UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Lecture 11 Support Vector Machine I

STAT 441/505: Applied Statistical Methods in Data Mining

Linglong Kong

Department of Mathematical and Statistical Sciences
University of Alberta

Winter, 2016

Outline

Separable Hyperplanes

- Imagine a situation where you have a two class classification problem with two predictors X_1 and X_2.
- Suppose that the two classes are linearly separable i.e. one can draw a straight line in which all points on one side belong to the first class and points on the other side to the second class.
- Then a natural approach is to find the straight line that gives the biggest separation between the classes i.e. the points are as far from the line as possible
- This is the basic idea of a support vector classifier.

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

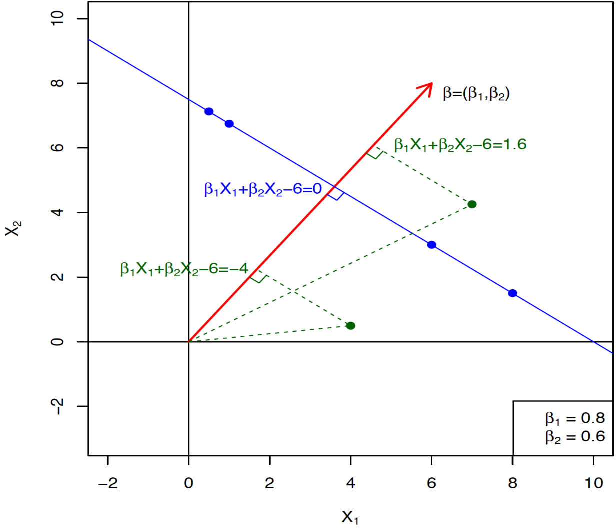
Hyperplane

- A hyperplane in p dimensions is a flat affine subspace of dimension $p-1$.
- In general the equation for a hyperplane has the form

$$\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p = 0.$$

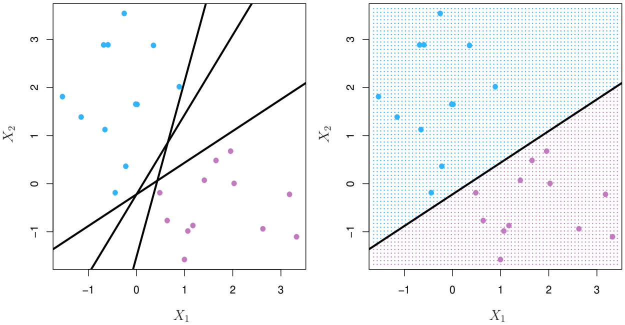
- In $p = 2$ dimensions a hyperplane is a line.
- If $\beta_0 = 0$, the hyperplane goes through the origin, otherwise not.
- The vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ is called the normal vector — it points in a direction orthogonal to the surface of a hyperplane.

Hyperplane in 2 Dimensions



UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Separating Hyperplane



- If $f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$, then $f(X) > 0$ for points on one side of the hyperplane, and $f(X) < 0$ for points on the other.
- If we code the colored points as $Y_i = +1$ as blue and $Y_i = -1$ as purple, then if $Y_i \cdot f(X_i) > 0$ for all i, $f(X) = 0$ defines a Separating Hyperplane.

Hard Margin

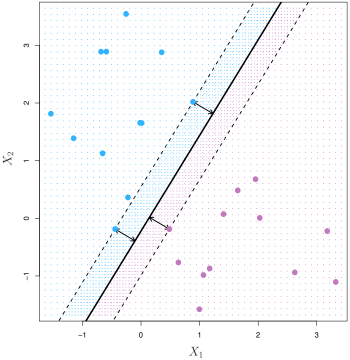
- Among all separating hyperplanes, find the one that makes the biggest gap or margin between the two classes.
- Constrained optimization problem

$$\text{maximize}_{\beta_0,\beta_1,\cdots,\beta_p} M$$

$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p) \geq M$$

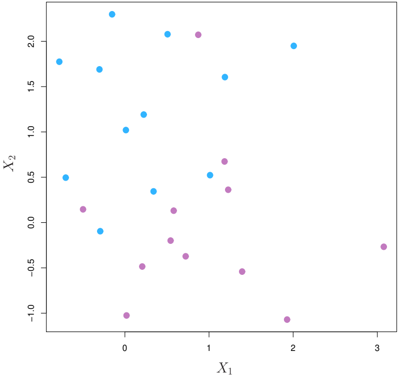
$$\text{for } i = 1, \cdots, n.$$



- This can be rephrased as a convex quadratic program, and solved efficiently. The function `svm()` in package `e1071` solves this problem efficiently.

