Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

NEUROCOMPUTING

Neurocomputing 64 (2005) 107–117

www.elsevier.com/locate/neucom

Evolutionary tuning of multiple SVM parameters

Frauke Friedrichs, Christian Igel*

Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany

Available online 7 January 2005

Abstract

The problem of model selection for support vector machines (SVMs) is considered. We propose an evolutionary approach to determine multiple SVM hyperparameters: The covariance matrix adaptation evolution strategy (CMA-ES) is used to determine the kernel from a parameterized kernel space and to control the regularization. Our method is applicable to optimize non-differentiable kernel functions and arbitrary model selection criteria. We demonstrate on benchmark datasets that the CMA-ES improves the results achieved by grid search already when applied to few hyperparameters. Further, we show that the CMA-ES is able to handle much more kernel parameters compared to grid-search and that tuning of the scaling and the rotation of Gaussian kernels can lead to better results in comparison to standard Gaussian kernels with a single bandwidth parameter. In particular, more flexibility of the kernel can reduce the number of support vectors.

Keywords: Support vector machines; Model selection; Evolutionary algorithms

1. Introduction

Support vector machines (SVMs, e.g., [5,22,24]) are learning machines based on two key elements: a general purpose linear learning algorithm and a problem specific kernel that computes the inner product of input data points in a feature space. The choice of the kernel function is the crucial step in handling a learning task with an SVM. For example, it is important to achieve a distribution of the data in the feature

*Corresponding author. Tel.: +49 2343225558; fax: +49 2343214209.
E-mail address: christian.igel@neuroinformatik.rub.de (C. Igel).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.11.022

space that reflects the affiliation to the class labels. Often a parameterized family of kernel functions is considered and the problem reduces to finding an appropriate parameter vector for the given problem. In case of noisy, non-separable data one also has to choose a regularization parameter, which controls the trade-off between minimizing the training error and the complexity of the decision function. The kernel parameters together with the regularization parameter are called the hyperparameters of the SVM.

In practice the hyperparameters are usually determined by grid search. That is, the hyperparameters are varied with a fixed step-size through a wide range of values and the performance of every combination is assessed using some performance measure. Because of the computational complexity, grid search is only suitable for the adjustment of very few parameters. Perhaps, the most elaborate systematic technique for choosing multiple hyperparameters are gradient descent methods [3,4,8,15]. These algorithms iterate the following procedure: The SVM is trained using the current hyperparameter vector, the gradient of some generalization error bound w.r.t. the hyperparameters is calculated, and a step is performed in the parameter space based on this gradient. However, this approach has some significant drawbacks. The kernel function has to be differentiable. The score function for assessing the performance of the hyperparameters (or at least an accurate approximation of this function) also has to be differentiable with respect to kernel and regularization parameters. This excludes reasonable measures such as the number of support vectors. In [3, Section 6.2] separability of the dataset is assumed when computing the derivative, which is a very restrictive assumption. Iterative gradient-based algorithms, which usually rely on smoothed approximations of a score function, do not ensure that the search direction points exactly to an optimum of the original, often discontinuous generalization performance measure.

We propose an evolutionary method for hyperparameter selection that does not suffer from the limitations described above. Evolutionary algorithms have been successfully applied to model selection for neural networks [11,18,25]. This includes the recent applications of genetic algorithms for feature selection of SVMs [6,7,14,17]. We use the covariance matrix adaptation evolution strategy (CMA-ES, [10]) to search for an appropriate hyperparameter vector. The fitness function that is optimized directly corresponds to some generalization performance measure. We apply our method to tuning Gaussian kernels, where not only the scaling but also the orientation is adapted.

We give a short description of SVMs in Section 2 and of the CMA-ES in Section 3. The parameterization of general Gaussian kernels is introduced in Section 4. We present experimental results in Section 5 and draw our conclusions in Section 6.

2. Support vector machines

We consider L_1-norm soft margin SVMs for the discrimination of two classes. Let $(\boldsymbol{x}_i, y_i), 1 \leqslant i \leqslant \ell$, be the training examples, where $y_i \in \{-1, 1\}$ is the label associated with input pattern $\boldsymbol{x}_i \in X$. The main idea of SVMs is to map the input vectors to a