### Introduction

Support Vector Machines (SVMs) were first introduced by Vladimir Vapnik and represent a supervised classification technique. The model can be used for classification and regression.

### SVMs Advantages

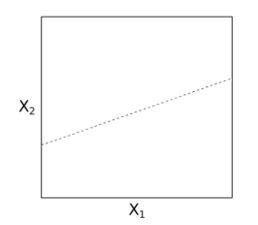
* High-Dimensionality
* Memory Efficiency
* Versatility

### SVMs Disadvantages

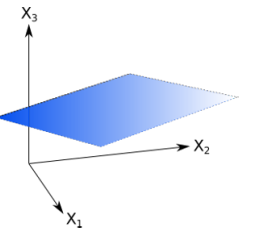
* Performs poorly when the number of training data samples is less than the number of features for each object
* Non-probabilistic algorithm – The classifier places objects above and bellow a hyperplane but there is no direct probabilistic interpretation of group membership

### Linear Separating Hyperplane

* Represents a key concept in SVM theory
* It is a one dimension object that divides the feature space in two regions
* Let’s consider a two dimensional feature space . In this case the hyperplane is just a straight line as shown bellow:



* The three-dimensional feature space:



* The general case . Let’s consider a p-dimensional feature space. . A hyperplane can be defined by the following equation:

(1)

(does not pass through the origin)

Equation (1) can be written as following:

(2)

Equation (2) can also be written:

(3)

Elements above the hyperplane satisfy:

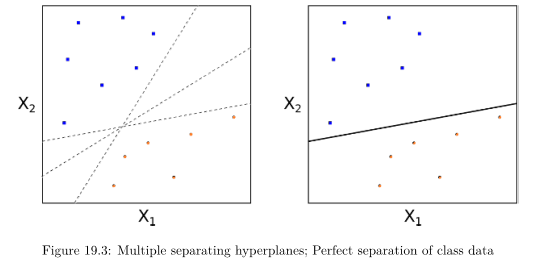
(4)

Elements bellow the hyperplane satisfy:

(5)

### Email classification using SVM

* Email spam filtering represents a classic example regarding classification using machine learning methods.
* Let’s consider 1,000 emails . A spam email is marked with +1 and non-spam is marked with -1.
* Email keywords represent *features*
* – the number of keywords. SVM allows a high number of features
* The next step is to transpose this problem into mathematic equations.
* – set of n training observations
* A class label is associated to each training observation. - class label where we consider if an email is span or non-spam.
* – n pairs for training observations representing features and class labels
* – test observations that can be used to test the performance of classifiers
* Goal: develop a classifier that can separate emails into spam or non-spam based on keywords.



if

if

– function with test observation

If then

If then

Goal: determine components of b.

### Maximal margin hyperplane

* There can be many hyperplanes obtained by translating or rotating a plane without touching any training observation
* The goal is to estimate the parameters of the optimal plane (maximal margin hyperplane)
* A maximal margin hyperplane (MMH) is the one is the furthest away from any training observations
* Margin – the smallest perpendicular distance to a training observation from the hyperplane
* The margin is the largest at MMH
* Overfitting problem can appear at MMH (the algorithm is a good fit for training data but performs poorly at testing data)

### Constructing the maximal margin hyperplane

training observations and class labels   
Optimization problem:

,

### Support Vector Machines

An inner product is defined for two-dimensional vectors as following:

For two observations, an inner product is defined as following:

A support vector classifier for a particular observation can be represented as a linear combination of inner products:

coefficients, one for each of the training observations.

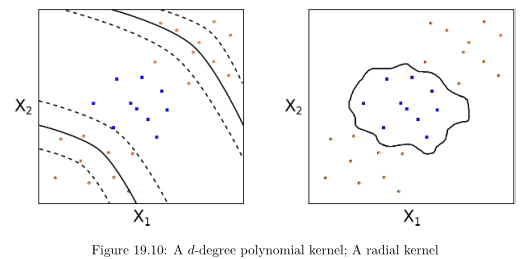
There is no need to calculate all the inner products between pairs of the training observations. We have to calculate only the inner products for the support vectors (subset of training observations). We call this subset .

if

– kernel function

Linear kernel:

Polynomial kernels of degree :



Popular radial kernel:

Standard Euclidean distance