Signal Processing 131 (2017) 364–375

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Approximating the Neyman–Pearson detector with 2C-SVMs. Application to radar detection

David de la Mata-Moya, María Pilar Jarabo-Amores, Jaime Martín de Nicolás, Manuel Rosa-Zurera *

Signal Theory and Communications Department, Polytechnic School, University of Alcalá, Campus Universitario, 28805 Alcalá de Henares (Madrid), Spain

ARTICLE INFO

Article history:
Received 26 April 2016
Received in revised form
27 July 2016
Accepted 15 August 2016
Available online 16 August 2016

Keywords:
C-SVM
2C-SVM
Neyman–Pearson detector

ABSTRACT

This paper presents a study about the possibility of implementing approximations to the Neyman–Pearson detector with C-Support Vector Machines and 2C-Support Vector Machines. It is based on obtaining the functions these learning machines approximate to after training to minimize the empirical risk, and on the possible implementation of the Neyman–Pearson detector with these approximated functions. The function approximated by a C-Support Vector Machine after perfect training is a binary function, with only two possible outputs. When the output of the C-Support Vector Machine is compared to a threshold, whose value is the intermediate between the possible outputs, an implementation of the Maximum-A-Posteriori classifier is obtained. On the other hand, the function approximated by a 2C-Support Vector Machine after perfect training is also a binary function, but this machine implements the Neyman–Pearson detector for a fixed probability of false alarm and probability of detection pair, that can be selected with the parameter γ which controls the costs of the error function. Some experiments about radar detection have been carried out, in order to confirm the theoretical results. The results of these experiments allow us to confirm that the 2C-Support Vector Machine can implement very good approximations to the Neyman–Pearson detector.

1. Introduction

This paper tackles the application of Support Vector Machines (SVMs) to approximate the Neyman–Pearson (NP) detector. This detector maximizes the probability of detection (P_D), for a fixed probability of false alarm (P_{FA}) [1], and it is especially useful in binary hypothesis tests when the assignment of costs and the knowledge of prior probabilities are difficult. The NP criterion has been widely used in radar applications such as optimum radar detector design in different radar scenarios [2,3,4–8], MIMO radars [9–11], and distributed radar sensor networks [12,13]. It has also been applied in some other topics: watermarking [14,15], fault-induced dips detection [16], biometric [17], or gravitational waves detection [18].

If $p(\mathbf{z}|H_0)$ and $p(\mathbf{z}|H_1)$ are the likelihood functions of the input random vector $\mathbf{z}$ for the null hypothesis (H_0) and the alternative hypothesis (H_1), respectively, a possible implementation of the NP detector consists in comparing the Likelihood Ratio (LR) ($\Lambda(\mathbf{z})$) to a threshold selected according to P_{FA} requirements ($\eta_{lr}(P_{FA})$), as expressed in (1) [19]:

$$\Lambda(\mathbf{z}) = \frac{p(\mathbf{z}|H_1)}{p(\mathbf{z}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{lr}(P_{FA}) \tag{1}$$

This approach requires a complete characterization of the likelihood functions. Usually, statistical models of interference and targets are assumed and their unknown parameters (e.g. the target strength, or the Doppler frequency) are modeled as random variables. When the distributions of these parameters are known a priori, the optimum detector in the NP sense can be implemented by comparing the average likelihood ratio (ALR) to a detection threshold fixed according to P_{FA} requirements [19]. The ALR can lead to intractable integrals that should be solved by numerical approximations.

Another possibility is to obtain the Maximum Likelihood estimators of the parameters, implementing the Generalized Likelihood Ratio Test (GLRT). An approximation of these parameters will affect the P_{FA}, and detection losses are to be expected when the interference or target statistical properties vary from those assumed in the design.

* Corresponding author.
E-mail addresses: david.mata@uah.es (D.d.l. Mata-Moya), mpilar.jarabo@uah.es (M.P. Jarabo-Amores), jaime.martinn@uah.es (J.o.s. Martín de Nicolás), manuel.rosa@uah.es (M. Rosa-Zurera).

http://dx.doi.org/10.1016/j.sigpro.2016.08.021
0165-1684/

A different approach to implement nearly optimum detectors in the NP sense consists in using a learning machine that approximates a discriminant function after training, which implements the NP detector if it is compared to a suitable threshold. The possibility of approximating the NP optimum detector using supervised learning machines trained to minimize the sum-of-squares and the cross-entropy errors has been previously studied in [20,21], respectively. Those studies concluded that the NP detector can be approximated by using a learning machine trained to minimize those error functions with a training set consisting of experimental data, but without prior knowledge of the likelihood functions and without statistical models assumed in the design. The main drawbacks of this approach are the difficulty of obtaining representative training samples, and the selection of the best learning machine architecture.

In [20,21], learning machines with only one output were considered. The learning machines outputs were compared to a threshold in order to decide in favor of the null or the alternative hypothesis. The threshold allows us to fix the desired P_{FA}. An equivalent implementation consists in varying the bias of the output neuron [22,23]. A different approach was used in [24], where a neural network (NN) with two outputs in interval $(0, 1)$ was used, and the subtraction of both outputs was compared to a threshold. This approach is equivalent to using a NN with only one output and desired outputs $(-1, 1)$ [25]. Radial Basis Function Neural Networks (RBFNNs), Second Order Neural Networks (SONNs) and committees of NNs have also been applied to approximate the NP detector [26–28].

Another intelligent agent is the Support Vector Machine (SVM). It is an approximate implementation of the method of structural risk minimization that can provide good generalization on detection and classification problems without incorporating problem-domain knowledge [29,30]. In [31], SVMs are applied to detection problems; the Receiver Operating Characteristic (ROC) curves of the original SVM-based detectors were obtained using two strategies for varying the P_{FA}, which consist in varying the bias of the original SVM formulation (C-SVM) or the costs assigned to the two types of error (missed detections and false alarms) using cost-sensitive SVMs (2C-SVMs), respectively. In [32], the design of support vector classifiers with respect to the Neyman–Pearson criterion is studied. It is extended in [33], where 2C-SVMs were used to approximate detectors based on the NP and minimax criteria. They consider the minimization of objective functions which depend on the desired P_{FA}. The best results have been obtained experimentally with greedy searches to find the best training parameters for obtaining the highest P_D for a given P_{FA}. Nevertheless, as far as the authors know, there is no theoretical study in the literature about the capabilities of SMVs to approximate the Neyman–Pearson detector for a wide range of P_{FA} values.

The analysis to follow is based upon that in [20], employing its methodology to study the possibility of using C-SVMs and 2C-SVMs for implementing good approximations of the NP detector. It was demonstrated in [33] that 2C-SVMs can be trained to obtain a desired P_{FA} when the detector is implemented with the learning machine, but it has not been demonstrated that at the same time the resulting P_D is maximum, as in the Neyman–Pearson test. Our approach is more general, and it considers the possibility of approximating a discriminant function $f_0(\mathbf{z})$ with the trained learning machine, which can implement approximations to the Neyman–Pearson detector by comparing its output to a threshold:

$$f_0(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0(P_{FA}) \tag{2}$$

A method and a sufficient condition for the discriminant function $f_0(\mathbf{z})$ to be used to implement the Neyman–Pearson detector are going to be used throughout. The method was proposed in [35], and consists in checking the fulfillment of the following requisites:

1. The discriminant function $f_0(\mathbf{z})$ approximated by the learning machine after training must depend on the problem likelihood functions.
2. If the discriminant function is compared to a threshold η_0, the likelihood ratio $\Lambda(\mathbf{z})$ can be isolated, obtaining an expression similar to (1).
3. For the expression (2) to be equivalent to the Neyman–Pearson detector, it is sufficient that the relation between η_{lr} and η_0 does not depend on $\mathbf{z}$.

In other words, if η_0 can be determined for each pair (η_{lr}, P_{FA}), and it is not a function of $\mathbf{z}$, the rule (2) is an implementation of the NP detector. In this paper, this sufficient condition is applied to determine the suitability of C-SVMs and 2C-SVMs to approximate the NP detector. To that end, the discriminant functions approximated by the C-SVM and 2C-SVM after training are obtained.

Besides, we demonstrate that the discriminant functions that are obtained after C-SVM and 2C-SVM perfect trainings are binary, making the selection of P_{FA} unfeasible by varying the threshold the discriminant function is compared with. At the same time, it is demonstrated that C-SVMs can be used to implement the Minimum Probability of Error binary classifier, and 2C-SVMs, can be used to implement the Neyman–Pearson detector for a given pair of (P_D, P_{FA}), that corresponds to one point of the ROC curve. The value of P_{FA} could be adjusted by varying one of the parameters during training.

The paper is structured as follows: in Section 2, the problem to be solved is presented, jointly with a study about SVMs. The study of the possibility of using C-SVMs to approximate the Neyman–Pearson detector is included in Section 3. An equivalent study with 2C-SVMs is carried out in Section 4. In Section 5, experiments and results are presented to illustrate the previous theoretical study. Finally, the main contributions of this paper are summarized in Section 6.

2. Background about SVM classifiers

Let us consider a learning machine with one output to classify input vectors $\mathbf{z} \in \mathbb{R}^L$ into two hypotheses or classes, H_0 and H_1. Let $\mathcal{Z}_i$ be the set of all possible input vectors generated under hypothesis H_i with probability density function $p(\mathbf{z}|H_i)$, and $\mathcal{Z}$ the ensemble of all possible input vectors ($\mathcal{Z}_0 \cup \mathcal{Z}_1 = \mathcal{Z}$). A training set $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$ ($\mathcal{X} \subset \mathcal{Z}$) is available, where $\mathcal{X}_i \subset \mathcal{Z}_i$ is composed of N_i pre-classified patterns, with desired outputs d_i, $i \in \{0, 1\}$, and $N = N_0 + N_1$. The output of the learning machine is denoted by $f(\mathbf{z}, \alpha)$, representing α the set of parameters to be adjusted during training. Once the learning machine is trained, it operates by comparing the output to a threshold, deciding in favor of H_1 if the output is higher than the threshold, and in favor of H_0 if the output is lower than the threshold. It is represented in Fig. 1.

In this section, SVM classifiers are described, introducing training as an optimization process subject to some constraints. The objective functions for training C-SVMs and 2C-SVMs are also presented, with the parameters that must be adjusted.

