

Study of MLP and RBFN performance for signal detection

P. Jarabo-Amores, R. Gil-Pita, M. Rosa-Zurera and F. López-Ferreras

Dpto. de Teoría de la Señal y Comunicaciones

Escuela Politécnica, Universidad de Alcalá

Ctra. Madrid-Barcelona, km. 33.600

28871, Alcalá de Henares - Madrid (SPAIN)

E-mail: {mpilar.jarabo,roberto.gil,manuel.rosa,francisco.lopez}@uah.es

Abstract. This paper deals with the application of Neural Networks to binary hypothesis tests based on multiple observations. The problem of detecting a desired signal in Additive-White-Gaussian-Noise is considered, assuming that the desired signal observations are also gaussian, independent and identically distributed random variables. The test statistic is then the squared magnitude of the observation vector and the optimum boundary is a hyper-sphere in the input space. The dependence of the neural network detector on the Training-Signal-to-Noise-Ratio and the number of hidden units is studied. Results show that Radial Basis Function Networks not only are more robust when varying the Training-Signal-to-Noise-Ratio and the number of hidden units, but the best approximation to the Neyman-Pearson detector is achieved with them.

1 Introduction

This paper deals with the application of Neural Networks (NN) to binary hypothesis test based on multiple observations. The problem of detecting a desired signal in Additive-White-Gaussian-Noise (AWGN) of zero mean and variance σ_n is considered, assuming that the desired signal observations are also gaussian, independent and identically distributed random variables with zero mean and variance σ_s . This signal and interference model is one of the most used in the analysis and design of many communications and radar systems. For example, it is used to model the received signal in pulsed scanning radar systems where the desired target consists of many independent scatterers of approximately equal echoing areas, and the received pulse magnitude varies from pulse to pulse [1]. In this case, the likelihood ratio detector (LRT) based on the Neyman-Pearson statistical hypothesis test is an optimum one [2], [3], but the performance of this detector relies on the knowledge of the interference and desired signal probability density functions (pdf) and power spectra. In actual situations, the received signal characteristics are unknown and differ from those supposed in the model. So, the resulting performance of the LRT detector based on this model will be worse [4].

Neural Networks (NN) are proposed as a solution, due to their ability to learn from their environment, and to improve performance in some sense through learning. NN can implement the Bayesian optimum detector [5] and it has been demonstrated [6],[7] that the back-propagation algorithm applied to a simple feed-forward network approximates the optimum Bayesian classifier, when using the mean square error criterion.

The objective of this work is the design of neural network based detectors capable of approximating the Neyman-Pearson detector performance in an unknown environment where the characteristics of the desired signal and the interference are unknown and may vary with time. Multi-Layer Perceptrons (MLPs) and Radial-Basis Function Networks (RBFNs) are considered, comparing their performance to the LRT detector for white gaussian signals in white gaussian interference. We analyze the influence of training parameters like variances of signal and interference (σ_n and σ_s) and network structure, in order to find the advantages and limitations of both detection schemes.

2 The optimum detector

Given a set of n observations, z_1, z_2, \dots, z_n which define a point in a n -dimensional space, $\mathbf{z} = [z_1, z_2, \dots, z_n]^T$, the detection system has to decide if they are originated either from noise only (the null hypothesis H_0) or from both noise and signal (the alternative hypothesis H_1). The objective is to minimize a risk function that is defined as the average cost [2]:

$$\bar{C} = \sum_{i=0}^1 \sum_{j=0}^1 C_{ij} P(D_i | H_j) P(H_j) \quad (1)$$

where:

- $P(D_i | H_j)$ is the probability of deciding H_i when H_j is the true hypothesis.
- $P(H_j)$ is the prior probability of hypothesis H_j .
- C_{ij} is the cost associated with deciding H_i when the true hypothesis is the hypothesis H_j .

To specify detector performance, the probability of detection (P_D) and the probability of false alarm (P_{FA}) are used. The probability of detection is the probability of deciding in favor of H_1 when it is the true hypothesis. The probability of false alarm is the probability of deciding in favor of H_1 when H_0 is true.

