# 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, we need to be able to represent the patterns as vectors in some dot product space (which need not coincide with ). To this end, we define the map:

(1.5)

Second, even if the original patterns exist in a dot product space, we may still want to consider more general similarity measures obtained by applying a map (1.5). In that case, will typically be a nonlinear map.

Here, the space is called a *feature space*.

To summarize, embedding the data into via has three benefits:

1 ) it lets us define a similarity measure from the dot product in ,

(1.6)

2 ) it allows us to study patterns geometrically

3 ) the freedom to choose the mapping will enable us to design a large variety of similarity measures and learning algorithms.

Example: A Simple Pattern Recognition Algorithm

We assume that our data are embedded into a dot product space . Using the dot product, we can measure distances in this space. Using the dot product, we can measure distances in this space.

The basic idea of the algorithm is to assign a previously unseen pattern to the class with closer mean.

We compute the means of the two classes in feature space as:

(1.7)

(1.7)

where and are the number of examples with positive and negative labels. We assume that both classes are non-empty, i.e. .

## Literature

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