

Layered Perceptron Versus Neyman-Pearson Optimal Detection

CHRISTOPHE F. BAS

IRESTE, University of Nantes, CP 3003, 44087 NANTES Cedex 03, FRANCE

ROBERT J. MARKS II

Department of Electrical Engineering, FT-10, University of Washington, Seattle, WA 98036, USA

Abstract

A layered perceptron artificial neural network (ANN) is trained to detect positive signals corrupted with noise which, for our test, is Laplacian. Comparison of the ANN performance is made with both Neyman-Pearson optimal and linear detectors. The ANN invariably outperforms the linear detector and is shown to be nearly optimal. The optimal detector requires knowledge of signal and noise parameters. The ANN does not.

I. Introduction

We consider the following classical problem from detection theory. When given M samples, we are to decide between the two following hypotheses:

$$H_0: X_i = N_i \quad \text{for } i = 1, 2, \dots, M$$

or

$$H_1: X_i = N_i + S \quad \text{for } i = 1, 2, \dots, M \text{ and } S > 0$$

where S is a known constant signal and N_i is a random variable.

We are to design a detector that will determine whether the M samples are pure noise (H_0) or are a constant signal accompanied by noise (H_1). Throughout this paper, we assume independent and identically distributed Laplace noise samples N_i with pdf:

$$p_N(n) = \frac{\gamma}{2} \exp(-\gamma|n|)$$

The ability of a detector to decide between H_0 and H_1 is measured by:

$$\alpha = \text{P[False alarm]} = \text{P}[H_1 \text{ is the decision} | H_0 \text{ is correct}]$$

and

$$\beta = \text{P[Detection]} = \text{P}[H_0 \text{ is the decision} | H_0 \text{ is correct}].$$

These two probabilities are used to draw the *receiver operating characteristics* (ROC) where β versus α is observed. The detector which maximizes β for a given α is the Neyman-Pearson optimal [1-2].

In our study we compare the performances of three detectors: a trained layered perceptron ANN, the Neyman-Pearson optimal detector and a linear detector (optimal in the presence of Gaussian noise). The relative performance among these detectors is determined by comparison of their ROC curves. The use of ROC curves as a measurement of the layered perceptron performance was first suggested by Eberhart and Dobbins [3]. Ferkinhoff et al. [4] have compared performance of an ANN to a modified Neyman-Pearson detector under a Gaussian (linear) assumption.

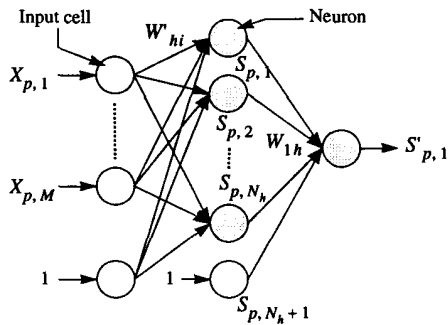


Figure 1: The ANN-detector structure where $X_{p,i}$ is an input sample, $S_{p,h}$ a hidden neuron state, $S'_{p,1}$ the output neuron state, W_{hi} the weights between the input cells and the hidden layer, and W_{1h} the weights between the hidden layer and the output.

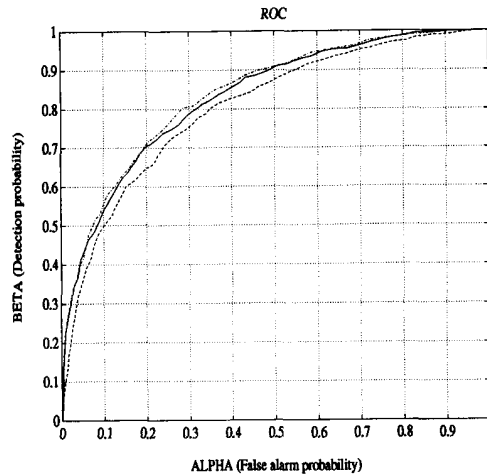


Figure 2: ROC of a 10 input cell ANN-detector for 200 (top), 50 (middle) and 10 (low) hidden neurons with $S=1$, $\gamma=0.5$ and 10^6 passes for each structure.

II. Structure, training and behavior of an ANN-detector

The neural network used is the standard feed forward single hidden layer one output ANN which is shown in Figure 1. Each neuron non-linear function is the sigmoid whose equation is:

$$F(A_j) = \frac{1}{1 + \exp(-A_j)}$$

where A_j is the weighted input summation of the previous neurons and cells linked to the j^{th} neuron. The neural network was trained using the conventional back-propagation algorithm [5] using a step size $\eta = 0.01$ and a momentum factor of 0.003.

