10 Implementation

This chapter gives an overview of methods for solving the optimization problems specific to Support Vector Machines. Algorithms specific to other settings, such as Kernel PCA and Kernel Feature Analysis (Chapter 14), Regularized Principal Manifolds (Chapter 17), estimation of the support of a distribution (Chapter 8), Kernel Discriminant Analysis (Chapter 15), or Relevance Vector Machines (Chapter 16) can be found in the corresponding chapters. The large amount of code and number of publications available, and the importance of the topic, warrants this separate chapter on Support Vector implementations. Moreover, many of the techniques presented here are prototypical of the solutions of optimization problems in other chapters of this book and can be easily adapted to particular settings.

Overview

Due to the sheer size of the optimization problems arising in the SV setting we must pay special attention to how these problems can be solved efficiently. In Section 10.1 we begin with a description of strategies which can benefit almost all currently available optimization methods, such as universal stopping criteria, caching strategies and restarting rules. Section 10.2 details low rank approximations of the kernel matrix, $K \in \mathbb{R}^{m \times m}$. These methods allow the replacement of K by the outer product $ZZ^\top$ of a "tall and skinny" matrix $Z \in \mathbb{R}^{m \times n}$ where $n \ll m$. The latter can be used directly in algorithms whose speed improves with linear Support Vector Machines (SMO, Interior Point codes, Lagrangian SVM, and Newton's method).

Subsequently we present four classes of algorithms; interior point codes, subset selection, sequential minimization, and iterative methods. Interior Point methods are explained in Section 10.3. They are some of the most reliable methods for moderate problem sizes, yet their implementation is not trivial. Subset selection methods, as in Section 10.4, act as meta-algorithms on top of a basic optimization algorithm by carving out sets of variables on which the actual optimization takes place. Sequential Minimal Optimization, presented in Section 10.5, is a special case thereof. Due to the choice of only two variables at a time the restricted optimization problem can be solved analytically which obviates the need for an underlying base optimizer. Finally, iterative methods such as online learning, gradient descent, and Lagrangian Support Vector Machines are described in Section 10.6. Figure 10.1 gives a rough overview describing under which conditions which optimization algorithm is recommended.

Prerequisites

This chapter is intended for readers interested in implementing an SVM themselves. Consequently we assume that the reader is familiar with the basic concepts of both optimization (Chapter 6) and SV estimation (Chapters 1, 7, and 9).