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

Paul Bosch [a], Julio López [a,*], Héctor Ramírez [b], Hugo Robotham [a]

[a] *Facultad de Ingeniería, Universidad Diego Portales, Ejército 441, Santiago, Chile*
[b] *Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático (CNRS UMI 2807), FCFM, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile*

ARTICLE INFO

ABSTRACT

In this work we apply multi-class support vector machines (SVMs) 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.

© 2013 Elsevier Ltd. All rights reserved.

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 Fernandes et al., 2006; Horne, 2000). 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 plankton acoustics.

A wide range of statistical techniques have been used to classify mono-specific fish-school. Fernandes (2009) 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, Pauly, Williams, and Sale (2009) use kernel methods to classify fish schools in single beam and multibeam acoustic data. In Robotham, Bosch, Gutierrez-Estrada, Castillo, and Pulido-Calvo (2010), support vector machine methods (tool for discriminating between two groups, see Cortes & Vapnik, 1995; Vapnik, 1995) 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 (see Simmonds & MacLennan, 2005. 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 (see Korneliussen, Heggelund, Eliassen, & Johansen, 2009).

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°25′S–43°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°, calibrated according to standard

procedures (Foote, Knudsen, Vestnes, MacLennan, & Simmonds, 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.n and Massé (1993) is/are cited in the text but not provided in the reference list. Please provide it/them in the reference list or delete these citations from the text.

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). 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 = 1 m, minimum candidate length = 1 m, maximum vertical linking distance = 1 m, and maximum horizontal linking distance = 15 m.

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 Scalabrin & Massé, 1993):

1. Morphological: mean height (H), length (L) and perimeter (P) with units in meter (m); area (m^2) and unidimensional descriptors:
   – Elongation $= \frac{\text{length}}{\text{mean height}}$.
   – 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 Fig. 1.

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 (Saketha Nath & Bhattacharyya, 2007; Shivaswamy, Bhattacharyya, & Smola, 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 $\mathcal{T} = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ of m training data, where $x_i \in \mathbb{R}^n$ representing the ith 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 Cortes & Vapnik, 1995):

$$\min_{w,b} \frac{1}{2}\|w\|^2 \qquad\qquad\qquad (1)$$
$$y_i(w^\top x_i - b) \geqslant 1, \quad i = 1, \ldots, m.$$

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:

$$\min_{w,b} \frac{1}{2}\|w\|^2$$
$$\Pr\{w^\top \mathbf{X}_1 - b \leqslant -1\} \leqslant \eta_1, \qquad (2)$$
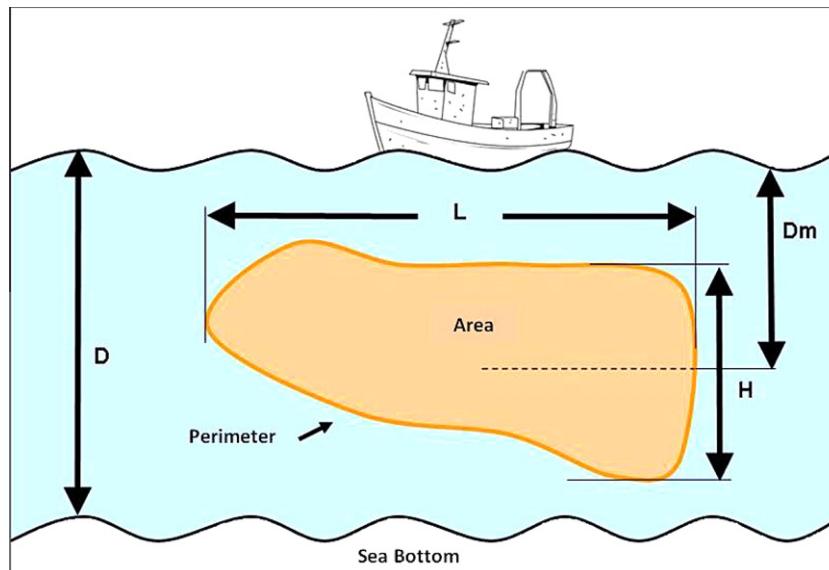$$\Pr\{w^\top \mathbf{X}_2 - b \geqslant 1\} \leqslant \eta_2.$$

Fig. 1. Representation of bathymetric and morphological descriptors.

