Erasmus School of Economics

Machine Learning

FEM31002

Support Vector Machines

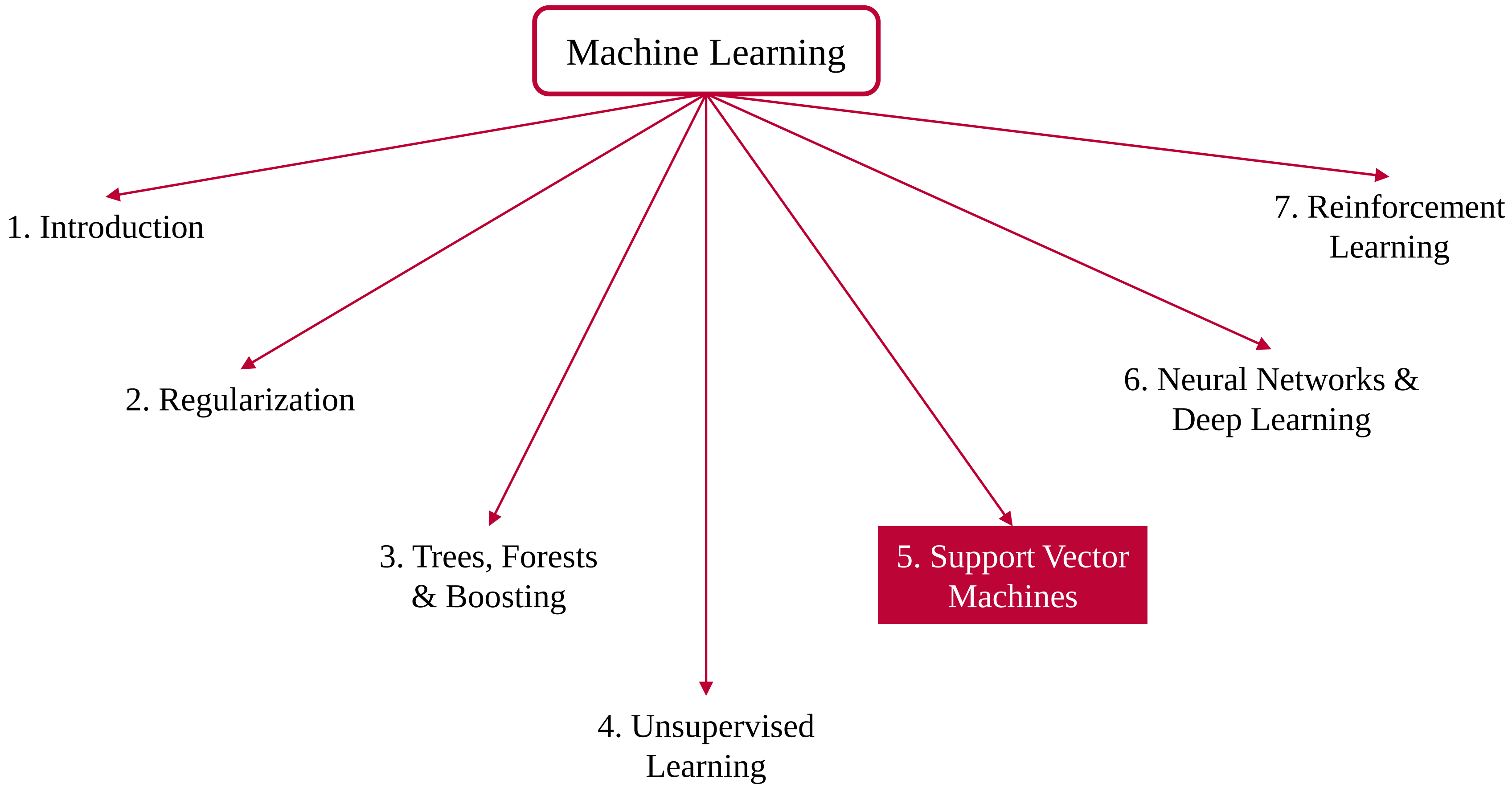
Part 1

Ilker Birbil

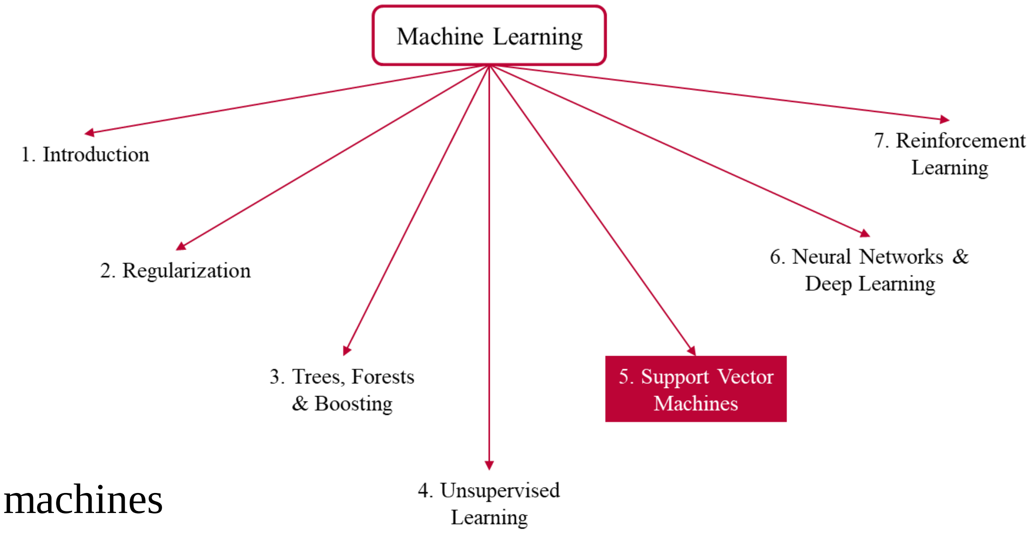
birbil@ese.eur.nl

ERASMUS UNIVERSITEIT ROTTERDAM

Outline


Machine Learning
1. Introduction
2. Regularization
3. Trees, Forests & Boosting
4. Unsupervised Learning
5. Support Vector Machines
6. Neural Networks & Deep Learning
7. Reinforcement Learning

Outline



- Geometry of support vector machines
- Dual problem
- Path to kernels
- Simple kernels

