

## RESEARCH

## Open Access

# Radar detection with the Neyman–Pearson criterion using supervised-learning-machines trained with the cross-entropy error

María-Pilar Jarabo-Amores, David de la Mata-Moya, Roberto Gil-Pita and Manuel Rosa-Zurera\*

## Abstract

The application of supervised learning machines trained to minimize the Cross-Entropy error to radar detection is explored in this article. The detector is implemented with a learning machine that implements a discriminant function, which output is compared to a threshold selected to fix a desired probability of false alarm. The study is based on the calculation of the function the learning machine approximates to during training, and the application of a sufficient condition for a discriminant function to be used to approximate the optimum Neyman–Pearson (NP) detector. In this article, the function a supervised learning machine approximates to after being trained to minimize the Cross-Entropy error is obtained. This discriminant function can be used to implement the NP detector, which maximizes the probability of detection, maintaining the probability of false alarm below or equal to a predefined value. Some experiments about signal detection using neural networks are also presented to test the validity of the study.

**Keywords:** Radar detection, Supervised learning machines, Error function, Cross-entropy

## 1 Introduction

In this article, an extension of the theoretical study presented in [1] about the capability of supervised learning machines to approximate the Neyman–Pearson (NP) detector is presented. This detector can be implemented by comparing the likelihood ratio,  $\Lambda(\mathbf{z})$ , to a detection threshold fixed taking into account Probability of False Alarm ( $P_{FA}$ ) requirements, as stated in expression (1) [2,3], being  $f(\mathbf{z}|H_i)$ ,  $i \in \{0, 1\}$ , the likelihood functions under both, the null ( $H_0$ ) and the alternative ( $H_1$ ) hypothesis.

$$\Lambda(\mathbf{z}) = \frac{f(\mathbf{z}|H_1)}{f(\mathbf{z}|H_0)} \underset{H_0}{\overset{H_1}{\geq}} \eta(P_{FA}) \quad (1)$$

The NP criterion has been widely used in radar applications. The robustness of the likelihood ratio detector for moderately fluctuating targets was studied in [4]. In the last years, considering also radar applications, the NP criterion has been applied in MIMO radars [5,6], distributed radar sensor networks [7], and for the detection of ships in

marine environments [8]. The NP criterion has also been applied in some other topics: watermarking [9,10], fault-induced dips detection [11], detection in sensor networks [12], disease diagnosis [13,14], biometric [15], or gravitational waves detection [16]. Great efforts are being made nowadays to solve a number of problems related to signal detection in noise [17].

For the NP detector to be implemented, both likelihood functions must be known. Usually, statistical models of interference and targets are assumed and their parameters are estimated using the available data. Obviously, detection losses are expected when the interference or target statistical properties vary from those assumed in the design. In addition, when some of the parameters are random, and the composite hypothesis tests must be used, the average likelihood ratio can lead to intractable integrals that should be solved by numerical approximations. The usage of learning machines based detectors, allows us to approximate the NP detector, just using training data obtained experimentally, without knowledge of the likelihood functions. The main advantage of this approach is that no statistical models have to be assumed during the design, and if a suitable error function is used during training, a good approximation to the optimal NP detector

\*Correspondence: manuel.rosa@uah.es

Department of Signal Theory and Communications, University of Alcalá, Campus Universitario s/n, 28805 Alcalá de Henares, (Madrid), Spain

is obtained [1]. The main drawback is the difficulty of obtaining representative training samples, and the definition of the most suitable learning machine architecture.

The application of supervised learning machines to approximate the NP detector has already been studied. The easiest way is to use a learning machine with only one output, which is compared to a threshold in order to decide in favor of the null or the alternative hypothesis. The threshold is used to fix the desired  $P_{FA}$ . This scheme has been used previously in several works [1,18-20]. An equivalent implementation consists in varying the bias of the output neuron [21,22]. A different approach is used in [23]: a two-output NN with outputs in (0, 1) was used, comparing the subtraction of both outputs to a threshold. This approach is equivalent to using a NN with only one output and desired outputs  $\{-1, 1\}$  [24]. More recently, Radial Basis Function Neural Networks (RBFNN) have also been applied to approximate the NP detector [25-27]. Support Vector Machines (SVMs) have been applied to signal detection in background noise [28]. Finally, detectors based on committee machines have also been proposed [29-32].

The possibility of approximating the NP detector using adaptive systems trained in a supervised manner was studied in [33]. A sufficient condition for a discriminant function to be suitable for implementing the NP detector was obtained. In [1], those results were used to carry out a more general study about the capability of learning machines to approximate the NP detector when they are trained in a supervised manner to minimize an error function, demonstrating that the Sum-of-Squares error is suitable to approximate the NP detector, and the Minkowski error with  $R = 1$  is suitable to approximate the minimum probability of error classifier. With  $R = 1$ , the Minkowski error reduces to the mean absolute deviation.

The Sum-of-Squares error is optimal for training supervised learning machines in order to detect or to classify Gaussian signals. If non-Gaussian interference is assumed in the radar, probably there exists some other error functions which give rise to better results, motivating the study to know if they fulfil the sufficient condition established in [1,33]. In this article, one more error function is considered, the Cross-Entropy error. The study demonstrates that the Cross-Entropy error is also suitable to be used for training supervised learning machines in order to approximate the NP detector, even improving the performance of learning machines trained with the Sum-of-Squares error.

The article is structured as follows: In Section 2, the problem this article deals with is presented. The function that a learning classifier with one output approximates to, when has been trained to minimize the Cross-Entropy error, is calculated in Section 3. The condition stated in [1] is applied to demonstrate that the approximated function is useful to approximate the NP detector. In Section 4,

some experiments are presented to illustrate the previous theoretical studies. Finally, in Section 5 the main contributions of this article are summarized, and conclusions are extracted.

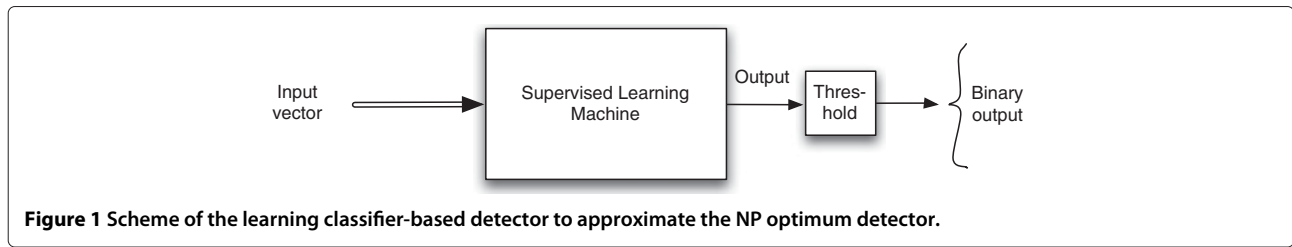
## 2 Problem statement

The usefulness of supervised learning machines trained to minimize the Sum-of-Squares error to approximate the NP detector, and the usefulness of the Minkowski error with  $R = 1$  to approximate the minimum probability of error classifier, have been demonstrated in [1]. In this article, we extend the study to the Cross-Entropy error. The discriminant function the learning machine approximates to after training is obtained, and the sufficient condition stated in [33] is applied. The detector is implemented by comparing the output of the discriminant function to a threshold. The final approximation error will depend on the selected error function, the selected training and validation sets, the system structure, and the training algorithm [34]. In order to obtain a good approximation, the training set must be a representative subset of the input space, the function implemented by the learning machine must be sufficiently general that there is a choice of adaptive parameters which makes the error function sufficiently small, and the learning algorithm must be able to find the appropriate minimum of the error function [35].