When the prior probabilities and the costs can not be determined, it is usual to seek a decision strategy that constraints P_{FA} to an acceptable value while maximizing P_D . The test is then said to be a Neyman-Pearson Test [2], also called "*the most powerful test*" (MP), since it achieves the largest P_D among all the test that have the same P_{FA} .

If the observations under each hypothesis are gaussian independent, identically distributed random variables, the conditional density functions $f(\mathbf{z}/H_0)$

and $f(\mathbf{z}/H_1)$ are multivariate normal probability density functions. Assuming that under H_0 , z_1, z_2, \dots, z_n have zero mean and variance σ_n , while under H_1 they have zero mean with variance $\sigma_n + \sigma_s$, the desired Signal-to-Noise Ratio (SNR) can be defined as (2):

$$SNR = 10 \log_{10}(snr) = 10 \log_{10}\left(\frac{\sigma_s}{\sigma_n}\right) \quad (2)$$

The likelihood functions $f(\mathbf{z}/H_0)$ and $f(\mathbf{z}/H_1)$ are given by (3) and (4).

$$f(\mathbf{z}/H_1) = \frac{1}{(2\pi(snr + 1))^{n/2}} \exp\left(-\frac{1}{2(snr + 1)} \mathbf{z}^T \mathbf{z}\right) \quad (3)$$

$$f(\mathbf{z}/H_0) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \mathbf{z}^T \mathbf{z}\right) \quad (4)$$

and the likelihood ratio is (5):

$$\frac{f(\mathbf{z}/H_1)}{f(\mathbf{z}/H_0)} = \frac{1}{(snr + 1)^{n/2}} \exp \frac{snr}{2(snr + 1)} \mathbf{z}^T \mathbf{z} \quad (5)$$

According to the Neyman-Pearson criterion, the optimum decision rule is given by (6), [2].

$$\frac{1}{(snr + 1)^{n/2}} \exp\left(\frac{snr}{2(snr + 1)} \mathbf{z}^T \mathbf{z}\right) \underset{H_0}{\overset{H_1}{\geq}} \eta_0 \quad (6)$$

where the threshold η_0 is determined from the specified value of P_{FA} . By taking logarithms on both sides and rearranging terms, (6) can be transformed into (7) showing that the test statistic is the squared magnitude of the observation vector \mathbf{z} .

$$\mathbf{z}^T \mathbf{z} \underset{H_0}{\overset{H_1}{\geq}} \frac{snr + 1}{snr} n \eta_0 \ln(snr + 1) \triangleq \eta \quad (7)$$

Figure (1) shows the P_D versus SNR curves for different values of P_{FA} , for the LRT detector based on the Neyman-Pearson statistical hypothesis test. These will be the desired results to be obtained using the NN based detector.

3 Neural network based approach

In this approach, Multi-Layer Perceptrons (MLPs) and Radial-Basis-Function Networks (RBFNs) are designed for approximating the LRT detector based on the Neyman-Pearson statistical hypothesis test. The structures of the detectors are shown in figure 2. In both cases, the NN is finished with a hard threshold detector. If the output of the network is greater than the threshold, T , we decide that hypothesis H_1 is true. On the other hand, if the output of the network is lower than T , we decide that hypothesis H_0 is true.

Cybenko theorem [8] states that a single hidden layer is sufficient for a MLP to compute a uniform ε approximation to a given training set represented by