During the ANN training, a new set of Laplace random variables is withdrawn for each pass. The use of fresh random variables at each training step induces the neural network to learn rather than to memorize. Typically, an increase in the number of hidden neurons allows greater flexibility in the choices of training boundaries, at the potential cost of over determination (memorization). Here, however, the use of new noise random variables for each training sample set removes the potential for memorization.

Figure 2 shows the ROC curves of a 10 input cell ANN trained with: 10 (dashed), 50 (solid) and 200 (dash-dotted) hidden neurons, $\gamma = 0.5$ for 10^6 passes and tested on 2000 signals. If the confidence interval of the (α, β) curves is assumed to be 95%, then the maximal deviation on both α and β is 3.1% (α and β are Bernoulli random variables).

III. Comparison

The Neyman-Pearson optimal detector for our constant signal in a Laplace noise detection problem is shown in Figure 3. The M Random Variables are subjected to a memoryless nonlinearity, $g(x_i)$, and summed. The sum is compared to a threshold T to determine which hypothesis to announce. For Laplace noise, this memoryless nonlinearity is shown in Figure 4 [1-2,6-7]. For the Gaussian noise, the "nonlinearity" is $g(X_i) = X_i$. When noise characteristics are unknown, the linear detector is commonly used. On the other hand, the same Neyman-Pearson detector is used when the cardinality of samples is changed. Performance of the linear and Neyman-Pearson optimal curves in the presence of Laplace noise is well studied [6-7].

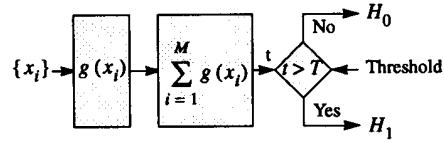


Figure 3: The Neyman-Pearson optimal detector.

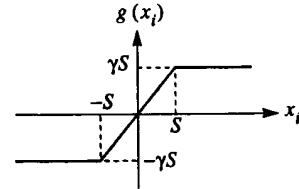


Figure 4: The memory-less nonlinearity for Laplace noise.

Let us now compare the two last detectors with the ANN-detector. All examples are normalized to $S = 1$. The ROC curves of Figure 5 are drawn for an ANN with 5 input cells, 20 hidden neurons, 10^6 learning passes and for different values of γ .

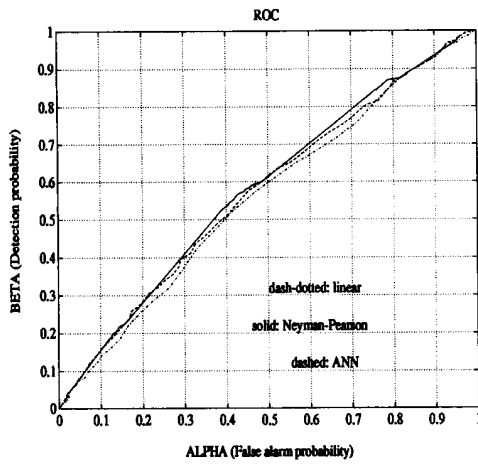


Figure 5a: Comparison for $\gamma = 0.1$.

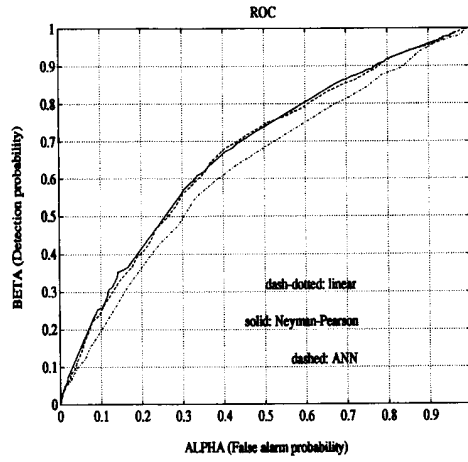


Figure 5b: Comparison for $\gamma = 0.3$.

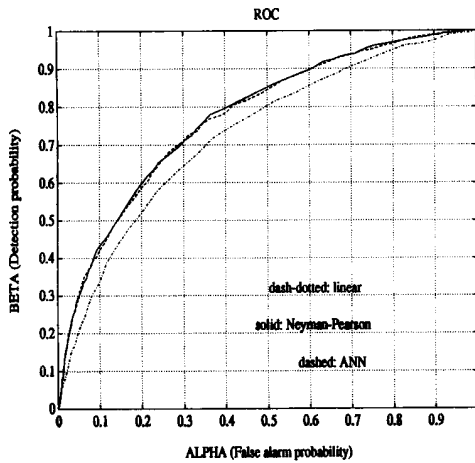


Figure 5c: Comparison for $\gamma = 0.5$.

