Maximum Margin Hyperplane

6.036 Introduction to Machine Learning

Online Algorithms

<u>Prediction Game</u>

L: $\theta^{(0)} = 0\ (vector)$

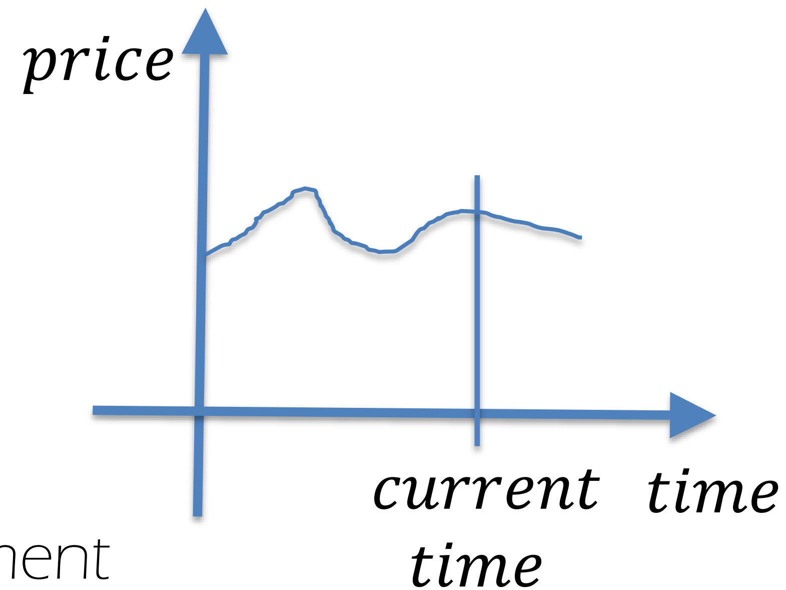
Round k

N: selects $x^{(i)} \in R^d$

L: predicts $sign(\theta^{(k-1)} \cdot x^{(i)})$

N: reveals $y^{(i)}$

L: suffers Loss $\left(\overbrace{y^{(i)}\theta^{(k-1)} \cdot x^{(i)}}^{\text{agreement}}\right)$, updates θ