In our study, a learning machine with one output is considered, that is used to classify input vectors  $\mathbf{z} = [z_1, z_2, \dots, z_L]^T$  into two hypotheses or classes,  $H_0$  and  $H_1$ , which stand for the absence of target and for its presence, respectively, in radar detection problems. The basic detector scheme is represented in Figure 1.

Given a decision rule, let  $\mathcal{Z}_i$  be the set of all possible input vectors that will be assigned to hypothesis  $H_i$ , and  $\mathcal{Z}$  the ensemble of all possible input vectors ( $\mathcal{Z}_0 \cup \mathcal{Z}_1 = \mathcal{Z}$ ). The output of the learning machine is represented by  $F(\mathbf{z})$ , and the desired output by  $t_{H_i}$ . A training set,  $Z = Z_0 \cup Z_1$ , where  $Z_1$  is composed of  $N_1$  training patterns from hypothesis  $H_1$ , and  $Z_0$  is composed of  $N_0$  training patterns from hypothesis  $H_0$  ( $N = N_1 + N_0$ ), is available.

In order to study the suitability of the error function to be used to approximate the NP detector, the same strategy applied in [1] is used. The function the learning machine approximates to after training is obtained, as a function of the likelihood functions and the prior probabilities. The implemented detector compares the learning machine output to a threshold  $\eta_0$ , which varies to fix the  $P_{FA}$ . The NP detector is usually implemented by comparing the likelihood ratio to a threshold  $\eta_{lr}$ , fixed according to the required  $P_{FA}$ . A sufficient condition has been established in [1,33], which states that for a learning machine to approximate the NP detector, the relation between  $\eta_{lr}$  and  $\eta_0$  doesn't depend on the input vector.



**Figure 1** Scheme of the learning classifier-based detector to approximate the NP optimum detector.

In the following sections, in order to obtain the function the learning machine approximates to after training, the strong law of large numbers is going to be applied [36]. It asserts that if  $\langle X_i \rangle$  is a sequence of independent and identically distributed random variables which has an expectation  $\mu$ , then:

$$\mu = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i \quad (2)$$

### 3 Discriminant function approximated by a learning-machine trained to minimize the cross-entropy error

The error function to be studied is the Cross-Entropy Error [34,37], defined in the following expression, when a one output learning machine is considered, the desired outputs are one and zero, and  $F(\mathbf{z}) : \mathcal{Z} \rightarrow (0, 1)$  (the function implemented by the system maps  $\mathcal{Z}$  into the interval  $(0, 1)$ ):

$$E = -\frac{1}{N} \left[ \sum_{\mathbf{z} \in H_1} \ln[F(\mathbf{z})] + \sum_{\mathbf{z} \in H_0} \ln[1 - F(\mathbf{z})] \right] \quad (3)$$

If the number of patterns tends to infinity ( $N \rightarrow \infty$ ), the error can be expressed as follows:

$$E_m = -\lim_{N \rightarrow \infty} \left[ \frac{N_1}{N} \frac{1}{N_1} \sum_{\mathbf{z} \in H_1} \ln[F(\mathbf{z})] + \frac{N_0}{N} \frac{1}{N_0} \sum_{\mathbf{z} \in H_0} \ln[1 - F(\mathbf{z})] \right] \quad (4)$$

Applying the strong law of large numbers, expression (5) is obtained:

$$E_m = - \int_{\mathcal{Z}} (P(H_1)f(\mathbf{z}|H_1) \ln(F(\mathbf{z})) + P(H_0)f(\mathbf{z}|H_0) \ln(1 - F(\mathbf{z}))) dz \quad (5)$$

The function  $F(\mathbf{z})$  that minimizes  $E_m$ , which will be denoted by  $F_0(\mathbf{z})$ , is obtained using *calculus of variations*, and particularly the Euler-Lagrange differential equation [38,39]. The calculus of variations can be used to find the

function  $F(\mathbf{z})$  that minimizes the functional  $J(F)$  defined as follows:

$$J(F) = \int_{\mathcal{Z}} I \left( \mathbf{z}, F(\mathbf{z}), \frac{\partial F(\mathbf{z})}{\partial z_1}, \frac{\partial F(\mathbf{z})}{\partial z_2}, \dots, \frac{\partial F(\mathbf{z})}{\partial z_L} \right) dz, \quad (6)$$

where  $\mathbf{z} = [z_1, z_2, \dots, z_L]^T$ ,  $I$  is twice differentiable with respect to the indicated arguments, and  $F$  is a function in  $C^2(\mathcal{Z})$  that assumes prescribed values at all points of the boundary  $\delta\mathcal{Z}$  of the domain  $\mathcal{Z}$ . The function  $F$  that minimizes  $J(F)$  can be obtained by solving the Euler-Lagrange equation (7), where  $F'_k = \frac{\partial F}{\partial z_k}$ .

$$\frac{\partial I}{\partial F} - \sum_{k=1}^L \frac{\partial}{\partial z_k} \left( \frac{\partial I}{\partial F'_k} \right) = 0 \quad (7)$$

In our problem,  $J(F) = E_m(F)$ , and  $I(\mathbf{z}, F(\mathbf{z})) = -P(H_1)f(\mathbf{z}|H_1) \ln(F(\mathbf{z})) - P(H_0)f(\mathbf{z}|H_0) \ln(1 - F(\mathbf{z}))$ , which does not depend on the first derivatives of  $F$ . Therefore,  $F$  only needs to be defined in  $C^0(\mathcal{Z})$  and the Euler-Lagrange equation reduces to:

$$\frac{\partial}{\partial F} (-P(H_1)f(\mathbf{z}|H_1) \ln(F(\mathbf{z})) - P(H_0)f(\mathbf{z}|H_0) \ln(1 - F(\mathbf{z}))) = 0 \quad (8)$$

The function  $F_0(\mathbf{z})$  that minimizes  $E_m$  is given in (9) and the detection rule is obtained by comparing  $F_0(\mathbf{z})$  to  $\eta_0$  (10).