Support Vector Machines

$$(x_i, y_i), i = 1, 2, \ldots, n$$

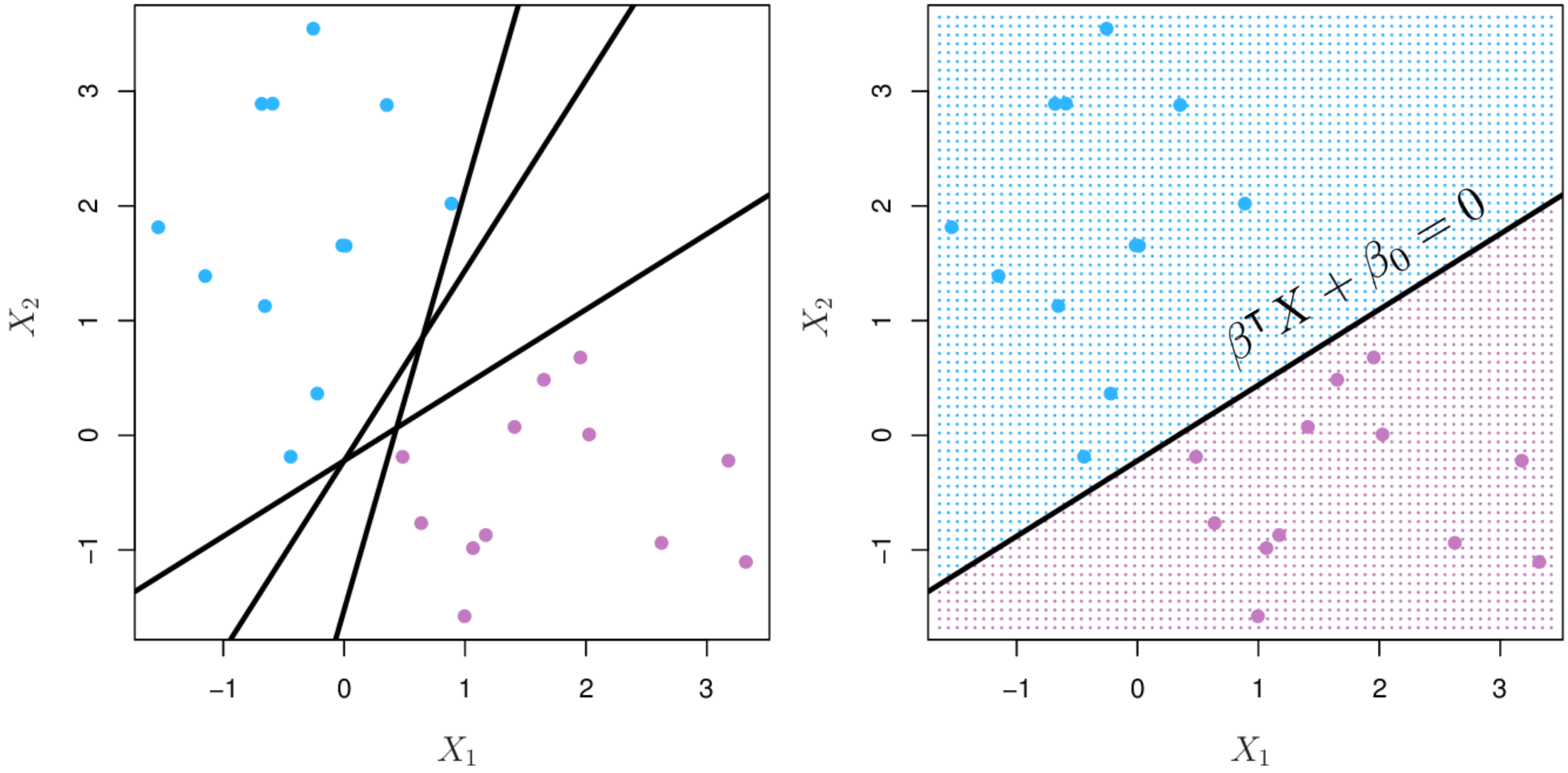
$$x_i \in \mathbb{R}^p \qquad\qquad y_i \in \{-1, +1\}$$

Idea: Come up with a hyperplane so that the data points are classified according to the side of the hyperplane (halfspace) that they reside