2.1. C-SVM classifiers

Let us suppose a learning machine designed to learn the mapping $\mathbf{z}_j \to d_i$, $\forall\, \mathbf{z}_j \in \mathcal{X}_i$. The machine actually performs the mapping $\mathbf{z}_j \to f(\mathbf{z}_j, \alpha)$, where α defines the adjustable parameters of the learning machine. In this context, the expected risk [29], or

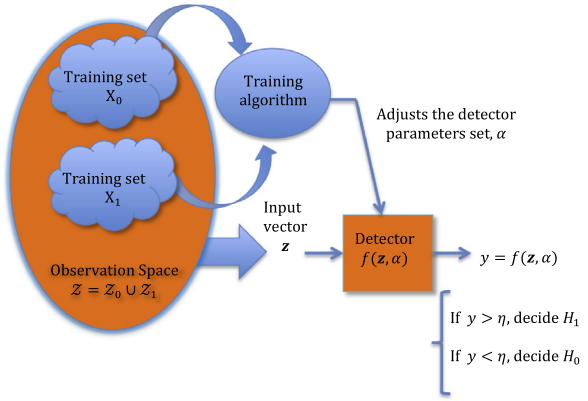

Fig. 1. Scheme of the elements of the tackled detection problem.

just the risk, is defined as:

$$R(\alpha) = \int \frac{1}{2}|y - f(\mathbf{z}, \alpha)|p(\mathbf{z}, y)d\mathbf{z}dy \tag{3}$$

where $p(\mathbf{z}, y)$ is the joint density of the input and the desired output, y. In a binary classification problem, the empirical risk R_{emp} is defined to be just the mean error measured on the training set [34], and it is expressed in (4), being d_i, $i \in \{0, 1\}$ the desired output for any pattern $\mathbf{z} \in \mathcal{X}_i$:

$$R_{emp}(\alpha) = \frac{1}{2N} \sum_{\forall \mathbf{z} \in \mathcal{X}_0} |d_0 - f(\mathbf{z}, \alpha)| + \frac{1}{2N} \sum_{\forall \mathbf{z} \in \mathcal{X}_1} |d_1 - f(\mathbf{z}, \alpha)| \tag{4}$$

In C-SVMs, the function implemented by the learning machine is a linear function of the results of mapping the input pattern $\mathbf{z}$ into a higher dimensional space $\mathcal{H}$ with the functions $\Phi(\mathbf{z})$, that are known as "kernel functions". The parameters of the learning machine are the weights vector $\mathbf{w}$, the bias constant, and the parameters the functions $\Phi(\mathbf{z})$ depend on. The function implemented by the learning machine is going to be represented as $f(\mathbf{z})$ for simplicity:

$$f(\mathbf{z}) = \mathbf{w}^T \Phi(\mathbf{z}) + b \tag{5}$$

In the simplest implementation of the detector, for each training example $\mathbf{z}$, the function should provide $f(\mathbf{z}) \geq 0$, if $\mathbf{z} \in \mathcal{X}_1$ and $f(\mathbf{z}) \leq 0$, if $\mathbf{z} \in \mathcal{X}_0$, that is, training examples from the two hypotheses are separated in $\mathcal{H}$ by the hyperplane $\mathbf{w}^T\Phi(\mathbf{z}) + b = 0$. The SVM is based on the hyperplane which maximizes the separating margin between the two classes that can be obtained mathematically by minimizing the following cost function, subject to the separability constraints [29]:

$$J(\mathbf{w}) = \frac{1}{2} \| \mathbf{w} \|^2 \tag{6}$$

subject to:

$$d_i f(\mathbf{z}_i) \geq 1; \quad i = 1, 2, \ldots, N \tag{7}$$

In practice, the training data may not be completely separable by any hyperplane in space $\mathcal{H}$. To relax the separability constraints in expression (7), slack variables denoted by ξ_i are introduced as follows:

$$d_i f(\mathbf{z}_i) \geq 1 - \xi_i; \quad \xi_i \geq 0; \quad i = 1, 2, \ldots, N \tag{8}$$

For an error to occur $d_i f(\mathbf{z}_i) < 0$; therefore, the corresponding slack variables $\xi_i \geq 1$, so that $\sum_i \xi_i$ is an upper bound on the number of training errors. The objective function can be changed to assign an extra cost for errors, and the optimization problem to solve in C-SVM becomes the following [32,33]:

$$\min_{f,\xi} \left\{ \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{N} \xi_i \right\} \tag{9}$$

subject to:

$$d_i f(\mathbf{z}_i) \geq 1 - \xi_i; \quad i = 1, \ldots, N$$
$$\xi_i \geq 0, \quad i = 1, \ldots, N$$

The constraint $d_i f(\mathbf{z}_i) \geq 1 - \xi_i$, together with $\xi_i \geq 0$, is equivalent to $\xi_i = \max(0, 1 - d_i f(\mathbf{z}_i))$. Hence, the learning problem is equivalent to the unconstrained optimization problem defined in expression (10) for C-SVMs:

$$\min_{f} \left\{ \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{N} \max(0, 1 - d_i f(\mathbf{z}_i)) \right\} \tag{10}$$

or equivalently, $u(\cdot)$ being the Heaviside step function, and normalizing by the number of training patterns:

$$\min_{f} \left\{ \frac{1}{2N} \| \mathbf{w} \|^2 + \frac{C}{N} \sum_{i=1}^{N} (1 - d_i f(\mathbf{z}_i)) u(1 - d_i f(\mathbf{z}_i)) \right\} \tag{11}$$

where $\frac{C}{N} \sum_{i=1}^{N} (1 - d_i f(\mathbf{z}_i)) u(1 - d_i f(\mathbf{z}_i))$ is the soft-margin loss function, and $\frac{1}{2N} \| \mathbf{w} \|^2$ is the regularization term.

2.2. 2C-SVM classifier

The C-SVM formulation implicitly penalizes errors in both classes alike. As an alternative, 2C-SVMs allow the control of the cost associated with the two possible errors: C_{10} associated with false alarms and C_{01} associated with detection losses $\left(0 \leq C_{ij} \leq 1; \; i, j \in \{0, 1\}; \; i \neq j\right)$.

C_{10} and C_{01} can be related, using an additional parameter γ $(0 \leq \gamma \leq 1)$, obtaining $C\gamma = C_{10}$ and $C(1 - \gamma) = C_{01}$. The primal formulation of the 2C-SVM is [32,33]:

$$\min_{f,\xi,\gamma} \left\{ \frac{1}{2} \| \mathbf{w} \|^2 + C\gamma \sum_{i \in \mathcal{X}_0} \xi_i + C(1 - \gamma) \sum_{i \in \mathcal{X}_1} \xi_i \right\} \tag{12}$$

subject to:

$$d_i f(\mathbf{z}_i) \geq 1 - \xi_i; \quad i = 1, \ldots, N \xi_i \geq 0, \quad i = 1, \ldots, N$$

Considering again that the constraint $d_i f(\mathbf{z}_i) \geq 1 - \xi_i$, together with $\xi_i \geq 0$ is equivalent to $\xi_i = \max(0, 1 - d_i f(\mathbf{z}_i))$, the equivalent expression (13) for 2C-SVMs is obtained:

$$\min_{f,\gamma} \left\{ \frac{1}{2} \| \mathbf{w} \|^2 + C\gamma \sum_{i \in \mathcal{X}_0} \max\left(0, 1 - d_i f\left(\mathbf{z}_i\right)\right) + C\left(1 - \gamma\right) \sum_{i \in \mathcal{X}_1} \max\left(0, 1 - d_i f\left(\mathbf{z}_i\right)\right) \right\} \tag{13}$$

or equivalently:

$$\min_{f,\gamma} \left\{ \frac{1}{2N} \| \mathbf{w} \|^2 + \frac{C}{N}\gamma \sum_{i \in \mathcal{X}_0} \left(1 - d_i f\left(\mathbf{z}_i\right)\right) u\left(1 - d_i f\left(\mathbf{z}_i\right)\right) + \frac{C}{N}\left(1 - \gamma\right) \sum_{i \in \mathcal{X}_1} \left(1 - d_i f\left(\mathbf{z}_i\right)\right) u\left(1 - d_i f\left(\mathbf{z}_i\right)\right) \right\} \tag{14}$$

which is again formulated as the summation of a loss function, and a regularization term.

3. C-SVMs capability to implement discriminant functions useful to approximate the NP detector

In this section, the function $f_0(\mathbf{z})$ which minimizes the unconstrained objective function defined in expression (11) is obtained, under some general conditions. Let us consider that the conditions for perfect training are fulfilled, understanding that "perfect training" is achieved when the learning machine is a "sufficiently powerful model" that is able to approximate the discriminant function to any degree of accuracy, and the minimum of the error function is indeed reached (the training set must be a representative subset of the input space, and the learning algorithm must be able to find the appropriate minimum).

In the context of the detection problem this paper deals with, it is supposed that the desired outputs of the learning machine are $d_i = -1$ and $d_i = 1$, for $\mathbf{z}_i \in \mathcal{X}_0$ and $\mathbf{z}_i \in \mathcal{X}_1$, respectively. The number of training patterns is supposed high $(N \to \infty)$, and representative (patterns are generated according to the likelihood functions $p(\mathbf{z}|H_i)$). Therefore, the training process is equivalent to the following optimization problem:

$$\min_f \left\{ \lim_{N\to\infty} \left\{ \frac{1}{2N} \| \mathbf{w} \|^2 + \frac{C}{N} \sum_{i=1}^{\infty} (1 - d_i f(\mathbf{z}_i)) u(1 - d_i f(\mathbf{z}_i)) \right\} \right\} \tag{15}$$

In these conditions, the regularization term $\left(\frac{1}{2N} \| \mathbf{w} \|^2\right)$ tends to zero, and the functional to be minimized only depends on the loss function of expression (11) [36]. Using the Strong Law of Large Numbers, the optimization problem can be formulated with expression (16):

$$\min_f \left\{ C \int_{\mathcal{Z}} \left(P(H_0)\left(1 + f(\mathbf{z})\right) u\left(1 + f(\mathbf{z})\right) p\left(\mathbf{z}\middle|H_0\right) + P(H_1) \right.\right.$$
$$\left.\left. \left(1 - f(\mathbf{z})\right) u\left(1 - f(\mathbf{z})\right) p(\mathbf{z}|H_1) \right) d\mathbf{z} \right\} \tag{16}$$

