

# Support vector machine under uncertainty: An application for hydroacoustic classification of fish-schools in Chile

Paul Bosch, Julio López, Héctor Ramírez, Hugo Robotham

► **To cite this version:**

Paul Bosch, Julio López, Héctor Ramírez, Hugo Robotham. Support vector machine under uncertainty: An application for hydroacoustic classification of fish-schools in Chile. *Expert Systems with Applications*, Elsevier, 2013, 40, pp.4029 - 4034. <10.1016/j.eswa.2013.01.006>. <hal-01088228>

**HAL Id: hal-01088228**

**<https://hal.inria.fr/hal-01088228>**

Submitted on 27 Nov 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Support vector machine under uncertainty: an application for hydroacoustic classification of fish-schools in Chile

Paul Bosch\*    Julio López\*    Héctor Ramírez†    Hugo Robotham\*

December 6, 2012

## Abstract

In this work we apply multi-class support vector machines (SVM) and a multi-class stochastic SVM formulation to the classification of fish schools of three species: anchovy, common sardine, and jack mackerel, and we compare their performance. The data used come from acoustic measurements in southern-central Chile. These classifications were carried out by using a diver set of descriptors including morphology, bathymetry, energy, and space positions. In both type of formulations, the deterministic and the stochastic one, the strategy used to classify multi-class SVM consists in employing the criterion one-species-against-the-Rest. We thus provide an empirical way to adjust the parameters involved in the stochastic classifiers with the aim of improving its performance. When this procedure is applied to the classification of fish schools we obtain a classifier with a better performance than the deterministic classifier.

**Key words:** Support vector machines, multi-class classification, robust chance constraints, second-order cone programming, species identification.

## 1 Introduction

Acoustics surveys are used to estimate the abundance and to study the behavior of many species. The last decades shows a great progress in the development of acoustics devices, and of new methods and techniques for automatically identifying species (see Horne (2000), [12] and Fernandes et al.(2006), [10]). The echograms obtained with acoustics devices provide information of size, location and echo intensity of fish schools. However, its classification is still not detectable by modern devices. Indeed, according to Horne (2000), the species identification is the big challenge of fisheries and plancton acoustics.

A wide range of statistical techniques have been used to classify mono-specific fish-school. Fernandes (2009)(see [11]) presents a complete list of studies including principal component analysis, discriminant-function analysis, artificial neural networks, nearest-neighbor analysis,  $k$ -means clustering and mixture models. Buelens et al. (2009) (see [4]) use kernel

---

\*Facultad de Ingeniería, Universidad Diego Portales, Ejército 441, Santiago, Chile (paul.bosch@udp.cl, julio.lopez@udp.cl, hugo.robatham@udp.cl).

†Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático (CNRS UMI 2807), FCFM, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile (hramirez@dim.uchile.cl).

methods to classify fish schools in single beam and multibeam acoustic data. In Robotham et al. (2010) (see [19]), support vector machine methods (tool for discriminating between two groups, see [8]) are used for the automatic classification of small pelagic fish species from acoustic surveys data. The results obtained by this approach was compared with an artificial neural networks methods.

Typically, the methods used to classify species of fishes require a scrutiny process of the echograms, the guide of expert criteria and additional information from trawl sampling (Simmonds and MacLennan (2005), see [22]). This procedure incorporates some grade of subjectivity. Thus, other approaches have been developed based only on the information provided by the echosounders. For instance, the identification of species based on school “descriptors parameters” from single-frequency acoustic data represents one type of approach. A second type of approach, which will not be exploded in this work, makes use of multifrequency acoustic data (Korneliussen et al. (2009), see [15]).

The purpose of this article is to compare the results of support vector machines (deterministic SVM) and classification when the uncertainty of data is incorporated for automatic acoustic identification of small pelagic fish species; anchovy (*Engraulis ringens*), common sardine (*Strangomera bentincki*) and jack mackerel (*Trachurus murphyi*) in southern-central Chile. The latter makes no precise assumption on the class-conditional densities with given mean and covariance matrix, that is, it is a “worst-case” setting.

## 1.1 Data collection and descriptors

School data were obtained from 11 acoustic assessment surveys performed with the R/V Abate Molina in northern and south-central Chile ( $18^{\circ}25'S-43^{\circ}50'S$ ) between 1991 and 2006. The data were collected using a scientific echosounder (SIMRAD EK-500) with a split-beam transducer (ES38 38 kHz) with a nominal -3 dB beam with of  $7^{\circ}$ , calibrated according to standard procedures (Foote et al., 1987). The ping rate of the echosounder in the surveys was 1 vs-1, the pulse duration was 1 ms and a minimum threshold of -65 dB. An Engel pelagic trawl with a 14-*m* vertical opening and 14-mm mesh size in the codend as used to identify the species in the acoustic survey. The flotation line of this net was adapted for fishing near the surface.