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 η_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) with their robust counterparts

$$\sup_{\mathbf{X}_1 \sim (\mu_1, \Sigma_1)} \text{Prob}\{w^\top \mathbf{X}_1 - b \leqslant -1\} \leqslant \eta_1, \quad \sup_{\mathbf{X}_2 \sim (\mu_2, \Sigma_2)} \text{Prob}\{w^\top \mathbf{X}_2 - b \geqslant 1\} \leqslant \eta_2.$$

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 Lobo, Vandenberghe, Boyd, & Lebret, 1998; Alizadeh and Goldfarb, 2003), which is a deterministic formulation of (2) (see Alvarez, López, & Ramírez, 2010; Saketha Nath & Bhattacharyya, 2007 for all details):

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{3}$$
$$w^\top \mu_1 - b \geqslant 1 + \kappa_1 \|S_1^\top w\|$$
$$b - w^\top \mu_2 \geqslant 1 + \kappa_2 \|S_2^\top w\|,$$

where $\Sigma_i = S_i S_i^\top$ (for instance, Cholesky factorization) and η_i and κ_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 (3) to be feasible. Since $\kappa_i \to 0$ when $\eta_i \to 1$, the problem (3) can be made feasible whenever $\mu_1 \neq \mu_2$ by choosing appropriate values for η_1 and η_2. The election of different values of η_1 and η_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., 2012). This is the case of medical diagnosis of cancer (e.g. Nahar, Imam, Tickle, Shawkat Ali, & Chen, 2012; Kononenco, 2001), where the cost of misclassifying a cancer patient is higher than the cost of misclassifi of a healthily patient.

Finally, since $\kappa_i \to \infty$ when $\eta_i \to 0$, we note that problem (3) becomes unfeasible for small values of η_1 and/or η_2.

2.2.1. Practical implementation

So far we have assumed that the mean-covariance pairs (μ_i, Σ_i) are known. However, in many practical situations we only have the training data set $\mathcal{T} = \{(x_1, y_1), \ldots, (x_m, y_m)\}$. 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 (μ_i, Σ_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 $\mathbf{T}_1 = \left[x_1^1, \ldots, x_{m_1}^1\right] \in \mathbb{R}^{n \times m_1}$ a $n \times m_1$ data matrix for positive class and by $\mathbf{T}_2 = \left[x_1^2, \ldots, x_{m_2}^2\right] \in \mathbb{R}^{n \times m_2}$ a $n \times m_2$ data matrix for negative class. The empirical estimates of the mean and covariance are given by

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

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 (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

$$\mathbf{g}_1 = \frac{1}{m_1}[\mathbf{K}_{11}\mathbf{1}_{m_1}; \mathbf{K}_{21}\mathbf{1}_{m_1}], \quad \mathbf{g}_2 = \frac{1}{m_2}[\mathbf{K}_{12}\mathbf{1}_{m_2}; \mathbf{K}_{22}\mathbf{1}_{m_2}],$$

$$\mathbf{G}_1 = \frac{1}{m_1}[\mathbf{K}_{11}; \mathbf{K}_{21}]\left(\mathbf{I}_{m_1} - \frac{1}{m_1}\mathbf{1}_{m_1}\mathbf{1}_{m_1}^\top\right)[\mathbf{K}_{11}, \mathbf{K}_{12}]$$

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

$$(\mathbf{K}_{12})_{ij} = \left\langle x_i^1, x_j^2 \right\rangle. \tag{4}$$

Thus, the formulation (3) can be written as:

$$\min_{s,b} \frac{1}{2}s^\top \mathbf{K} s \tag{5}$$
$$s^\top \mathbf{g}_1 - b \geqslant 1 + \kappa_1 \sqrt{s^\top \mathbf{G}_1 s}$$
$$b - s^\top \mathbf{g}_2 \geqslant 1 + \kappa_2 \sqrt{s^\top \mathbf{G}_2 s},$$

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 $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying the Mercer conditions (see Mercer, 1909). Thus, for instance, the inner product (4) is replacing by $(\mathbf{K}_{12})_{ij} = k\left(x_i^1, x_j^2\right)$. 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 (5) is rewritten as follows

$$\min_{v,b} \frac{1}{2}\|v\|^2 \tag{6}$$
$$v^\top \mathbf{h}_1 - b \geqslant 1 + \kappa_1 \sqrt{v^\top \mathbf{H}_1 v}$$
$$b - v^\top \mathbf{h}_2 \geqslant 1 + \kappa_2 \sqrt{v^\top \mathbf{H}_2 v},$$