$$F_0(\mathbf{z}) = \frac{P(H_1)f(\mathbf{z}|H_1)}{P(H_1)f(\mathbf{z}|H_1) + P(H_0)f(\mathbf{z}|H_0)} \quad (9)$$

$$\frac{P(H_1)f(\mathbf{z}|H_1)}{P(H_1)f(\mathbf{z}|H_1) + P(H_0)f(\mathbf{z}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \quad (10)$$

Dividing the numerator and denominator of the left side of rule (10) by  $f(\mathbf{z}|H_0)$ , an equivalent rule can be obtained, which is a function of the likelihood ratio:

$$\frac{P(H_1)\Lambda(\mathbf{z})}{P(H_1)\Lambda(\mathbf{z}) + P(H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \quad (11)$$

Extracting the likelihood ratio, a new equivalent rule (12) can be derived, which compares the likelihood ratio

to a new threshold,  $\eta_{lr}$ . The expression which relates  $\eta_{lr}$  and  $\eta_0$  is presented in (13):

$$\Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)\eta_0}{P(H_1)(1 - \eta_0)} = \eta_{lr} \quad (12)$$

$$\eta_0 = \frac{\eta_{lr}P(H_1)}{P(H_0) + \eta_{lr}P(H_1)} \quad (13)$$

The relation between  $\eta_{lr}$  and  $\eta_0$  does not depend on the input vector,  $\mathbf{z}$ . Thus, according to the sufficient condition proposed in [1,33], the detection rule (10) is an implementation of the NP detector.

#### 4 Experiments

In this section, detectors based on Multilayer Perceptrons (MLPs) are designed for three cases studies: detection of colored Gaussian signals in white Gaussian interference, detection of colored Gaussian signals in correlated Gaussian clutter plus white Gaussian noise, and detection of non-fluctuating targets in  $K$ -distributed interference [40,41]. In practical situations, the statistical properties of the interference can be estimated and tracked to some degree, but the target parameters are very difficult to estimate. When the target parameters are unknown, the NP detector is built with the Average-Likelihood ratio, that is compared to a threshold. The three cases study have been selected because the optimum NP detector can be easily approximated using a Maximum Likelihood estimator of the Average-Likelihood ratio.

Two strategies are followed to check the performance of the proposed detectors:

- First, the supervised learning machine-based detectors are compared with an approximation of the NP detector. The Average-Likelihood ratio is approximated by a Maximum-Likelihood estimator, based on the Constrained Generalized Likelihood Ratio (CGLR) [42]. The CGLR is built with a number of filters equal to the dimension of the input vector ( $L$ ). An increase in the number of filter does not produce a significant improvement in the performance of the CGLR.
- Second, the detectors obtained after training the MLPs with the Cross-Entropy error, are compared with the equivalent obtained after training with the Sum-of-Squares error, and the Minkowski error with  $R = 1$ . These comparisons are only performed for the first case study, due to space limitations, but similar comparative results are obtained in the other two cases study. The comparison is completed with a representation of  $P_D$  versus SNR for the best detectors obtained with the different error functions, for the first case study.

The Receiver Operating Characteristic (ROC) curves of all the considered detectors are represented, to show the validity of our approach.

#### 4.1 Considered detection problems

The following detection problems are studied in this article, to assess the capability of learning machines trained to minimize the Cross-Entropy error to implement good approximations of the NP optimum detector:

- **Case study 1:** Detection of Gaussian fluctuating targets in presence of Additive White Gaussian Noise (AWGN). This case corresponds to target detection in the clear conditions. This case study has been subdivided into two:
  - Detection of Gaussian targets with unknown correlation coefficient.
  - Detection of Swerling I (SWI) targets with unknown Doppler shift.
- **Case study 2:** Detection of Gaussian fluctuating targets in presence of correlated Gaussian clutter and AWGN. This model can be used for target detection in AWGN and sea/land clutter with low resolution radar systems, or high resolution radar systems with incidence angle higher than 10 degrees. Again, this case study has been subdivided into two:
  - Detection of Gaussian targets with unknown correlation coefficient.
  - Detection of SWI targets with unknown Doppler shift.
- **Case study 3:** Detection of non fluctuating targets in presence of spiky  $K$ -distributed clutter ( $\nu = 0.5$ , where  $\nu$  is the shape parameter of the  $K$ -distribution). This model is suitable for target detection in sea/land clutter with high resolution radar systems and low grazing angles. In this case, the problem of detecting Swerling V (SWV) targets with unknown Doppler shift is considered.

#### 4.2 Detectors architecture, training and test parameters

Multilayer Perceptrons with real arithmetic are designed. Each MLP has an input layer, one hidden layer, and one output. In these examples, a pulsed radar which provides eight complex-valued echoes in each exploration, due to antenna rotation and beamwidth, is considered (this is the usual case in air traffic control radars). Each complex valued echo consist of the in-phase and quadrature components. Considering that the input vector is composed of eight complex valued echoes,  $L = 16$  real inputs are required. The dependence of performance on the number of neurons in the hidden layer ( $M$ ) is studied. The output

is compared to a hard threshold, selected according to  $P_{FA}$  (Figure 1). According to this, the architectures of the MLPs are labeled with MLP L/M/1. The activation function of the processing units is the sigmoidal one.

For training the different MLPs, a training set composed of 50,000 patterns has been used. The training set consists of patterns belonging to both hypothesis, that are considered with equal prior probabilities (the same number of patterns for  $H_0$  and  $H_1$ ). A cross-validation strategy has been used during training to avoid over-fitting, following the  $k$ -fold approach with  $k = 5$ . The validation set is composed of 10,000 patterns, 5,000 from each hypothesis.

For testing, a different set of patterns has been used. The number of patterns under hypothesis  $H_0$  is  $2 \cdot 10^7$ , to estimate  $P_{FA}$  values higher than  $10^{-6}$  using conventional Monte-Carlo simulation with a relative error lower than 10%, while the number of patterns under hypothesis  $H_1$  is  $5 \cdot 10^4$ , to estimate  $P_D$ .

The algorithm used for training with the Cross-Entropy error is the one described in [43], while the algorithm used for training with the Sum-of-Squares error and the Minkowski error ( $R = 1$ ), is the Conjugate Gradient method [44].

### 4.3 Results

In this subsection the results obtained for the above mentioned detection problems are presented. Different Signal to Interference Ratios (SIR) are considered in the experiments to obtain  $P_D > 0.8$  for  $P_{FA}$  values of interest in radar applications. The SIR becomes the Signal to Noise Ratio (SNR), when the interference is only noise, and Signal to Clutter Ratio (SCR), when the interference is only clutter. For the MLPs, the influence of the number of hidden neurons, and therefore, the influence of the learning machine architecture is studied too. In some of the cases, the results are better when the number of hidden neurons is higher, but for detecting targets in spiky  $K$ -distributed clutter, very good results are obtained even with a low number of neurons in the hidden layer. This is because the surface which separates the acceptance regions of both hypothesis in the optimum detector can be approximated with a simpler architecture. In all the cases, there exists an architecture of the neural network that guarantees a very good approximation to the optimum detector, when the neural network is trained with the Cross-Entropy error.

