

Optimal Hyperplanes

CS4780 – Machine Learning
Fall 2009

Thorsten Joachims
Cornell University

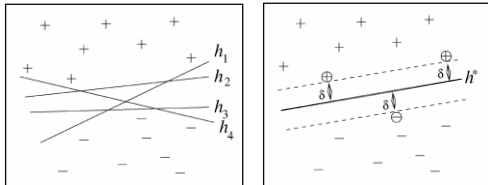
Reading: Schoelkopf/Smola Chapter 7.1-7.3, 7.5 (online)

Outline

- Optimal hyperplanes and margins
- Hard-margin Support Vector Machine
- Primal optimization problem
- Soft-margin Support Vector Machine

Optimal Hyperplanes Linear Hard-Margin Support Vector Machine

Assumption: Training examples are linearly separable.

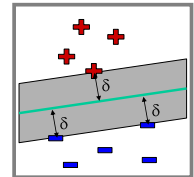


Hard-Margin Separation

Goal: Find hyperplane with the largest distance to the closest training examples.

Optimization Problem (Primal):

$$\begin{aligned} \min_{\vec{w}, b} \quad & \frac{1}{2} \vec{w} \cdot \vec{w} \\ \text{s.t.} \quad & y_1(\vec{w} \cdot \vec{x}_1 + b) \geq 1 \\ & \dots \\ & y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1 \end{aligned}$$

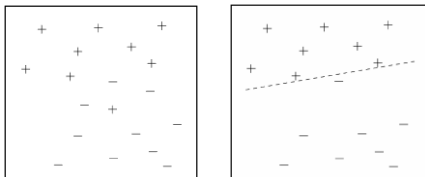


Support Vectors: Examples with minimal distance (i.e. margin).

Non-Separable Training Data

Limitations of hard-margin formulation

- For some training data, there is no separating hyperplane.
- Complete separation (i.e. zero training error) can lead to suboptimal prediction error.



Soft-Margin Separation

Idea: Maximize margin and minimize training error.

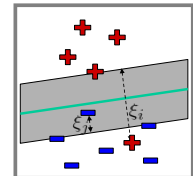
Hard-Margin OP (Primal):

$$\begin{aligned} \min_{\vec{w}, b} \quad & \frac{1}{2} \vec{w} \cdot \vec{w} \\ \text{s.t.} \quad & y_1(\vec{w} \cdot \vec{x}_1 + b) \geq 1 \\ & \dots \\ & y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1 \end{aligned}$$

Soft-Margin OP (Primal):

$$\begin{aligned} \min_{\vec{w}, \xi, b} \quad & \frac{1}{2} \vec{w} \cdot \vec{w} + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_1(\vec{w} \cdot \vec{x}_1 + b) \geq 1 - \xi_1 \wedge \xi_1 \geq 0 \\ & \dots \\ & y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1 - \xi_n \wedge \xi_n \geq 0 \end{aligned}$$

- Slack variable ξ_i measures by how much (x_i, y_i) fails to achieve margin δ
- $\sum \xi_i$ is upper bound on number of training errors
- C is a parameter that controls trade-off between margin and training error.



Controlling Soft-Margin Separation

- $\sum \xi_i$ is upper bound on number of training errors
- C is a parameter that controls trade-off between margin and training error.

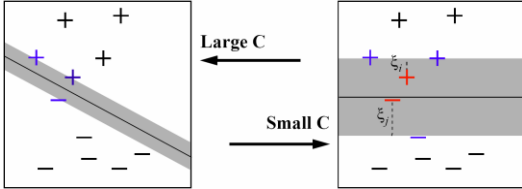
Soft-Margin OP (Primal):

$$\min_{\vec{w}, \xi, b} \frac{1}{2} \vec{w} \cdot \vec{w} + C \sum_{i=1}^n \xi_i$$

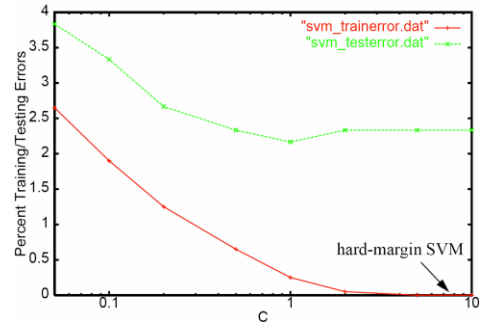
$$s.t. \ y_1(\vec{w} \cdot \vec{x}_1 + b) \geq 1 - \xi_1 \wedge \xi_1 \geq 0$$

$$\dots$$

$$y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1 - \xi_n \wedge \xi_n \geq 0$$



Example Reuters “acq”: Varying C



Example: Margin in High-Dimension

Training Sample S_{train}	\vec{x}							y
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
(\vec{x}_1, y_1)	1	0	0	1	0	0	0	1
(\vec{x}_2, y_2)	1	0	0	0	1	0	0	1
(\vec{x}_3, y_3)	0	1	0	0	0	1	0	-1
(\vec{x}_4, y_4)	0	1	0	0	0	0	1	-1
	\vec{w}							b
	w_1	w_2	w_3	w_4	w_5	w_6	w_7	
Hyperplane 1	1	1	0	0	0	0	0	2
Hyperplane 2	0	0	0	1	1	-1	-1	0
Hyperplane 3	1	-1	1	0	0	0	0	0
Hyperplane 4	0.5	-0.5	0	0	0	0	0	0
Hyperplane 5	1	-1	0	0	0	0	0	0
Hyperplane 6	0.95	-0.95	0	0.05	0.05	-0.05	-0.05	0