The acoustic records of the fish schools detected by the echosounder were processed with Echoview 3.0 software (Sonardata). Schools were identified by fishing hauls carry out using a midwater trawl. The parameters of the fish schools were determined automatically by the algorithm SHAPES programmed into the software Echoview and described in Barange (1994), and Coetzee (2000), (see [3] and [7], respectively). Each aggregation was manually marked in a region on the image of the echogram, and each case was individually analyzed. The parameters used were minimum candidate height= $1m$ , minimum candidate length = $1m$ , maximum vertical linking distance = $1m$ , and maximum horizontal linking distance = $15m$ .

The input data for a classification of fish-school is a collection of acoustics records. Each acoustics record, is characterized by a duple  $(x, y)$ , where  $x \in \mathbb{R}^n$  is the  $n$ -dimensional descriptors vector and  $y$  is a category (class) species which belong this descriptor. We used 12 descriptors for each school detected, which were grouped in four categories (see Scalabrín and Massé (1993), [21]):

1. Morphological: mean height ( $H$ ), length ( $L$ ) and perimeter ( $P$ ) with units in meter ( $m$ ); area ( $m^2$ ) and unidimensional descriptors:

- $Elongation = \frac{length}{mean\ height}$ .

- $Fractal\ dimension = 2 \frac{\ln 0.25P}{\ln area}$ .

2. Bathymetric: Bottom depth ( $D$ ) and mean school depth ( $D_m$ ) with units in meter and unidimensional descriptor:

- $School\ altitude\ index = 100 \frac{D-D_m}{D}$ .

3. Energetic: Acoustic energy ( $m^2/mn^2$ ) and acoustic density ( $dB$ ).

4. Space position: School-shore distance ( $mn$ ).

Some bathymetric and morphological descriptors are presented in Figure 1.

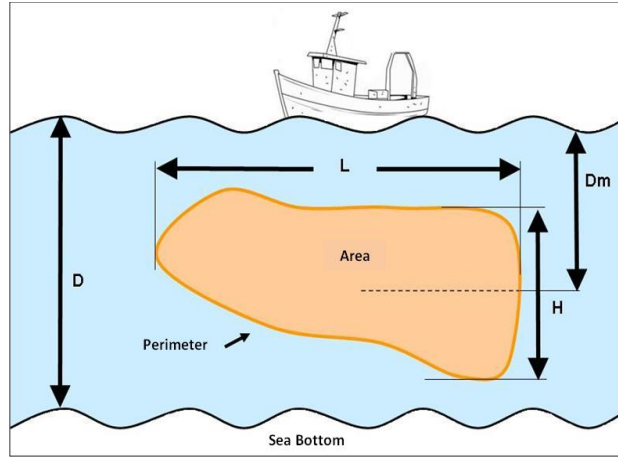


Figure 1: Representation of bathymetric and morphological descriptors

## 2 Support vector machines

In this section we describe the support vector machines (SVMs) developed by Cortes and Vapnik (1995), and second-order cone programming SVMs formulation based on the first two moments of each class, the mean and covariance (Nath and Bhattacharyya (2007); Shivaswamy et al. (2006)).

### 2.1 $l_2$ -Support vector machines

Support vector machines (SVMs) is a statistical classification method originally designed for binary classification. Given a set  $\mathcal{T} = \{(x_1, y_1), \dots, (x_m, y_m)\}$  of  $m$  training data, where  $x_i \in \mathbb{R}^n$  representing the  $i$ th training data and  $y_i \in \{-1, 1\}$  the class label of  $x_i$ , SVM

provides the optimal hyperplane  $f(x) = w^\top x - b$  that separates two classes. When the training data are linearly separable, this hyperplane separates two classes with no training error, and maximizes minimum distance from the training data to the hyperplane. In order to maximize this minimum distance, we need to classify correctly the vectors  $x_i$  of the training set into two different classes  $y_i$ , using the smallest norm of coefficients  $w$ . The maximum hyperplane problem can be formulated as the following Quadratic Programming (QP) optimization problem (see [8]):

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2 \\ & y_i(w^\top x_i - b) \geq 1, \quad i = 1, \dots, m. \end{aligned} \tag{2.1}$$

## 2.2 Support vector machines under uncertainty

Let  $\mathbf{X}_1$  and  $\mathbf{X}_2$  be two random vector variables associated to the positive and negative classes, respectively. In order to construct a maximum margin linear classifier such that the false-negative and false-positive error rates do not exceed  $\eta_1 \in (0, 1]$  and  $\eta_2 \in (0, 1]$ , respectively, we consider the following *Quadratic Chance-Constrained Programming* (QCCP) problem:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2 \\ & \Pr\{w^\top \mathbf{X}_1 - b \leq -1\} \leq \eta_1, \\ & \Pr\{w^\top \mathbf{X}_2 - b \geq 1\} \leq \eta_2. \end{aligned} \tag{2.2}$$