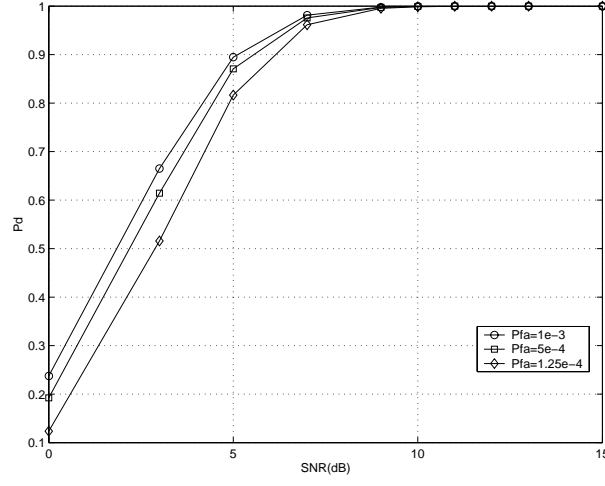


Fig. 1. LRT detector performance for different values of P_{FA}

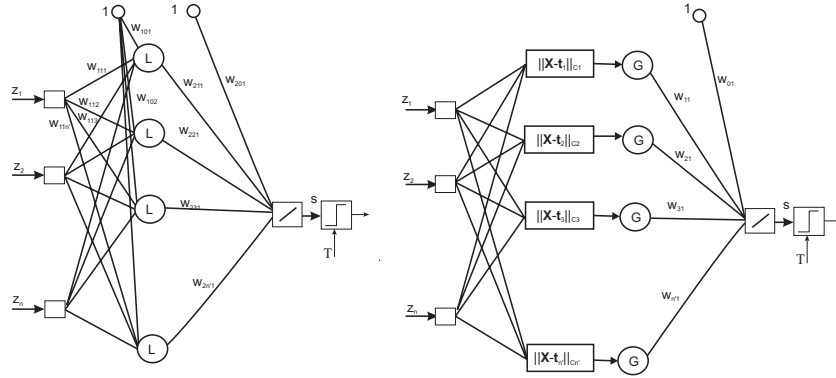


Fig. 2. MLP (left) and RBFN (right) structures

the set of inputs and the desired target output. Because of that, an MLP with a hidden layer has been considered. The activation function of all the neurons of the hidden layer is the logistic one given in (8).

$$L(x) = \frac{1}{1 + \exp(-x)} \quad (8)$$

The activation function of the output neuron is lineal, because when using a logistic function the network output ranges from 0 to 1 producing a saturation effect and a minimum value of P_{FA} for which the threshold equals 1. For lower values of P_{FA} , the threshold can not be increased and P_{FA} and P_D remain constant. This minimum value of P_{FA} depends on the Training-Signal-to Noise Ratio (TSNR) and the network structure.

In Radial-Basis Function Networks, the function associated to the hidden units (*radial-basis function*) is usually the multivariate normal function. For the i -th hidden unit it can be expressed using (9).

$$G_i(\mathbf{z}) = \frac{|\mathbf{C}_i|^{\frac{-1}{2}}}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{(\mathbf{z} - \mathbf{t}_i)^T \mathbf{C}_i^{-1} (\mathbf{z} - \mathbf{t}_i)}{2}\right\} \quad (9)$$

where \mathbf{C}_i is the covariance matrix, which controls the smoothness properties of the function (is a $n \times n$ real matrix).

Using the weighted norm [9], whose squared form is defined in (10), equation (9) can be transformed into equation (11).

$$\|\mathbf{x}\|_C^2 = \frac{1}{2} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} \quad (10)$$

$$G_i(\mathbf{z}) = \frac{|\mathbf{C}_i|^{\frac{-1}{2}}}{(2\pi)^{\frac{n}{2}}} \exp(-\|\mathbf{x} - \mathbf{t}_i\|_{\mathbf{C}_i}^2) \quad (11)$$

Taking into consideration that the hidden unit output is multiplied by a weight that will be adjusted during training, the RBF function $G(\cdot)$ in figure 2 is:

$$G(x) = \exp(-x^2) \quad (12)$$

The matrices \mathbf{C}_i can be set to a scalar multiple of the unit matrix, to a diagonal matrix with different diagonal elements or to a non-diagonal matrix. In this case we have set them to a scalar multiple of the unit matrix. The experiments we have carried out are described in the next sub-section.

3.1 Description of experiments

Training sets composed of 2,000 16-length patterns of interference and desired signal-plus-interference randomly distributed, have been built for different values of Training-Signal-to-Noise Ratio (TSNR) ranging from $-3dB$ to $25dB$, in order to study the dependence of the network performance on this parameter.