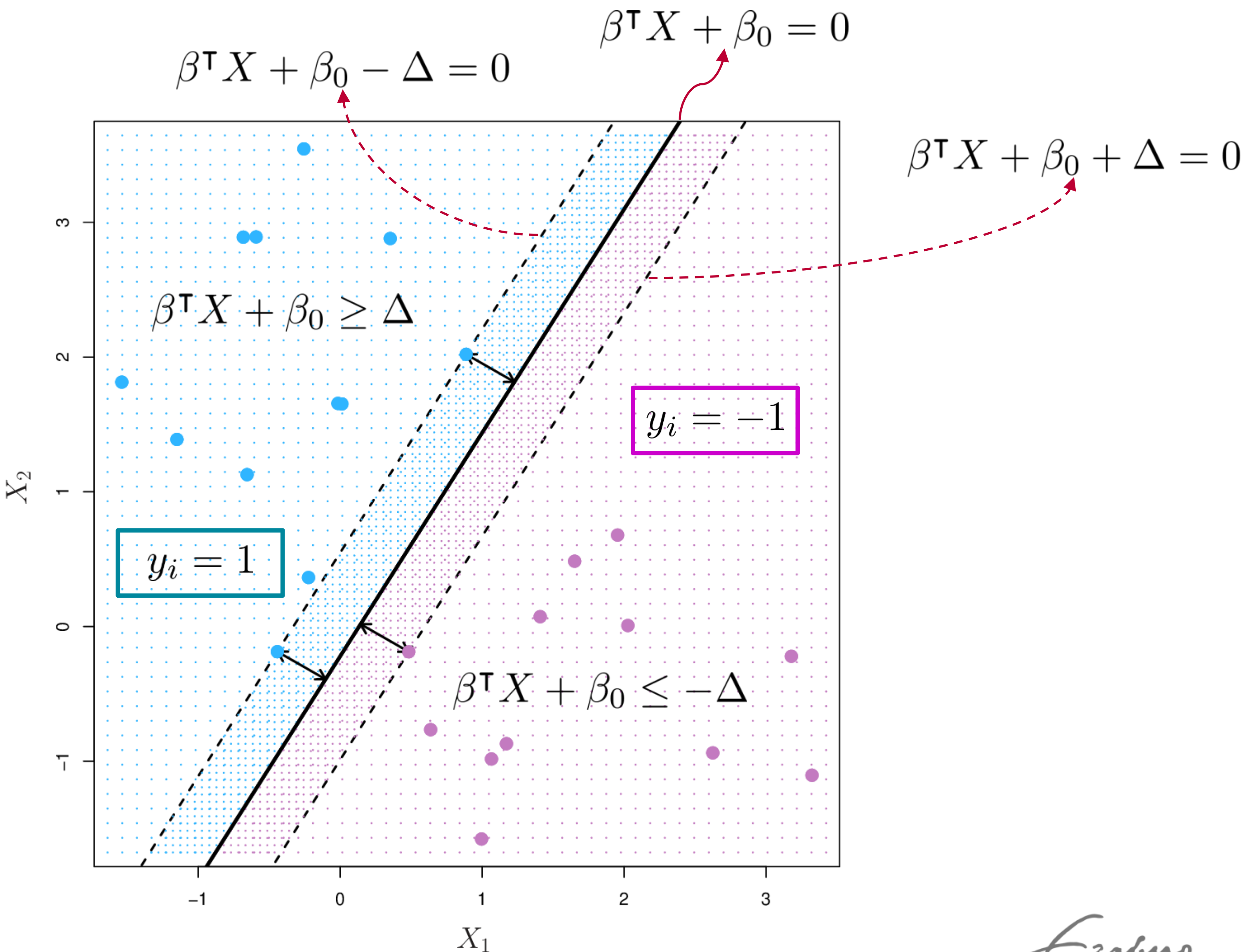
Hyperplane

$$\{X \in \mathbb{R}^p : \beta^\intercal X + \beta_0 = 0 \text{ for some } \beta \in \mathbb{R}^p, \beta_0 \in \mathbb{R}\}$$