In other words, we require that the random variable  $\mathbf{X}_i$  lies on the correct side of the hyperplane with probability greater than  $1 - \eta_i$  for  $i = 1, 2$ . Assume that for  $i = 1, 2$  we only know the mean  $\mu_i \in \mathbb{R}^n$  and covariance matrix  $\Sigma_i \in \mathbb{R}^{n \times n}$  of the random vector  $\mathbf{X}_i$ . In this case, for each  $i = 1, 2$  we want to be able to classify correctly, up to the rate  $\eta_i$ , even for the worst distribution in the class of distributions which have common mean and covariance  $\mathbf{X}_i \sim (\mu_i, \Sigma_i)$ , replacing the probability constraints in (2.2) with their robust counterparts

$$\sup_{\mathbf{X}_1 \sim (\mu_1, \Sigma_1)} \Pr\{w^\top \mathbf{X}_1 - b \leq -1\} \leq \eta_1, \quad \sup_{\mathbf{X}_2 \sim (\mu_2, \Sigma_2)} \Pr\{w^\top \mathbf{X}_2 - b \geq 1\} \leq \eta_2.$$

Thanks to an appropriate application of the multivariate Chebyshev inequality, this worst distribution approach leads to the following quadratic second-order cone programming (QSOCP) problem (see [1]), which is a deterministic formulation of (2.2) (see [20, 2] for all details):

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2 \\ & w^\top \mu_1 - b \geq 1 + \kappa_1 \|S_1^\top w\| \\ & b - w^\top \mu_2 \geq 1 + \kappa_2 \|S_2^\top w\|, \end{aligned} \tag{2.3}$$

where  $\Sigma_i = S_i S_i^\top$  (for instance, Cholesky factorization) and  $\eta_i$  and  $\kappa_i$  are related via the formula  $\kappa_i = \sqrt{\frac{1-\eta_i}{\eta_i}}$ , for  $i = 1, 2$ .

Note that any feasible hyperplane must separate the means, hence the natural condition  $\mu_1 \neq \mu_2$  is necessary for (2.3) to be feasible. Since  $\kappa_i \rightarrow 0$  when  $\eta_i \rightarrow 1$ , the problem (2.3) can be made feasible whenever  $\mu_1 \neq \mu_2$  by choosing appropriate values for  $\eta_1$  and  $\eta_2$ . The election of different values of  $\eta_1$  and  $\eta_2$  leads to classify with *preferential bias* towards a particular class. For instance, in many binary medical classification problems, the cost of misclassifying one category is higher than the other, and in these applications it is desirable to employ a classifier with selective sensitivity or specificity (Cismondi et al [6]). This is the case of medical diagnosis of cancer (e.g. [14, 18]), where the cost of misclassifying a cancer patient is higher than the cost of misclassifying of a healthily patient.

Finally, since  $\kappa_i \rightarrow \infty$  when  $\eta_i \rightarrow 0$ , we note that problem (2.3) becomes unfeasible for small values of  $\eta_1$  and/or  $\eta_2$ .

## 2.2.1 Practical implementation

So far we have assumed that the mean-covariance pairs  $(\mu_i, \Sigma_i)$  are known. However, in many practical situations we only have the training data set  $\mathcal{T} = \{(x_1, y_1), \dots, (x_m, y_m)\}$ . Assuming that  $\mathcal{T}$  consists of two samples of independent observations of the random vectors  $\mathbf{X}_1$  for  $y = 1$  and  $\mathbf{X}_2$  for  $y = -1$ , the idea is to replace  $(\mu_i, \Sigma_i)$  with a statistical estimator  $(\hat{\mu}_i, \hat{\Sigma}_i)$ ; this can be done by computing the sample mean and covariance for each class from the available observations.

We denote by  $\mathbf{T}_1 = [x_1^1, \dots, x_{m_1}^1] \in \mathbb{R}^{n \times m_1}$  a  $n \times m_1$  data matrix for positive class and by  $\mathbf{T}_2 = [x_1^2, \dots, x_{m_2}^2] \in \mathbb{R}^{n \times m_2}$  a  $n \times m_2$  data matrix for negative class. The empirical estimates of the mean and covariance are given by

$$\mu_i = \hat{\mu}_i = \frac{1}{m_i} \mathbf{T}_i \mathbf{1}_{m_i}, \quad \Sigma_i = \hat{\Sigma}_i = S_i S_i^\top \quad \text{with} \quad S_i = \frac{1}{\sqrt{m_i}} (\mathbf{T}_i - \mu_i \mathbf{1}_{m_i}^\top),$$

for  $i = 1, 2$ , where  $\mathbf{1}_{m_i}$  denotes a vector of ones of dimension  $m_i$ . Since  $w \in \mathbb{R}^n$ , it can be written as  $w = [\mathbf{T}_1, \mathbf{T}_2]s + \mathbf{M}r$ , where  $\mathbf{M}$  is a matrix with its columns as vectors orthogonal to training data points and  $s, r$  are vectors of combining coefficients. The columns of  $\mathbf{T}_1$ ,  $\mathbf{T}_2$  and  $\mathbf{M}$  together span entire  $\mathbb{R}^n$ . Now, the terms involving  $w$  in the constraints of (2.3) can be written as

