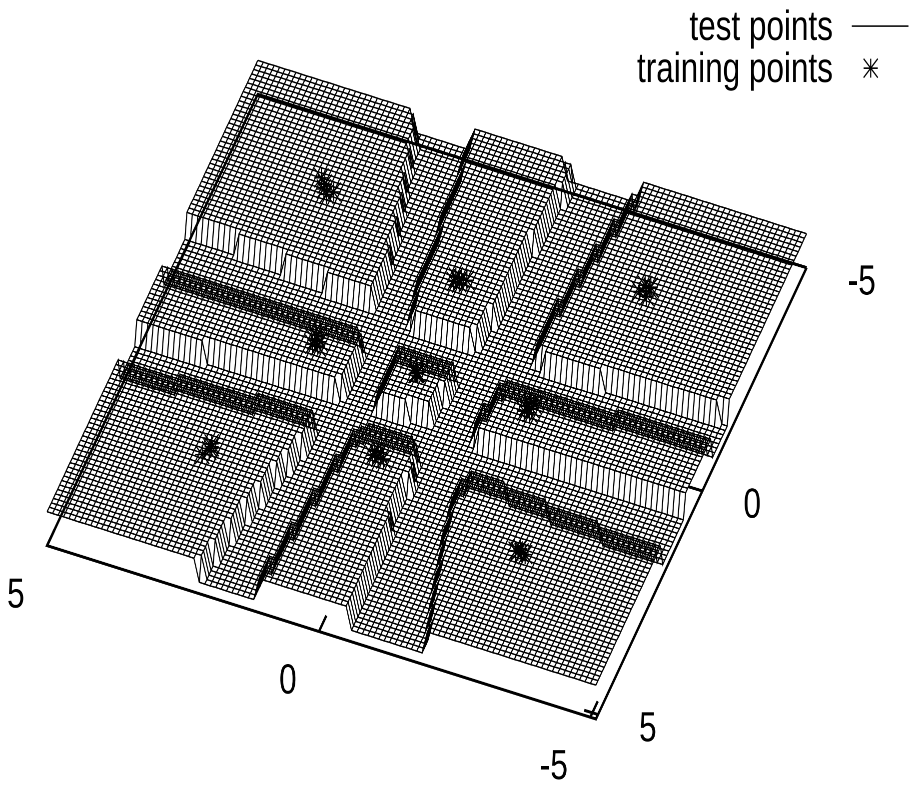

Figure 6: Decision regions for the MLP network (9 classes)

	Digit Error Rate				Digit Rejection Rate			
confidence	MLP	MLP^{neg}	GMLP	DMLP	MLP	MLP^{neg}	GMLP	DMLP
0	8.0	9.5	9.0	7.9	0	0	0	0
0.15	5.3	5.7	3.8	5.1	5.3	7.1	8.7	6.6
0.25	5.3	4.3	2.9	3.4	6.5	9.5	10.4	10.9
0.35	3.7	3.5	2.0	2.2	11.9	16.2	14.5	14.9
0.45	2.6	2.2	1.4	1.6	16.6	18.4	21.0	20.0
0.55	3.0	2.6	0.9	1.1	15.7	17.2	27.7	25.3
0.65	1.7	1.7	0.6	0.8	25.4	19.9	31.9	32.7

