

Support vector machines



Friday, July 17, 2009

Support vector machine

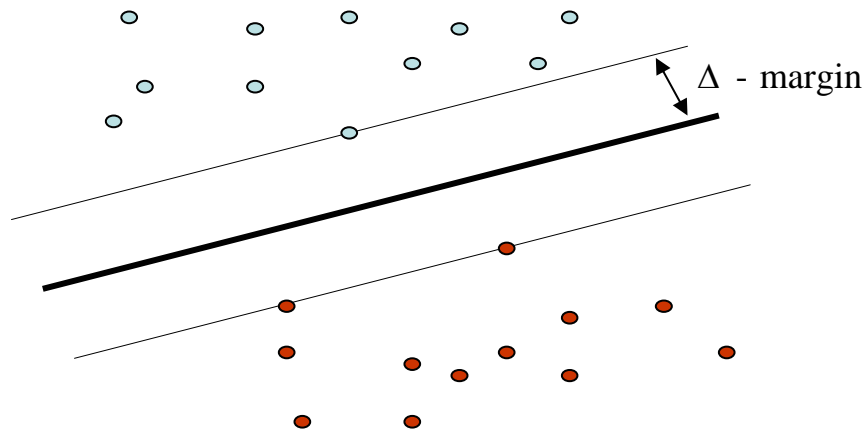
Given a set of training data

$$(x_1, y_1), \dots, (x_l, y_l), x_i \in \mathcal{R}^n, y_i \in \{+1, -1\}$$

find a function that can estimate

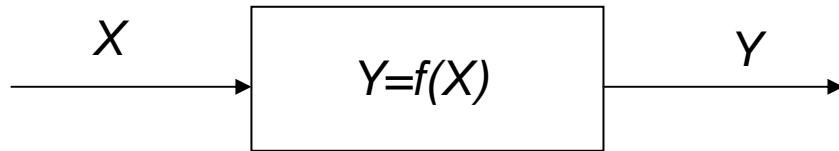
$$y_j^* \in \{+1, -1\} \text{ given new } x_j^* \in \mathcal{R}^n$$

and minimize the frequency of the future error.



Support vector machine

*based on fundamentals of statistical learning theory
(Vapnik-Chervonenkis theory)*



Instead of identifying the unknown function (what classical statistics does), the main goal of VC theory is to imitate the unknown function.

The key discovery of VC theory:

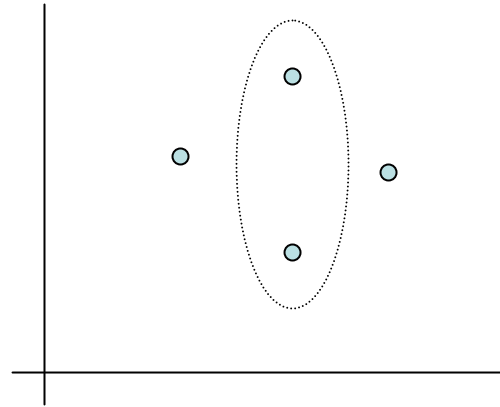
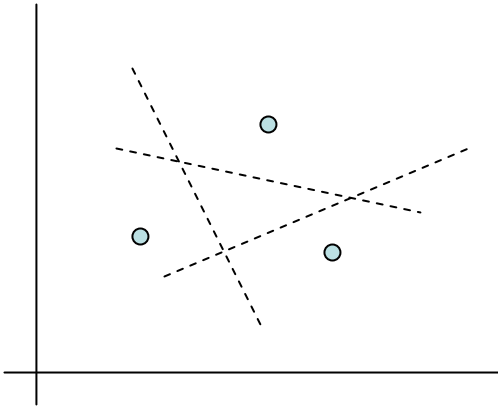
- Two and only two factors are responsible for generalization:
 - One (empirical loss) defines how well the function approximates data*
 - Another (capacity, VC dimension) defines the diversity of the set of functions from which one chooses an approximation function**
- If VC dimension is finite, then one can achieve a good generalization. If it is not finite the generalization is impossible.*

