



Chapter I

Kernel Methods: A Paradigm for Pattern Analysis

Nello Cristianini, University of Bristol, UK

John Shawe-Taylor, University College London, UK

Craig Saunders, University of Southampton, UK

Introduction

During the past decade, a major revolution has taken place in pattern-recognition technology with the introduction of rigorous and powerful mathematical approaches in problem domains previously treated with heuristic and less efficient techniques. The use of convex optimisation and statistical learning theory has been combined with ideas from functional analysis and classical statistics to produce a class of algorithms called kernel methods (KMs), which have rapidly become commonplace in applications. This book, and others, provides evidence of the practical applications that have made kernel methods a fundamental part of the toolbox for machine learning, statistics, and signal processing practitioners. The field of kernel methods has not only provided new insights and therefore new algorithms, but it has also created much discussion on well-established techniques such as Parzen windows and Gaussian processes, which use essentially the same technique but in different frameworks.

This introductory chapter will describe the main ideas of the kernel approach to pattern analysis, and discuss some of the reasons for their popularity. Throughout the chapter, we will assume that we have been given a set of data (be it made of vectors, sequences, documents, or any other format) and that we are interested in detecting relations existing within this data set. The ideas presented here have been introduced by many different researchers, but we will not point out the history behind these ideas, preferring to add pointers to books that address this field in a more coherent way.

The fundamental step of the kernel approach is to embed the data into a (Euclidean) space where the patterns can be discovered as linear relations. This capitalizes on the fact that over the past 50 years, statisticians and computer scientists have become very good at detecting linear relations within sets of vectors. This step therefore reduces many complex problems to a class of well-understood problems.

The embedding and subsequent analysis are performed in a *modular* fashion, as we will see below. The embedding map is defined implicitly by a so-called *kernel function*. This function depends on the specific data type and domain knowledge concerning the patterns that are to be expected in the particular data source. The second step is aimed at detecting relations within the embedded data set. There are many pattern-analysis algorithms that can operate with kernel functions, and many different types of kernel functions, each with different properties. Indeed, one reason for the popularity of the kernel approach is that these algorithms and kernels can be combined in a modular fashion. This strategy suggests a *software engineering approach* to learning systems' design through the breakdown of the task into subcomponents and the reuse of key modules.

In this introductory chapter, through the example of least squares linear regression, we will introduce all of the main ingredients of kernel methods. Though this example means that we will have restricted ourselves to the particular task of supervised regression, four key aspects of the approach will be highlighted.

1. Data items are embedded into a Euclidean space called the feature space.
2. Linear relations are sought among the images of the data items in the feature space.
3. The algorithms are implemented in such a way that the coordinates of the embedded points are not needed, only their pairwise inner products.
4. The pairwise inner products can be computed efficiently directly from the original data items using a kernel function.

These four observations will imply that, despite restricting ourselves to algorithms that optimise linear functions, our approach will enable the development of a rich toolbox of efficient and well-founded methods for discovering nonlinear relations in data through the use of nonlinear embedding mappings.

KMs offer a very general framework for performing pattern analysis on many types of data. The main idea of kernel methods is to embed the data set $S \subseteq X$ into a (possibly high-dimensional) vector space $\mathbb{R}^N$, and then to use linear pattern-analysis algorithms to detect relations in the embedded data. Linear algorithms are extremely efficient and well understood, both from a statistical and computational perspective. The embedding map is denoted here by ϕ , and it is understood that $\phi : X \rightarrow \mathbb{R}^N$ can be any set.

38 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/kernel-methods-paradigm-pattern-analysis/24817

Related Content

Moving Object Classification Under Illumination Changes Using Binary Descriptors

S. Vasavi, Ayesha Farha Shaikand Phani chaitanya Krishna Sunkara (2019). *Optoelectronics in Machine Vision-Based Theories and Applications* (pp. 188-232). www.irma-international.org/chapter/moving-object-classification-under-illumination-changes-using-binary-descriptors/209833

Medical Video Summarization using Central Tendency-Based Shot Boundary Detection

G. G. Lakshmi Priyaand S. Domnic (2013). *International Journal of Computer Vision and Image Processing* (pp. 55-65). www.irma-international.org/article/medical-video-summarization-using-central/78371

Key Frame Extraction from Video: Framework and Advances

Sergii Mashtalirand Olena Mikhnova (2014). *International Journal of Computer Vision and Image Processing* (pp. 68-79). www.irma-international.org/article/key-frame-extraction-from-video/115840

Evaluation Approach of Arabic Character Recognition

Hanan Aljuaid, Dzulkifli Mohamadand Muhammad Sarfraz (2013). *Intelligent Computer Vision and Image Processing: Innovation, Application, and Design* (pp. 128-145). www.irma-international.org/chapter/evaluation-approach-arabic-character-recognition/77037

Defect Detection Approach Based on Combination of Histogram Segmentation and Probabilistic Estimation Technique

Shervan Fekri Ershad (2013). *Intelligent Computer Vision and Image Processing: Innovation, Application, and Design* (pp. 230-238). www.irma-international.org/chapter/defect-detection-approach-based-combination/77043