Multi-objective Model Selection for Support Vector Machines

Christian Igel

Institute for Neurocomputing,
Ruhr-University Bochum,
44780 Bochum, Germany
christian.igel@neuroinformatik.rub.de

Abstract. In this article, model selection for support vector machines is viewed as a multi-objective optimization problem, where model complexity and training accuracy define two conflicting objectives. Different optimization criteria are evaluated: Split modified radius margin bounds, which allow for comparing existing model selection criteria, and the training error in conjunction with the number of support vectors for designing sparse solutions.

1 Introduction

Model selection for supervised learning systems requires finding a suitable trade-off between at least two objectives, especially between model complexity and accuracy on a set of noisy training examples (→ bias vs. variance, capacity vs. empirical risk). Usually, this multi-objective problem is tackled by aggregating the objectives into a scalar function and applying standard methods to the resulting single-objective task. However, this approach can only lead to satisfactory solutions if the aggregation (e.g., a linear weighting of empirical error and regularization term) matches the problem. Thus, choosing an appropriate aggregation itself is an optimization task. A better way is to apply multi-objective optimization (MOO) to approximate the set of Pareto-optimal trade-offs and to choose a final solution afterwards, as discussed in the context of neural networks in [1, 2, 3, 4]. A solution is Pareto-optimal if it cannot be improved in any objective without getting worse in at least one other objective [5, 6, 7].

In the following, we consider MOO of the kernel and the regularization parameter of support vector machines (SVMs). We show how to reveal the trade-off between different objectives to guide the model selection process, e.g., for the design of sparse SVMs. One advantage of SVMs is that theoretically well founded bounds on the expected generalization performance exist, which can serve as model selection criteria.[1] However, in practice heuristic modifications of these bounds—e.g., corresponding to different weightings of capacity and empirical

[1] When used for model selection in the described way, the term "bound" is slightly misleading.

C. A. Coello Coello et al. (Eds.): EMO 2005, LNCS 3410, pp. 534–546, 2005.

risk—can lead to better results [8]. The MOO approach enables us to compare model selection criteria proposed in the literature after optimization.

As we consider only kernels from a parameterized family of functions, our model selection problem reduces to multidimensional real-valued optimization. We present a multi-objective evolution strategy (ES) with self-adaptation for real-valued MOO. The basics of MOO using non-dominated sorting [6, 9] and the ES are presented in the next section. Then SVMs and the model selection criteria we consider are briefly described in section 3.2 and the experiments in section 4.

2 Evolutionary Multi-objective Optimization

Consider an optimization problem with M objectives $f_1, \ldots, f_M : X \to \mathbb{R}$ to be minimized. The elements of X can be partially ordered using the concept of Pareto dominance. A solution $\boldsymbol{x} \in X$ dominates a solution $\boldsymbol{x}'$ and we write $\boldsymbol{x} \prec \boldsymbol{x}'$ iff $\exists m \in \{1, \ldots, M\} : f_m(\boldsymbol{x}) < f_m(\boldsymbol{x}')$ and $\nexists m \in \{1, \ldots, M\} : f_m(\boldsymbol{x}) > f_m(\boldsymbol{x}')$. The elements of the (Pareto) set $\{\boldsymbol{x} \,|\, \nexists \boldsymbol{x}' \in X : \boldsymbol{x}' \prec \boldsymbol{x}\}$ are called Pareto-optimal. Without any further information, no Pareto-optimal solution can be said to be superior to another. The goal of multi-objective optimization (MOO) is to find in a single trial a diverse set of Pareto-optimal solutions, which provide insights into the trade-offs between the objectives. When approaching a MOO problem by linearly aggregating all objectives into a scalar function, each weighting of the objectives yields only a limited subset of Pareto-optimal solutions. That is, various trials with different aggregations become necessary—but when the Pareto front (see below) is not convex, even this inefficient procedure does not help (cf. [6, 7]). Evolutionary multi-objective algorithms have become the method of choice for MOO [5, 6]. The most popular variant is the non-dominated sorting genetic algorithm NSGA-II, which shows fast convergence to the Pareto-optimal set and a good spread of solutions [6, 9]. On the other hand, evolution strategies (ES) are among the most elaborated and best analyzed evolutionary algorithms for real-valued optimization. Therefore, we propose a new method that combines ES with concepts from the NSGA-II.

2.1 Non-dominated Sorting

We give a concise description of the non-dominated sorting approach used in NSGA-II. For more details and efficient algorithms realizing this sorting we refer to [9]. First of all, the elements in a finite set $A \subseteq X$ of candidate solutions are ranked according to their level of non-dominance. Let the non-dominated solutions in A be denoted by $\mathrm{ndom}(A) = \{a \in A \,|\, \nexists a' \in A : a' \prec a\}$. The Pareto front of A is then given by $\{(f_1(a), \ldots, f_M(a)) \,|\, a \in \mathrm{ndom}(A)\}$. The elements in $\mathrm{ndom}(A)$ have rank 1. The other ranks are defined recursively by considering the set without the solutions with lower ranks. Formally, let $\mathrm{dom}_n(A) = \mathrm{dom}_{n-1}(A) \setminus \mathrm{ndom}_n(A)$ and $\mathrm{ndom}_n(A) = \mathrm{ndom}(\mathrm{dom}_{n-1}(A))$ for $n \in \{1, \ldots\}$ with $\mathrm{dom}_0 = A$. For $a \in A$ we define the level of non-dominance $\mathrm{r}(a, A)$ to be i iff $a \in \mathrm{ndom}_i(A)$.