# Partially Parametric SVM

José M. Matías

Dpto. de Estadística, Universidad de Vigo, 36310 Vigo, Spain
jmmatias@uvigo.es

**Abstract.** In this paper we propose a simple and intuitive method for constructing partially linear models and, in general, partially parametric models, using support vector machines for regression and, in particular, using regularization networks (splines). The results are more satisfactory than those for classical nonparametric approaches. The method is based on a suitable approach to selecting the kernel by relaying on the properties of positive definite functions. No modification is required of the standard SVM algorithms, and the approach is valid for the $\varepsilon$-insensitive loss. The approach described here can be immediately applied to SVMs for classification and to other methods that use the kernel as the inner product.

## 1 Introduction

The support vector machine (SVM) approach ([10], [7]) is one of the most relevant non-parametric curve-smoothing methods of recent years. Based on essentially linear techniques, its formal elegance and the intuition lying behind the concepts of margin ($\varepsilon$-insensitive loss in regression), support vectors, feature space and kernel as inner product in the feature space, all make this family of techniques a highly attractive option, not to mention the unique solution to the optimization problem in hand and the sparsity of its final expression.

However, the SVMs, like many techniques arising in the machine learning field, have one major drawback in terms of modeling, and that is the difficulty in interpreting their final expression. Yet the SVMs do permit difficult points to be identified (support vectors), and these are undoubtedly of relevance to the analyst. That said, interpretative capacity is still restricted to this set of important points, and fails to identify the relationships between input and output variables.

This paper is intended to palliate this problem by means of the construction of partially linear models ([6], [9]) based on the SVMs (PL–SVMs) and, with greater generality, partially parametric SVMs (PP–SVMs), which include as particular cases, partially polynomial SVMs.

Recently, Espinoza et al. [3] proposed the PL–LSSVMs as an adaptation of SVMs with quadratic loss (LSSVM), in order to implement a partially linear model.

Unlike the PL–LSSVMs, our method results in a standard SVM, and so it exploits two of the most powerful elements of the SVMs, namely, the feature space and associated kernel, and the concept of the support vector.

C. Bento, A. Cardoso, and G. Dias (Eds.): EPIA 2005, LNAI 3808, pp. 67–75, 2005.

The method is thus valid for the $\varepsilon$-insensitive loss (linear or quadratic) and can be used immediately since none of the SVM algorithms need modification – all that is needed is a suitably intuitive selection of the kernel.

Moreover, our approach not only enables partially linear models to be formulated, like the PL–LSSVMs, but also enables more general models to be formulated for the parametric part, and in particular, polynomials of any degree.

## 2 Partially Parametric SVM

Let the model be as follows:

$$\begin{aligned} \mathsf{Y}_i &= h(\mathbf{Z}_i;\beta) + g(\mathbf{T}_i) + \varepsilon_i,\ i = 1, ..., n \qquad (1) \\ &= \psi(\mathbf{X}_i;\beta) + \varepsilon_i,\ i = 1, ..., n \qquad (2) \end{aligned}$$

where for $i = 1, ..., n$, $\varepsilon_i$ is i.i.d. zero-mean random noise, $\mathbf{Z}_i$ is the ith observation of the random vector $\mathbf{Z} = (\mathsf{Z}_1, ..., \mathsf{Z}_{d_1})^t$, $h(\cdot;\beta) : \mathbb{R}^{d_1} \to \mathbb{R}$ is a known and predetermined function (which will serve as the parametric term in the model), parameterized by a vector $\beta \in \mathbb{R}^m$ of parameters to be estimated, $g : \mathbb{R}^{d_2} \to \mathbb{R}$ is an unknown function, and $\mathbf{T}_i$ is the ith observation of the random vector $\mathbf{T} = (\mathsf{T}_1, ..., \mathsf{T}_{d_2})^t$. Likewise, we denote as $\mathbf{X} = (\mathbf{Z}^t\ \mathbf{T}^t)^t$ the vector for all the covariables (where $t$ denotes transpose), with observations $\mathbf{X}_i = (\mathbf{Z}_i^t\ \mathbf{T}_i^t)^t \in \mathcal{S} \subset \mathbb{R}^d$, with $d = d_1 + d_2$, $i = 1, ..., n$, and with $\psi$ as the regression function $\mathbb{E}(\mathsf{Y}|\mathbf{X} = \mathbf{x})$.

The model in (1) includes as a particular case the partially linear model:

$$\mathsf{Y}_i = \mathbf{Z}_i^t\beta + g(\mathbf{T}_i) + \varepsilon_i,\ i = 1, ..., n \qquad (3)$$

– requiring only a definition of $h(\mathbf{Z}_i;\beta) = \mathbf{Z}_i^t\beta$, – but also other parametric models (e.g. higher degree polynomials).

The model in (2) is a typical application of SVM to regression. Thus, by means of a transformation $\phi : \mathcal{S} \subset \mathbb{R}^d \to \mathbb{R}^r$ in which $r$ may be infinite, a feature space is defined $\mathcal{F} = \{\phi(\mathbf{x}) : \mathbf{x} \in \mathcal{S} \subset \mathbb{R}^d\}$ as equipped with an inner product (defined in turn by means of a positive definite function (kernel)[1] $\langle\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle_{\mathcal{F}} = k(\mathbf{x}_i, \mathbf{x}_j)$) in which the problem is posed of determining the optimum hyperplane $f_{\mathbf{w},b}(\mathbf{x}) = \langle\mathbf{w}, \phi(\mathbf{x})\rangle + b$ for the problem:

$$\min_{\mathbf{w},b}\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\ell_p(y_i, f(\mathbf{x}_i))\}$$

where $\ell_p(y, f(\mathbf{x})) = |y - f(\mathbf{x})|_\varepsilon^p = \max\{0, (|y - f(\mathbf{x})| - \varepsilon)^p\}$ is the Vapnik $\varepsilon$-insensitive loss in which $p = 1$ or 2.

[1] Henceforth, wherever we consider it superfluous, we will omit any reference to the space $\mathcal{F}$ in its inner product $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ and in its norm.