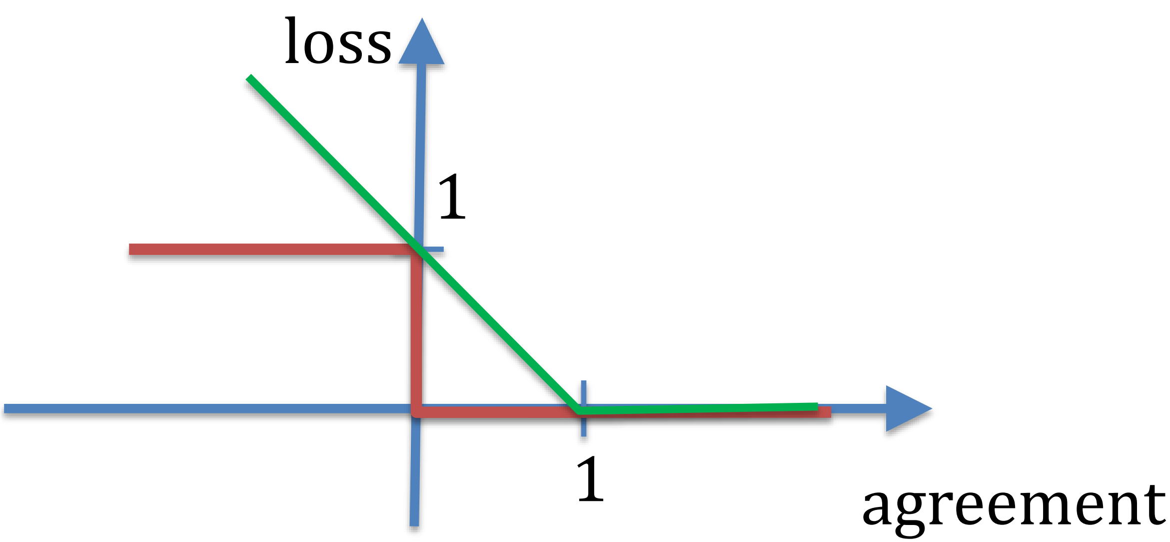


Goal: to minimize the overall loss

Loss Functions

Zero-one loss $$L_{01}(y\theta \cdot x) = [[y\theta \cdot x \leq 0]]$$

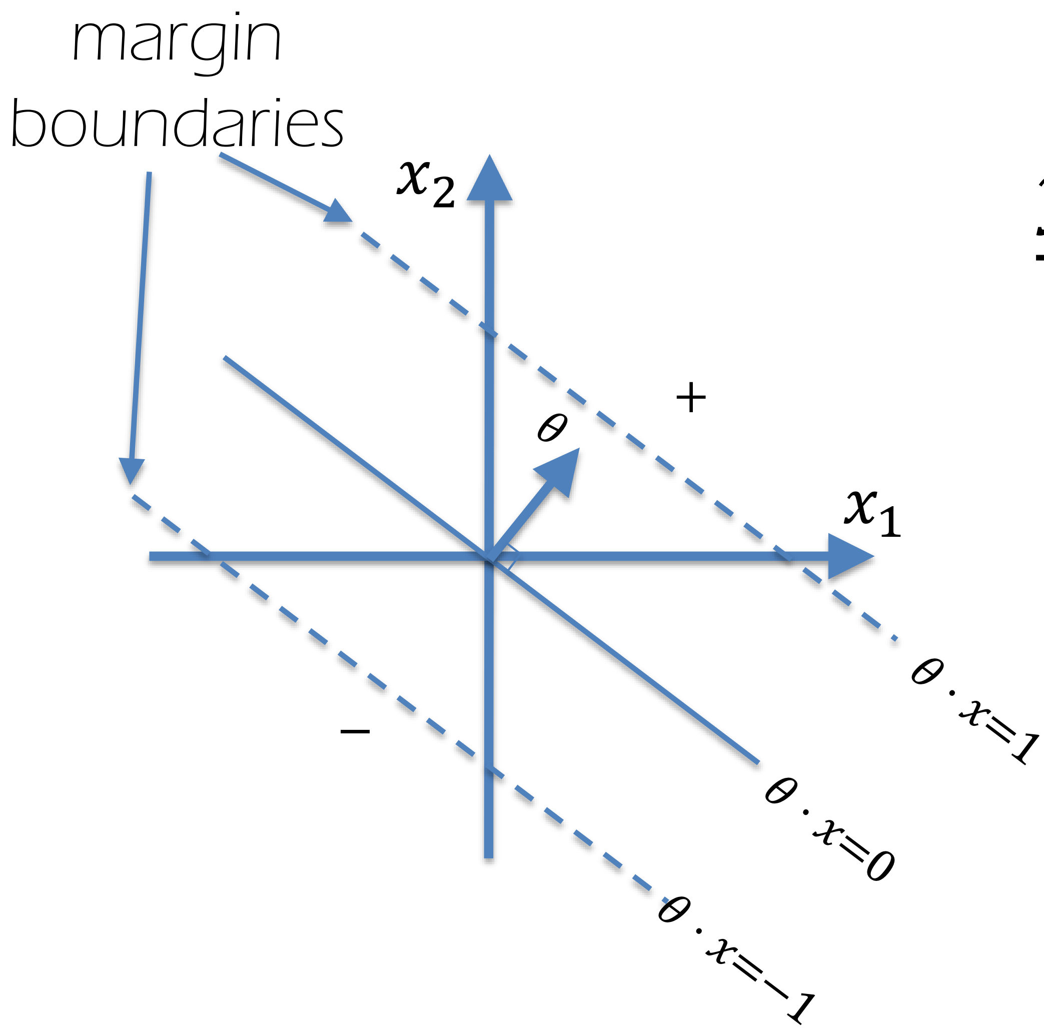
Hinge loss $$L_h(y\theta \cdot x) = \begin{cases} 1 - y\theta \cdot x, & y\theta \cdot x \leq 1 \\ 0, & otherwise \end{cases}$$



Hinge loss

$$\underbrace{\frac{y\theta \cdot x}{\|\theta\|}}_{\text{distance to decision boundary}}$$