For training the MLPs, the Levenberg-Marquardt algorithm has been used [10]. For early stopping the MLP training process, a validation set with the same characteristics has been built for each TSNR. For training the RBFNs, we have applied a three learning phases strategy [9][11][12]:

1. The centres of the radial basis functions are determined by fitting a gaussian mixture model with circular covariances using the EM algorithm. The mixture model is initialized using a small number of iterations of the k-means algorithm.
2. The basis function widths are set to the maximum inter-centre squared distance.
3. The hidden to output weights that give rise to the least squares solution can be determined using the LMS algorithm.

Once the MLPs and the RBFNs have been trained, the P_D has been evaluated for different Signal-to-Noise-Ratios (SNRs) and different values of P_{FA} , in order to assess the networks generalization capability. These experiments have been repeated using different number of hidden units, ranging from 8 to 56 in steps of 8, in order to examine the dependence of performance on network size. The most relevant results are presented in the next sub-section.

3.2 Results

Due to the limitation in space, only the most relevant results are presented. Curves of P_D versus SNR values ranging from $0dB$ to $15dB$ are presented for two values of P_{FA} , $5 \cdot 10^{-4}$ and $1.25 \cdot 10^{-4}$. Figure 3 shows the performance of a MLP with 8 hidden units for a P_{FA} of $5 \cdot 10^{-4}$ and $1.25 \cdot 10^{-4}$, respectively, and different TSNRs. Figure 4 shows the performance of a MLP of 32 hidden units and the same values of P_{FA} and TSNR.

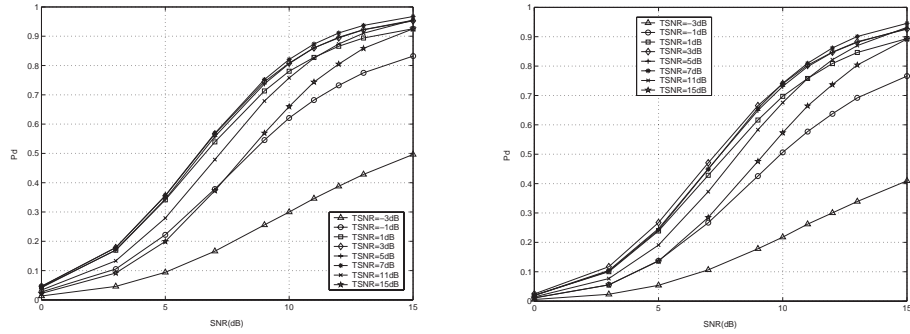


Fig. 3. MLP with 8 hidden units and $P_{FA} = 5 \cdot 10^{-4}$ (left) and $P_{FA} = 1.25 \cdot 10^{-4}$ (right)

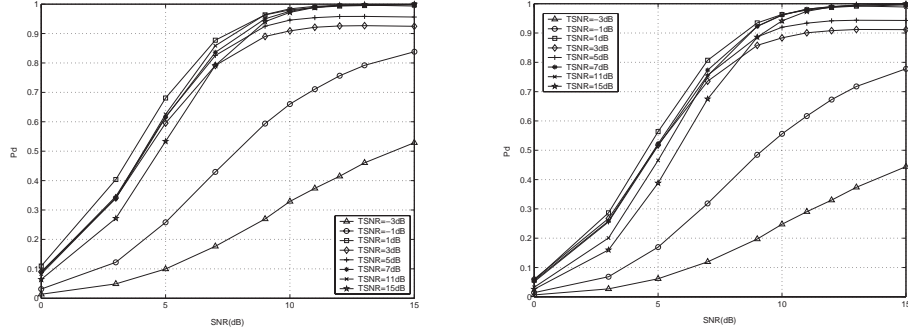


Fig. 4. MLP with 32 hidden units and $P_{FA} = 5 \cdot 10^{-4}$ (left) and $P_{FA} = 1.25 \cdot 10^{-4}$ (right)

The first conclusion that can be extracted from the results is the high dependence of the MLP performance on TSNR. Networks trained with very low TSNR values can not determine a suitable boundary. On the other hand, MLPs trained with high values of TSNR present better performance than in the former case, but not the best. The best performance for the MLP with 8 hidden units and a $P_{FA} = 5 \cdot 10^{-4}$ is obtained for a TSNR of 7dB. When considering lower P_{FA} values, it is more difficult to determine the best value of TSNR for the set of SNR values. A trade-off value of TSNR=7dB can be determined although in a practical situation low values of SNR are more likely and a TSNR value of 3dB should be selected.