#### 4.3.1 Detection of Gaussian fluctuating targets in AWGN

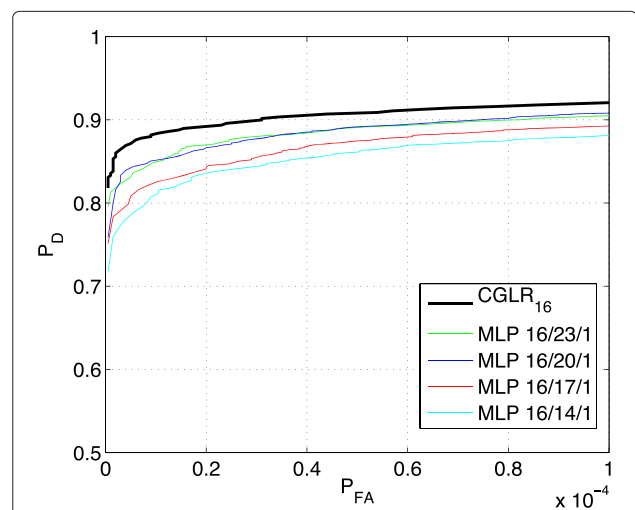
First, the detection of Gaussian fluctuating targets with unknown correlation coefficient in AWGN is studied. In this case, the SNR is 7 dB. MLPs with a number of hidden neurons which varies from 14 to 23, in steps of 3, have been trained and tested. The ROC curves of the detectors based on MLPs trained with the Cross-Entropy error are

presented in Figure 2. As higher the number of neurons, better the approximation to the CGLR detector, demonstrating that the optimum detector can be approximated with a MLP if the number of freedom parameters of the architecture to be fitted during training is high enough for the considered detection problem.

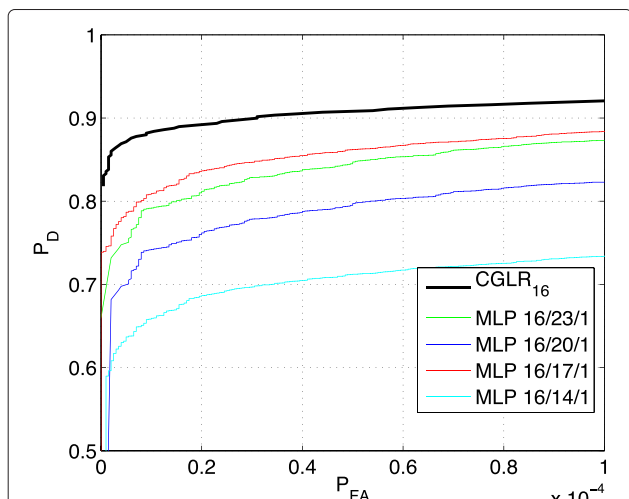
For comparison purposes, MLPs have been trained to minimize the Sum-of-Squares error and the Minkowski error ( $R = 1$ ). The ROC curves are presented in Figures 3 and 4, respectively. The results training with the Sum-of-Squares error show high variability, and the results training with the Minkowski error ( $R = 1$ ) are clearly worse, as expected from the theoretical study presented in [1], and with high variability too. To show the dependence with SNR, the best MLP-based detectors have been selected, and tested with different SNR values. The results are presented in Figure 5.

In a second experiment, the SWI targets with unknown Doppler shift in AWGN is considered. The Doppler shift is modeled as a uniform random variable in the interval  $[0, 2\pi)$ . In this case, the SNR is 10 dB. Again, MLPs with different number of hidden neurons have been trained and tested with the Cross-Entropy error, the Sum-of-Squares error, and the Minkowski error ( $R = 1$ ), to study the dependence of performance on the network architecture. The number of hidden neurons that has been considered varies from 14 to 23, in steps of 3. The ROC curves are presented in Figures 6, 7 and 8.

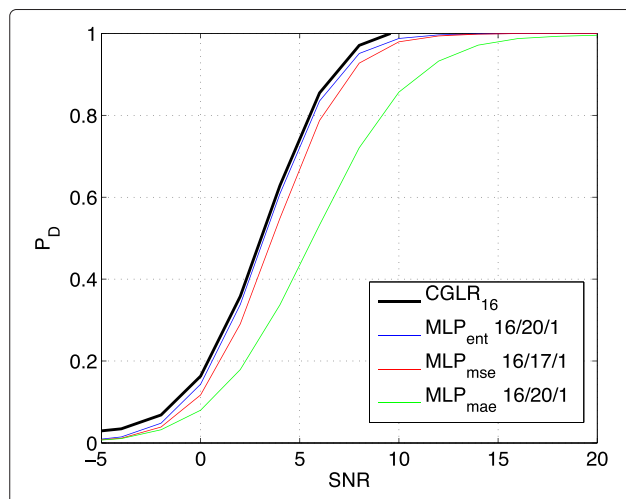
The results obtained training with the Cross-Entropy error are clearly the best. The results obtained training with the Sum-of-Squares error show higher variability and are slightly worse. The results obtained training with



**Figure 2** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting Gaussian targets with unknown correlation coefficient in AWGN.



**Figure 3** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Sum-of-Squares error for detecting Gaussian targets with unknown correlation coefficient in AWGN.



**Figure 5** Variation of  $P_D$  versus SNR, for the best detectors obtained training with the Cross-Entropy error ( $MLP_{ent}$ ), the Sum-of-Squares error ( $MLP_{mse}$ ), and the Minkowski error ( $R = 1$ ) ( $MLP_{mae}$ ).

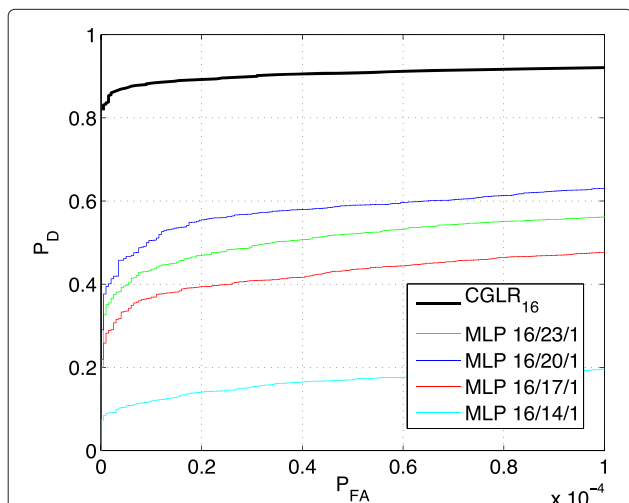
the Minkowski error ( $R = 1$ ) are clearly the worst, as expected.