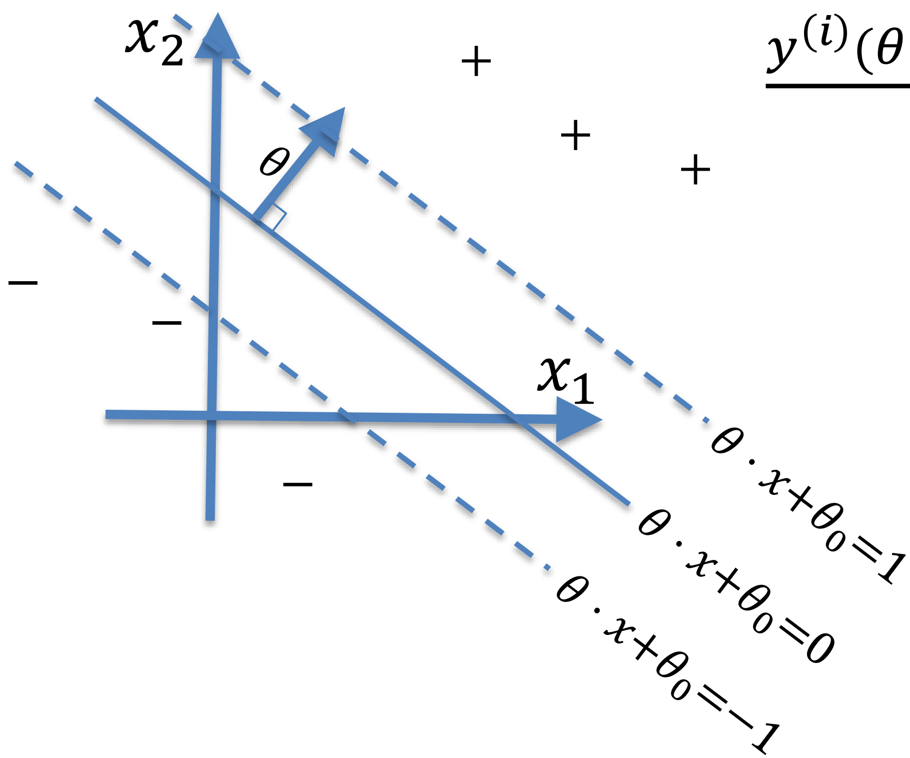
$$\underbrace{\frac{y\theta \cdot x}{\|\theta\|} = \frac{1}{\|\theta\|}}_{\text{margin}}$$



Max-margin Hyperplane

distance to the decision boundary

$$\frac{y^{(i)}(\theta \cdot x^{(i)} + \theta_0)}{\|\theta\|} \geq \frac{1}{\|\theta\|}, i = 1 \ldots n$$



Max-margin separator
(support vector machine)