$$w^\top \mu_i = s^\top \mathbf{g}_i, \quad w^\top \Sigma_i w = s^\top \mathbf{G}_i s, \quad i = 1, 2,$$

where

$$\begin{aligned} \mathbf{g}_1 &= \frac{1}{m_1} [\mathbf{K}_{11} \mathbf{1}_{m_1}; \mathbf{K}_{21} \mathbf{1}_{m_1}], & \mathbf{g}_2 &= \frac{1}{m_2} [\mathbf{K}_{12} \mathbf{1}_{m_2}; \mathbf{K}_{22} \mathbf{1}_{m_2}], \\ \mathbf{G}_1 &= \frac{1}{m_1} [\mathbf{K}_{11}; \mathbf{K}_{21}] (\mathbf{I}_{m_1} - \frac{1}{m_1} \mathbf{1}_{m_1} \mathbf{1}_{m_1}^\top) [\mathbf{K}_{11}, \mathbf{K}_{12}] \end{aligned}$$

and

$$\mathbf{G}_2 = \frac{1}{m_2} [\mathbf{K}_{12}; \mathbf{K}_{22}] (\mathbf{I}_{m_2} - \frac{1}{m_2} \mathbf{1}_{m_2} \mathbf{1}_{m_2}^\top) [\mathbf{K}_{21}, \mathbf{K}_{22}]$$

with  $\mathbf{I}_{m_i}$  denoting the identity matrix of dimension  $m_i \times m_i$ ,  $\mathbf{K}_{11} = \mathbf{T}_1^\top \mathbf{T}_1$ ,  $\mathbf{K}_{12} = \mathbf{K}_{21} = \mathbf{T}_1^\top \mathbf{T}_2$ ,  $\mathbf{K}_{22} = \mathbf{T}_2^\top \mathbf{T}_2$  matrices whose elements are inner products of data points. For

instance, the entry  $(i, j)$  for the matrix  $\mathbf{K}_{12}$  is

$$(\mathbf{K}_{12})_{ij} = \langle x_i^1, x_j^2 \rangle. \quad (2.4)$$

Thus, the formulation (2.3) can be written as:

$$\begin{aligned} \min_{s,b} \quad & \frac{1}{2} s^\top \mathbf{K} s \\ s^\top \mathbf{g}_1 - b & \geq 1 + \kappa_1 \sqrt{s^\top \mathbf{G}_1 s} \\ b - s^\top \mathbf{g}_2 & \geq 1 + \kappa_2 \sqrt{s^\top \mathbf{G}_2 s}, \end{aligned} \quad (2.5)$$

where  $\mathbf{K} = [\mathbf{K}_{11}, \mathbf{K}_{12}; \mathbf{K}_{21}, \mathbf{K}_{22}]$ . Note that in order to solve the above problem, we need to know only the dot products of training data points. Thus, one can solve the above problem in any feature space as long as the inner products in that space are available.

Many practical situations cannot directly solved with the former approach. Consequently, one usually introduces a kernel formulation. This consists of considering a kernel function  $k : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  satisfying the Mercer conditions (see [17]). Thus, for instance, the inner product (2.4) is replaced by  $(\mathbf{K}_{12})_{ij} = k(x_i^1, x_j^2)$ . The typical kernel functions include the Gaussian kernel defined by  $k(u, v) = \exp(-\gamma \|u - v\|^2)$  with  $\gamma \in \mathbb{R}_+$ , and the polynomial function  $k(u, v) = (u^\top v + 1)^d$  with  $d \in \mathbb{N}$ . After choosing this kernel function, the quantities  $\mathbf{g}_1, \mathbf{g}_2, \mathbf{G}_1, \mathbf{G}_2$  and  $\mathbf{K}$  can be computed. In the case when  $\mathbf{K}$  is positive definite, we can use the Cholesky factorization  $\mathbf{K} = \mathbf{L}^\top \mathbf{L}$  to obtain a full rank matrix  $\mathbf{L} \in \mathbb{R}^{m \times m}$ . Thus, introducing a new variable  $v = \mathbf{L}s$ , the formulation (2.5) is rewritten as follows

$$\begin{aligned} \min_{v,b} \quad & \frac{1}{2} \|v\|^2 \\ v^\top \mathbf{h}_1 - b & \geq 1 + \kappa_1 \sqrt{v^\top \mathbf{H}_1 v} \\ b - v^\top \mathbf{h}_2 & \geq 1 + \kappa_2 \sqrt{v^\top \mathbf{H}_2 v}, \end{aligned} \quad (2.6)$$

where  $\mathbf{h}_i = \mathbf{L}^{-\top} \mathbf{g}_i$  and  $\mathbf{H}_i = \mathbf{L}^{-\top} \mathbf{G}_i \mathbf{L}^{-1}$ , for  $i = 1, 2$ .

