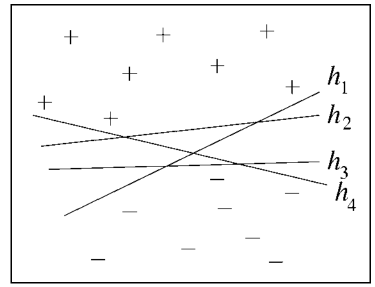
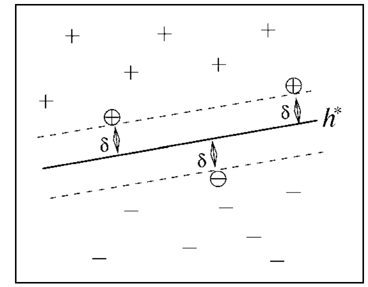
Optimal Hyperplanes

Outline:
- What is an optimal hyperplane / Support Vector Machine?
- Hard-margin separation.
- What to do, if the training set is not linearly separable?
- Soft-margin separation.

Optimal Hyperplane
Linear Hard-Margin Support Vector Machine

Assumption: The training examples are linearly separable.





Hard-Margin Separation

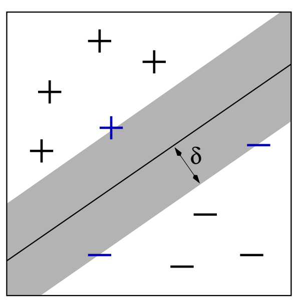
Goal: Find hyperplane with the largest distance to closest examples.

Hard Margin Optimization Problem (Primal):

minimize $J(\vec{w}, b) = \frac{1}{2}\vec{w} \cdot \vec{w}$

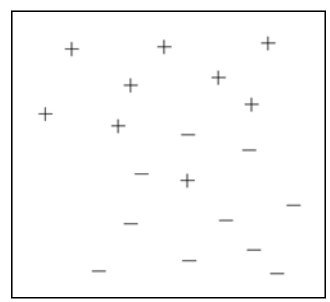
s. t. $y_i[\vec{w} \cdot \vec{x}_i + b] \geq 1$

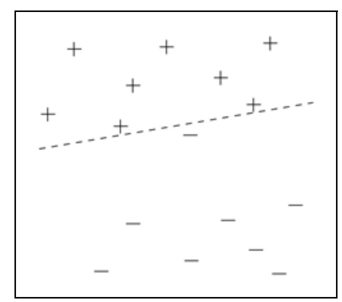
Support Vectors: Examples with minimal distance.



Non-Separable Training Samples

- For some training samples there is no separating hyperplane!
- Complete separation is suboptimal for many training samples!





=> minimize trade-off between margin and training error.

Soft-Margin Separation

Idea: Maximize margin and minimize training error simultanously.

Hard Margin:

minimize $J(\vec{w}, b) = \frac{1}{2}\vec{w} \cdot \vec{w}$

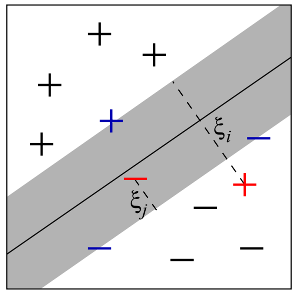
s. t. $y_i[\vec{w} \cdot \vec{x}_i + b] \geq 1$

Soft Margin:

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

s. t. $y_i[\vec{w} \cdot \vec{x}_i + b] \geq 1 - \xi_i$ *and* $\xi_i \geq 0$

- slack variable ξ_i measures by how much example $(\vec{x}_i, y_i)$ fails to achieve a target margin of δ.
- $\sum \xi_i$ is an upper bound on the number of training errors.
- C is a parameter that controls trade-off between margin and error.

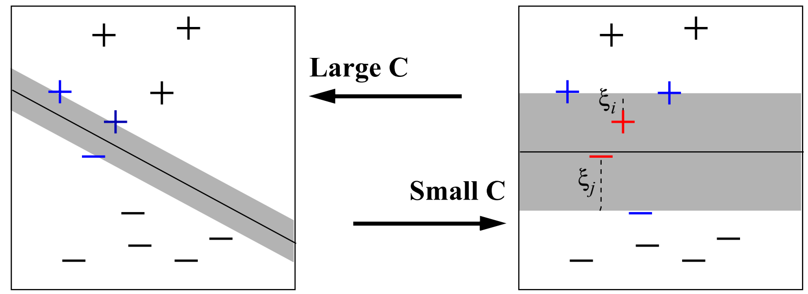


Controlling Soft-Margin Separation

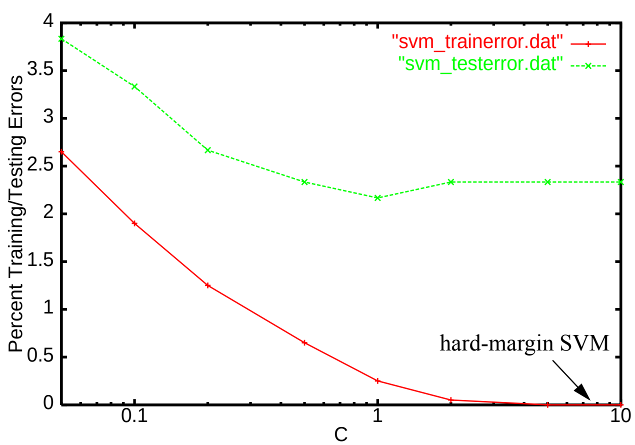
Soft Margin: minimize $P(\vec{w}, b, \vec{\xi}) = \frac{1}{2}\vec{w} \cdot \vec{w} + C\sum_{i=1}^{n}\xi_i$

s. t. $y_i[\vec{w} \cdot \vec{x}_i + b] \geq 1 - \xi_i$ *and* $\xi_i \geq 0$

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



Example Reuters “ acq” : Varying C



Observation: Typically no local optima, but not necessarily...