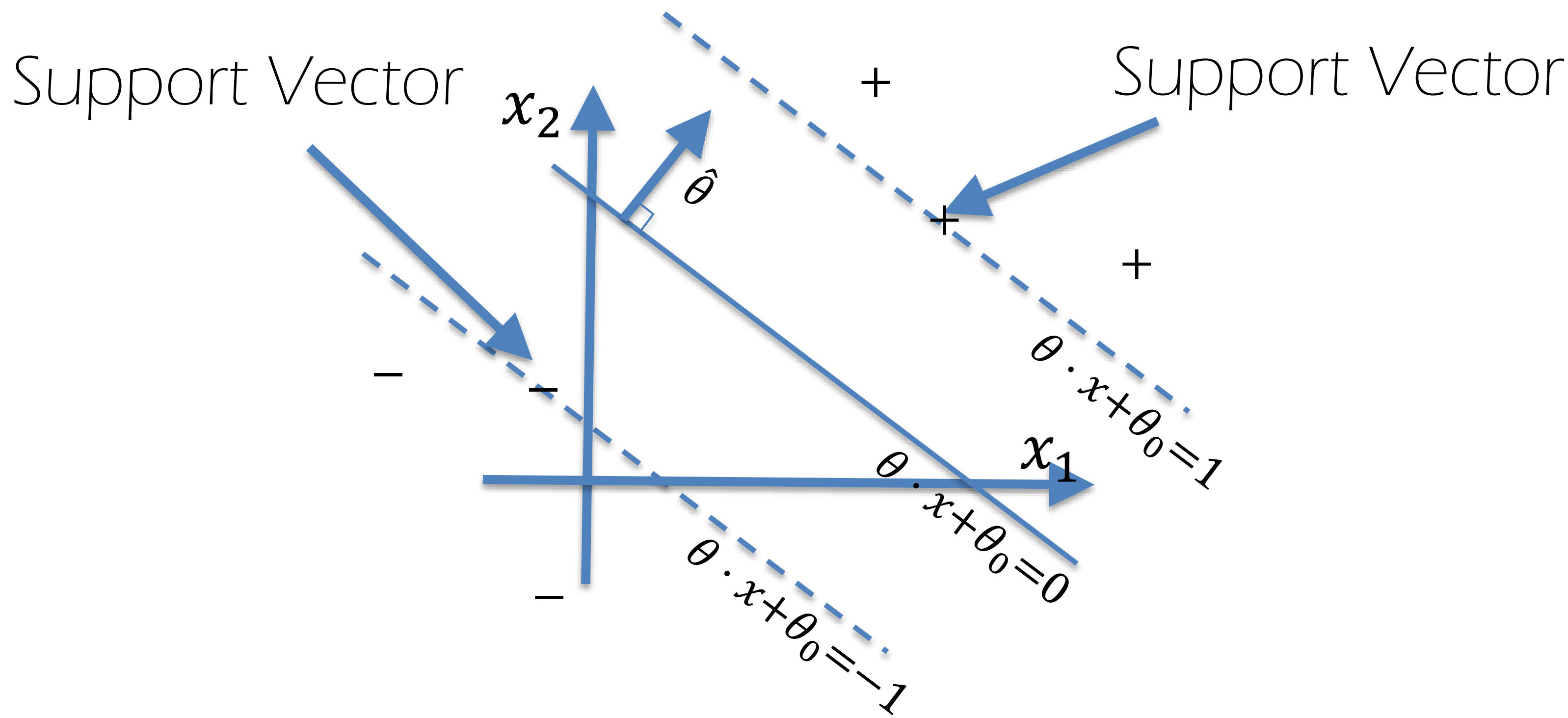
$$\max \frac{1}{\|\theta\|}$$

$$\text{s.t.} \qquad \frac{y^{(i)}(\theta \cdot x^{(i)} + \theta_0)}{\|\theta\|} \geq \frac{1}{\|\theta\|}, i = 1 \ldots n$$

$$\min \ \frac{1}{2}\|\theta\|^2$$

$$\text{s.t.} \quad y^{(i)}(\theta \cdot x^{(i)} + \theta_0) \ \geq 1 \ , i = 1 \ldots n$$

Support Vector Machine



SVM properties

1) Unique (at least one pos. & one neg. example)

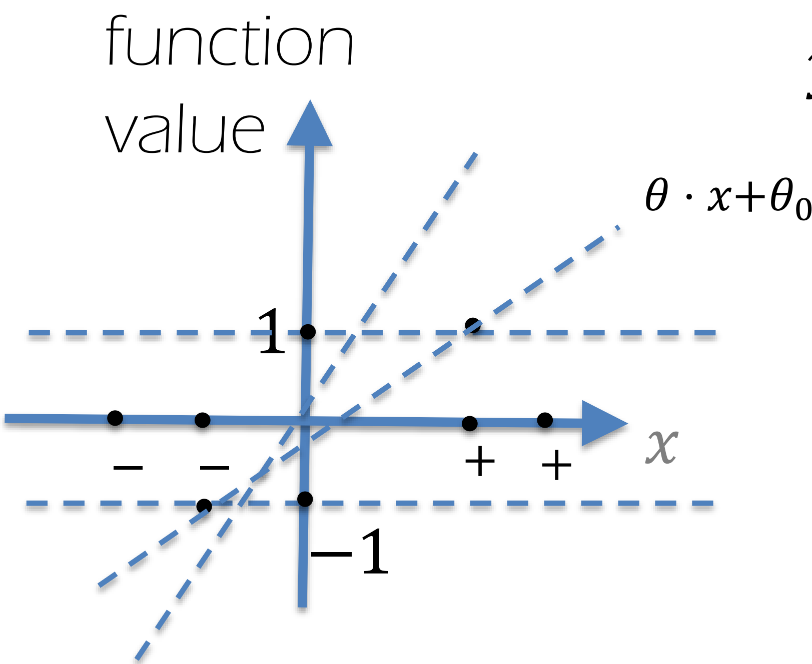
2) Sparse (few support vectors)