Perfectly Separable Data



Perfectly Separable Data



Perfectly Separable Data

Objective: Choose β and Δ such that the distance between the two hyperplanes

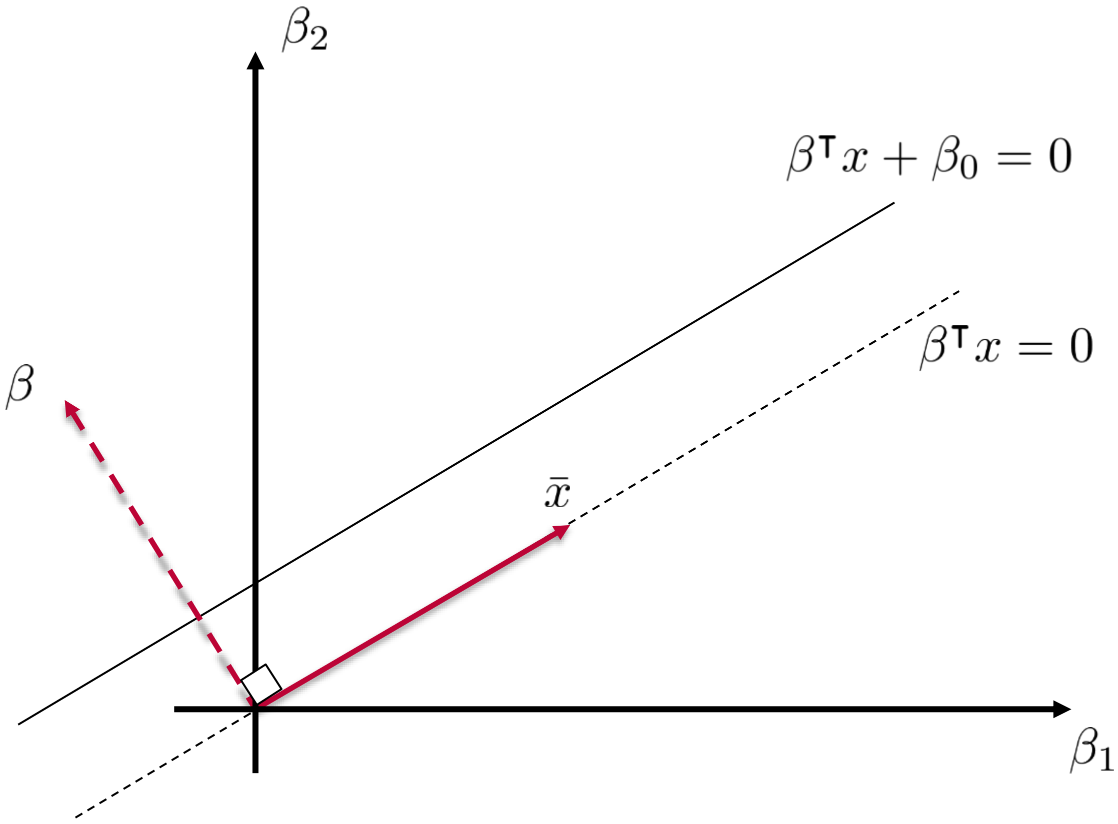
$$\beta^\intercal X + \beta_0 - \Delta = 0 \quad \text{and} \quad \beta^\intercal X + \beta_0 + \Delta = 0$$

is as large as possible (with scaling we can set $\Delta = 1$)

$$(x_i, y_i) \begin{cases} y_i = 1 & \longrightarrow \quad \beta^\intercal x_i + \beta_0 - 1 \geq 0 \\ y_i = -1 & \longrightarrow \quad \beta^\intercal x_i + \beta_0 + 1 \leq 0 \end{cases}$$