When the number of hidden units is increased, the generalization capability of the network clearly improves for all the TSNR values. For the MLP with 32 hidden units the best TSNR for all the considered SNRs and P_{FA} values is 1dB.

Taking into consideration that the test statistic is the squared magnitude of the observation vector, the optimum boundary is a hyper-sphere in the input space. Each hidden unit of a MLP determines a hyper-plane in the input space, so the optimum boundary has to be estimated from n hyper-planes. It seems clear that the more hidden units the better approximation can be obtained. But the training of big MLPs is very complex due to problems related to local minima and over-fitting.

Figures 5 and 6 show the performance of RBFNs for different values of TSNR and P_{FA} . When they are compared to figures 3 and 4, the following conclusions can be extracted:

1. The dependence on TSNR is less important.
2. The dependence on network size is also smaller.
3. The best results are usually obtained for the lower values of TSNR.

This better behavior of the RBFNs can be explained taking into consideration the optimum boundary to be approximated. In this case, each hidden unit determines a hyper-sphere centered in representative points. So it is easier to

approximate a hyper-sphere and the influence of the number of hidden units is less important.

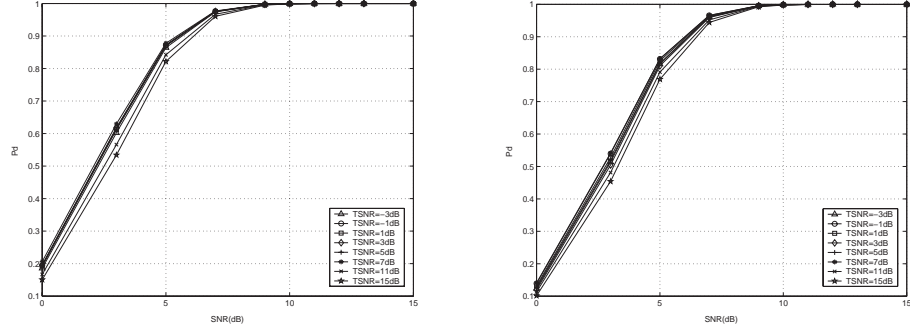


Fig. 5. RBFN with 8 hidden units and $P_{FA} = 5 \cdot 10^{-4}$ (left) and $P_{FA} = 1.25 \cdot 10^{-4}$ (right)

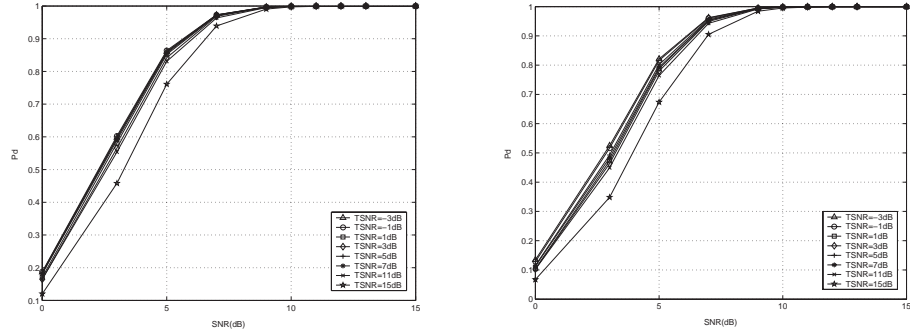


Fig. 6. RBFN with 32 hidden units and $P_{FA} = 5 \cdot 10^{-4}$ (left) and $P_{FA} = 1.25 \cdot 10^{-4}$ (right)

Finally, the best neural detectors are compared with the LRT detector based on the Neyman-Pearson statistical hypothesis test. Figure 7 shows that the RBFN detector closely approximates the LRT one, but the MLP performance is clearly far from this optimum detector, as it can be expected from the previous reasonings.