#### 4.3.2 Detection of Gaussian fluctuating targets in presence of correlated Gaussian clutter and AWGN

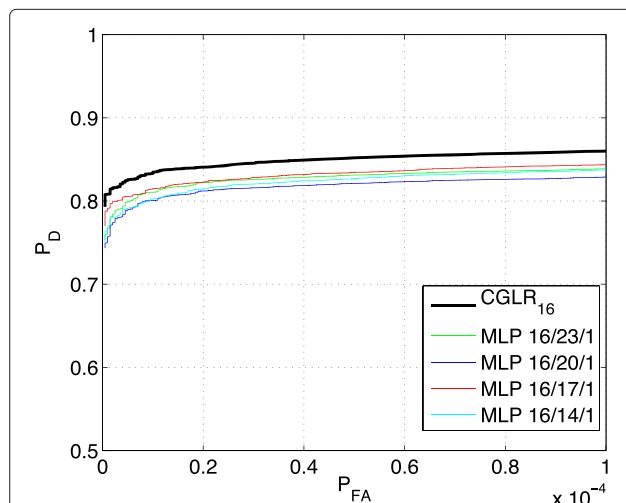
In this section, we focus on the study of the Cross-Entropy error function. The results obtained with the other two above mentioned error functions are not presented, due to space limitations, but similar conclusions could be extracted in this case. Now, the number of hidden neurons varies from 14 to 20, in steps of 3, because a good

approximation to the NP detector can be obtained with a simpler architecture.

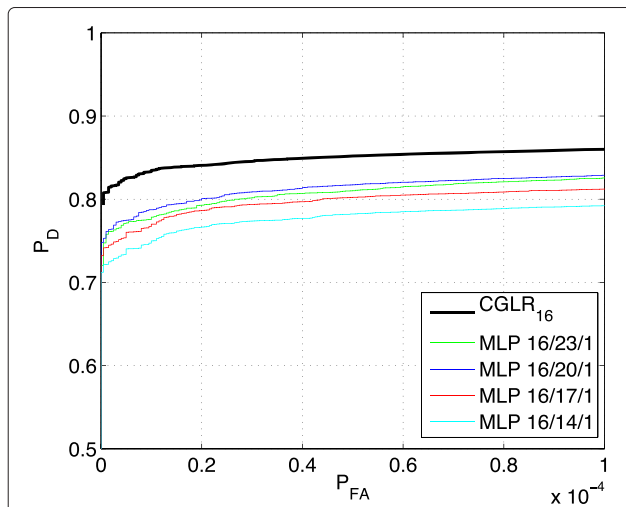
In this case, the detection of Gaussian fluctuating targets in correlated Gaussian clutter and AWGN is considered. The level of interference is expressed with the SIR, but the clutter to noise ratio (CNR) should also be known. Again, two different kinds of targets are considered: those with unknown correlation coefficient ( $\rho_t$ ), and Swerling I targets with unknown Doppler shift (the Doppler shift is modeled as a uniform random variable in the interval  $[0, 2\pi)$ ).



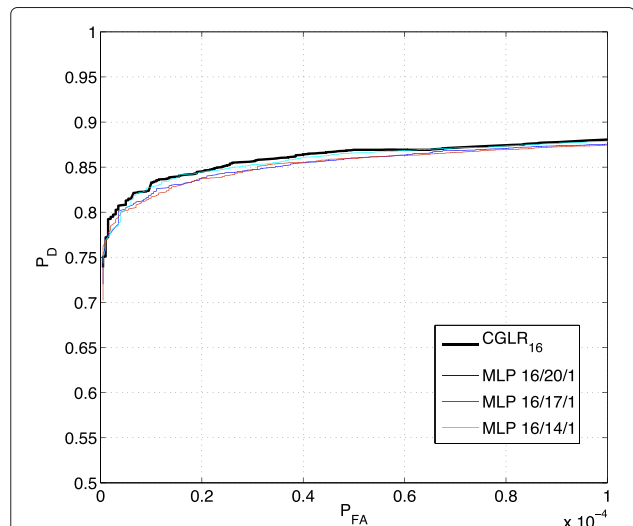
**Figure 4** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Minkowski error ( $R = 1$ ) for detecting Gaussian targets with unknown correlation coefficient in AWGN.



**Figure 6** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting Swerling I targets with unknown Doppler shift in AWGN.



**Figure 7** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Sum-of-Squares error for detecting Swerling I targets with unknown Doppler shift in AWGN.



**Figure 9** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting Gaussian targets with unknown  $\rho_t$  in Gaussian clutter plus AWGN (CNR = 20 dB,  $\rho_c = 0.7$ , SIR = 0 dB).

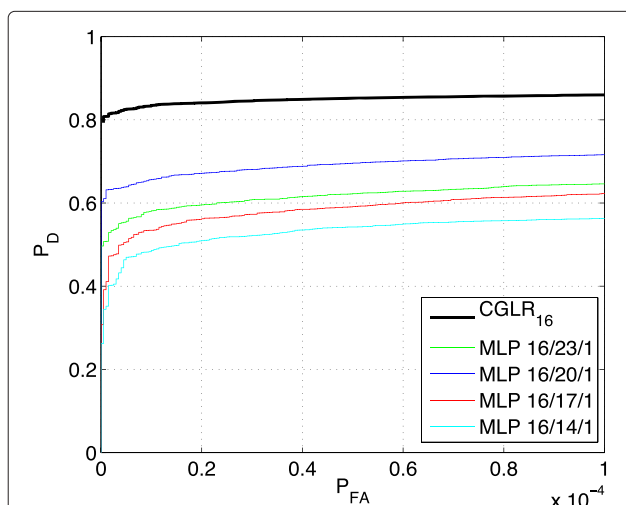
The results for the detection of correlated Gaussian targets in correlated Gaussian clutter and AWGN are presented in Figures 9 and 10. In both figures, the CNR = 20dB, and  $\rho_t$  is unknown, modeled as a uniform random variable in the interval [0, 1]. In Figure 9,  $\rho_c = 0.7$  and SIR = 0dB. In Figure 10,  $\rho_c = 0.995$  and SIR = -10dB.