Table 2: Digit error rate vs digit rejection rate

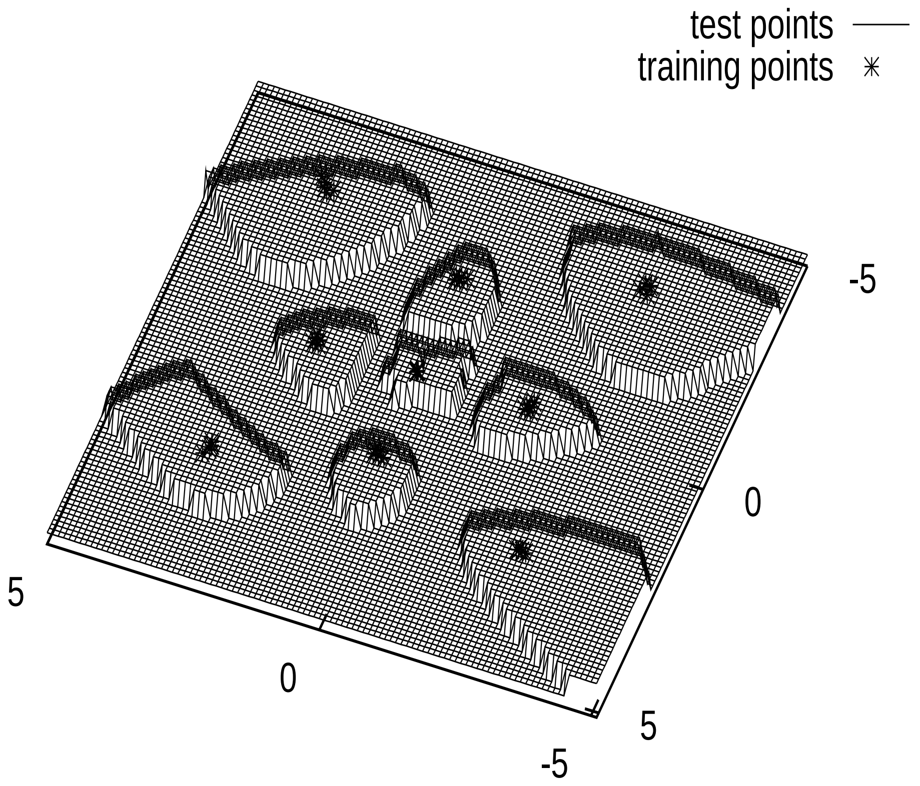

Figure 7: Decision regions for the DMLP network (9 classes)

	Digit Recognition Rate				Letter Rejection Rate			
confidence	MLP	GMLP	DMLP	RBF	MLP	GMLP	DMLP	RBF
0	92.0	91.0	92.2	91.1	0	0	0	0
0.15	89.4	87.5	88.3	85.7	17	36	29	44
0.25	88.2	86.7	85.7	80.3	25	44	42	63
0.35	84.4	83.5	82.9	74.9	32	55	53	74
0.45	80.8	77.6	78.4	66.7	46	65	62	82
0.55	81.3	71.4	73.6	55.8	48	76	71	89