4 Conclusions

In this work, the design of Multi-Layer Perceptrons and Radial-Basis Function Networks capable of approximating the Neyman-Pearson detector performance

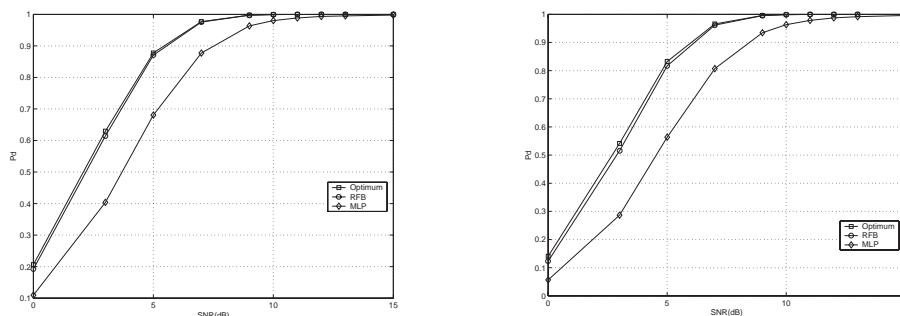


Fig. 7. Comparison of the best NN detectors and the LRT detector

are considered. For comparison purposes the detection of a desired signal in Additive-White-Gaussian-Noise is considered, assuming that the desired signal observations are also gaussian independent and identically distributed random variables. For such a problem, the test statistic is the squared magnitude of the observation vector, so the optimum boundary is a hyper-sphere in the input space.

The dependence of the neural network detector on the Training-Signal-to-Noise-Ratio and the number of hidden units is studied. Results show that RBFNs not only are more robust when varying the TSNR and the number of hidden units, but the best approximation to the Neyman-Pearson detector is achieved with them. This is due to the fact that RBFNs approximate the detection boundary by a combination of hyper-spheres, while MLPs approximate it from the hyper-planes determined by the hidden layer.

References

1. M. Skolnik: Radar handbook. Second edition. McGraw-Hill, Inc., 1990.
2. M.D. Srinath, P.K. Rajasekaran, R. Viswanathan: Introduction to statistical signal processing with applications. Prentice-Hall, Inc., 1996.
3. V. Aloisio, A. di Vito, G. Galati: Optimum detection of moderately fluctuating radar targets. IEE Proceedings on Radar, Sonar and Navigation, Vol. 141, No. 3, pp. 164-170, June 1994.
4. A. di Vito, M. Naldi: Robustness of the likelihood ratio detector for moderately fluctuating radar targets. IEE Proceedings on Radar, Sonar and Navigation, Vol. 146, No. 2, pp. 107-112, April 1999.
5. J. W. Watterson: An optimum multilayer perceptron neural receiver for signal detection. IEEE Trans. on Neural Networks, Vol. 1, No. 4, pp. 298-300, December 1990.
6. D.W. Ruck, S.K. Rogers, M. Kabrisky, M.E. Oxley, B.W. Suter: The multilayer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Transactions on Neural Networks, Vol. 1, No. 4, pp. 296-298, April 1990.
7. E.A. Wan: Neural network classification: A bayesian interpretation. IEEE Transactions on Neural Networks, Vol. 1, No. 4, pp. 303-305, December 1990.

8. G. Cybenko: Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, Vol. 2, pp. 303-314, 1989.
9. S. Haykin: *Neural networks. A comprehensive foundation* (second edition). Prentice-Hall Inc., 1999.
10. M.T. Hagan, M.B. Menhaj: Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, pp. 989-993, November 1994.
11. C.M. Bishop: *Neural networks for pattern recognition*. Oxford University Press Inc., 1995.
12. F. Schwenker, H.A. Kestler, G. Palm: Three learning phases for radial-basis-function networks. *Neural Networks*, Vol. 14, Issue 4-5, pp. 439-458, 2001.