Multi-Classification by Using Tri-Class SVM

CECILIO ANGULO†*, FRANCISCO J. RUIZ1, LUIS GONZÁLEZ2, and JUAN ANTONIO ORTEGA3
1Grup de Recerca en Enginyeria del Coneixement, Universitat Politècnica de Catalunya, Av. Víctor Balaguer sn. 08800 – Vilanova i la Geltrú, Spain. e-mail: cecilio.angulo@upc.edu
2Departamento de Economía Aplicada I, Universidad de Sevilla, Avenida Ramón y Cajal, l. 41018 – Sevilla, Spain
3Escuela Técnica Superior de Ingeniería Informática, Universidad de Sevilla, Avenida Reina Mercedes, sn. 41012 – Sevilla, Spain

Abstract. The standard form for dealing with multi-class classification problems when bi-classifiers are used is to consider a two-phase (decomposition, reconstruction) training scheme. The most popular decomposition procedures are pairwise coupling (one versus one, 1-v-1), which considers a learning machine for each Pair of classes, and the one-versus-all scheme (one versus all, 1-v-r), which takes into consideration each class versus the remaining classes. In this article a 1-v-1 tri-class Support Vector Machine (SVM) is presented. The expansion of the architecture of this machine into three categories specifically addresses the decomposition problem of how to prevent the loss of information which occurs in the usual 1-v-1 training procedure. The proposed machine, by means of a third class, allows all the information to be incorporated into the remaining training patterns when a multi-class problem is considered in the form of a 1-v-1 decomposition. Three general structures are presented where each improves some features from the precedent structure. In order to deal with multi-classification problems, it is demonstrated that the final machine proposed allows ordinal regression as a form of decomposition procedure. Examples and experimental results are presented which illustrate the performance of the new tri-class SV machine.

Key words. bi-classifier, multi-classification, ordinal regression, Support Vector Machine

Abbreviations. 1-v-1 – one versus one; all versus all; pairwise coupling; 1-v-r – one versus the rest; one versus all; s.t. – subject to; SV – Support Vector; SVM – Support Vector Machine

1. Introduction

Support Vector Machines (SVMs) are learning machines which implement the structural risk minimization inductive principle to obtain good generalization on a limited number of learning patterns. This theory was originally developed by Vapnik on the basis of a separable binary classification problem with signed outputs ±1 [21].

The SVM presents good theoretical properties and behaviour in problems of binary classification [9]. There are several papers which generalize the original
bi-class approach to multi-classification problems [16, 17, 1] through different algorithms, such as 1-v-r SVM or 1-v-1 SVM (see [15] for a comparison of SVM multi-class methods). In this work it is assumed that problems with more than 2 classes will be considered, hence the original bi-class SVM is extended to a more general tri-class SVM approach. The proposed final tri-class machine is presented in a three-stage procedure: first the original idea of a third class is introduced which was developed by Angulo and Català [3, 2]; secondly a more specific machine, as proposed by Angulo and González [5] is presented; finally, the proposed novel tri-class SVM is explained, which implies a huge computational cost reduction with respect to the former proposals, and a meeting point for both classification and ordinal regression techniques.

The rest of the article is organized as follows: in Section 2, the standard SVM classification learning paradigm is briefly presented in order to introduce some notation. Section 3 is devoted to a short introduction about SVMs for multi-classification. In Section 4, the 1-v-1 tri-class SV Machine is presented, and its faster computational counterpart is derived in Section 5. Examples and experimental results are displayed in Section 6 to illustrate its behaviour and strengths. Finally, some conclusions are drawn and future research suggested.

2. Bi-Class SV Machine Learning

The SV Machine is an implementation of a more general regularization principle known as the large margin principle. Let

\[\mathcal{Z} = \{(x_1, y_1), \ldots, (x_n, y_n)\} = \{z_1, \ldots, z_n\} \in (\mathcal{X} \times \mathcal{Y})^n\]  

be a training set, where \(\mathcal{X}\) is the input space and

\[\mathcal{Y} = \{\theta_1, \theta_2\} = \{-1, +1\}\]  

the output space. Let

\[\phi: \mathcal{X} \to \mathcal{F} \subseteq \mathbb{R}^d\]  

be a feature mapping, with \(\phi = (\phi_1, \ldots, \phi_d)\), for the usual ‘kernel trick’. \(\mathcal{F}\) is named feature space. Let

\[x \equiv \phi(x) \in \mathcal{F}\]  

be the representation of \(x \in \mathcal{X}\). A binary linear classifier,

\[f_w(x) = \langle \phi(x), w \rangle + b = \langle x, w \rangle + b\]  

is sought in the space \(\mathcal{F}\), with \(f_w: \mathcal{X} \to \mathcal{F} \to \mathbb{R}, \ b \in \mathbb{R}\), and where outputs are obtained by thresholding the classifier, \(h_w(x) = \text{sign}(f_w(x))\). According to [12], the