Again,  $\mathbf{H}_i$  being positive semidefinite can be written as  $\mathbf{H}_i = \mathbf{D}_i \mathbf{D}_i^\top$ . Then, (2.6) can be written in the standard QSOCP form

$$\begin{aligned} \min_{v,b} \quad & \frac{1}{2} \|v\|^2 \\ v^\top \mathbf{h}_1 - b & \geq 1 + \kappa_1 \|\mathbf{D}_1^\top v\|, \\ b - v^\top \mathbf{h}_2 & \geq 1 + \kappa_2 \|\mathbf{D}_2^\top v\|. \end{aligned} \quad (2.7)$$

Clearly, by introducing a new variable  $t$  and a constraint  $\|v\| \leq t$ , (2.7) can be casted as the following linear SOCP with three blocks (cf. [1])

$$\begin{aligned} \min_{t,v,b} \quad & t \\ t & \geq \|v\|, \\ v^\top \mathbf{h}_1 - b & \geq 1 + \kappa_1 \|\mathbf{D}_1^\top v\|, \\ b - v^\top \mathbf{h}_2 & \geq 1 + \kappa_2 \|\mathbf{D}_2^\top v\|. \end{aligned} \quad (2.8)$$

Hence, one can classify a new data point  $x$  using the following decision function

$$f(x) = \text{sign}(w^\top x - b) = \text{sign}(s^\top h(x) - b), \quad (2.9)$$

where  $s = \mathbf{L}^{-1}v$  and  $h(x) = \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix}$  with  $(h_i(x))_l = k(x_l^i, x)$  for  $l = 1, \dots, m_i$  and  $i = 1, 2$ .

### 3 Numerical results and discussion

Support vector machines (SVM) were originally designed for binary classification. However, they can be adapted for multi-class problems. According to (Weston and Watkins, 1999 and Hsu and Lin, 2002) two strategies are used to approach the multi-class SVM problem. A first strategy consist of solving a series of binary classifications using two possible approaches: one-species-against-one (1-vs-1) and one-species-against-the-Rest (1-vs-R). A second strategy directly considers all the data in a single optimization formulation. The latter leads to an optimization problem which is much more difficult to solve numerically (Weston and Watkins, 1999). In this paper, we use the first strategy together with criterion 1-vs-R. This constructs  $M$  binary SVM classifiers, each of which separates one class from all the rest. Throughout this article this set of machines is called *multi-class 1-vs-R classifier*, in contrast with each *binary 1-vs-R classifier* used to construct it.

Thus, the *deterministic  $i$ th binary 1-vs-R classifier* is a SVM trained with all the training examples of the  $i$ th class with positive labels, and all the others with negative labels. That is, the  $i$ th SVM solves problem (2.1) where the decision function is given by

$$f_i(x) = w_i^\top x - b_i. \quad (3.1)$$

So, it solves

$$\min_{w_i, b_i} \frac{1}{2} \|w_i\|^2; \quad \tilde{y}_j (w_i^\top x_j - b_i) \geq 1, \quad j = 1, \dots, m, \quad (3.2)$$

where  $\tilde{y}_j = 1$  if  $y_j = i$  and  $\tilde{y}_j = -1$  otherwise. Then, at the classification phase, a sample  $x$  is classified in the class which attains the largest value of  $f_i(x)$ , that is,  $x$  is in the  $i^*$ th class when  $f_{i^*}(x) = \max\{f_i(x) : i = 1, \dots, M\}$ . In the (exceptional) case when this maximum is attained in more than one class sample  $x$  is (by convention) classified in the class associated with the lowest index  $i^*$ .

Based on this idea and on the stochastic approach described in Section 2, we can similarly construct *stochastic binary 1-vs-R classifiers* and *stochastic multi-class 1-vs-R classifiers*. More precisely, in order to construct a *stochastic binary 1-vs-R classifier*, we solve (2.7) instead of (3.2), where now the variable  $v$  is replaced by  $w_i$ , and function  $f$ , defined in (2.9), is replaced by  $f_i$ , defined in (3.1).

In the next section we construct several multi-class 1-vs-R classifiers and we use the average rate of correct classifications in order to compare their performance. Each multi-class 1-vs-R classifier is composed by three binary 1-vs-R classifiers, one for each fish species. All our numerical experiment were solved using different codes in Matlab. Indeed, we have used the SVM-Light software (<http://svmlight.joachims.org/>) for solving deterministic SVM and the SeDuMi software for solving SOCP problems of the form (2.8) (<http://sedumi.ie.lehigh.edu/>).



### 3.1 Deterministic and stochastic formulation with kernel

A well-known variant of the deterministic binary 1-vs-R classifier is to consider *margins* to classify each class. Roughly speaking, the relevance of these margins in the SVM is driven by a penalty positive parameter denoted by  $C$  (the penalization is stronger when  $C$  is larger). In (Robotham et al [19]) the estimation of the parameters  $C$  and  $\gamma$  was realized for divers machines for classifying fish-schools using acoustic data (all of them with Gaussian kernel). In the following table, these parameters appear together with the classification rate of each one of the three deterministic binary 1-vs-R classifier.