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, (6) can be written in the standard QSOCP form

$$\min_{v,b} \frac{1}{2}\|v\|^2 \tag{7}$$
$$v^\top \mathbf{h}_1 - b \geqslant 1 + \kappa_1 \|\mathbf{D}_1^\top v\|,$$
$$b - v^\top \mathbf{h}_2 \geqslant 1 + \kappa_2 \|\mathbf{D}_2^\top v\|.$$

Clearly, by introducing a new variable t and a constraint $\|v\| \leqslant t$, (7) can be casted as the following linear SOCP with three blocks (cf. Alizadeh & Goldfarb, 2003)

$$\min_{t,v,b} t \tag{8}$$
$$t \geqslant \|v\|,$$
$$v^\top \mathbf{h}_1 - b \geqslant 1 + \kappa_1 \|\mathbf{D}_1^\top v\|,$$
$$b - v^\top \mathbf{h}_2 \geqslant 1 + \kappa_2 \|\mathbf{D}_2^\top v\|.$$

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

$$f(x) = \text{sign}(w^\top x - b) = \text{sign}(s^\top h(x) - b), \qquad (9)$$

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

3. Numerical results and discussion

Support vector machines (SVMs) were originally designed for binary classification. However, they can be adapted for multi-class problems. According to Weston and Watkins (1999), 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 & 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 ith binary 1-vs-R classifier* is a SVM trained with all the training examples of the ith class with positive labels, and all the others with negative labels. That is, the ith SVM solves problem (1) where the decision function is given by

$$f_i(x) = w_i^\top x - b_i. \qquad (10)$$

So, it solves

$$\min_{w_i, b_i} \frac{1}{2}\|w_i\|^2; \quad \tilde{y}_j(w_i^\top x_j - b_i) \geqslant 1, \quad j = 1, \ldots, m, \qquad (11)$$

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 $f_{i^*}(x) = \max\{f_i(x) : i = 1, \ldots, M\}$. 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 (7) instead of (11), where now the variable v is replaced by w_i, and function f, defined in (9), is replaced by f_i, defined in (10).

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 (8) (http://sedumi.i.e.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. (2010) the estimation of the parameters C and γ was realized for divers machines for classifying fish-schools using acoustic data (all of them with Gaussian kernel).

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

Species	Classification rate (%)		Parameters (C, γ)
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)

In Table 1, these parameters appear together with the classification rate of each one of the three deterministic binary 1-vs-R classifier.

Next, in Table 2 we present the *confusion matrix*[1] obtained for a deterministic multi-class classification based on the last three deterministic binary 1-vs-R classifiers.

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 η_1 and η_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 η_1 and η_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:

$$\eta_1^A = \frac{\text{Total Anchovy}}{\text{Total fish}} = \frac{98}{228} = 0.43 \quad \text{and} \quad \eta_2^A = \frac{\text{Total Rest}}{\text{Total fish}} = \frac{130}{228} = 0.57.$$

Tables 3 and 4 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.

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 Table 3). In particular, in the stochastic binary Jack Mackerel vs. the rest classifier, the decision function seems to became more important than in the deterministic case. Indeed, there are three anchovy and one common sardine misclassified as Jack Mackerel (see Tables 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 divers stochastic multi-class 1-vs-R classifiers when η_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 η_2 from

$$\eta_2^J = \frac{\text{Total Rest}}{\text{Total fish}} = \frac{201}{228} = 0.88.$$

Since η_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 affect 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.

---

[1] This matrix represents the count of classifier's class predictions with respect to the actual outcome on some labeled learning set.

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	

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

Species	Classification rate (%)		Parameters	
			γ	(η_1,η_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	

Table 5
Confusion matrices for different values of η_2^I.

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

Table 6
Confusion matrices for different values of η_2^A.

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

Table 7
Confusion matrix for our final experiment.

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

Table 5 above presents the confusion matrices for the stochastic multi-class 1-vs-R classifiers obtained for different values η_2^J in the stochastic binary 1-vs-R Jack Mackerel vs. the rest classifier. As expected, lower values of η_2^J imply better classification rates for this particular classifier. This has also led to a improvement of the average classification research 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 η_2^J, which is non-monotonic and seems to attains its best performance for η_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 situations suggest us to proceed our experiment by varying the values η_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 η_2^A. We have inspected several values of η_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 Table 6 via their confusion matrices.