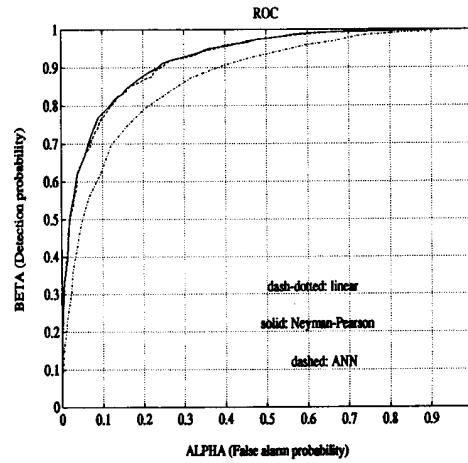


Figure 5d: Comparison for $\gamma = 1$.

Figure 6 compares the ROC curves when the number of input samples M changes for 5 and 10 for $\gamma = 0.5$. All the ANN curves shown here are the result of a training on 10^6 passes. Here, and in our other simulations, the ANN-detector ROC curve lies very close to the optimal detector when it has been trained with enough patterns and has sufficient hidden neurons.

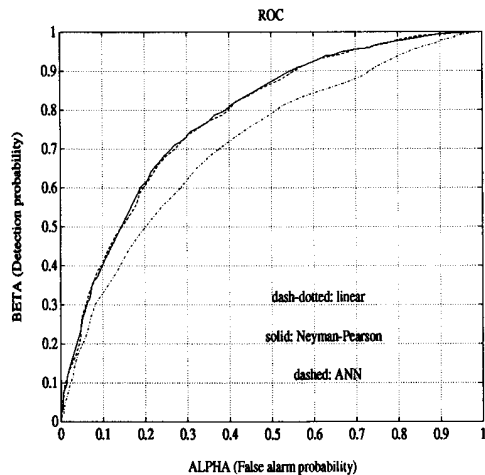


Figure 6a: Comparison for $M=5$.

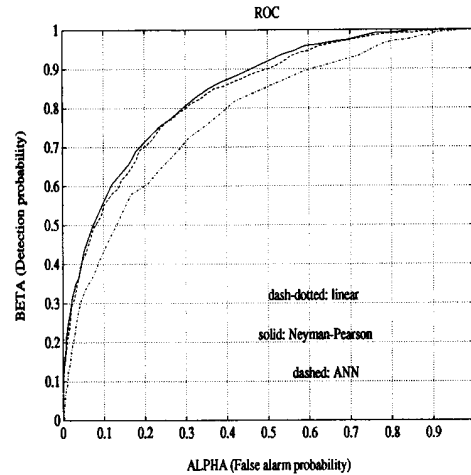


Figure 6b: Comparison for $M=10$.

IV. Conclusion

Invariably, we found that, for a sufficiently large number of hidden neurons, the layered perceptron ROC curve lies just below the optimal detector curve and significantly above the linear detector one. The ANN thus successfully learned the characteristics of the Laplace noise.

The advantage of the neural network is that it does not require the noise and constant signal parameters (e.g. S and γ for the Neyman-Pearson detector in Laplace noise) that the optimal detector design requires. When at least one characteristic changes (either on the signal or on the ANN structure) the ANN training process has to be repeated.

Alternately, S and γ can be estimated from the training data for purposes of the Neyman-Pearson type. This approach is not considered here.

References

- [1] J.H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-18, pp.241-250, March 1972.
- [2] S. A. Kassam, **Signal Detection in Non-Gaussian Noise**, Springer Verlag, 1988.
- [3] Eberhart and Dobbins, **Neural Network PC Tools: a Practical Guide**, Academic Press 1990.
- [4] D. J. Ferkinhoff, S. C. Nardone, J. G. Baylog and K. F. Gong, "Feature Extraction and Interpretation for Dynamic System Model Resolution," **Twenty-fourth Asilomar Conference on Signals, Systems & Computers**, vol. 1, pp. 156-162, Maple Press, 1990.
- [5] D.E. Rumelhart and J. L. McClelland, **Parallel distributed processing**, vol. 1, p.327, MIT Press 1986.
- [6] R. J. Marks, G.L. Wise, D.G. Haldeman and J. L. Whited, "Detection in Laplace noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-14, pp. 866-872, November 1978.
- [7] M. I. Dadi and R. J. Marks, "Detector relative efficiencies in the presence of Laplace noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, pp. 568-582, July 1987.