The results for the SWI targets with unknown Doppler shift in correlated Gaussian clutter and AWGN are presented in Figures 11 and 12. In both figures, the CNR = 20 dB, and the Doppler shift is unknown, modeled as a uniform random variable in the interval  $[0, 2\pi)$ . In

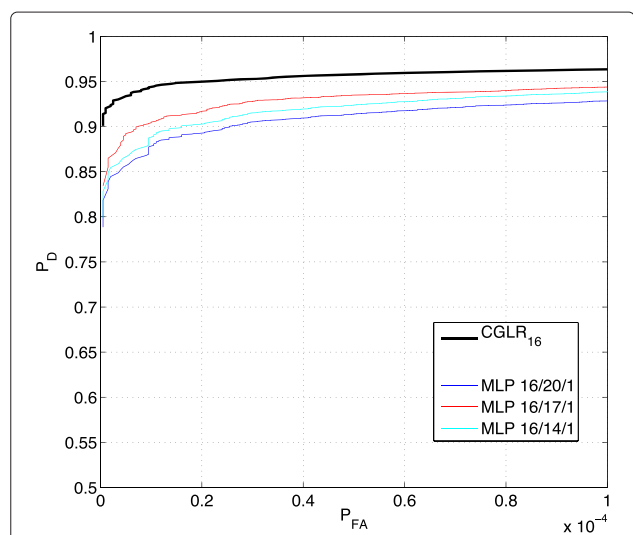
Figure 11,  $\rho_c = 0.7$  and SIR = 13 dB. In Figure 12,  $\rho_c = 0.995$  and SIR = 1 dB. Again, training with the Cross-Entropy error, as higher the number of hidden neurons, better the obtained approximation to the CGLR detector taken as reference.

### 4.3.3 Detection of non fluctuating targets in presence of spiky K-distributed clutter

In this case, the results obtained with MLPs with different number of hidden neurons for detecting non-fluctuating

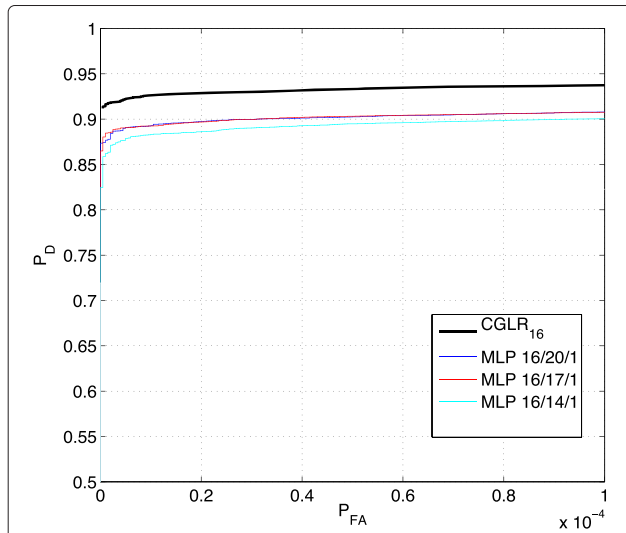


**Figure 8** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Minkowski error (R=1) for detecting Swerling I targets with unknown Doppler shift in AWGN.

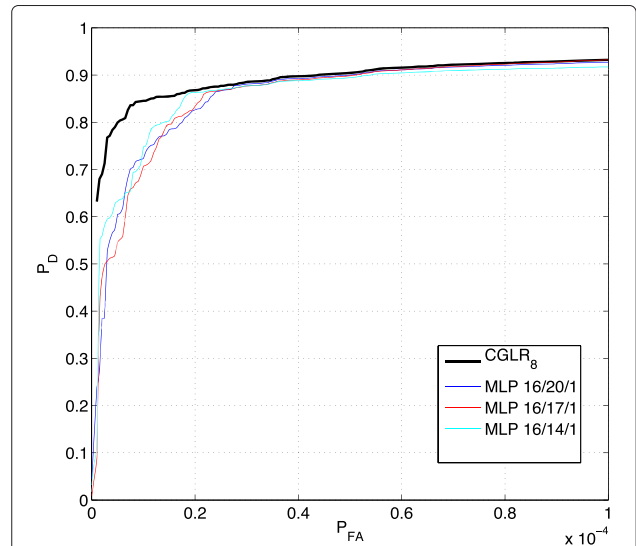


**Figure 10** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting Gaussian targets with unknown  $\rho_t$  in Gaussian clutter plus AWGN (CNR = 20 dB,  $\rho_c = 0.995$ , SIR = -10 dB).





**Figure 11** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting Swerling I targets with unknown Doppler shift in Gaussian clutter plus AWGN (CNR = 20 dB,  $\rho_c = 0.7$ , SIR = 13 dB).



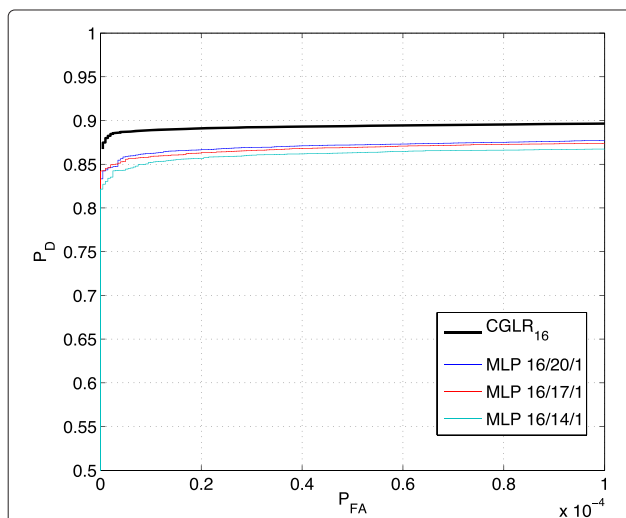
**Figure 13** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting non-fluctuating targets in spiky  $K$ -distributed clutter (SCR = 9 dB,  $\rho_c = 0$ ).

targets in spiky  $K$ -distributed clutter are presented. These experiments have been included to show the utility of our approach for detection purposes with high resolution radars and low grazing angles. In this case, the considered interference is only clutter ( $\rho_c = 0$  and SCR = 9 dB in Figure 13, and  $\rho_c = 0.9$  and SCR = -3 dB in Figure 14). The good approximation to the reference detector can be observed in all cases, even for very low  $P_{FA}$  values, and with a reduced number of neurons in the hidden layer.

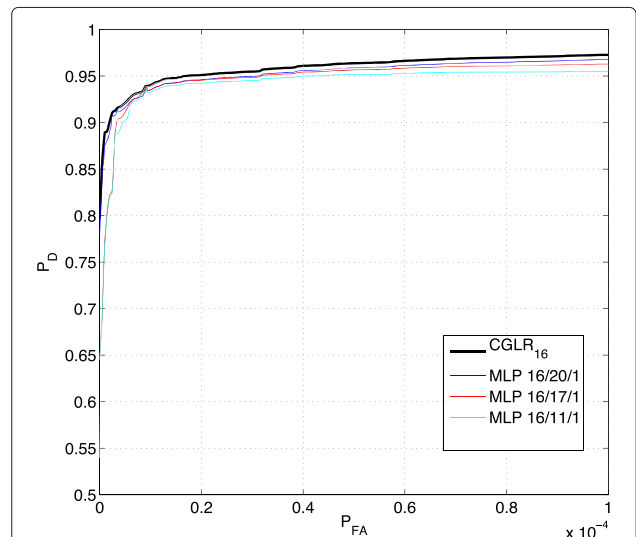
## 5 Conclusions

In this article, the possibility of approximating the NP detector using learning machines trained in a supervised manner to minimize the Cross-Entropy error has been studied.