Finally, we made a last experiment where only the false-positive rate of the stochastic binary common sardine vs. the rest classifier, η_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.

In Table 7 above 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 η_2.

References

Alizadeh, F., & Goldfarb, D. (2003). Second-order cone programming. *Mathematics Programs Series B, 95*(1), 3–51.

Alvarez, F., López, J., & Ramírez, C. H. (2010). Interior proximal algorithm with variable metric for second-order cone programming: Applications to structural optimization and support vector machines. *Optimization Methods Software, 25*(6), 859–881.

Barange, M. (1994). Acoustic identification, classification and structure of biological patchiness on the edge of the Agulhas Bank and its relation to frontal features. *South African Journal of Marine Science, 14*, 333–347.

Buelens, B., Pauly, T., Williams, R., & Sale, A. (2009). Kernel methods for the detection and classification of fish schools in single beam and multibeam acoustic data. *ICES Journal of Marine Science, 66*, 1130–1135.

Cismondi, F., Horna, A. L., Fialhoa, A. S., Vieirab, S. M., Retic, S. R., Sousab, J. M. C., et al. (2012). Multi-stage modeling using fuzzy multi-criteria feature selection to improve survival prediction of ICU septic shock patients. *Expert Systems with Applications, 39*(16), 12332–12339.

Coetzee, J. (2000). Use of a shoal analysis and patch estimation system (SHAPES) to characterize sardine schools. *Aquatic Living Resources, 13*, 1–10.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

Foote, K., Knudsen, H., Vestnes, G., MacLennan, D., & Simmonds, J. (1987). Calibration of acoustics instruments for sh density estimation: a practical guide. *ICES Cop.Res. Rep., 144*, 57.

Fernandes, P. G. (2009). Classification trees for species identification of fish-school echotraces. *ICES Journal of Marine Science, 66*, 1073–1080.

Fernandes, P. G., Kornielussen, R. J., Lebourges-Dhaussy, A., Masse, J., Iglesias, M., Diner, N., & Ona, E., et al. (2006). The SIMFAMI project: Species identification methods for acoustic multifrequency information. Final Report to the EC. Q5RS-2001-02054.

Horne, J. K. (2000). Acoustic approaches to remote species identification: A review. *Fisheries Oceanography, 94*, 356–371.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Network, 13*, 415–425.

Kononenco, I. (2001). Machine learning for medical diagnosis: History, state of art and perspective. *Artificial Intelligence in Medicine, 23*, 89–109.

Korneliussen, R. J., Heggelund, Y., Eliassen, I. K., & Johansen, G. O. (2009). Acoustic species identification of schooling fish. *ICES Journal of Marine Science, 66*, 1111–1118.

Lobo, M. S., Vandenberghe, L., Boyd, S., & Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra Applications, 284*(1–3), 193–228.

Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society, London, 209*, 415–446.

Nahar, Jesmin, Imam, Tasadduq, Tickle, Kevin S., Shawkat Ali, A. B. M., & Chen, Yi-Ping Phoebe (2012). Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer. *Expert Systems with Applications, 39*(16), 12371–12377.

Robotham, H., Bosch, P., Gutierrez-Estrada, J. C., Castillo, J., & Pulido-Calvo, I. (2010). Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks. *Fisheries Research*, 115–122.

Saketha Nath, J., & Bhattacharyya, C. (2007). Maximum margin classifiers with specified false positive and false negative error rates. In *Proceedings of the seventh SIAM international conference on data mining*, April 26–28. Minneapolis, Minnesota. <http://www.siam.org/meetings/proceedings/2007/datamining/papers/004Jagarlapudi.pdf>.

Scalabrin, C., & Massé, J. (1993). Acoustic detection of the spatial and temporal distribution of fish shoals in the Bay of Biscay. *Aquatic Living Resources, 6*, 269–283.

Shivaswamy, P. K., Bhattacharyya, C., & Smola, A. J. (2006). Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research, 7*, 1283–1314.

Simmonds, E. J., & MacLennan, D. N. (2005). *Fisheries Acoustics: Theory and Practice. Fish and Aquatic Resources Series 10* (2nd ed.,). Oxford: Blackwell Science.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.

Weston, J., & Watkins, C. (1999). Support vector machines for multi-class pattern recognition. In *Proceedings of the seventh european symposium on artificial neural networks* (pp. 219–224).