Examples

The VC dimension of linear indicator functions

$$I(x) = \text{sgn}(x^T w + b), \quad x \in \mathfrak{R}^n, w \in \mathfrak{R}^n$$

is equal $n + 1$

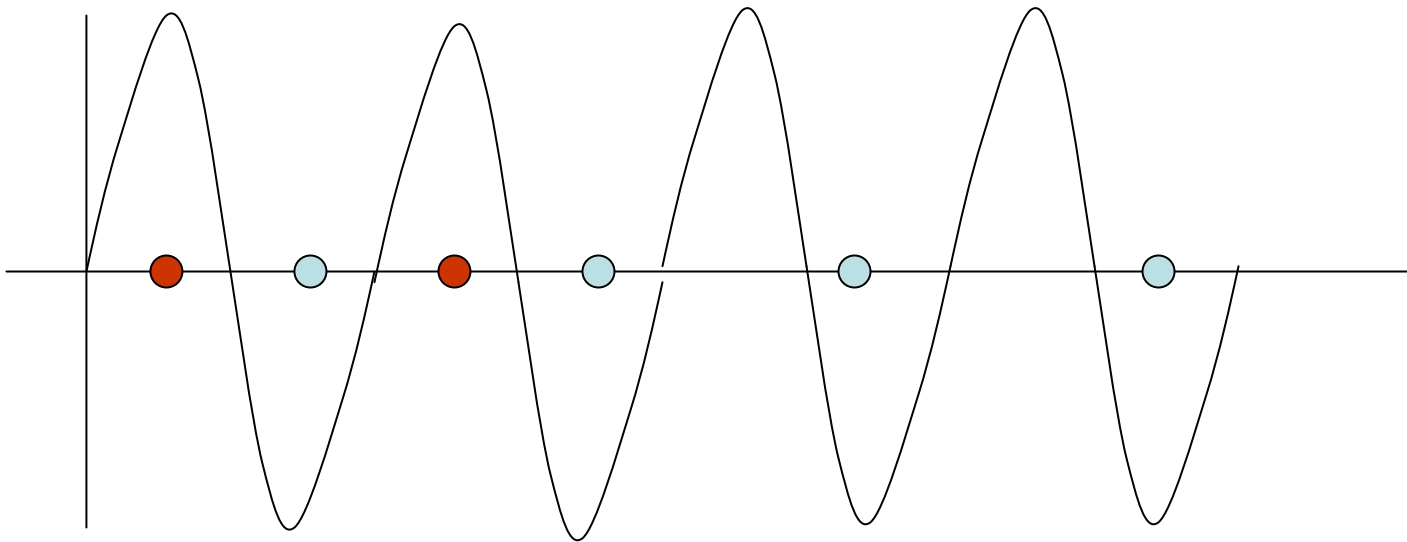


Examples

The VC dimension of the set of functions

$$I(x) = \text{sgn}(\sin ax), \quad x \in \mathcal{R}^1, w \in \mathcal{R}^1$$

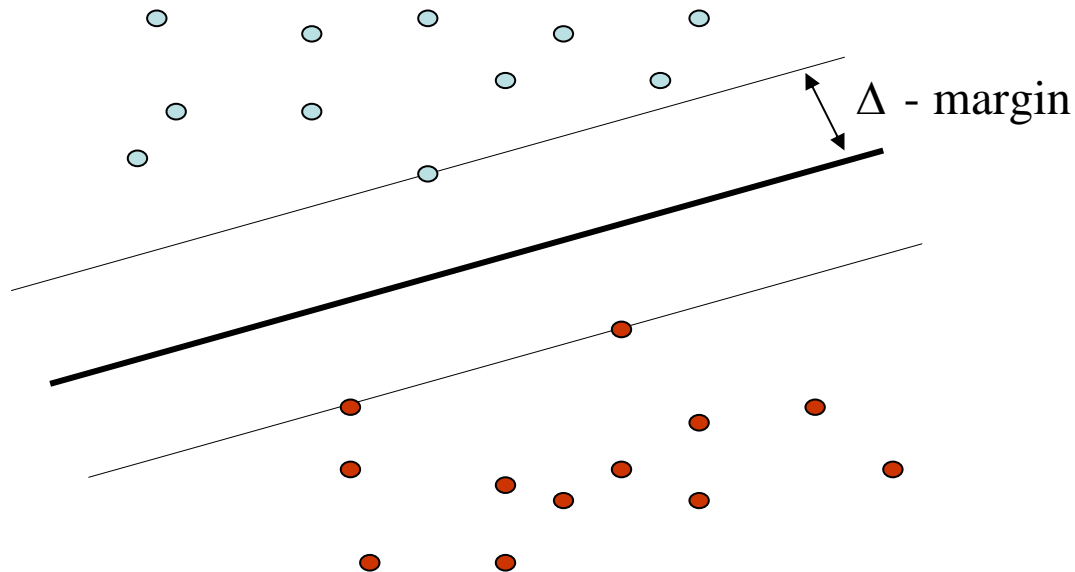
is infinity



Support vector machine

Let the vector $x \in \mathfrak{R}^n$ belong to a sphere of radius R . Then the set of Δ - margin separating hyperplanes has a VC dimension bounded as follows