Table 1: Classification rates (%) using deterministic binary 1-vs-R classifiers.

Species	Classification rate (%)		Parameters ( $C, \gamma$ )
Anchovy vs rest	86.7	88.5	(150, 0.14)
Jack mackerel vs rest	81.5	99.5	(110, 0.12)
Common sard. vs rest	90.3	94.4	(117, 0.15)

Next, we present the *confusion matrix*<sup>1</sup> obtained for a deterministic multi-class classification based on the last three deterministic binary 1-vs-R classifiers.

Table 2: Confusion matrix for deterministic multi-class 1-vs.-R classifier.

Species	Anchovy	Jack mackerel	Common sardine	Testing data	Classification rate	Average rate
Anchovy	89	0	9	98	90.8%	
Jack mackerel	4	23	0	27	85.2%	88.4%
Common sard.	11	0	92	103	89.3%	

Our first idea is to use some attributes of this confusion matrix (for instance, classification rate, average rate) to compare the performance of the deterministic multi-class 1-vs.-R classifier presented above with the performances of several stochastic multi-class 1-vs.-R classifiers. They are constructed for different sets of parameters  $\eta_1$  and  $\eta_2$  (each classifier contains six parameters, one pair for each of the three binary 1-vs.-R classifiers) For this, we first set the values of parameters  $\eta_1$  and  $\eta_2$  as the proportion of each class among the total. For instance, in the machine anchovy vs. the rest, the values of these parameters are given by:

$$\eta_1^A = \frac{\text{Total Anchovy}}{\text{Total fish}} = \frac{98}{228} = 0,43 \quad \text{and} \quad \eta_2^A = \frac{\text{Total Rest}}{\text{Total fish}} = \frac{130}{228} = 0,57.$$

The next tables illustrate the performances of each stochastic binary 1-vs-R classifiers and of the stochastic multi-class 1-vs-R classifier. The latter via its confusion matrix.

<sup>1</sup>This matrix represents the count of classifier's class predictions with respect to the actual outcome on some labeled learning set.

Table 3: Classification rates (%) using stochastic binary 1-vs-R classifiers.

Species	Classification		Parameters	
	rate (%)		$\gamma$	$(\eta_1, \eta_2)$
Anchovy vs rest	87.8	80.0	0.14	(0.43, 0.57)
Jack mackerel vs rest	74.1	80.6	0.12	(0.12, 0.88)
Common sard. vs rest	91.3	86.4	0.15	(0.45, 0.55)

Table 4: Classification rates (%) using stochastic multi-class 1-vs-R classifier.

Species	Anchovy	Jack mackerel	Common sardine	Testing data	Classification rate	Average rate
Anchovy	86	3	9	98	87.76%	
Jack mackerel	4	22	1	27	81.48%	84.57%
Common sard.	15	1	87	103	84.47%	

Notice that this choice of parameter leads to a worst performance than the deterministic classifier. This is reflected in lower classification rates for the three binary 1-vs-R classifiers (see Table 1 and compare with 3). In particular, in the stochastic binary jack mackerel vs. the rest classifier, the decision function seems to become more important than in the deterministic case. Indeed, there are three anchovy and one common sardine misclassified as jack mackerel (see Table 2 and 4). However, in the deterministic SVM neither anchovies nor common sardine were classified as jack mackerel.

### 3.2 Sensitivity analysis of stochastic multi-class 1-vs-R classifiers

In this section, we evaluate the average classification rate of diverse stochastic multi-class 1-vs-R classifiers when  $\eta_2$  (false-positive rate) varies. We start this analysis from the results established in Table 4, it shows the relative importance of the stochastic binary 1-vs-R jack mackerel vs. the rest classifier in the performance of the stochastic multi-class 1-vs-R classifiers. Then, we think it is interesting to start our sensitivity analysis by decreasing the value of  $\eta_2$  from

$$\eta_2^J = \frac{\text{Total Rest}}{\text{Total fish}} = \frac{201}{228} = 0,88.$$

Since  $\eta_2$  represents the false-positive rate, lower values of this value should improve the accuracy of this particular stochastic binary 1-vs-R jack mackerel vs. the rest classifier. However, the performance of the whole stochastic multi-class 1-vs-R classifiers could be affected if the rest of the stochastic binary 1-vs-R classifiers behaves badly. The remaining parameters remain unaltered (see Table 3) as well as the number of species used for testing: 98 anchovy, 27 jack mackerel and 103 sardines.

Table 5: Confusion matrices for different values of  $\eta_2^J$ .