Generally, the function $f_0(\mathbf{z})$ that minimizes expression (16) should be obtained using Calculus of Variations and particularly the Euler–Lagrange differential equation [37,38]. The calculus of variations can be used to find the function $f(\mathbf{z})$ that minimizes the functional $J(f)$ defined as follows:

$$J(f) = \int_{\mathcal{Z}} I\left(\mathbf{z}, f(\mathbf{z}), \frac{\partial f(\mathbf{z})}{\partial z_1}, \frac{\partial f(\mathbf{z})}{\partial z_2}, \ldots, \frac{\partial f(\mathbf{z})}{\partial z_L}\right) d\mathbf{z} \tag{17}$$

where 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 (18), where $f'_k = \frac{\partial f}{\partial z_k}$:

$$\frac{\partial I}{\partial f} - \sum_{k=1}^{L} \frac{\partial}{\partial z_k}\left(\frac{\partial I}{\partial f'_k}\right) = 0 \tag{18}$$

In our problem, the functional to be minimized only depends on $f(\mathbf{z})$, and after avoiding unnecessary constants can be expressed as follows:

$$\int_{\mathcal{Z}} I\left(f\left(\mathbf{z}\right)\right) d\mathbf{z} = \int_{\mathcal{Z}} \left(P\left(H_0\right)\left(1 + f\left(\mathbf{z}\right)\right) u\left(1 + f\left(\mathbf{z}\right)\right) p\left(\mathbf{z}\middle|H_0\right) \right.$$
$$\left. + P\left(H_1\right)\left(1 - f\left(\mathbf{z}\right)\right) u\left(1 - f\left(\mathbf{z}\right)\right) p\left(\mathbf{z}\middle|H_1\right) \right) d\mathbf{z} \tag{19}$$

$I(f(\mathbf{z}))$ depends on the Heaviside step function, making the application of the Euler–Lagrange equation unfeasible. In order to find the solution of (16), it is important to consider the fact that $I(f(\mathbf{z}))$ is a continuous piece-wise function, which can be expressed as follows:

$$I(f(\mathbf{z})) = \begin{cases} P(H_1)p(\mathbf{z}|H_1)(1 - f(\mathbf{z})) & \text{if} \quad f(\mathbf{z}) < -1 \\ P(H_0)p(\mathbf{z}|H_0)(1 + f(\mathbf{z})) + & \\ + P(H_1)p(\mathbf{z}|H_1)(1 - f(\mathbf{z})) & \text{if} \quad -1 < f(\mathbf{z}) < 1 \\ P(H_0)p(\mathbf{z}|H_0)(1 + f(\mathbf{z})) & \text{if} \quad 1 < f(\mathbf{z}) \end{cases} \tag{20}$$

After some simple manipulations, expression (21) is obtained, where $A = P(H_1)p(\mathbf{z}|H_1)$, $D = P(H_0)p(\mathbf{z}|H_0)$, $B = D + A$, and $C = D - A$:

$$I(f(\mathbf{z})) = \begin{cases} A - Af(\mathbf{z}) & \text{if} \quad f(\mathbf{z}) < -1 \\ B + Cf(\mathbf{z}) & \text{if} \quad -1 < f(\mathbf{z}) < 1 \\ D + Df(\mathbf{z}) & \text{if} \quad 1 < f(\mathbf{z}) \end{cases} \tag{21}$$

According to (21), $I(f(\mathbf{z}))$ is linear in $f(\mathbf{z})$ and strictly non-negative in each considered interval. Note that $A, D > 0, \forall\, \mathbf{z} \in \mathcal{Z}$, so that the slopes of the first and the last stretches of $I(f(\mathbf{z}))$ are negative and positive, respectively. The slope of the second stretch can be positive or negative, depending of the values of A and D. It is easy to reason that the minimum is reached at $f(\mathbf{z}) = -1$ or $f(\mathbf{z}) = 1$, depending on the slope of $I(f(\mathbf{z}))$ in the interval $(-1, 1)$. Therefore, the function $f_0(\mathbf{z})$ which minimizes the functional of expression (16) is given by

$$f_0(\mathbf{z}) = \begin{cases} -1 & \text{if} \quad P(H_1)p(\mathbf{z}|H_1) < P(H_0)p(\mathbf{z}|H_0) \\ 1 & \text{if} \quad P(H_1)p(\mathbf{z}|H_1) > P(H_0)p(\mathbf{z}|H_0) \end{cases} \tag{22}$$

It means that after training, the C-SVM will provide an output close to -1 for those input vectors for which $P(H_1)p(\mathbf{z}|H_1) < P(H_0)p(\mathbf{z}|H_0)$, and an output close to 1 for those input vectors for which $P(H_1)p(\mathbf{z}|H_1) > P(H_0)p(\mathbf{z}|H_0)$. If a decision is taken in favor of hypothesis H_0 when the output is close to -1 and in favor of H_1 when the output is close to 1, the Maximum-A-Posteriori (MAP) classifier is implemented, that is equivalent to the Minimum Probability of Error classifier in the binary case. The decision rule of the Minimum Probability of Error classifier is:

$$P(H_1)p(\mathbf{z}|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P(H_0)p(\mathbf{z}|H_0) \tag{23}$$

This result is equivalent to the one obtained in [20] for the Minkowski Error with $R = 1$, and it is interpreted as follows: since the function approximated by the C-SVM is binary, the selection of the threshold in order to achieve a required P_{FA} is not possible, concluding that this learning machine is not suitable to approximate the NP detector, and can only approximate the Minimum Probability of Error classifier.

4. 2C-SVMs capability to implement discriminant functions useful to approximate the NP detector

It has been demonstrated in Section 3 that C-SVM does not allow implementing the Neyman–Pearson detector by just modifying the detection threshold, or some parameters of the objective function. To achieve this objective, the use of 2C-SVMs is proposed. The objective of this section is to find the function $f_0(\mathbf{z})$ which minimizes the unconstrained objective function defined in expression (14), supposing again that the conditions for perfect training are fulfilled.

The number of training patterns is supposed high $(N \to \infty)$ and representative (patterns are generated according to the likelihood functions $p(\mathbf{z}|H_i)$). In this case, the training process is equivalent to the following optimization problem:

$$\min_f\left\{\lim_{N\to\infty}\left\{\frac{1}{2N}\parallel \mathbf{w}\parallel^2+\frac{C}{N}\gamma\sum_{i\in\mathcal{X}_0}\left(1-d_i f\left(\mathbf{z}_i\right)\right)u\left(1-d_i f\left(\mathbf{z}_i\right)\right)\right.\right.$$
$$\left.\left.+\frac{C}{N}\left(1-\gamma\right)\sum_{i\in\mathcal{X}_1}\left(1-d_i f\left(\mathbf{z}_i\right)\right)u\left(1-d_i f\left(\mathbf{z}_i\right)\right)\right\}\right\} \tag{24}$$

Also assume, without loss of generality, that the desired outputs of the learning machine are $d_i=-1$ and $d_i=1$, for $\mathbf{z}_i\in\mathcal{X}_0$ and $\mathbf{z}_i\in\mathcal{X}_1$, respectively. Applying the Strong Law of Large Numbers to (24), the optimization problem can be formulated with expression (25):

$$\min_f\left\{C\int_{\mathcal{Z}}\left(\gamma P(H_0)p(\mathbf{z}|H_0)(1+f(\mathbf{z}))u(1+f(\mathbf{z}))\right.\right.$$
$$\left.\left.+(1-\gamma)P(H_1)p(\mathbf{z}|H_1)(1-f(\mathbf{z}))u(1-f(\mathbf{z}))\right)d\mathbf{z}\right\} \tag{25}$$

The functional to be minimized is expressed in (26), where constant C has not been considered, because it does not have any influence on the solution of the minimization process:

$$\int_{\mathcal{Z}} I\left(f\left(\mathbf{z}\right)\right)d\mathbf{z}$$
$$=\int_{\mathcal{Z}}\left(\gamma P(H_0)p(\mathbf{z}|H_0)(1+f(\mathbf{z}))u(1+f(\mathbf{z}))\right.$$
$$\left.+(1-\gamma)P(H_1)p(\mathbf{z}|H_1)(1-f(\mathbf{z}))u(1-f(\mathbf{z}))\right)d\mathbf{z} \tag{26}$$

$I(f(\mathbf{z}))$ depends on the Heaviside step function, making the application of the Euler–Lagrange equation unfeasible. Following a procedure similar to the one applied in Section 3, $I(f(\mathbf{z}))$ is expressed as a piece-wise function:

$$\begin{cases}(1-\gamma)P(H_1)p(\mathbf{z}|H_1)(1-f(\mathbf{z})) & \text{if} \quad f(\mathbf{z})<-1\\ \gamma P(H_0)p(\mathbf{z}|H_0)(1+f(\mathbf{z}))+ & \\ +(1-\gamma)P(H_1)p(\mathbf{z}|H_1)(1-f(\mathbf{z})) & \text{if} \quad -1<f(\mathbf{z})<1\\ \gamma P(H_0)p(\mathbf{z}|H_0)(1+f(\mathbf{z})) & \text{if} \quad 1<f(\mathbf{z})\end{cases} \tag{27}$$

which can also be expressed as (28) to be analyzed easily:

$$I(f(\mathbf{z}))=\begin{cases}A-Af(\mathbf{z}) & \text{if} \quad f(\mathbf{z})<-1\\ B+Cf(\mathbf{z}) & \text{if} \quad -1<f(\mathbf{z})<1\\ D+Df(\mathbf{z}) & \text{if} \quad 1<f(\mathbf{z})\end{cases} \tag{28}$$

In expression (28), constants A, B, C, and D have the following values: $A=(1-\gamma)P(H_1)p(\mathbf{z}|H_1)$, $D=\gamma P(H_0)p(\mathbf{z}|H_0)$, $B=D+A$, and $C=D-A$. Considering that $A, D>0, \forall\,\mathbf{z}\in\mathcal{Z}$, and following the same reasoning that was used in Section 3, the minimum of $I(f(\mathbf{z}))$, and therefore, the solution of the minimization problem formulated in expression (25) is:

$$f_0(\mathbf{z})=\begin{cases}-1 & \text{if} \quad (1-\gamma)P(H_1)p(\mathbf{z}|H_1)<\gamma P(H_0)p(\mathbf{z}|H_0)\\ 1 & \text{if} \quad (1-\gamma)P(H_1)p(\mathbf{z}|H_1)>\gamma P(H_0)p(\mathbf{z}|H_0)\end{cases} \tag{29}$$

After training, the 2C-SVM will provide outputs close to +1 or −1, depending on the conditions expressed in (29). If the output of the 2C-SVM is compared to a threshold $\eta=0$, the intermediate value between +1 and −1, the decision rule is equivalent to:

$$\frac{p(\mathbf{z}|H_1)}{p(\mathbf{z}|H_0)}\underset{H_0}{\overset{H_1}{\gtrless}}\frac{\gamma P(H_0)}{(1-\gamma)P(H_1)} \tag{30}$$

Varying the value of γ we can select different thresholds of the likelihood ratio based detector. These thresholds will be denoted η_{lr}. By varying $\eta_{lr}=\frac{\gamma P(H_0)}{(1-\gamma)P(H_1)}$, we can implement detectors with pairs (P_{FA}, P_D) corresponding to different points of the Receiver Operating Characteristic (ROC) curve of the Neyman–Pearson detector.