$$VC_{\text{dim}} \leq \min \left\{ \frac{R^2}{\Delta^2}, n \right\} + 1$$



Support vector machine

Theorem. With probability $1 - \delta$ one can assert that the probability that a test example will not be separated correctly by the Δ - margin hyperplane has the bound

$$P_{error} \leq \frac{m}{l} + \frac{\kappa}{2} \left(1 + \sqrt{1 + \frac{4m}{l\kappa}} \right),$$

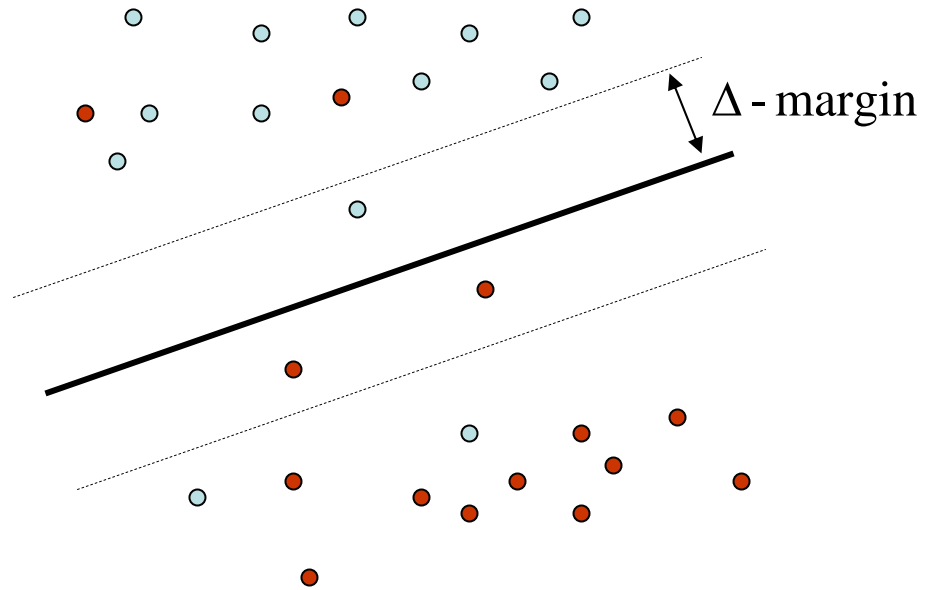
where

$$\kappa = 4 \frac{VC_{\dim} \left(\ln \frac{2l}{VC_{\dim}} + 1 \right) - \ln \frac{\delta}{4}}{l},$$

m is the number of training examples that are not separated correctly by the Δ - margin hyperplane and the VC dimension bounded as follows