Parameter	Species	Anchovy	Jack mackerel	Common sardine	Classification rate	Average rate
$\eta_2^J = 0.78$	Anchovy	87	2	9	88.78 %	84.91 %
	Jack mackerel	4	22	1	81.48 %	
	Common sard.	15	1	87	84.47 %	
$\eta_2^J = 0.68$	Anchovy	88	1	9	89.80 %	85.25 %
	Jack mackerel	4	22	1	81.48 %	
	Common sard.	15	1	87	84.47 %	
$\eta_2^J = 0.58$	Anchovy	88	1	9	89.80 %	85.25 %
	Jack mackerel	4	22	1	81.48 %	
	Common sard.	15	1	87	84.47 %	
$\eta_2^J = 0.48$	Anchovy	89	0	9	90.82 %	85.59 %
	Jack mackerel	4	22	1	81.48 %	
	Common sard.	15	1	87	84.47 %	
$\eta_2^J = 0.38$	Anchovy	89	0	9	90.82 %	85.59 %
	Jack mackerel	4	22	1	81.48 %	
	Common sard.	15	1	87	84.47 %	
$\eta_2^J = 0.28$	Anchovy	88	1	9	89.80 %	86.49 %
	Jack mackerel	4	23	0	85.19 %	
	Common sard.	15	1	87	84.47 %	
$\eta_2^J = 0.18$	Anchovy	86	3	9	87.76 %	87.04 %
	Jack mackerel	3	24	0	88.89 %	
	Common sard.	15	1	87	84.47 %	

Table 5 above presents the confusion matrices for the stochastic multi-class 1-vs-R classifiers obtained for different values  $\eta_2^J$  in the stochastic binary 1-vs-R jack mackerel vs. the rest classifier. As expected, lower values of  $\eta_2^J$  imply better classification rates for this particular classifier. This has also led to an improvement of the average classification rate of the stochastic multi-class 1-vs-R classifiers, even when this positive behavior is not ensured by this choice. Indeed, this average rate could be worse due to a poorer performance of one of the two other stochastic binary 1-vs-R classifiers. This effect can be seen for instance in the variation of the performance of the stochastic binary anchovy vs. the rest classifier with respect to  $\eta_2^J$ , which is non-monotonic and seems to attain its best performance for  $\eta_2^J$  between 0.38 and 0.48. Notice that during these experiments the classification of the stochastic common sardine vs. the rest remains unaltered. This situation suggests us to proceed our experiment by varying the values  $\eta_2^A$  corresponding to the false-positive rate of the stochastic binary anchovy vs. the rest classifier.

For this, we set  $\eta_2^J = 0.18$  and keep constant the rest of values of all parameters except for  $\eta_2^A$ . We have inspected several values of  $\eta_2^J$  but only two of them give meaningful results:  $\eta_2^A = 0.47$  and  $\eta_2^A = 0.17$ . They are reported in following Table 6 via their confusion matrices.

Table 6: Confusion matrices for different values of  $\eta_2^A$ .

Parameter	Species	Anchovy	Jack mackerel	Common sardine	Classification rate	Average rate
$\eta_2^A = 0.47$	Anchovy	86	2	10	87.76 %	88.66 %
	Jack mackerel	3	24	0	88.89 %	
	Common sard.	10	1	92	89.32 %	
$\eta_2^A = 0.17$	Anchovy	84	3	11	85.71 %	89.21 %
	Jack mackerel	2	25	0	92.59 %	
	Common sard.	10	1	92	89.32 %	

Finally, we made a last experiment where only the false-positive rate of the stochastic binary common sardine vs. the rest classifier,  $\eta_2^S$ , varies. The analysis is similar than the previous ones and they are thus omitted. The final parameters are then set as follows:  $(\eta_1^A, \eta_2^A) = (0.43, 0.17)$ ,  $(\eta_1^J, \eta_2^J) = (0.12, 0.18)$  and  $(\eta_1^S, \eta_2^S) = (0.45, 0.25)$ . The results are summarized in Table 7 here below.

Table 7: Confusion matrix for our final experiment.

Species	Anchovy	Jack mackerel	Common sardine	Classification rate	Average rate
Anchovy	85	3	10	86.73 %	89.55 %
Jack mackerel	2	25	0	92.59 %	
Common sard.	10	1	92	89.32 %	

Table 7 shows that the average rate obtained by our last stochastic multi-class 1-vs-R classifier is 89.55%. This value is better than the one obtained by the deterministic multi-class 1-vs-R classifier 88.4% (see Table 2). So, despite of the uncertainties on the data, we have construct a more performant machine via an inspection procedure regarding false-positive rates  $\eta_2$ .

## Acknowledgements

This research was supported by CONICYT-Chile, under Anillo project 10336: “Analysis of control problems & applications” (Julio López and Héctor Ramírez) and FONDECYT Grants 1090063 (Paul Bosch), 11110188 (Julio López), and 1110888 (Héctor Ramírez). The third author is also supported by BASAL Project (Centro de Modelamiento Matemático, Universidad de Chile), and project BIONATURE of CIRIC, INRIA–Chile. The field data was recollected with the support of projects financed by the Chilean Fisheries Research Fund (FIP in Spanish) of the Undersecretary of Fisheries.