The implemented decision rule will maintain the constant false alarm property, provided that the likelihood functions and the priors of both hypotheses in the operating environment are equal to those considered in the design process (training sets).

In conventional radar detectors, Constant False Alarm Rate (CFAR) techniques are implemented for adaptively adjusting the detection threshold using estimations of the interference parameters, assuming a specific model of $p(\mathbf{z}|H_0)$. In our case, the function that is implemented by the 2C-SVM after training is binary, and there is not any parameter to be adjusted. Therefore, in a first approach to implement CFAR techniques, assuming a specific model of $p(\mathbf{z}|H_0)$ during training, a robustness study of the 2C-SVM based detector could be carried out by varying the parameters of the model in the test set, to identify the parameters variation interval where P_{FA} mismatch is bearable. After that, a bank of 2C-SVMs could be trained, one for each identified variation interval.

5. Experiments and results

The results of some experiments are presented in this section to demonstrate the capability of 2C-SVMs to approximate the optimum Neyman–Pearson detector. Different γ values are chosen to select different P_{FA}, and the objective is to confirm that the obtained P_D is close to the maximum value of the NP detector.

Some detection problems have been selected, for which the NP detector can be analytically formulated and used to be compared with the 2C-SVM-based detectors. More precisely, the detection of Swerling I (SW-I) and Swerling II (SW-II) targets [39] in Additive White Gaussian Noise (AWGN), and Swerling V (SW-V) targets with unknown Doppler shift in non-Gaussian interference, are considered [40]. The study is completed with results of detectors using real radar data of a public database.

For testing the detectors, P_{FA} and P_D values have been estimated using Monte-Carlo simulations with a number of test vectors high enough to guarantee a relative estimation error lower than 10%. Furthermore, a study of the influence of the training patterns number on the detection performance is carried out. Because of that, the polynomial kernel of expression (31) with $d=2$ has been selected, due to the better performance compared to radial basis for large training sets [41,42]:

$$G(\mathbf{z}_i,\mathbf{z}_j)=(1+\langle\mathbf{z}_i,\mathbf{z}_j\rangle)^d \tag{31}$$

$$\langle\mathbf{z}_i,\mathbf{z}_j\rangle=\sum_{l=1}^{L}z_{i,l}z_{j,l} \tag{32}$$

5.1. Detection of SW-I and SW-II targets in AWGN

In the experiments carried out in this subsection, the considered input vector $(\tilde{\mathbf{z}})$ consists of P complex samples, which are echoes provided by a pulse coherent radar. Echoes can represent SW-I or SW-II targets in AWGN (hypothesis H_1), or only AWGN (hypothesis H_0), and a value of $P=8$ has been used. The hypothesis are considered equiprobable for generating the training sets $(P(H_0)=P(H_1)=0.5)$.

SW-I and SW-II models have been selected because the formulation of the NP detector is known, and can be used for comparison purposes. The decision rules of the NP detector for SW-I and SW-II targets in AWGN are expressed in (33) and (34), respectively [40]:

$$\left(\sum_{i=1}^{P}\mathcal{R}\mathrm{e}\left(\widetilde{z}_i\right)\right)^2+\left(\sum_{i=1}^{P}\mathcal{I}\mathrm{m}\left(\widetilde{z}_i\right)\right)^2\underset{H_0}{\overset{H_1}{\gtrless}}\eta_s \tag{33}$$

$$\sum_{i=1}^{P}|\widetilde{z}_i|^2\underset{H_0}{\overset{H_1}{\gtrless}}\eta_s \tag{34}$$

where

- $\widetilde{z}_i$ are the complex-valued echoes of the input vector.
- $\eta_s(P_{FA})$ represents the threshold of the decision rules obtained after simplifying the likelihood ratio based detector. The threshold η_s is related to η_{lr}, and to the parameter γ used to adjust the values of P_{FA} and P_D.

2C-SVM based detectors have been designed to obtain predetermined values of P_{FA}, selected by varying the value of γ. To analyze the performance for a wide range of P_{FA} values, we have selected the set $P_{FA}\in\left\{10^{-2},10^{-3},10^{-4},10^{-6}\right\}$. The η_{lr}, γ, and P_D for those P_{FA} values when $SNR=7$ dB are presented in Table 1.

The resulting 2C-SVMs have been tested to know the actual P_D and P_{FA}, in order to determine how close the implemented detectors are to the NP optimum detectors. The results are presented in Tables 2 and 3. Although 2C-SVMs have the theoretical capability of approximating any desired point of the NP detector if the training sets have infinite number of elements, a finite number of patterns (N) has to be used. It is expected that the higher N, the better the approximation to the NP detector.

Results show that detectors based on 2C-SVMs trained with a γ value fixed according to P_{FA} requirements, are able to approximate the associated point of the NP detector ROC curve with high accuracy. Furthermore, the higher the number of patterns in the training set, the better the approximation, as expected from the theoretical analysis of Section 4 (a better approximation to the NP detector is obtained when the actual P_{FA} is closer to the theoretical one, and the P_D is higher).

5.2. Detection of SW-V targets with unknown Doppler shift in K-distributed clutter

This detection problem has been selected because the relation between γ and P_{FA} cannot be easily determined, being a good example to show how a general radar detection problem could be solved with 2C-SVMs. In general problems, the relation between η_{lr} and P_{FA} is unknown a priori, and the most suitable γ value would be determined by a grid search in a suitable interval.

The value of the regularization parameter C is also important, because if only a low number of training patterns is available, the regularization parameter should be high, in order to play down the importance of the regularization term in expression (24). On the contrary, if the number of training patterns is very high, the regularization parameter value has less importance. Usually, a grid search in the (C,γ) space could be carried out, training the 2C-SVM with each pair (C,γ), and evaluating its performance on a held-out validation set. Finally, the grid search algorithm outputs the settings that achieves the highest score in the validation procedure.

In this section, the design of robust detectors for SW-V targets with unknown Doppler shift (Ω) in spiky K-distributed clutter has been considered [4,5]. In this case study, the Average Likelihood Ratio (ALR) detector, which is expressed in (35), depends on an integral which involves the calculation of a modified Bessel function of the second kind:

$$\frac{\int_0^{2\pi}a(\Omega)^{\nu-P}K_{\nu-P}(a(\Omega))d\Omega}{b^{\nu-P}K_{\nu-P}(b)}\underset{H_0}{\overset{H_1}{\gtrless}}\eta_{ALR}(P_{FA}) \tag{35}$$

where

- a and b are defined in the following expressions:

$$a=\left(4\nu\left(\left(\widetilde{\mathbf{z}}-10^{\frac{SCR}{10}}\cdot\exp\left(j\left(\theta_R+\Omega\right)\right)\right)^T\boldsymbol{\Sigma}_{\widetilde{\mathbf{c}}\widetilde{\mathbf{c}}}^{-1}\cdot\left(\widetilde{\mathbf{z}}-10^{\frac{SCR}{10}}\cdot\exp\left(j\left(\theta_R+\Omega\right)\right)\right)^*\right)\right)^{\frac{1}{2}} \tag{36}$$

$$b=\sqrt{4\nu(\widetilde{\mathbf{z}}^T\boldsymbol{\Sigma}_{\widetilde{\mathbf{c}}\widetilde{\mathbf{c}}}^{-1}\widetilde{\mathbf{z}}^*)} \tag{37}$$

- P is the number of target echoes collected in a scan.
- ν is the shape parameter of the K-distribution, which is set to 0.5 in this section.
- $(\boldsymbol{\Sigma}_{\widetilde{\mathbf{c}}\widetilde{\mathbf{c}}})_{h,k}=\rho_c^{|h-k|^2}(h,k\in\{1,2,\ldots,P\})$; ρ_c is the clutter one-lag correlation coefficient, and clutter power is set to unity as a normalization criterion.
- SCR is the Signal to Clutter Ratio expressed in dB.
- θ_R is the received phase of the target echo.
- Ω is the unknown Doppler shift, that is modeled as an uniform random variable in the interval $[\Omega_1,\Omega_2]$.

Table 1
η_{lr} and γ values associated with $P_{FA}\in\{10^{-2},10^{-3},10^{-4},10^{-6}\}$ for Swerling I and II in AWGN. The probabilities of detection of the NP detectors $SNR=7$ dB is also presented.

P_{FA}	Parameter	Target model	
		Swerling I	Swerling II
10^{-2}	η_{lr}	2.1674	0.3638
	γ	0.6843	0.2667
	P_D	0.8944	0.9936
10^{-3}	η_{lr}	20.4604	7.4175
	γ	0.9534	0.8812
	P_D	0.8455	0.9815
10^{-4}	η_{lr}	197.1525	125.5957
	γ	0.9950	0.9921
	P_D	0.7988	0.9585
10^{-6}	η_{lr}	$2.0038\cdot10^4$	$2.5036\cdot10^4$
	γ	0.99995	0.99996
	P_D	0.7135	0.8776

Table 2
P_D ($SNR=7$ dB) values obtained by 2C-SVMs based detectors for Swerling I and II in AWGN, trained with the γ parameters determined in Table 1.

Desired P_{FA}	Regularization parameter C	Swerling I		Swerling II	
		$N=200$	$N=1000$	$N=200$	$N=1000$
10^{-2}	$C=10^3$	0.8885	0.8909	0.9892	0.9916
	$C=10^5$	0.8930	0.8936	0.9916	0.9933
10^{-3}	$C=10^3$	0.8428	0.8433	0.9726	0.9794
	$C=10^5$	0.8431	0.8452	0.9770	0.9812
10^{-4}	$C=10^3$	0.7919	0.7958	0.9325	0.9478
	$C=10^5$	0.7977	0.7982	0.9328	0.9578
10^{-6}	$C=10^3$	0.6488	0.6958	0.7995	0.8618
	$C=10^5$	0.6907	0.7055	0.8547	0.8710

Table 3
P_{FA} values obtained by 2C-SVMs based detectors for Swerling I and II in AWGN, trained with the γ parameters determined in Table 1.

