

Overview

This chapter describes the central ideas of Support Vector (SV) learning in a nutshell. Its goal is to provide an overview of the basic concepts.

One such concept is that of a kernel. Rather than going immediately into mathematical detail, we introduce kernels informally as similarity measures that arise from a particular representation of patterns (Section 1.1), and describe a simple kernel algorithm for pattern recognition (Section 1.2). Following this, we report some basic insights from statistical learning theory, the mathematical theory that underlies SV learning (Section 1.3). Finally, we briefly review some of the main kernel algorithms, namely Support Vector Machines (SVMs) (Sections 1.4 to 1.6) and kernel principal component analysis (Section 1.7).

Prerequisites

We have aimed to keep this introductory chapter as basic as possible, whilst giving a fairly comprehensive overview of the main ideas that will be discussed in the present book. After reading it, readers should be able to place all the remaining material in the book in context and judge which of the following chapters is of particular interest to them.

As a consequence of this aim, most of the claims in the chapter are not proven. Abundant references to later chapters will enable the interested reader to fill in the gaps at a later stage, without losing sight of the main ideas described presently.

1.1 Data Representation and Similarity

Training Data

One of the fundamental problems of learning theory is the following: suppose we are given two classes of objects. We are then faced with a new object, and we have to assign it to one of the two classes. This problem can be formalized as follows: we are given empirical data

$$(x_1, y_1), \dots, (x_m, y_m) \in \mathcal{X} \times \{\pm 1\}. \quad (1.1)$$

Here, $\mathcal{X}$ is some nonempty set from which the *patterns* x_i (sometimes called *cases*, *inputs*, *instances*, or *observations*) are taken, usually referred to as the *domain*; the y_i are called *labels*, *targets*, *outputs* or sometimes also *observations*.¹ Note that there are

1. Note that we use the term pattern to refer to individual observations. A (smaller) part of the existing literature reserves the term for a generic *prototype* which underlies the data. The