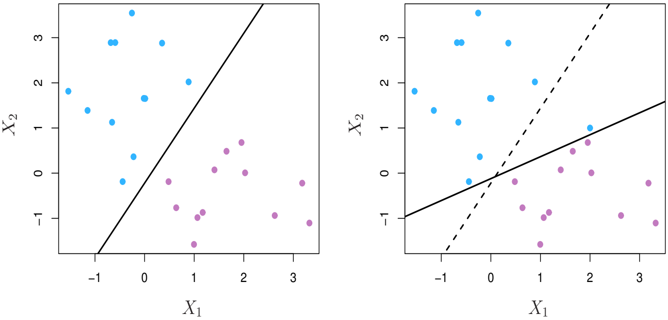
Hard Margin

- The data on the left are not separable by a linear boundary.
- In general it is true for $n < p$.



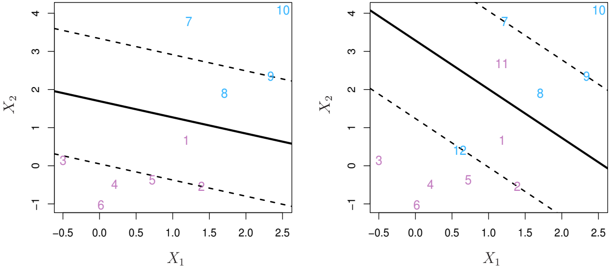
UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Hard Margin



- Sometimes the data are separable, but noisy. This can lead to a poor solution for the maximal-margin (hard margin) classifier. boundary.

Soft Margin



- The support vector classifier maximizes a soft margin.

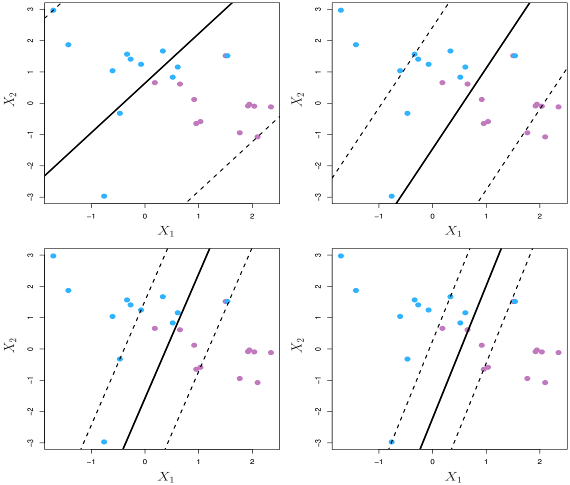
$$\text{maximize}_{\beta_0,\beta_1,\cdots,\beta_p;\epsilon_1,\cdots\epsilon_n} M; \text{ subject to } \sum_{j=1}^{p} \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p) \geq M(1 - \epsilon_i)$$

$$\epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C.$$

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

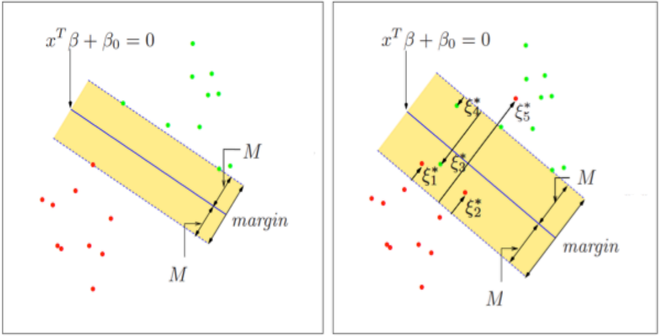
C is a regularization parameter



- C is a regularization parameter and represent the price we need to pay to separate the two classes.

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

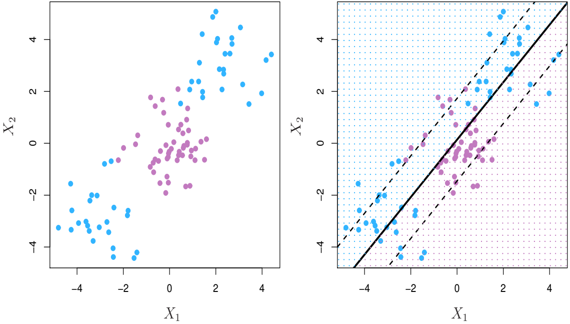
Support Vectors



- Only those support vectors determine the optimization solution for both hard margin and soft margin.

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Linear boundary can fail



- Sometime a linear boundary simply won't work, no matter what value of C.
- For example, in the situation shown above.
- What do we do? the kernel trick!!!

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Summary and Remark

- Hyperplane
- Support vector classifier
- Read textbook Chapter 12 and `R` code
- Do `R` lab