1D Example

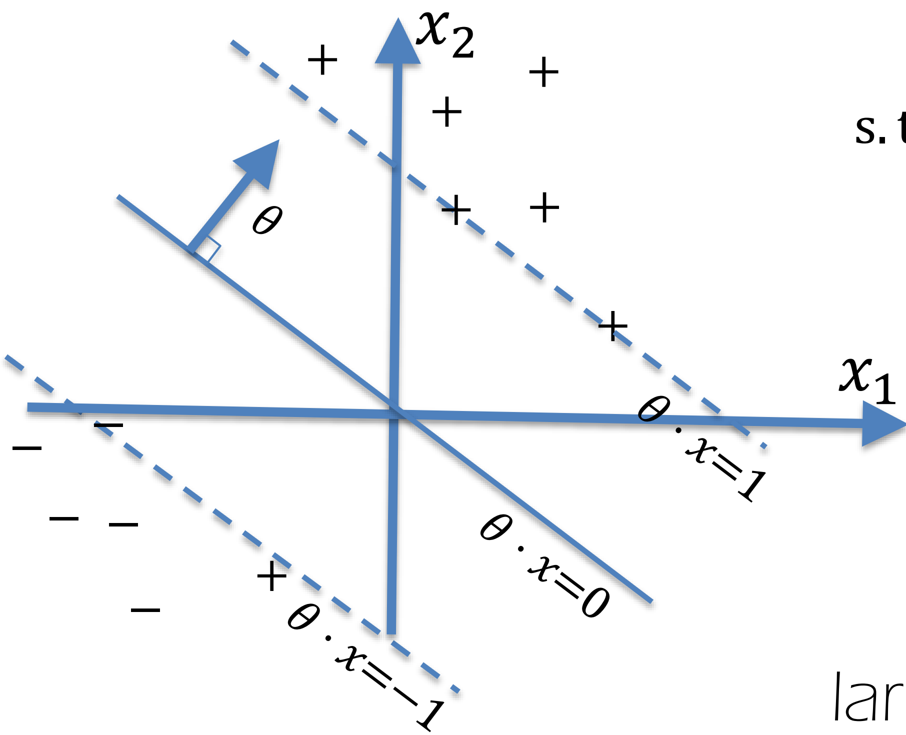
$x \in R$

$y^{(i)}(\theta \cdot x^{(i)} + \theta_0) = 1, i \in SV$

$y^{(i)}(\theta \cdot x^{(i)} + \theta_0) \geq 1, i \notin SV$



SVM: quadratic program



$$\min \; \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t.} \quad y^{(i)}(\theta \cdot x^{(i)}) \geq 1 - \xi_i \;, i = 1 \ldots n$$