Conventional coherent radar detectors usually apply Doppler processors (MTD, Moving Target Detectors, or MTI, Moving target Indicators) to reduce the clutter in the received signal. Most of these approaches assume



**Figure 12** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting Swerling I targets with unknown Doppler shift in Gaussian clutter plus AWGN (CNR = 20 dB,  $\rho_c = 0.995$ , SIR = 1 dB).



**Figure 14** ROC curves of the MPLs with different number of hidden neurons, trained to minimize the Cross-Entropy error for detecting non-fluctuating targets in spiky  $K$ -distributed clutter (SCR = -3 dB,  $\rho_c = 0.9$ ).



Gaussian statistics, and are implemented with linear filters. The modulus of the filtered observation vector is obtained, and finally compared to a detection threshold. Due to clutter residuals, Constant False Alarm (CFAR) techniques are applied to fulfil  $P_{FA}$  requirements. Many of the proposed solutions, assume a Gaussian distributed background. In the literature, the detection of radar targets in non-Gaussian clutter has also been addressed, but most of the approaches are based on the design of CFAR detectors that assume a specific probability density function of the clutter, and try to estimate the detection threshold for maintaining the desired  $P_{FA}$ . In this article, the learning capabilities of supervised learning machines are exploited to approximate the NP detector in cases where not only the clutter but also target parameters, are unknown. This is the general case study in a radar problem, where detection is formulated as a composite hypothesis test. Instead of using the Sum-of-Squares error, the Cross-Entropy error is considered for training, in order to exploit its better properties with respect to the sensitivity to the presence of outliers in the training set.

The function approximated by the learning machine after training has been calculated using the calculus of variations, with the objective of finding the function that minimises the formulated Cross-Entropy error.

Once the function the supervised learning machines approximates to after training has been obtained, the method proposed in [33] has been applied to demonstrate that this function can be used to implement the NP detector by comparing the trained learning machine output to a threshold, selected according to  $P_{FA}$  requirements.

This theoretical result has been assessed with some experiments. Neural networks based detectors have been considered for detecting different types of signals in different types of interferences. Results prove that an MLP trained to minimize the Cross-Entropy error, can implement a very good approximation to the NP detector for the considered detection problems, even for low  $P_{FA}$  values.

Different experiments have been performed for detecting fluctuating radar targets (Gaussian targets with unknown correlation coefficient and Swerling I target with unknown Doppler shift) in AWGN, and in correlated Gaussian clutter plus AWGN. Also, the detection of non-fluctuating targets in spiky  $K$ -distributed clutter has been studied. In all the cases, the capability of learning machines (particularly, MLPs) trained to minimize the Cross-Entropy error, to approximate the optimum detector in the NP sense, has been demonstrated. To obtain a good approximation, the number of hidden neurons must be high enough. This number is related to the minimum number of hyperplanes necessary to enclose the acceptance regions of the detection problem, but this theoretical study is beyond the objective of this article.

For comparison purposes, some experiments have been carried out with MLPs trained to minimize the Sum-of-Squares error, and the Minkowski error with  $R = 1$ . The results obtained training with the Cross-Entropy error are better than the results obtained training with the Sum-of-Squares error, but both can be used to approximate the NP detector. The detectors trained with the Minkowski error with  $R = 1$  are the worst, and this detector is very far from the NP optimum detector.

Compared with conventional radar detectors, this approach has the following advantages:

- Good approximations to the optimal NP detector can be obtained if a suitable error function is selected, if a representative training set is available, if the learning machine architecture has a high enough number of free parameters, and finally, if a good training algorithm is used.
- No statistical models have to be assumed during the design. On the contrary, most of the CFAR detectors that can be found in the literature assume some statistical model for the interference that is used to adjust the detection threshold to maintain the desired  $P_{FA}$ .
- The implementation of the NP detector based on the Average-Likelihood ratio, when some of the parameters or the statistical models assumed in the design are random, can lead to intractable integrals. When supervised learning machines are used to approximate the NP detector, only training data obtained experimentally are necessary, without knowledge of the statistical distributions, and without solving those integrals.

Obviously, supervised learning machines are not an ideal solution in radar detection problems. The main drawback is the difficulty of obtaining representative training samples, and the definition of the most suitable learning machine architecture.

#### Competing interests

The author declare that they have no competing interests.

#### Acknowledgements

This work has been partially funded by the Spanish Ministry of Economy and Competitiveness with project TEC2012-38701.

Received: 31 July 2012 Accepted: 5 February 2013

Published: 9 March 2013

#### References

1. J Jarabo-Amores, M Rosa-Zurera, R Gil-Pita, F López-Ferreras, Study of two error functions to approximate the Neyman–Pearson detector using supervised learning machines. *IEEE Trans. Signal Proces.* **57**(11), 4175–4181 (2009)
2. J Neyman, K Pearson, On the problem of the most efficient test of statistical hypotheses. *Philosop. Trans. Roy. Soc. Lond. A* **231**, 492–510 (1933)
3. HV Trees, *Detection, estimation, and modulation theory*, vol. 1. (Wiley, New York, 1968)

