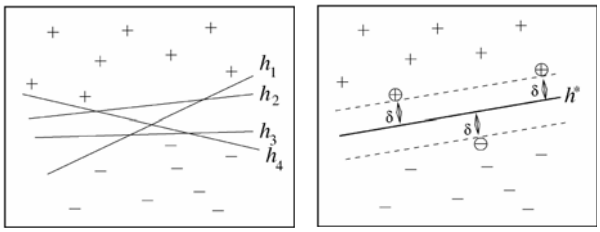
Optimal Hyperplanes

Assumption: Training examples are linearly separable.



Definition: *For a linear classifier $h_{\vec{w},b}$, the* **margin** δ *of an example $(\vec{x}, y)$ is $\delta = y(\vec{w} \cdot \vec{x} + b)$.*

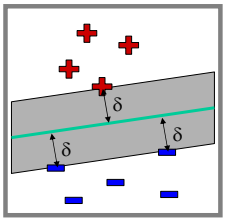
Definition: *The margin is called* geometric margin, *if $||\vec{w}|| = 1$. Otherwise,* functional margin.

Hard-Margin Separation

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

Optimization Problem (Primal):

$$\min_{\vec{w},b} \frac{1}{2}\vec{w} \cdot \vec{w}$$
$$s.t. \quad y_1(\vec{w} \cdot \vec{x}_1 + b) \geq 1$$
$$\ldots$$
$$y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1$$



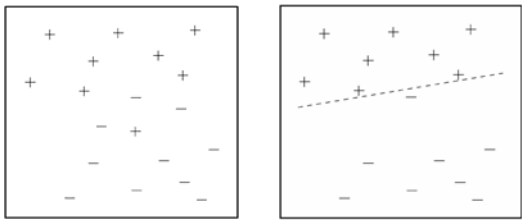
Support Vectors: Examples with minimal distance (i.e. margin).

Definition: *The (hard) margin of a linear classifier $h_{\vec{w},b}$ on data D is $\delta = min_{(\vec{x},y)\in D}\{y(\vec{w} \cdot \vec{x} + b)\}$.*

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.

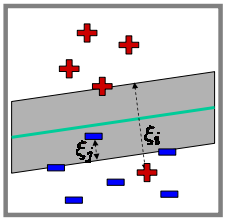
Hard-Margin OP (Primal):

$$\min_{\vec{w},b} \frac{1}{2}\vec{w} \cdot \vec{w}$$
$$s.t. \quad y_1(\vec{w} \cdot \vec{x}_1 + b) \geq 1$$
$$\ldots$$
$$y_n(\vec{w} \cdot \vec{x}_n + b) \geq 1$$

Soft-Margin OP (Primal):

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

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

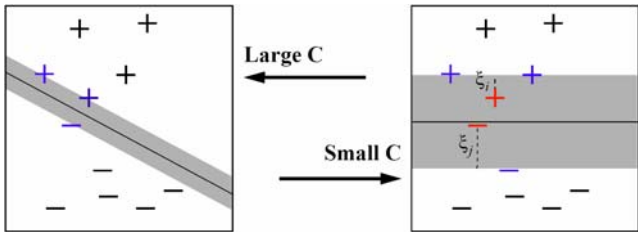


Controlling Soft-Margin Separation

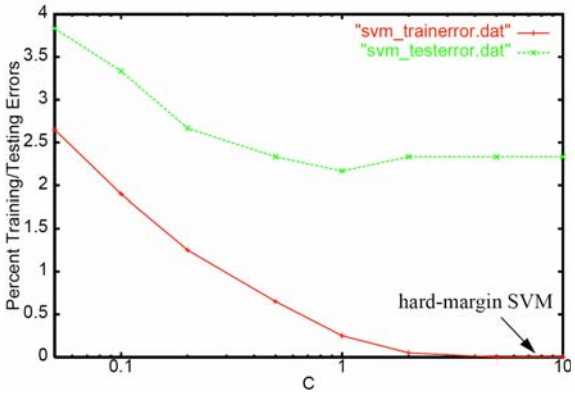
- $\Sigma\xi_i$ is upper bound on number of training errors
- C is a parameter that controls trade-off between margin and training error.

Soft-Margin OP (Primal):

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



Example Reuters "acq": Varying C



Example: Margin in High-Dimension

D_{train}	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
Example 1	1	0	0	1	0	0	0	1
Example 2	1	0	0	0	1	0	0	1
Example 3	0	1	0	0	0	1	0	-1
Example 4	0	1	0	0	0	0	1	-1
	w_1	w_2	w_3	w_4	w_5	w_6	w_7	**b**
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	1	-1	0	0	0	0	0	0
Hyperplane 5	0.95	-0.95	0	0.05	0.05	-0.05	-0.05	0