Desired P_{FA}	Regularization parameter C	Swerling II		Swerling I	
		$N=200$	$N=1000$	$N=200$	$N=1000$
10^{-2}	$C=10^3$	$1.2756\cdot10^{-2}$	$1.1229\cdot10^{-2}$	$2.7617\cdot10^{-2}$	$1.9478\cdot10^{-2}$
	$C=10^5$	$1.0189\cdot10^{-2}$	$1.0074\cdot10^{-2}$	$1.1795\cdot10^{-2}$	$1.0510\cdot10^{-2}$
10^{-3}	$C=10^3$	$1.1470\cdot10^{-3}$	$1.1400\cdot10^{-3}$	$2.6970\cdot10^{-3}$	$1.8420\cdot10^{-3}$
	$C=10^5$	$1.1300\cdot10^{-3}$	$1.0043\cdot10^{-3}$	$1.2470\cdot10^{-3}$	$1.0302\cdot10^{-3}$
10^{-4}	$C=10^3$	$1.2200\cdot10^{-4}$	$0.9700\cdot10^{-4}$	$2.7800\cdot10^{-4}$	$1.6000\cdot10^{-4}$
	$C=10^5$	$0.9600\cdot10^{-4}$	$1.0001\cdot10^{-4}$	$1.504\cdot10^{-4}$	$1.0047\cdot10^{-4}$
10^{-6}	$C=10^3$	$2.8443\cdot10^{-6}$	$1.3000\cdot10^{-6}$	$3.4100\cdot10^{-6}$	$1.8547\cdot10^{-6}$
	$C=10^5$	$1.9701\cdot10^{-6}$	$1.1022\cdot10^{-6}$	$2.5087\cdot10^{-6}$	$1.2010\cdot10^{-6}$

The ALR approach usually gives rise to integrals without analytical solution. As an alternative, sub-optimum solutions based on the Constrained Generalized Likelihood Ratio (CGLR) are usually used. In both cases, the estimation of the detection threshold for the desired P_{FA} can be really difficult, requiring the application of numerical integration techniques, such as Monte-Carlo simulations. In addition, the estimated threshold will be valid only for the assumed parameters of the clutter model. In radar scenarios with variable clutter parameters, a table of detection thresholds versus P_{FA}, for each considered parameter value, should be created. A CFAR technique based on the estimation of clutter parameters and the selection of the detection threshold from the table could be designed.

In this paper, the sub-optimum detector based on the CGLR has been used as reference, in order to make comparison with the proposed 2C-SVM based detector easier. It is expressed in (38):

$$\max_{\Omega_k}\Lambda\left(\Omega_k\right)\underset{H_0}{\overset{H_1}{\gtrless}}\eta_{CGLR}\left(P_{FA}\right),\quad k=1,\ldots,K \tag{38}$$

where K is the finite number of LR-based detectors, designed for equispaced discrete values in the variation range of Ω, and the corresponding CGLR detector will be denoted $CGLR_K$. Assuming $P=8$, $\theta_R=\pi/4$, and Ω is uniformly distributed in $[\pi/2-\pi/10;\ \pi/2+\pi/10]$, ROC curves of $CGLR_K$ detectors in uncorrelated and correlated ($\rho_c=0.9$) clutter have been obtained. The Ω variation interval has been proposed in order to obtain useful radar coverages and SCR is set to 9 dB and −3 dB, for uncorrelated and correlated clutter respectively, to obtain P_D higher than 80% for P_{FA} of interest in radar applications.

In Fig. 2, the ROC curves of $CGLR_k$ detectors for different values of k are depicted. It is showed that the $CGLR_P$ detector obtains almost the best results for SW-V targets in uncorrelated clutter. However, the $CGLR_{2P}$ detector is a good choice for SW-V targets in correlated K-distributed clutter (Fig. 3). Therefore, the $CGLR_P$ and $CGLR_{2P}$ are chosen for comparison purposes, for detecting SW-V targets in uncorrelated and correlated clutter, respectively.

In order to design the 2C-SVMs based solutions, a grid search in the C and γ space has been carried out to determine the values that guarantee a desired P_{FA}. The 2C-SVMs have been trained with $N=500$ training patterns, which is a compromise between detection performance and computational cost. The detectors are implemented by comparing the 2C-SVMs output to a fixed threshold, $\eta_0=0$. The estimated P_{FA} values are presented in logarithmic scale for the different pairs (γ, C) in Figs. 4 and 5, where the detection of SW-V targets in uncorrelated spiky K-distributed clutter, and in correlated K-distributed clutter are considered, respectively. In both cases, results confirm that the higher the γ, the lower the P_{FA}.

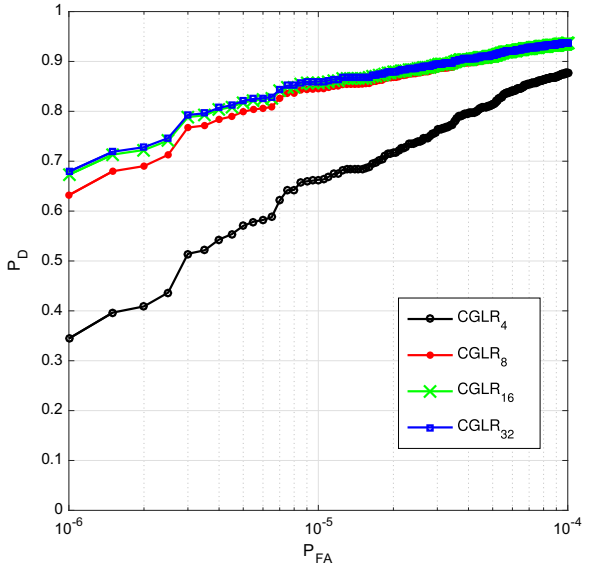

Fig. 2. ROC curves of $CGLR_K$ detector for SW-V targets with SCR = 9 dB, $\theta_R=\pi/4$ and Ω uniformly distributed in $[\pi/2-\pi/10;\ \pi/2+\pi/10]$, in spiky uncorrelated K-distributed clutter.

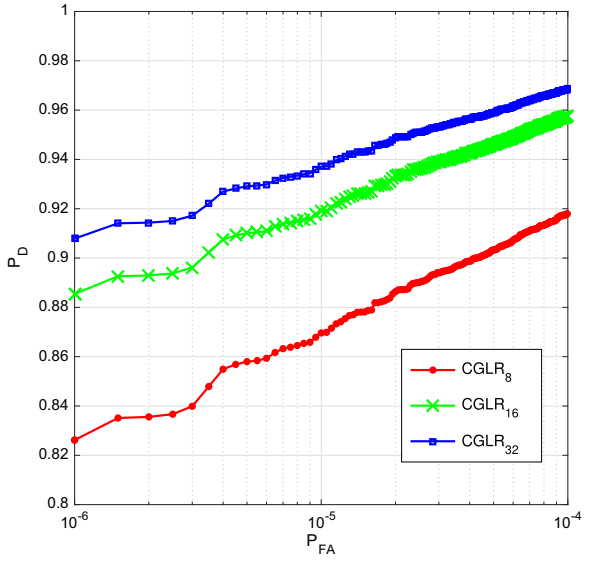

Fig. 3. ROC curves of $CGLR_K$ detector for SW-V targets with $SCR=-3$ dB, $\theta_R=\pi/4$ and Ω uniformly distributed in $[\pi/2-\pi/10;\ \pi/2+\pi/10]$, in spiky correlated ($\rho_c=0.9$) K-distributed clutter.

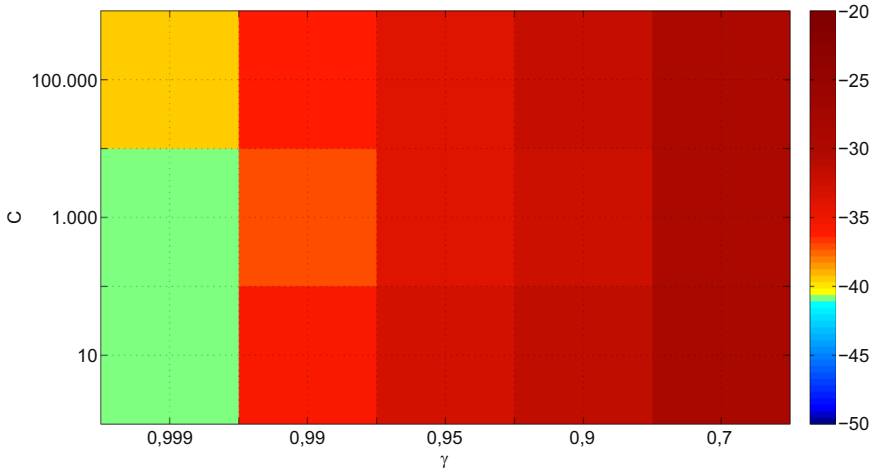

Fig. 4. $10\log_{10}(P_{FA})$ obtained with a grid search $[C;\ \gamma]$ of 2C-SVMs based detectors for SW-V targets (training $SCR = 9$ dB) in uncorrelated spiky K-distributed clutter.

5.3. Application to real radar data

In this section, examples of radar detection with real data using 2C-SVMs are presented. Data were acquired by a pulsed Doppler X-band radar deployed on Signal Hill by the Council for Scientific and Industrial Research (CSIR) [43]. The datasets used in this study are available to the international radar research community on [44].

Signal Hill location (Fig. 6) provided 140° azimuth coverage. A large sector spanned open sea whilst the remainder looked towards the West Coast coastline from the direction of the open sea. Grazing angles ranging from 10° at the coastline to 0.3° at the radar instrumented range coverage of 37.28 NM (Nautical Miles) were obtained. The pulse repetition frequency (PRF) was 2 kHz and the range resolution was 15 m. A collaborative 4.2 m inflatable rubber boat, that can be considered as a point target, was used during some measurements.

Datasets were recorded with different local wind conditions. The average wind speed varied between 0 knots and 40 knots and the significant sea wave height ranged between 1 m and 4.5 m. A file of this public database has been selected for the experiments. It has the reference *Dataset 10-104. TTrFA*, and the main parameters of the acquisition are summarized in Table 4 [44].

2C-SVMs have been trained for detection purposes, and the obtained results have been compared to the results of the reference detectors. For designing the reference detectors, a statistical characterization of real radar data has been carried out.