4. A di Vito, M Naldi, Robustness of the likelihood ratio detector for moderately fluctuating radar targets. *IEE Proc. Radar Sonar Navig.* **146**(2), 107–122 (1999)
5. Q He, in *Proceedings of IEEE-ICASSP*. MIMO radar diversity with Neyman–Pearson signal detection in non-Gaussian circumstance with non-orthogonal waveforms (Prague, Czech Republic, 2011), pp. 2764–2767
6. Q He, R Blum, Diversity gain for MIMO Neyman–Pearson signal detection. *IEEE Trans. Signal Process.* **59**(3), 869–881 (2011)
7. Y Yang, R Blum, B Sadler, A distributed and energy-efficient framework for Neyman–Pearson detection of fluctuating signals in large-scale sensor networks. *IEEE J. Sel. Areas Commun.* **28**(7), 1149–1158 (2010)
8. R Vicen-Bueno, R Carrasco-Alvarez, M Jarabo-Amores, J Nieto-Borge, E Alexandre-Cortizo, Detection of ships in marine environments by square integration mode and multilayer perceptrons. *IEEE Trans. Instrum. Meas.* **60**(3), 712–724 (2011)
9. T Furon, J Josse, S le Squin, Some theoretical aspects of watermarking detection. *Proc. SPIE.* **6072**, 553–564 (2006)
10. T Ng, H Garg, Maximum-likelihood detection in DWT domain image watermarking using Laplacian modeling. *IEEE Signal Process. Lett.* **12**, 285–288 (2005)
11. I Gu, N Ernberg, E Styvaktakis, M Bollen, A statistical-based sequential method for fast online detection of fault-induced voltage dips. *IEEE Trans. Power Del.* **19**, 497–504 (2004)
12. J Chamberland, V Veeravalli, Decentralized detection in sensor networks. *IEEE Trans. Signal Process.* **139**, 407–416 (2003)
13. P Sebastiani, E Gussoni, I Kohane, M Ramoni, Statistical challenges in functional genomics. *Stat. Sci.* **18**, 33–60 (2003)
14. S Dudoit, J Fridlyand, T Speed, Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* **97**, 77–87 (2002)
15. K Nandakumar, Y Chen, S Dass, A Jain, Likelihood ratio-based biometric score fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 342–347 (2008)
16. R Prix, B Krishnan, Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics. *Class. Quant. Grav.* **26**(20) (2009). doi:10.1088/0264-9381/26/20/204013
17. Y Sung, L Tong, H Poor, Neyman–Pearson detection of Gauss–Markov signals in noise: closed-form error exponent and properties. *IEEE Trans. Inf. Theory.* **52**, 1354–1365 (2006)
18. D Andina, J Sanz-González, J Jiménez-Pajares, in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 1. A comparison of criterion functions for a neural network applied to binary detection (Perth, WA, 1995), pp. 329–333
19. D Andina, J Sanz-González, O Rodríguez-Martín, in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 1. Performance improvements for a neural network detector (Perth, WA, Australia, 1995), pp. 492–495
20. A Burian, P Kuosmanen, J Saarinen, in *Proc. ISCAS*, vol. 5. Neural detectors with variable threshold (Orlando, FL, USA, 1999), pp. 599–602
21. V Ramamurti, SS Rao, PP Gandhi, in *Proc. ICASSP*, vol. 1. Neural detectors for signals in non-gaussian noise (Minneapolis, MN, USA, 1993), pp. 481–484
22. P Gandhi, V Ramamurti, Neural networks for signal detection in non-gaussian noise. *IEEE Trans. Signal Process.* **45**, 2846–2851 (1997)
23. D Munro, O Ersoy, M Bell, J Sadowsky, Neural network learning of low-probability events. *IEEE Trans. Aerosp. Electron. Syst.* **32**, 898–910 (1996)
24. D Ruck, S Rogers, M Kabrisky, M Oxley, B Suter, The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Trans. Neural Netw.* **1**, 296–298 (1990)
25. D Casasent, X Chen, New training strategies for RBF neural networks for X-ray agricultural product inspection. *Pattern Recogn.* **36**, 535–547 (2003)
26. D Casasent, X Chen, Radial basis function neural networks for nonlinear Fisher discrimination and Neyman Pearson classification. *Neural Netw.* **16**, 529–535 (2003)
27. P Jarabo-Amores, R Gil-Pita, M Rosa-Zurera, F López-Ferreras. MLP and RBFN for detecting white Gaussian signals in white Gaussian interference. *Lecture Notes in Computer Science* vol. 2687 (Maó, Spain, 2003), pp. 790–797
28. M Davenport, R Baraniuk, C Scott, in *Proc. of ICASSP*, vol. 5. Controlling false alarms with support vector machines (Toulouse, France, 2006), pp. 589–592
29. S Haykin, D Thomson, Signal detection in a nonstationary environment reformulated as an adaptive pattern classification problem. *IEEE Trans. Signal Process.* **85**(11), 2325–2344 (1998)
30. AM Fernandes, AB Utkin, AV Lavrov, RM Vilar, Development of neural network committee machines for automatic forest fire detection using lidar. *Pattern Recogn.* **37**, 2039–2047 (2004)
31. D Mata-Moya, P Jarabo-Amores, R Vicen-Bueno, M Rosa-Zurera, F López-Ferreras. Neural network detectors for composite hypothesis tests. *Lecture Notes in Computer Science*, vol. 4224 (Burgos, Spain, 2006), pp. 298–305
32. D Mata-Moya, P Jarabo-Amores, M Rosa-Zurera, J Nieto-Borge, F López-Ferreras, Combining MLPs and RBFNNs to detect signals with unknown parameters. *IEEE Trans. Instrum. Meas.* **58**(9), 2989–2995 (2009)
33. P Jarabo-Amores, M Rosa-Zurera, R Gil-Pita, F López-Ferreras, in *2005 IEEE/SP 13th Workshop on Statistical Signal Processing*. Sufficient condition for an adaptive system to approximate the Neyman–Pearson detector (Bordeaux, France, 2005), pp. 295–230
34. CM Bishop, *Neural Networks for Pattern Recognition*, 1edn. (Oxford University Press, Oxford, UK, 1995)
35. M Saerens, P Latinne, C Decaestecker, Any reasonable cost function can be used for a posteriori probability approximation. *IEEE Trans. Neural Netw.* **13**, 1204–1210 (2002)
36. J Slivka, N Severo, On the strong law of large numbers. *Proc. Am. Math. Soc.* **24**, 729–734 (1970)
37. J Hampshire, B Pearlmutter, in *Proc. of 1990 Connectionist Models Summer School*, ed. by DS Touretzky, JL Elman, TJ Sejnowski, and GE Hinton. Equivalence proofs for multilayer perceptron classifiers and the Bayesian discriminant function (San Diego, CA, USA, 1990), pp. 159–172
38. I Gelfand, S Fomin, *Calculus of Variations*. (Courier Dover Publications, Mineola, New York, 2000)
39. R Weinstock, *Calculus of Variations with Applications to Physics and Engineering*. (McGraw Hill, New York, 1952)
40. E Conte, M Longo, M Lops, S Ullo, Radar detection of signals with unknown parameters in  $K$ -distributed clutter. *IEE Proc. F Radar Signal Process.* **138**(2), 131–138 (1991)
41. E Conte, M Lops, G Ricci, Radar detection in  $K$ -distributed clutter. *IEE Proc. Radar Sonar Navig.* **141**(2), 116–118 (1994)
42. MM Nayebi, MR Aref, MH Bastani, Detection of coherent radar signals with unknown Doppler shift. *IEE Proc. Radar Sonar Navig.* **143**(2), 79–86 (1996)
43. A El-Jaroudi, J Makhoul, in *Proc. of Int. Conf. on Neural Networks IJCNN*, vol. 1. A new error criterion for posterior probability estimation with neural nets (San Diego, CA, USA, 1990), pp. 185–192
44. C Charalambous, Conjugate gradient algorithm for efficient training of artificial neural networks. *IEEE Proc. Circ. Dev. Syst.* **51**(3), 301–310 (1992)

doi:10.1186/1687-6180-2013-44

**Cite this article as:** Jarabo-Amores et al.: Radar detection with the Neyman–Pearson criterion using supervised-learning-machines trained with the cross-entropy error. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:44.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)