Table 3: Digit recognition rate vs letter rejection rate including RBF networks

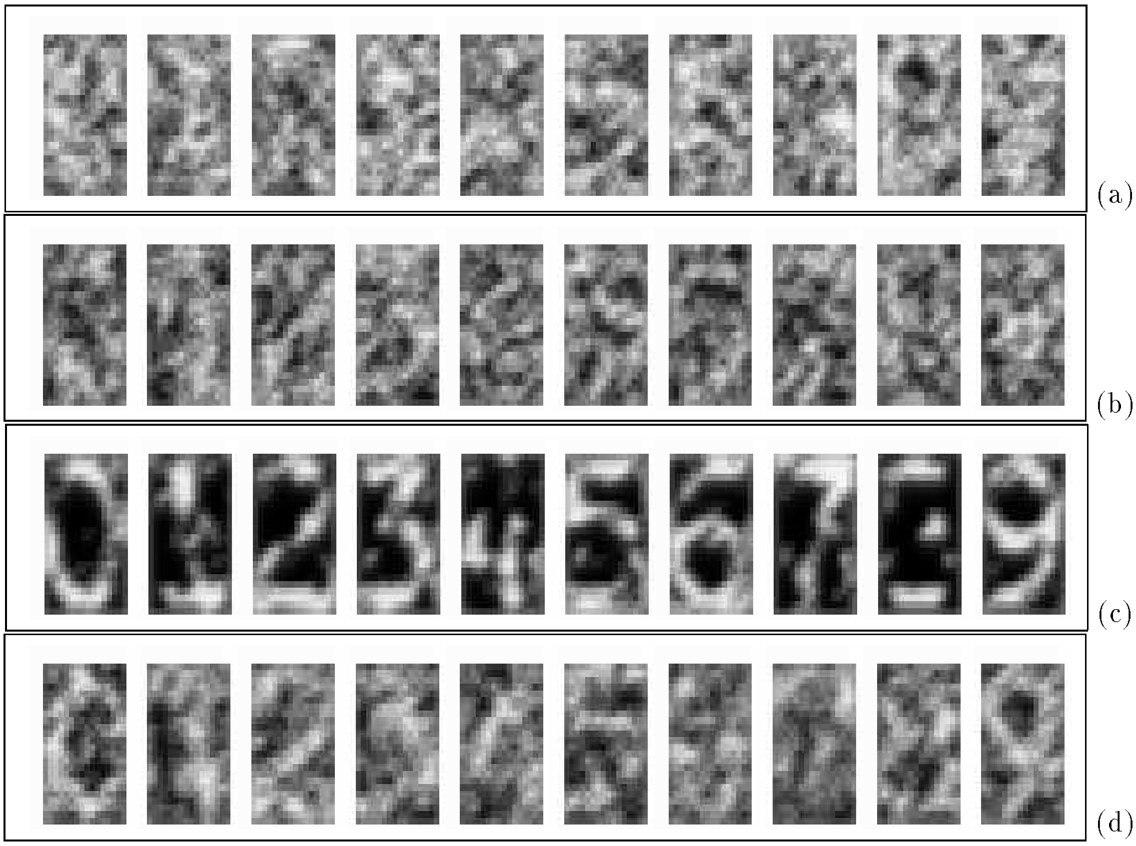

Figure 8: Inverted patterns : (a) MLP, (b) MLP^{neg}, (c) DMLP, (d) GMLP

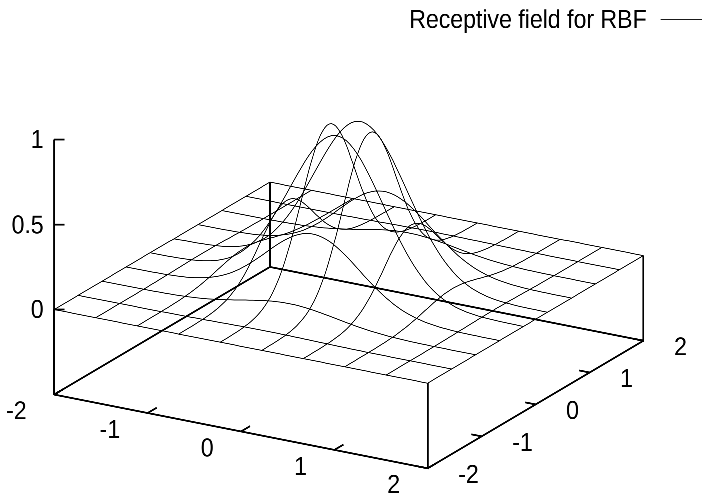

Figure 9: Combination of the Euclidian distance metric with the Gaussian activation function

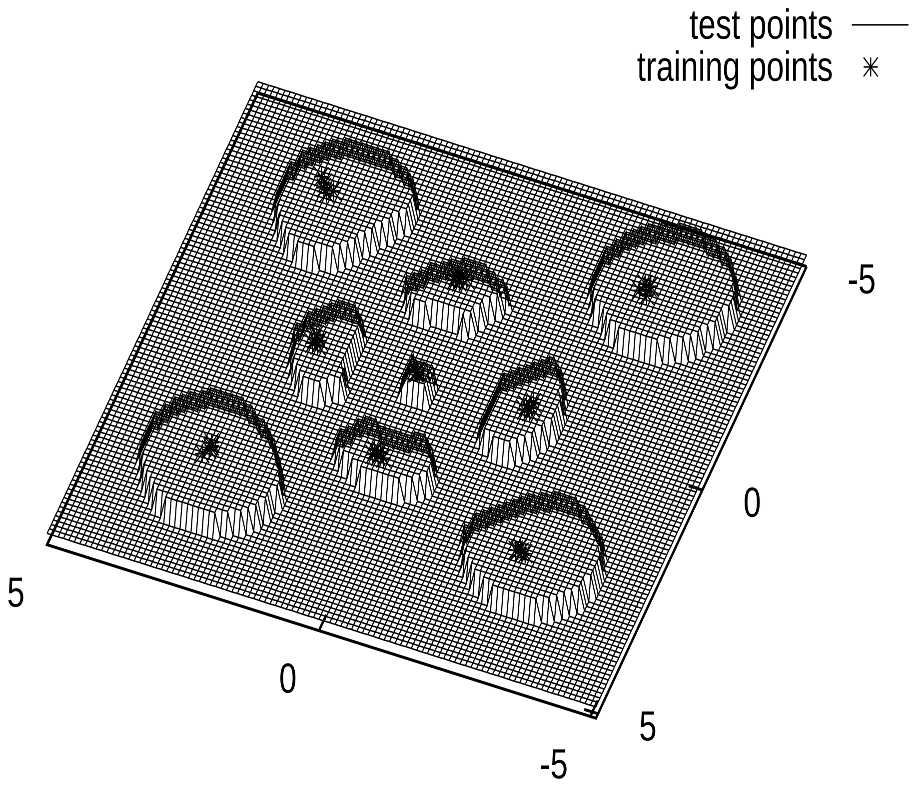

Figure 10: Decision regions for the RBF network (9 classes)

	Digit Error Rate				Digit Rejection Rate			
confidence	MLP	GMLP	DMLP	RBF	MLP	GMLP	DMLP	RBF
0	8.0	9.0	7.9	8.8	0	0	0	0
0.15	5.3	3.8	5.1	3.4	5.3	8.7	6.6	10.9
0.25	5.3	2.9	3.4	1.4	6.5	10.4	10.9	18.3
0.35	3.7	2.0	2.2	0.5	11.9	14.5	14.9	24.6
0.45	2.6	1.4	1.6	0.4	16.6	21.0	20.0	32.9
0.55	3.0	0.9	1.1	0.2	15.7	27.7	25.3	44.0

Table 4: Digit error rate vs digit rejection rate including RBF networks