Available online at www.sciencedirect.com
SCIENCE DIRECT®

NEUROCOMPUTING

ELSEVIER

Neurocomputing 64 (2005) 137–159

www.elsevier.com/locate/neucom

Building sparse representations and structure determination on LS-SVM substrates

K. Pelckmans*, J.A.K. Suykens, B. De Moor

K.U. Leuven, ESAT-SCD/SISTA, Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium

Available online 20 January 2005

Abstract

This paper proposes a new method to obtain sparseness and structure detection for a class of kernel machines related to least-squares support vector machines (LS-SVMs). The key method is to adopt an hierarchical modeling strategy. Here, the first level consists of an LS-SVM substrate which is based upon an LS-SVM formulation with additive regularization trade-off. This regularization trade-off is determined at higher levels such that sparse representations and/or structure detection are obtained. Using the necessary and sufficient conditions for optimality given by the Karush–Kuhn–Tucker conditions, one can guide the interaction between different levels via a well-defined set of hyper-parameters. From a computational point of view, all levels can be fused into a single convex optimization problem. Furthermore, the principle is applied in order to optimize the validation performance of the resulting kernel machine. Sparse representations as well as structure detection are obtained, respectively, by using an L_1 regularization scheme and a measure of maximal variation at the second level. A number of case studies indicate the usefulness of these approaches both with respect to interpretability of the final model as well as generalization performance.

Keywords: Least-squares support vector machines; Regularization; Structure detection; Model selection; Convex optimization; Sparseness

*Corresponding author. Tel.: +32 16 32 86 58; fax: +32 16 32 19 70.
E-mail addresses: kristiaan.pelckmans@esat.kuleuven.ac.be (K. Pelckmans), johan.suykens@esat.kuleuven.ac.be (J.A.K. Suykens).
URL: http://www.esat.kuleuven.ac.be/sista/lssvmlab/.

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.11.029

1. Introduction

The problem of inference of a model based on a finite set of observational data is often ill-posed [19]. To address this problem, typically a form of capacity control is introduced which is often expressed mathematically in the form of regularization [28]. Regularized cost functions have been applied successfully, e.g. in splines, multilayer perceptrons, regularization networks [18], support vector machines (SVM) and related methods (see, e.g. [11]). SVM [29] is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation which has also led to many other recent developments in kernel-based learning methods in general [22]. SVMs have been introduced within the context of statistical learning theory and structural risk minimization. In the methods, one solves convex optimization problems, typically quadratic programs. Least-Squares Support Vector Machines (LS-SVMs) [21,26] are reformulations to standard SVMs which lead to solving linear Karush–Kuhn–Tucker (KKT) systems for classification tasks as well as regression. Primal–dual LS-SVM formulations have also been given for KFDA, KPCA, KCCA, KPLS, recurrent networks and control [27].[1] Recently, LS-SVM methods were studied in combination with additive models [10] resulting in so-called componentwise LS-SVMs [15] which are suitable for component selection. So-called additive models consisting of a sum of lower dimensional nonlinearities per component (variable) have become one of the widely used non-parametric techniques as they offer a compromise between the somewhat conflicting requirements of flexibility, dimensionality and interpretability (see e.g. [11]).

The relative importance between the *smoothness* of the solution (as defined in different ways) and the norm of the residuals in the cost function involves a tuning parameter, usually called the regularization constant. The determination of regularization constants is important in order to achieve good generalization performance with the trained model and is an important problem in statistics and learning theory (see e.g. [11–13,22,25]). Several model selection criteria have been proposed in the literature to tune this constant. Special attention was given in the machine learning community to cross-validation and leave-one-out-based methods [24] and fast implementations were studied in the context of kernel machines (see e.g. [2]). In the following paper, the performance on an independent validation data-set is considered. The optimization of the regularization constant in LS-SVMs with respect to this criterion can be non-convex (and even non-smooth) in general. In order to overcome this difficulty, a re-parameterization of the regularization trade-off has been recently introduced in [16] referred to as *additive regularization trade-off*. When applied to the LS-SVM formulation, this leads to LS-SVM substrates. In [16], it was illustrated how to employ these LS-SVM substrates to obtain models which were optimal in training and validation or cross-validation sense.

[1] The Internet portal for LS-SVM related research and software can be found at http://www.esat.kuleuven.ac.be/sista/lssvmlab.