$$y_i(\beta^\intercal x_i + \beta_0) \geq 1$$

Perfectly Separable Data



$$\beta = (\beta_1, \beta_2)^\intercal$$

β is normal to the subspace

$$\beta^\intercal x = 0$$

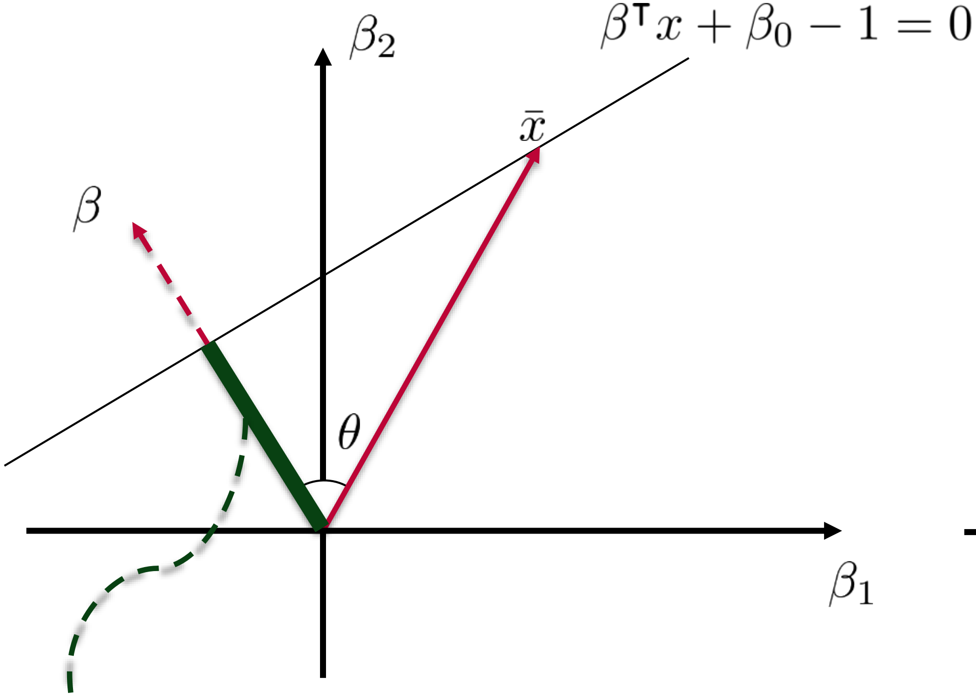
β is also normal to

$$\beta^\intercal x + \beta_0 = 1$$

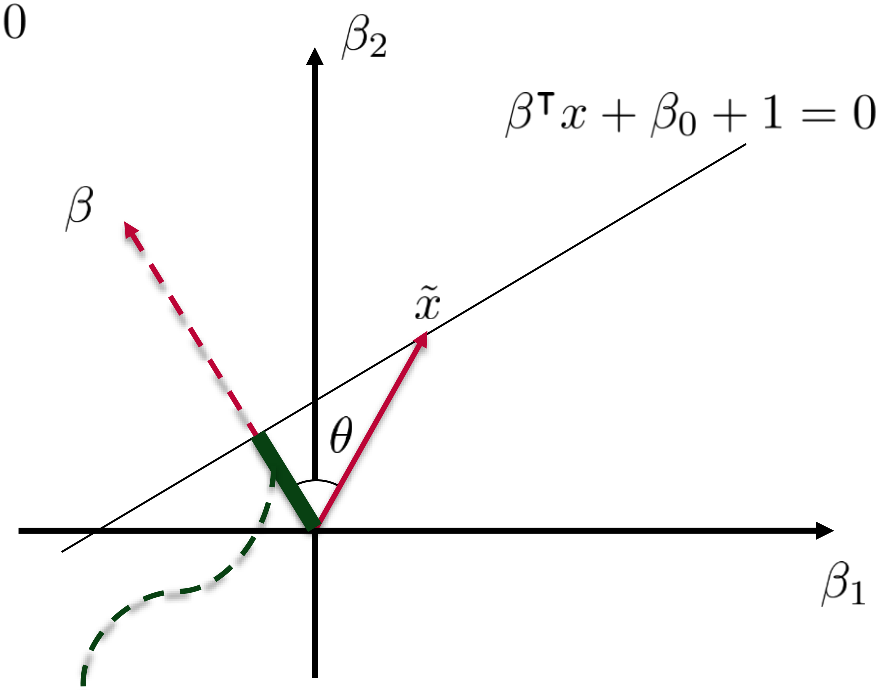
and

$$\beta^\intercal x + \beta_0 = -1$$

Perfectly Separable Data



$$\|\bar{x}\| \cos\theta = \frac{\beta^\intercal \bar{x}}{\|\beta\|} = \frac{1-\beta_0}{\|\beta\|}$$



$$\|\tilde{x}\| \cos\theta = \frac{\beta^\intercal \tilde{x}}{\|\beta\|} = \frac{-1-\beta_0}{\|\beta\|}$$

The distance between the two hyperplanes: $\dfrac{1-\beta_0}{\|\beta\|} - \dfrac{-1-\beta_0}{\|\beta\|} = \dfrac{2}{\|\beta\|}$

Perfectly Separable Data

$$
\begin{array}{ll}
\text{maximize} & \dfrac{2}{||\beta||} \\
\text{subject to} & y_i(\beta^\intercal x_i + \beta_0) \geq 1, \quad i = 1, 2, \ldots, n
\end{array}
$$

$$
\equiv \begin{array}{ll}
\text{minimize} & \frac{1}{2}||\beta|| \\
\text{subject to} & y_i(\beta^\intercal x_i + \beta_0) \geq 1, \quad i = 1, 2, \ldots, n
\end{array}
$$

$$
\equiv \begin{array}{ll}
\text{minimize} & \frac{1}{2}\beta^\intercal \beta \\
\text{subject to} & y_i(\beta^\intercal x_i + \beta_0) \geq 1, \quad i = 1, 2, \ldots, n
\end{array}
$$

This is an optimization problem with convex quadratic objective function and linear constraints. This type of problem is known as *quadratic programming*.

Perfectly Separable Data