## References

- [1] Alizadeh, F., D. Goldfarb, *Second-order cone programming*, Math. Program., vol. 95 (2003), no. 1, Ser. B, pp. 3–51.
- [2] Alvarez, F., López, J., Ramírez C., H., *Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and support vector machines*, Optim. Methods Softw., vol. 25 (2010), no. 6, pp. 859–881.
- [3] Barange, M., *Acoustic identification, classification and structure of biological patchiness on the edge of the Agulhas Bank and its relation to frontal features*, S. Afr. J. Marine Sci. 14 (1994), pp. 333-347.
- [4] Buelens, B., Pauly, T., Williams, R., Sale, A., *Kernel methods for the detection and classification of fish schools in single beam and multibeam acoustic data*, ICES J. Marine Sci. 66 (2009), pp. 1130-1135.
- [5] Canters, F., W.D. Genst and H. Dufourmont, *Assessing effects of input uncertainty in structural landscape classification*, International Journal of Geographical Information Science, Vol. 16, 2 (2002), pp. 129–149.
- [6] Cismondi, F., A.L. Horna, A. S. Fialhoa, S. M. Vieirab, S. R. Retic, J. M. C. Sousab and S. Finkelstein, *Multi-stage modeling using fuzzy multi-criteria feature selection to improve survival prediction of ICU septic shock patients*. Expert Systems with Applications, Volume 39, Issue 16 (2012), pp. 12332-12339.
- [7] Coetzee, J., *Use of a shoal analysis and patch estimation system (SHAPES) to characterize sardine schools*, Aquat. Living Resour. 13 (2000), pp. 1-10.
- [8] Cortes, C. and V. Vapnik. *Support-vector networks*, Machine Learning, vol. 20 (1995), pp. 273-297.
- [9] Cullen, A.C. and H.C. Frey, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Springer, N.Y., 1st edition, 1999.
- [10] Fernandes, P.G., Kornielussen, R.J., Lebourges-Dhaussy, A., Masse, J., Iglesias, M., Diner, N., Ona, E., et al. *The SIMFAMI Project: Species Identification Methods for Acoustic Multifrequency Information*. Final Report to the EC. Q5RS-2001-02054, (2006).
- [11] Fernandes, P.G., *Classification trees for species identification of fish-school echotraces*, ICES J. Marine Sci. 66 (2009), pp. 1073-1080.
- [12] Horne, J.K., *Acoustic approaches to remote species identification: a review*, Fish. Oceanogr. 94 (2000), pp. 356-371.
- [13] Hsu, C.W. and Lin, C.J., *A comparison of methods for multiclass support vector machines*, IEEE Trans. Neural Netw. 13 (2002), pp. 415–425.

- [14] Kononenko, I., *Machine learning for medical diagnosis: History, State of Art and Perspective*, Artificial Intelligence in Medicine, Vol. 23 (2001), pp. 89–109.
- [15] Korneliussen, R.J., Heggelund, Y., Eliassen, I.K., Johansen, G.O., *Acoustic species identification of schooling fish*, ICES J. Marine Sci. 66 (2009), pp 1111-1118.
- [16] Lobo, M.S., Vandenberghe, L., Boyd, S., Lebet, H. *Applications of second-order cone programming*, Linear Algebra Appl. vol. 284 (1998), no 1-3 pp. 193–228.
- [17] Mercer, J., Functions of positive and negative type and their connection with the theory of integral equations, Philosophical Transactions of the Royal Society, London, vol. 209 (1909), pp. 415–446.
- [18] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, A.B.M. Shawkat Ali, Yi-Ping Phoebe Chen, *Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer*, Expert Systems with Applications, Volume 39, Issue 16, 15 November 2012, pp. 12371-12377.
- [19] Robotham, H., Bosch, P., Gutierrez-Estrada, J.C., Castillo, J. and Pulido-Calvo, I., *Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks*, Fisheries Research (2010), pp. 115-122.
- [20] Saketha Nath, J. and C. Bhattacharyya, *Maximum margin classifiers with specified false positive and false negative error rates*, Proceedings of the Seventh SIAM International Conference on Data Mining, April 26-28, 2007, Minneapolis, Minnesota. <http://www.siam.org/meetings/proceedings/2007/datamining/papers/004Jagarlapudi.pdf>.
- [21] Scalabrin, C., Massé, J., *Acoustic detection of the spatial and temporal distribution of fish shoals in the Bay of Biscay*, Aquat. Living Resour. 6 (2003), pp. 269-283.
- [22] Simmonds, E.J., MacLennan, D.N., 2005. Fisheries Acoustics: Theory and Practice, 2nd ed. Fish and Aquatic Resources Series 10. Blackwell Science, Oxford.
- [23] Shivaswamy, P.K., C. Bhattacharyya and A.J. Smola, *Second order cone programming approaches for handling missing and uncertain data*, J. Mach. Learn. Res., vol. 7 (2006), pp. 1283-1314.
- [24] Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, 1995.
- [25] Weston, J., Watkins, C., *Support vector machines for multi-class pattern recognition*, In Proceedings of the Seventh European Symposium on Artificial Neural Networks, (1999), pp. 219–224.