$$VC_{\dim} \leq \min \left\{ \frac{R^2}{\Delta^2}, n \right\} + 1$$

Support vector machine



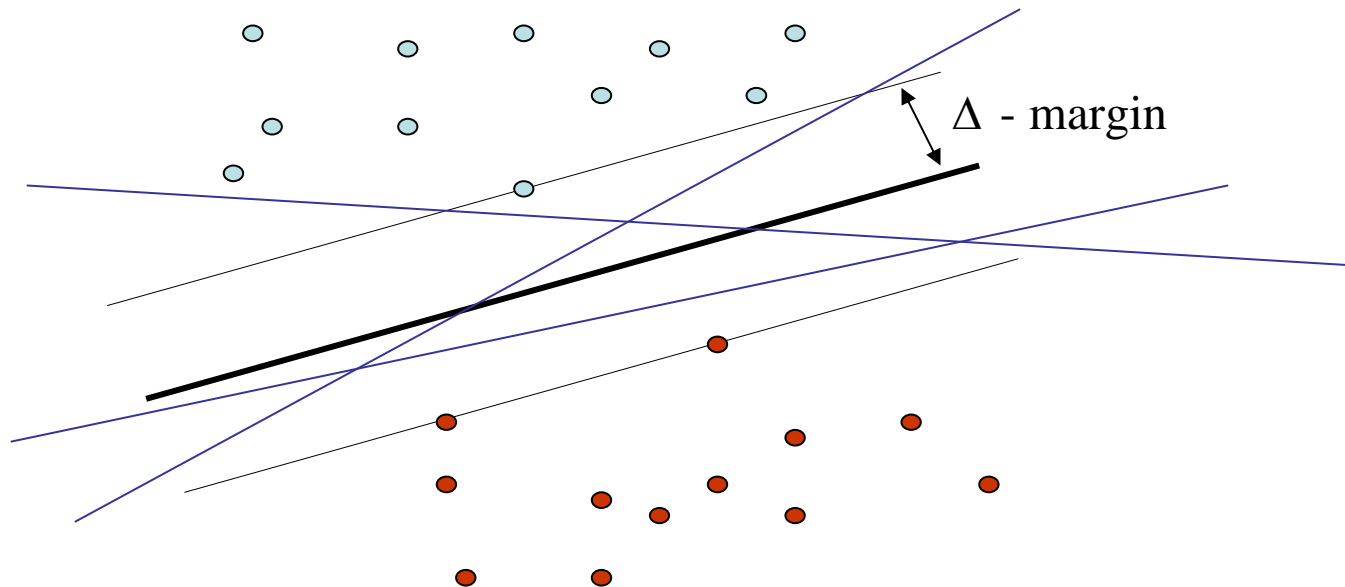
Support vector machine

Suppose that the data

$$(y_1, x_1), \dots, (y_l, x_l) \quad x \in \mathbb{R}^n \quad y \in \{+1; -1\}$$