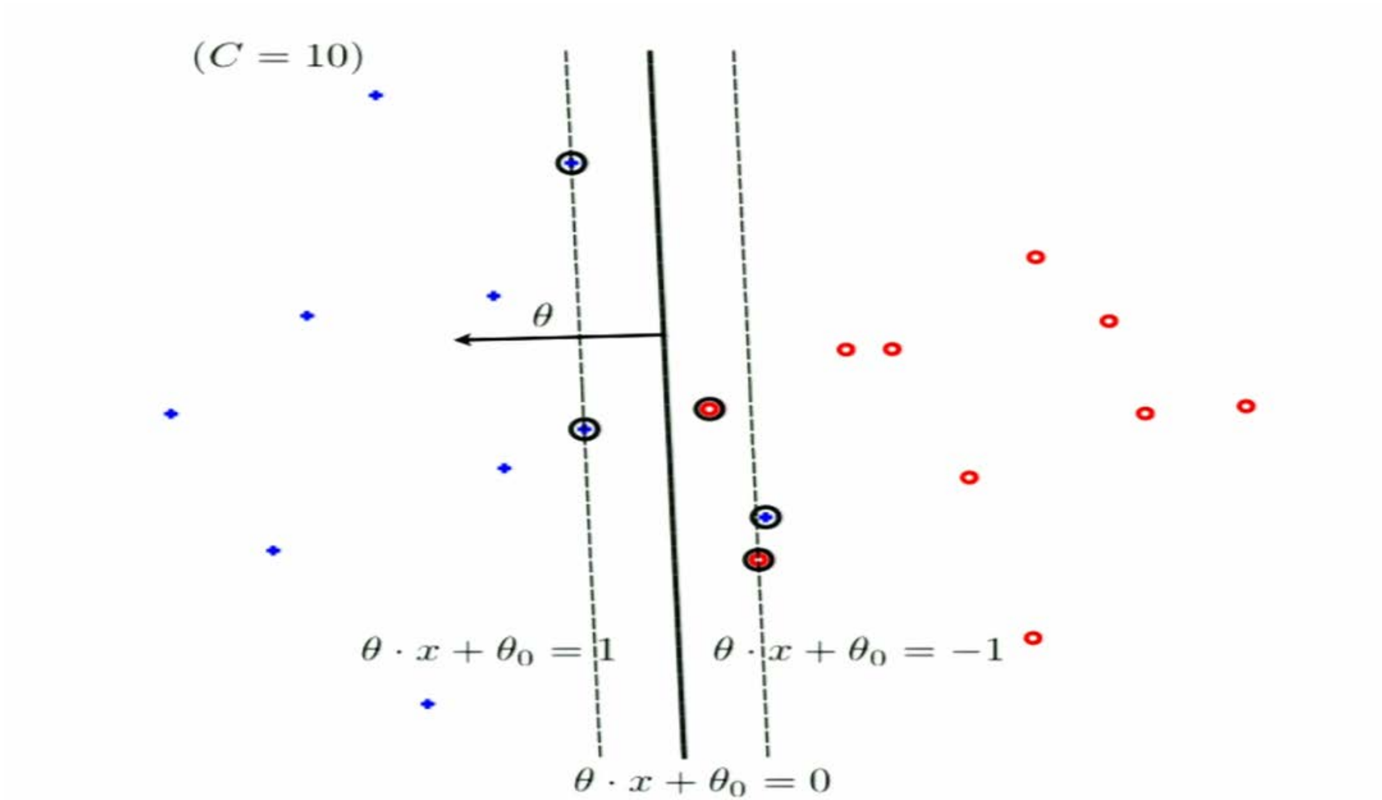
$$\xi_i \geq 0$$

slack

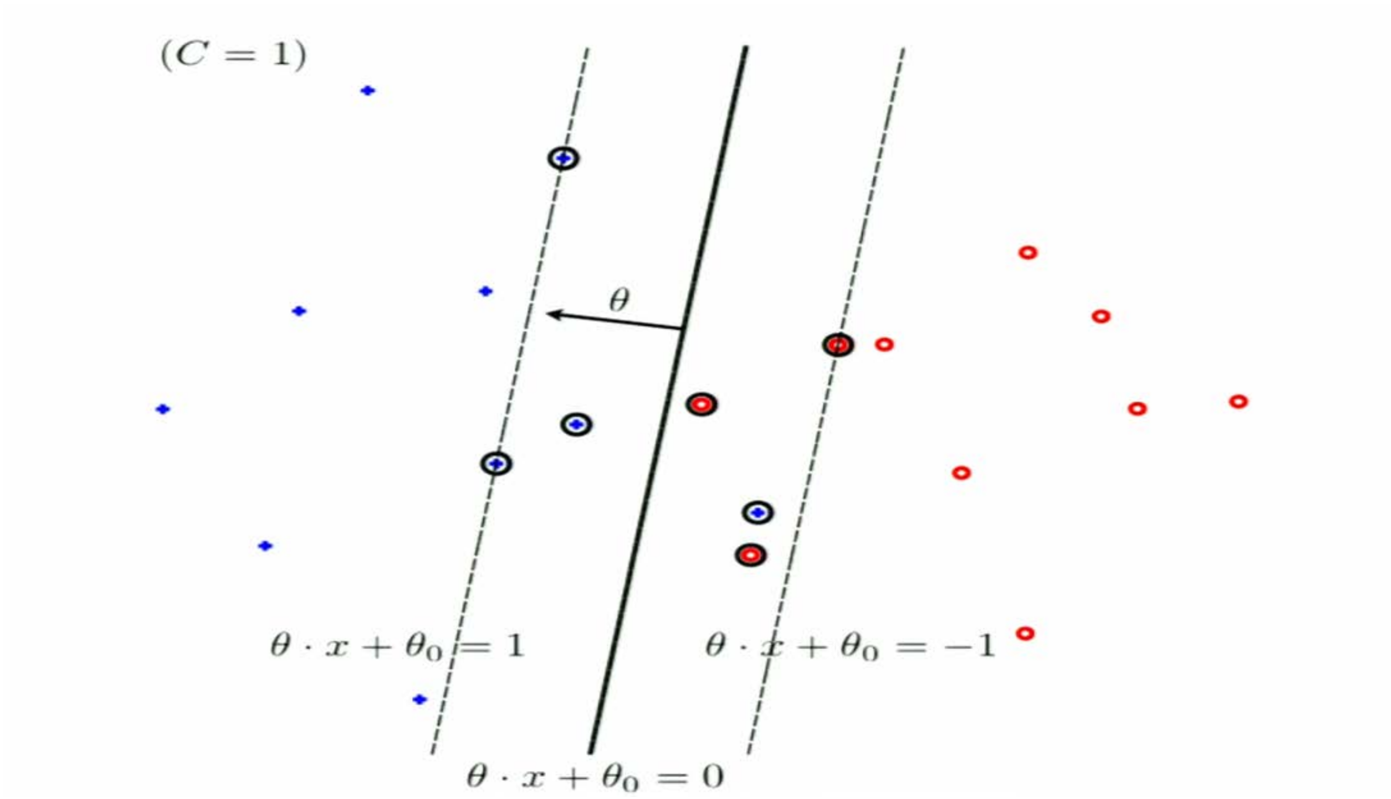
large C value ⇔ few violations

small C value ⇔ allow violations

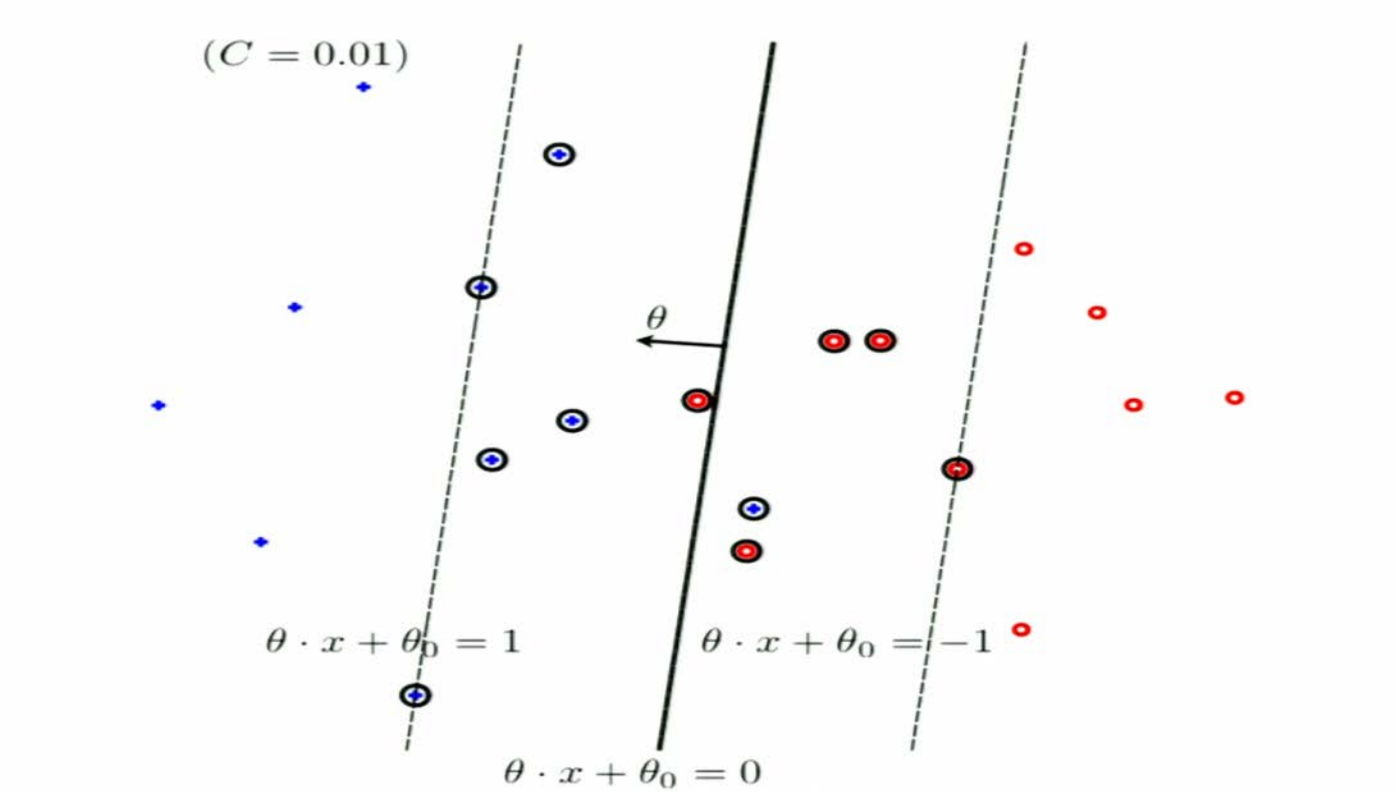
SVM: 2D Example



SVM: 2D Example



SVM: 2D Example



Optimization Problem

$$y^{(i)}(\theta \cdot x^{(i)}) \geq 1 - \xi_i, \qquad i = 1 \ldots n$$
$$\xi_i \geq 1 - y^{(i)}(\theta \cdot x^{(i)}), \qquad i = 1 \ldots n$$
$$\xi_i \geq L_h(y\theta \cdot x), \qquad i = 1 \ldots n$$

