# Notes on Learning with Kernels

compiled by D. Gueorguiev 3/26/2024

## Data Representation and Similarity

Fundamental problem of Learning theory:

Suppose we are given two classes of objects. We are faced with a new object, and we have to assign it to one of two classes. We are given empirical data

(1.1)

Here, is some nonempty set from which the *inputs* or *patterns* are taken, usually referred to as the *domain*; the data points are known as *labels* or *observations*. Note that there are only two classes of patterns. This case is referred to *binary pattern recognition / binary classification*.

In order to study the problem of learning we need an additional type of structure. We want to be able to generalize to unseen data points. In the case of pattern recognition this means that given some new pattern , we want to predict the corresponding . Loosely speaking, we choose such that is in some sense similar to the training examples (1.1). To this end we need a notion of similarity in and in .

Let us consider a similarity measure of the form

(1.2)

that is, a function that, given two patterns and , returns a real number characterizing their similarity. We will assume that k is symmetric. This similarity function is known as a *kernel*.

Simple type of similarity measure is the dot product:

(1.3)

Note that the dot product is not sufficiently general to deal with important problems.

First,

## Literature

[Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, Bernhard Schoelkopf, Alexander J. Smola, MIT, 2002](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/books/scholkopf2002learning_with_kernels.pdf)