can be separated by a hyperplane

$$(w \cdot x) - b = 0$$

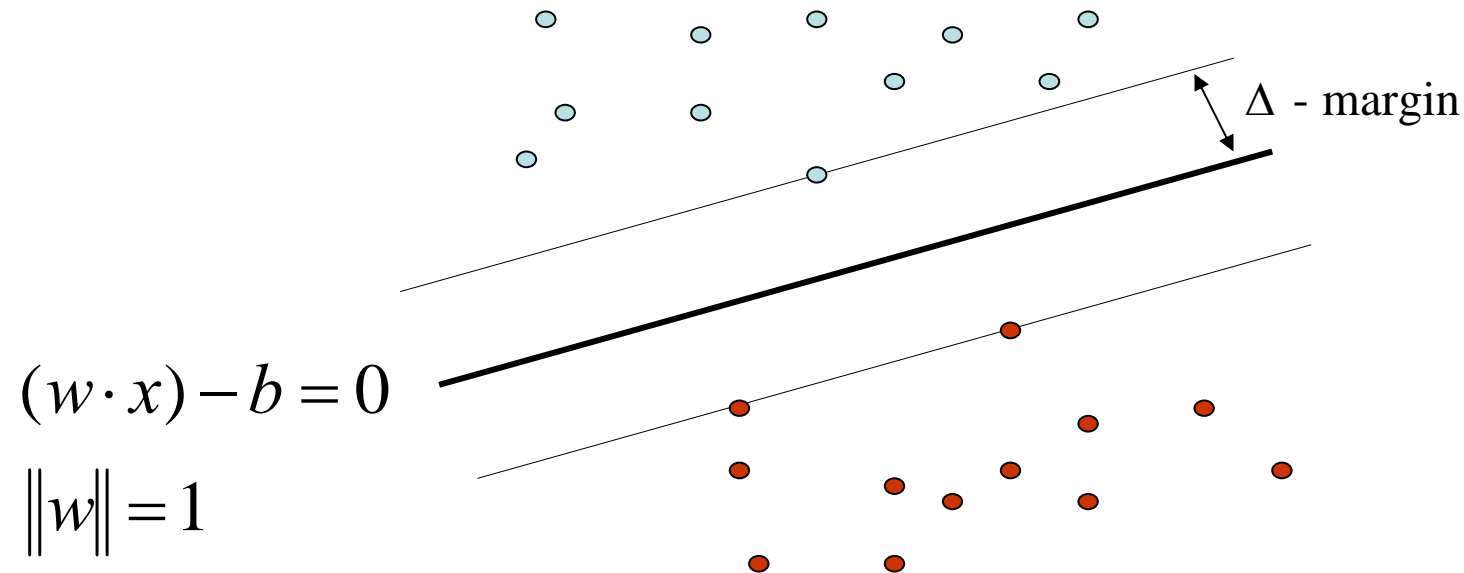


Support vector machine

Blue dots: $(w \cdot x_i) - b \geq 0, y_i = +1$

Red dots: $(w \cdot x_i) - b \leq 0, y_i = -1$

Combined: $y_i [(w \cdot x_i) - b] \geq 0, \forall i$ Variables: w and b



Support vector machine

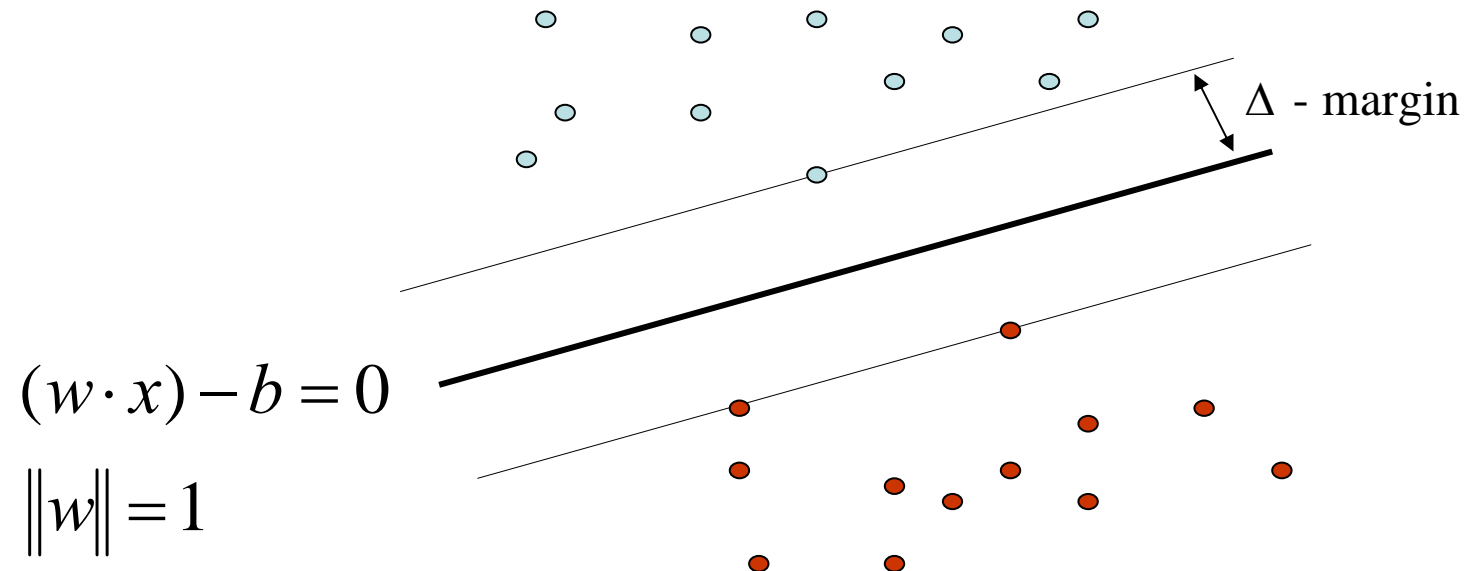
Blue dots: $(w \cdot x_i) - b \geq +\Delta, y_i = +1$

