Chapter 53

Positive Definite Kernels

This chapter is an introduction to positive definite kernels and the use of kernel functions in machine learning.

Let X be a nonempty set. If the set X represents a set of highly nonlinear data, it may be advantageous to map X into a space F of much higher dimension called the *feature space*, using a function $\varphi\colon X \to F$ called a *feature map*. This idea is that φ "unwinds" the description of the objects in F in an attempt to make it linear. The space F is usually a vector space equipped with an inner product $\langle -, - \rangle$. If F is infinite dimensional, then we assume that it is a Hilbert space.

Many algorithms that analyze or classify data make use of the inner products $\langle \varphi(x), \varphi(y) \rangle$, where $x, y \in X$. These algorithms make use of the function $\kappa\colon X \times X \to \mathbb{C}$ given by

$$\kappa(x, y) = \langle \varphi(x), \varphi(y) \rangle, \qquad x, y \in X,$$

called a *kernel function*.

The kernel trick is to pretend that we have a feature embedding $\varphi\colon X \to F$ (actually unknown), but to only use inner products $\langle \varphi(x), \varphi(y) \rangle$ that can be evaluated using the original data through the known kernel function κ. It turns out that the functions of the form κ as above can be defined in terms of a condition which is reminiscent of positive semidefinite matrices (see Definition 53.2). Furthermore, every function satisfying Definition 53.2 arises from a suitable feature map into a Hilbert space; see Theorem 53.8.

We illustrate the kernel methods on kernel PCA (see Section 53.4).

53.1 Feature Maps and Kernel Functions

Definition 53.1. Let X be a nonempty set, let H be a (complex) Hilbert space, and let $\varphi\colon X \to H$ be a function called a *feature map*. The function $\kappa\colon X \times X \to \mathbb{C}$ given by

$$\kappa(x, y) = \langle \varphi(x), \varphi(y) \rangle, \qquad x, y \in X,$$

is called a *kernel function*.