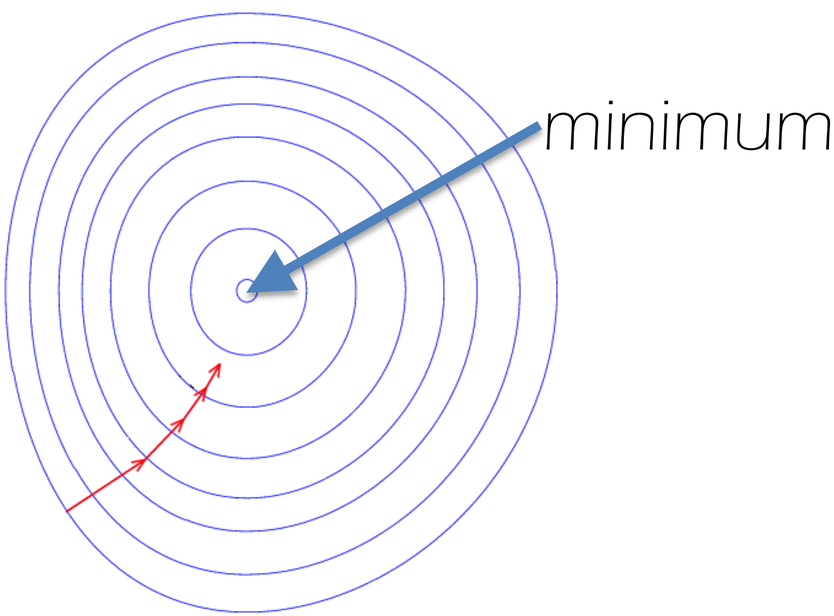
Objective function $\quad \min \quad \frac{\lambda}{2}\|\theta\|^2 + \frac{1}{n}\sum_{i=1}^{n} L_h\big(y^{(i)}(\theta \cdot x^{(i)})\big)$

Gradient Descent Algorithm

initialize $\boldsymbol{\theta}$

until converged

$$\theta := \theta - \eta \nabla_\theta F(\theta)$$



Stochastic Gradient Descent

SGD Algorithm

initialize $\boldsymbol{\theta}$

repeat until approx. solution found

select an example $\boldsymbol{i}$ at random

select $\boldsymbol{\eta}$

$\theta := \theta \; - \eta \nabla_\theta \mathrm{F}_i(\theta)$

$$\theta \coloneqq \theta - \eta \nabla_\theta \left[\frac{\lambda}{2} \|\theta\|^2 + L_h\big(y^{(i)}(\theta \cdot x^{(i)})\big) \right]$$

$$= \theta - \eta \nabla_\theta \left[\frac{\lambda}{2} \|\theta\|^2 \right] - \eta \nabla_\theta \left[L_h\big(y^{(i)}(\theta \cdot x^{(i)})\big) \right]$$

$$= \theta - \lambda \eta \theta - \eta \nabla_\theta \left[L_h\big(y^{(i)}(\theta \cdot x^{(i)})\big) \right]$$

$$= (1 - \lambda\eta)\theta + \eta \begin{cases} y^{(i)} x^{(i)} & \text{if } y^{(i)}(\theta \cdot x^{(i)}) \le 1 \\ 0 & \text{o.w.} \end{cases}$$

Pegasos Algorithm

Pegasos ($\{(x^{(i)}, y^{(i)}), i = 1, \ldots, n\}, \lambda, T$)

$\theta = 0$

for $\mathrm{t} = 1, \ldots \mathrm{T}$, **do**

select i at random

$\eta = 1/t$

if $y^{(i)}(\theta \cdot x^{(i)}) \leq 1$ then

$\theta \ := (1 - \lambda\eta)\theta + \eta y^{(i)} x^{(i)}$

else

$\theta \ := (1 - \lambda\eta)\theta$

return θ