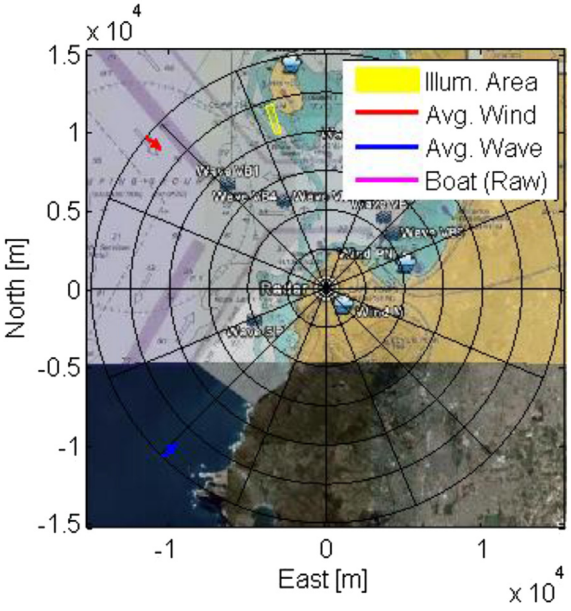

Fig. 6. Plan overview of radar deployment site [44].

5.3.1. Statistical characterization of real radar data

For a sea state higher than 2 (sea wave height higher than 0.5 m), low grazing angles and high resolutions, the K-distribution

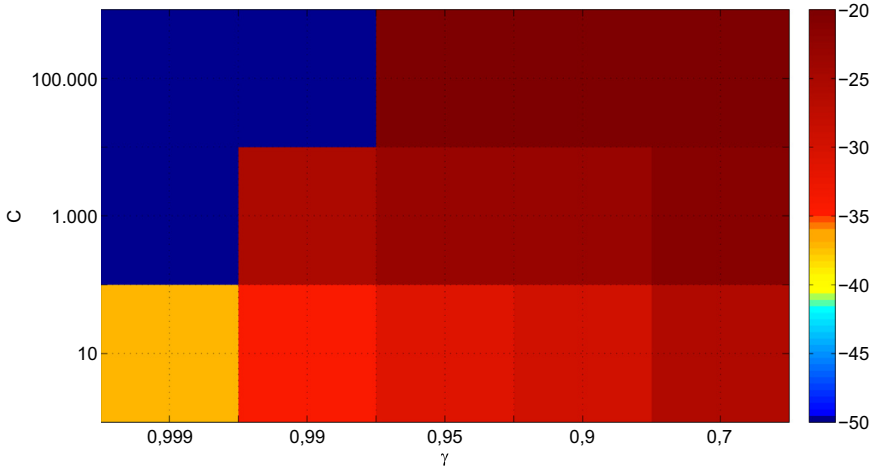

Fig. 5. $10\log_{10}(P_{FA})$ obtained with a grid search $[C;\ \gamma]$ of 2C-SVMs based detectors for SW-V targets (training $SCR = -3$ dB) in correlated ($\rho_c = 0.9$) spiky K-distributed clutter.

Table 4
Acquisition specifications of *Dataset 10-104. TTrFA* [44].

Acquisition time: 64.48 s (1984 patterns)	Range extend: 1903.7 m (128 gates)
Patterns of 64 complex samples Grazing angle: 0.93°–1.06° Tx frequency 8.8 GHz	Pattern repetition interval 0.0325 s Antenna azimuth: 354.8–357.6 N Significant wave height: 1.68 m, 244.8 N

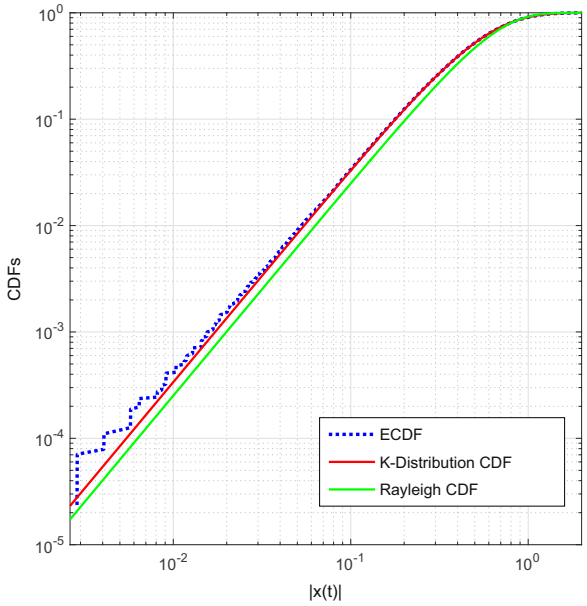

Fig. 7. Comparison of the empirical cumulative distribution function for the amplitude of 27th range gate of *Dataset 10-104.TTrFA*, with the *K*-distribution and Rayleigh cumulative distribution functions.

can be used for modeling sea clutter amplitude [6,45–47]. The significant wave height of the selected file is 1.68 m, corresponding to moderate (code 4) sea state. The grazing angles are low, and the resolution can be considered high. Under these conditions, *K*-Distribution can be used for modeling sea clutter amplitude.

The Empirical Cumulative Distribution Function (ECDF) of the samples of the 27th range gate (1630.47 m), associated with only sea clutter returns, has been estimated. This ECDF has been compared with the *K*-distribution and the Rayleigh Cumulative Distribution Functions (CDF), defined in expressions (39) and (40), respectively. The estimated and the theoretical CDFs for the amplitude of the data samples are shown in Fig. 7. The visual inspection shows a very good agreement with the K-Distribution (the ECDF and the *K*-distribution CDF overlap):

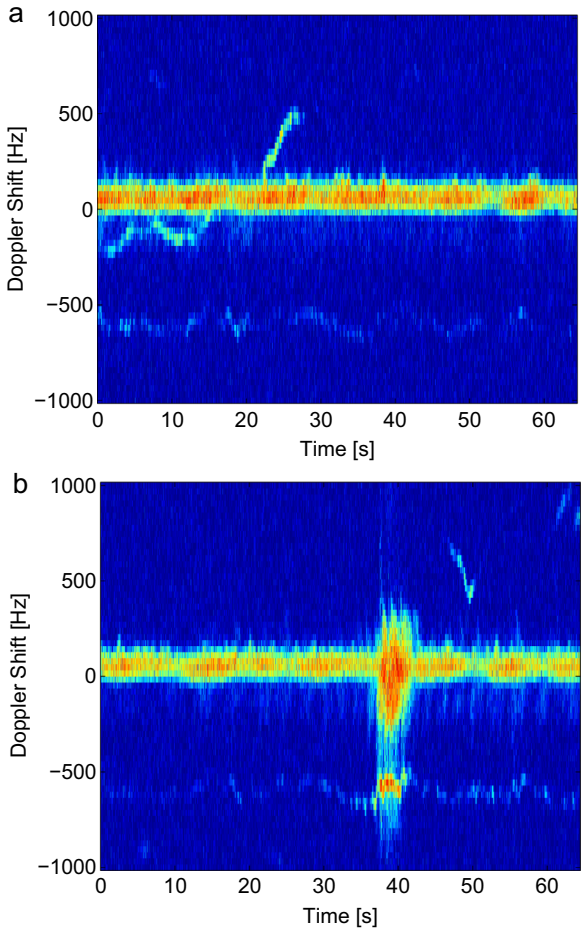

Fig. 9. Doppler Shift of 27th range gate (a) and 65th range gate (b) of *Dataset 10-104.TTrFA* (RCS [dB m^2/ Hz]).

$$F_x(x) = 1 - \exp^{-x^2/2\sigma^2}; \quad x \geq 0 \tag{39}$$

$$F_x(x) = 1 - \left(\frac{2}{2^{\nu} \cdot \Gamma(\nu)} \left(2 \cdot x \sqrt{\frac{\nu}{\mu}} \right)^{\nu} \cdot K_{\nu} \left(2 \cdot x \sqrt{\frac{\nu}{\mu}} \right) \right); \quad x \geq 0 \tag{40}$$

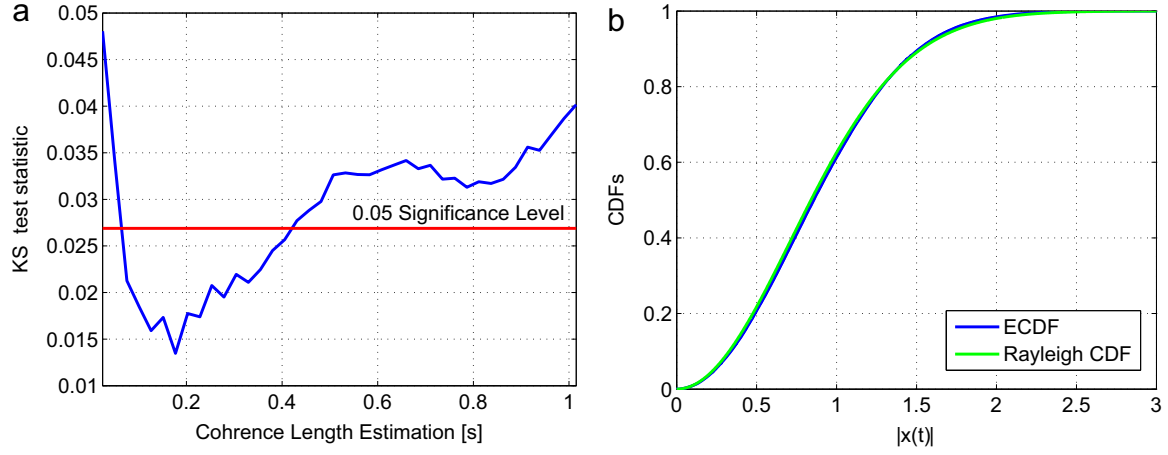

Fig. 8. KS-test statistic of the texture for different L_c (left) and CDFs for the amplitude of speckle component (L_c = 0.18 s(right)).

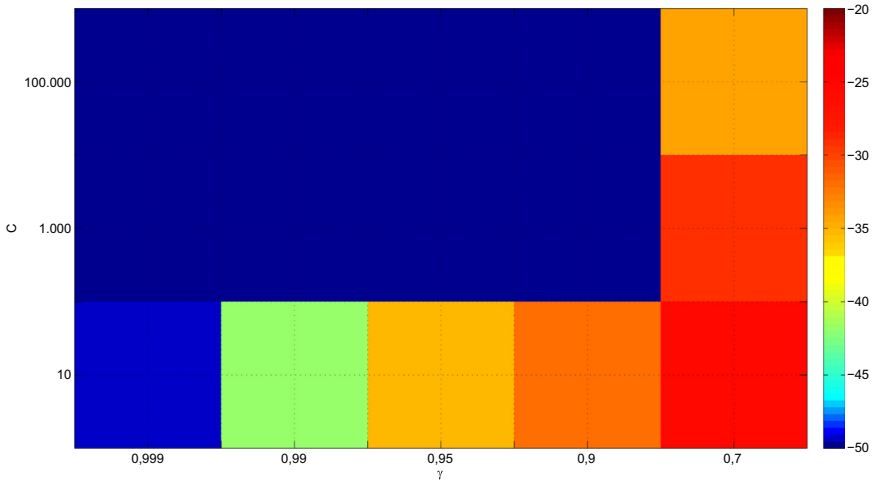

Fig. 10. 10log(P_{FA}) resulted from a grid search (γ, C) of 2C-SVMs based detector designed for target and clutter models that fit to real radar data.

Table 5
Comparison of detection capabilities of 2C-SVM, $CGLR_{16}$ and envelope detectors for *Dataset 08-028. TStFA.*

Detector				P_{FA}	P_D
Envelope detector				$1.0244{\cdot}10^{-4}$	0.1750
$CGLR_{16}$				$1.0183{\cdot}10^{-4}$	0.5101
SVM	C	γ	N_{sv}		
	10	0.99	969	$1.1024{\cdot}10^{-4}$	0.4879
	10	0.95	989	$20.1065{\cdot}10^{-4}$	0.8372
	10^5	0.7	997	$23.3584{\cdot}10^{-4}$	0.5111

To increase the rigor of our study, a goodness-of-fit test has also been carried out to assess whether the proposed distributions are suited to the dataset. The Kolmogorov–Smirnov (KS) statistic has been used, which quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. We have obtained that the K-distribution with shape parameter $\nu = 3.9357$ and scale parameter $\mu = 0.3969$ is the one that best fits with the data.

The K-distribution is formed by compounding two separate probability distributions, one representing the radar cross-section (RCS) and the other representing speckle. The component representing the RCS is a slowly varying non-negative Gamma process that introduces a power modulation of the local backscatter, consequence of longer wavelength sea waves (texture), $\tau[n]$. The component representing speckle is modeled as a complex Gaussian random process, $g[n]$. As the power modulation is slower than the speckle component, it is possible to approximate the received clutter sequence by the following expression:

$$z_k[n] = \sqrt{\tau[k]}\,g[n], \quad n = k, \ldots, k + L_c - 1 \tag{41}$$

where L_c is the coherence length of sea texture, defined as the number of samples for which the texture can be considered constant. Expressions (42) and (43) can be used to estimate the texture and the hypothetical speckle sequences, respectively, for each possible L_c:

$$\tilde{\tau}[n] = \frac{1}{L_c} \sum_{k=n-\frac{L_c}{2}}^{k=n+\frac{L_c}{2}-1} |z_k[n]|^2 \tag{42}$$

$$\tilde{g}[n] = \frac{z[n]}{\sqrt{\tilde{\tau}[n]}} \tag{43}$$

In Fig. 8, the KS-test values, obtained by comparing the texture sequence CDF to a gamma distribution for different L_c are presented. The threshold of a 5% significance level is also depicted. It is possible to conclude that the coherence length of sea texture is between 0.08 and 0.43 s (the distance determined by the KS-test is below the threshold). As the 64-pulses-patterns time is 0.0325 s, a constant texture can be considered in each pattern. Additionally, the real and imaginary parts fulfill the Jarque–Bera goodness-of-fit test (it is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution), so the Gaussian distribution fits the real and imaginary parts of speckle (the magnitude is Rayleigh). The speckle amplitude ECDF is depicted on the right side of Fig. 8.

The Autocorrelation Function (ACF) of the pulses associated to the same pattern was also studied. The estimated one-lag correlation coefficient is equal to 0.966. In Fig. 9, the spectra for the 27th and 65th range gates are presented. 27th range gate is associated with sea clutter correlated returns that are localized close to the zero Doppler. 65th range cells corresponds to target echoes which presented a normalized Doppler shift frequency variable $\Omega \in [-0.6; 0.6]$.

After that, CGLR and SVM detectors were designed for detecting SW-V targets with unknown Doppler shift in K-distributed clutter. In Section 5.2, the analytical expression of the CGLR in the case study is detailed and, using the same parameters, P=8 integrated pulses were considered. The estimated target power and angle were p_s=2.0068 and $\theta_R = 0.0157$ rad, respectively. As $p_c = \mu = 0.3969$, SCR equal to 7.04 dB was considered. According to the available data in 1984 patterns and 128 range cells, a $P_{FA} = 10^{-4}$, with estimation error lower than 20% was expected.

A grid search in the C and γ space was carried out to design the 2C-SVM detector. The associated P_{FA} were estimated using the same fixed threshold, $\eta_0 = 0$. Results are depicted in logarithmic units in Fig. 10. The detection performances of the 2C-SVMs designed with the selected pairs (γ, C) with the real radar database are compared with the P_D obtained with the $CGLR_{16}$ and the envelope detectors for $P_{FA} = 10^{-4}$. The results are presented in Table 5.

The detectors based on CGLR16 and 2C-SVM approximate the Neyman–Pearson detector. Therefore, the objective is to maximize the P_D for a given P_{FA}. They obtain higher P_D than the envelope detector for almost the same value of P_{FA}. Thus, they present a

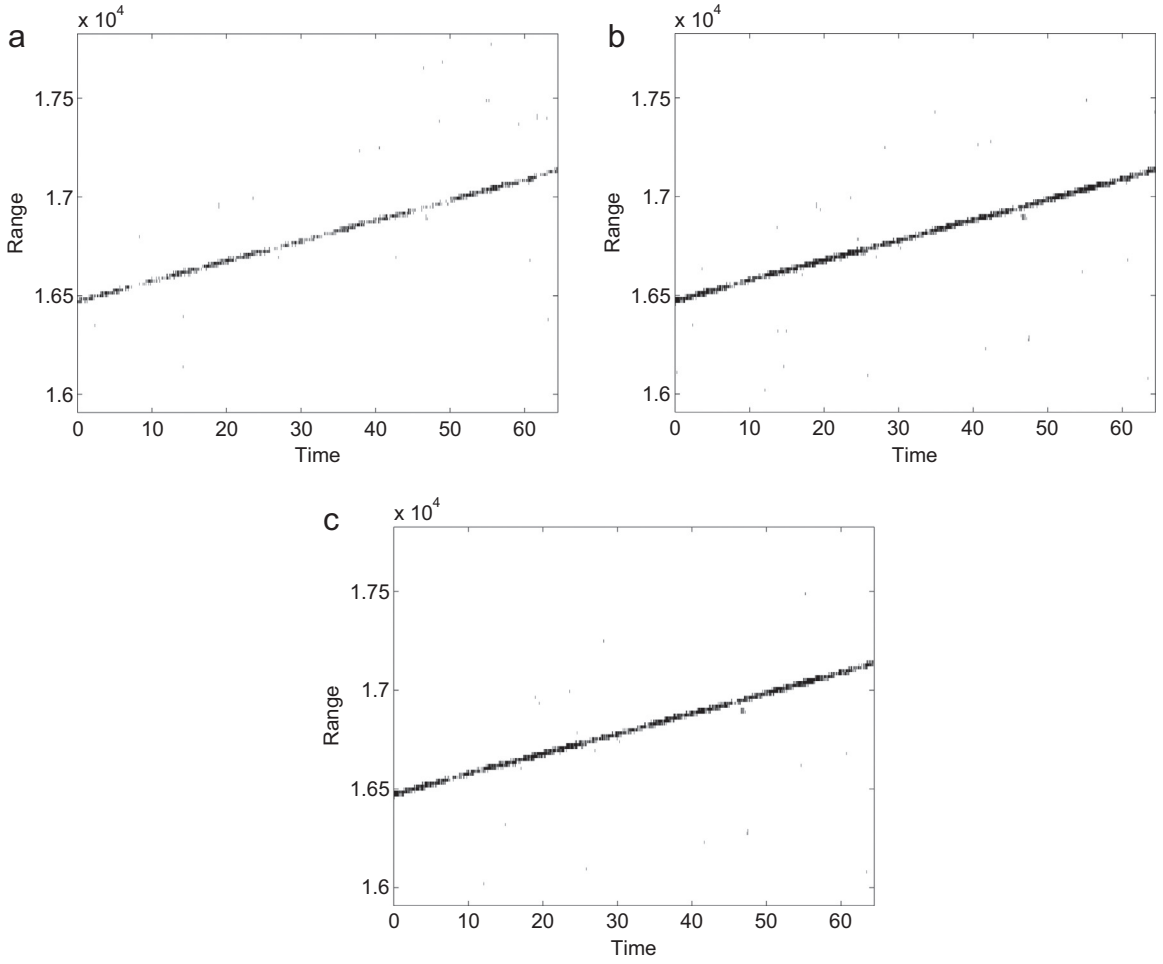

Fig. 11. Estimated centroids for the real radar matrix: envelope detector (a), $CGLR_{16}$ detector (b) and 2C-SVM designed with ($\gamma = 0.99$, $C = 10$) (c) detector.

significant improvement with respect to the envelope detector in the NP sense.

The 2C-SVM based detector designed with ($\gamma = 0.99$, $C = 10$) offers similar performance to the $CGLR_{16}$ detector for the P_{FA} closest to the desired one. Fig. 11 shows the considered detection schemes outputs for *Dataset 08-028. TStFA*. The estimated centroids are depicted and the target trajectory is clearly detected by the $CGLR_{16}$ and 2C-SVM detectors.

6. Conclusion

This paper deals with the capability of SVM-based radar detectors to approximate the NP detector. Supervised Learning machines can approximate the NP detector if they are trained with a suitable error function, such as the mean-squared error or the cross-entropy error [20]. In practice, for a good approximation to be obtained, 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 make the error function sufficiently small, and the learning algorithm must be able to find the appropriate minimum of the error function. But, from the theoretical point of view, the training error function has been demonstrated to be the key in order to obtain good approximations.

This paper extends the study to C-SVMs and 2C-SVMs, as learning machines that could be used to implement detectors close to the optimum NP one. The theoretical study is based on obtaining the function the learning machine converges to after training, if a sufficiently large number of training patterns is used. Training is carried out by solving the unconstrained optimization problems defined in expressions (11) and (14), for the C-SVM and 2C-SVM, respectively.

To obtain the functions the learning-machines converge to after training, the calculus of variations is used. It has been demonstrated that the C-SVM implements a function with only two values, which are obtained depending on which one is higher, $P(H_1)p(\mathbf{z}|H_1)$ or $P(H_0)p(\mathbf{z}|H_0)$. If the output of the C-SVM is compared to a threshold, with the intermediate value between the possible outputs, the implemented detector is equivalent to the Maximum-A-Posteriori (MAP) classifier.

Similarly, the 2C-SVM implements a function with only two values, but they are obtained depending on which one is higher $(1-\gamma)P(H_1)p(\mathbf{z}|H_1)$ or $\gamma P(H_0)p(\mathbf{z}|H_0)$. In this case, the detector that is implemented when the output of the 2C-SVM is compared to a threshold with an intermediate value between the possible outputs, is equivalent to the Neyman–Pearson detector for a fixed pair (P_{FA}, P_D). The values of P_{FA} and P_D vary with the parameter γ, which can be used to select different points in the ROC curve of the NP detector.

Some experimental results have been presented to demonstrate the theoretical study. First, 2C-SVMs have been trained to detect Swerling I and Swerling II targets in AWGN. The study has been carried out for three different desired P_{FA}, 10^{-2}, 10^{-3}, and 10^{-4}. The results confirm that very good approximations to the NP detectors are obtained if a sufficiently high number of training patterns is used, and the regularization parameter C is high.

A more complex detection problem has also been studied, for

which obtaining the value of γ for the desired P_{FA} is not easy. A grid search in the C-γ space has been carried out, demonstrating that the higher the value of γ, the lower the P_{FA}. Selecting the value of γ to approximate $P_{FA} = 10^{-4}$, it is demonstrated that the higher the number of support vectors, and the higher the value of C, the better the approximation to the NP detector.

The study has been completed with some experiments with real radar data, from a public database. To compare the performance of 2C-SVM based detectors with reference detectors, a statistical model of the data have been obtained. The parameters of the model are used to design the reference detectors, and the main conclusion is that 2C-SVMs performances are quite similar to the best reference detectors, which approximate the optimum NP detector.

From this study, we conclude that 2C-SVMs are quite good systems to approximate the NP detector for a given point of the ROC curve. On the contrary, C-SVMs only can approximate the MAP classifier, which is the optimum classifier from the point of view of minimum probability of error, if the hypotheses prior probabilities are equal. This study must be completed in the future to take into consideration the different alternatives for the basis functions of the learning machines, and the computational cost comparison to classical detectors with worse performance.

Acknowledgment

This work has been partially funded by the Ministry of Economy and Competitiveness of Spain under Projects TEC2012-38701 and TEC2015-71148-R, and by the University of Alcalá, under Project CCG2013/EXP-092.

References

[1] J. Neyman, E.S. Pearson, On the problem of the most efficient test of statistical hypotheses, Philos. Trans. R. Soc. Lond.: Ser. A 231 (1933) 492–510.
[2] M.M. Nayebi, M.R. Aref, M.H. Bastani, Detection of coherent radar signals with unknown Doppler shift, IEE Proc.—Radar Sonar Navig. 143 (1996), 79–86.
[3] V. Aloisio, A. di Vito, G. Galati, Optimum detection of moderately fluctuating radar targets, IEE Proc.—Radar Sonar Navig. 141 (1994), 164–170.
[4] F.A. Pentini, A. Farina, F. Zirilli, Radar detection of targets located in a coherent K distributed clutter background, IEE Proc. F—Radar Signal Process. 139 (1992), 239–245.
[5] E. Conte, M. Lops, G. Ricci, Radar detection in K-distributed clutter, IEE Proc.—Radar Sonar Navig. 141 (1994), 116–118.
[6] K.J. Sangston, F. Gini, M.S. Greco, Coherent radar target detection in heavy-tailed compound-Gaussian clutter, IEEE Trans. Aerosp. Electron. Syst. 48 (2012) 64–77.
[7] Y. Li, G. Zhang, R.J. Doviak, Ground clutter detection using the statistical properties of signals received with a polarimetric radar, IEEE Trans. Signal Process. 62 (2014) 597–606.
[8] E. Jay, J.P. Ovarlez, D. Declercq, P. Duvaut, BORD: Bayesian optimum radar detector, Signal Process. 83 (2003) 1151–1162.
[9] Q. He, MIMO radar diversity with Neyman-Pearson signal detection in non-Gaussian circumstance with non-orthogonal waveforms, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2011, pp. 2764–2767.
[10] Q. He, R.S. Blum, Diversity gain for MIMO Neyman–Pearson signal detection, IEEE Trans. Signal Process. 59 (2011) 869–881 (ssss).
[11] G.V. Moustakides, G.H. Jajamovich, A. Tajer, X. Wang, Joint detection and estimation: optimum tests and applications, IEEE Trans. Inf. Theory 58 (2012) 4215–4229.
[12] 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 (2010) 1149–1158.
[13] J. Liang, Y. Huo, C. Mao, Multitarget detection in heterogeneous radar sensor network with energy constraint, Signal Process. 126 (2016) 141–148.
[14] T. Furon, J. Josse, S. Le Squin, Some theoretical aspects of watermarking detection, Proc. SPIE (The International Society for Optical Engineering) 6072 (2006), 60721G.1–60721G.12.
[15] R. Cogranne, C. Zitzmann, F. Retraint, I. Nikiforov, L. Fillatre, P. Cornu, Statistical detection of LSB matching using hypothesis testing theory. in: Lecture Notes in Computer Science, vol. 7692, 2013, pp. 46–62.
[16] 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 Deliv. 19 (2004) 497–504.
[17] S. Acharya, A. Fridman, P. Brennan, P. Juola, R. Greenstadt, M. Kam, User authentication through biometric sensors and decision fusion, in: 2013 47th Annual Conference on Information Sciences and Systems (CISS), March 2013, pp. 1–6.
[18] S. Seader, P. Tenenbaum, J.M. Jenkins, C.J. Burke, χ^2 Discriminators for transiting planet detection in Kepler data, Astrophys. J. Suppl. Ser. 206 (2013) 25.
[19] H.L. Van Trees, Detection, Estimation, and Modulation Theory, 1, Wiley, New York, 1968.
[20] P. 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 Process. 57 (2009) 4175–4181.
[21] P. Jarabo-Amores, D. Mata-Moya, R. Gil-Pita, M. Rosa-Zurera, Radar detection with the Neyman–Pearson criterion using Supervised-Learning-Machines trained with the cross-entropy error, EURASIP J. Adv. Signal Process. (2013), http://dx.doi.org/10.1186/1687–6180-2013-44.
[22] V. Ramamurti, S.S. Rao, P.P. Gandhi, Neural detectors for signals in non-Gaussian noise, in: Proceedings of the International Conference on Acoustic, Speech, and Signal Processing (ICASSP), vol. 1, April 1993, pp. 481–484.
[23] P. Gandhi, V. Ramamurti, Neural networks for signal detection in non-Gaussian noise, IEEE Trans. Signal Process. 45 (1997) 2846–2851.
[24] D. Munro, O. Ersoy, M. Bell, J. Sadowsky, Neural network learning of low-probability events, IEEE Trans. Aerosp. Electron. Syst. 32 (1996) 898–910.
[25] 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 (1990) 296–298.
[26] P. Jarabo-Amores, R. Gil-Pita, M. Rosa-Zurera, F. López-Ferreras, MLP and RBFN for detecting white gaussian signals in white Gaussian interference, in: Lecture Notes in Computer Sciences, vol. 2687, 2003, pp. 790–797.
[27] D. Mata-Moya, P. Jarabo-Amores, M. Rosa-Zurera, J.C. Nieto-Borge, F. López-Ferreras, Combining MLPs and RBFNNs to detect signals with unknown parameters, IEEE Trans. Instrum. Meas. 58 (2009) 2989–2995.
[28] D. Mata-Moya, P. Jarabo-Amores, J. Martin de Nicolás-Presa, High order neural network based solution for approximating the average likelihood ratio, in: Proceedings of IEEE Workshop on Statistical Signal Processing, June 2011, pp. 657–660.
[29] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[30] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, in: Data Mining Knowl. Discov. 2 (1998), 121–167.
[31] F. Bach, D. Heckerman, E. Horvitz, Considering cost asymmetry in learning classifiers, J. Mach. Learn. Res. 7 (2006) 1713–1741.
[32] Mark A. Davenport, Richard G. Baraniuk, Clayton D. Scott, Controlling false alarms with support vector machines, in: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, May 2006, pp. 589–592.
[33] M.A. Davenport, R.G. Baraniuk, C.D. Scott, Tuning support vector machines for minimax and Neyman–Pearson classification, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1888–1898.
[34] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, Cambridge, UK, 2004.
[35] P. Jarabo-Amores, M. Rosa-Zurera, R. Gil-Pita, F. López-Ferreras, Sufficient condition for an adaptive system to approximate the Neyman–Pearson detector, in: Proceedings of IEEE Workshop on Statistical Signal Processing, July 2005, pp. 295–300.
[36] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Winston, Washington, DC, USA, 1977.
[37] I. Gelfand, S. Fomin, Calculus of Variations, Courier Dover, Mineola, NY, 2000.
[38] R. Weinstock, Calculus of Variations With Applications to Physics and Engineering, McGraw-Hill, New York, 1952.
[39] P. Swerling, Probability of detection for fluctuating targets, IRE Trans. Inf. Theory 6 (1960) 269–308.
[40] M.I. Skolnik, Introduction to Radar Systems, McGraw-Hill, 2002.
[41] A. Afifi, E.A. Zanaty, S. Ghoniemy, Improving the classification accuracy using support vector machines (SVMs) with new kernel, J. Glob. Res. Comput. Sci. 4 (2013) 1–7.
[42] M.G. Genton, Classes of kernels for machine learning: a statistical perspective, J. Mach. Learn. Res. 2 (2002) 299–312.
[43] P.L. Herselman, C.J. Baker, H.J. De Wind, Analysis of X-band calibrated sea clutter and small boat reflectivity at medium-to-low grazing angles, Int. J. Navig. Obs. 2008 (2008) 1–14.
[44] Small boat detection research at the CSIR, ⟨http://www.csir.co.za/small_boat_detection⟩ (accessed 17.07.2016).
[45] M.I. Skolnik, Introduction to Radar Systems, third ed., McGraw-Hill Science, 2002.
[46] K.D. Ward, R.J. Tough, S. Watts, Sea clutter: scattering, the k-distribution and radar performance, Inst. Eng. Technol. (IET) (2006).
[47] A.M. McDonald, H.J. de Wind, J.E. Cilliers, P. Herselman, Performance prediction for a coherent X-band radar in a maritime environment with K-distributed sea clutter, in: Proceedings of IEEE International Radar Conference, 2010, pp.1208–1213.