$$\Delta \geq 0$$

Red dots: $(w \cdot x_i) - b \leq -\Delta, y_i = -1$

Combined: $y_i [(w \cdot x_i) - b] \geq \Delta, \forall i$

Variables: w and b

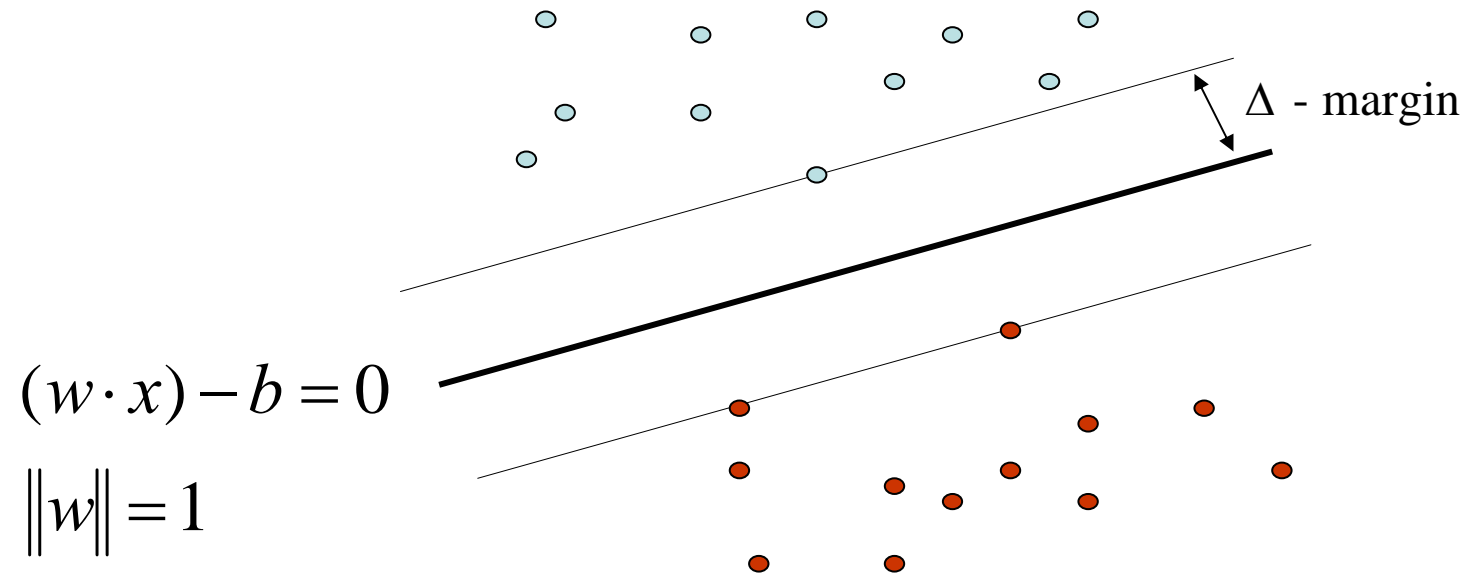


Support vector machine

Maximize the margin: $\max \Delta$

$$\text{s.t. } y_i[(w \cdot x_i) - b] \geq \Delta, \quad \forall i$$
$$w_1^2 + \dots + w_n^2 = 1, \quad \Delta \geq 0$$

Variables: w , Δ and b



Support vector machine

$$y_i \left[\left(\frac{w}{\Delta} \cdot x_i \right) - \frac{b}{\Delta} \right] \geq 1, \quad \forall i \quad \text{or:} \quad y_i \left[(\bar{w} \cdot x_i) - \bar{b} \right] \geq 1, \quad \forall i$$

Maximize the margin: $\min \|\bar{w}\|^2$

$$\text{s.t. } y_i \left[(\bar{w} \cdot x_i) - \bar{b} \right] \geq 1, \quad \forall i$$

Variables: \bar{w} and \bar{b}

Support vector machine

Maximize the margin: $\min \|w\|^2$

$$\text{s.t. } y_i [(w \cdot x_i) - b] \geq 1, \quad \forall i$$

Variables: w and b

Support vector machine

Maximize the margin: $\min (w \cdot w)$

$$\text{s.t. } y_i [(w \cdot x_i) - b] \geq 1, \quad \forall i$$

Variables: w and b

Support vector machine

Maximize the margin: $\min 0.5(w \cdot w)$

$$\text{s.t. } y_i [(w \cdot x_i) - b] \geq 1, \quad \forall i$$

Variables: w and b

Non separable case:

Maximize the margin: $\min 0.5(w \cdot w) + C \left(\sum_{i=1}^l \xi_i \right)$

$$\text{s.t. } y_i [(w \cdot x_i) - b] \geq 1 - \xi_i, \quad \forall i$$
$$\xi_i \geq 0$$

Variables: w, b and ξ

Support vector machine

$$(y_1, x_1), \dots, (y_l, x_l) \quad x \in \mathbb{R}^n$$

$$\text{Primal problem} \quad \min 0.5(w \cdot w) + C \sum_{i=1}^l \xi_i$$

$$\text{s.t.} \quad \xi_i \geq 0$$

$$\text{Variables: } w, b \text{ and } \xi \quad y_i [(x_i \cdot w) - b] \geq 1 - \xi_i, \quad i = 1, \dots, l$$

$$\text{Dual problem} \quad \max \sum_{i=1}^l \alpha_i - 0.5 \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{s.t.}$$

$$\text{Variables: } \alpha \quad \sum_{i=1}^l \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l$$

Support vector machine

*Optimization
problem for finding
support vectors*

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^l \alpha_i - 0.5 \sum_{i,j}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & \\ & \sum_{i=1}^l \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \end{aligned}$$

Kernels

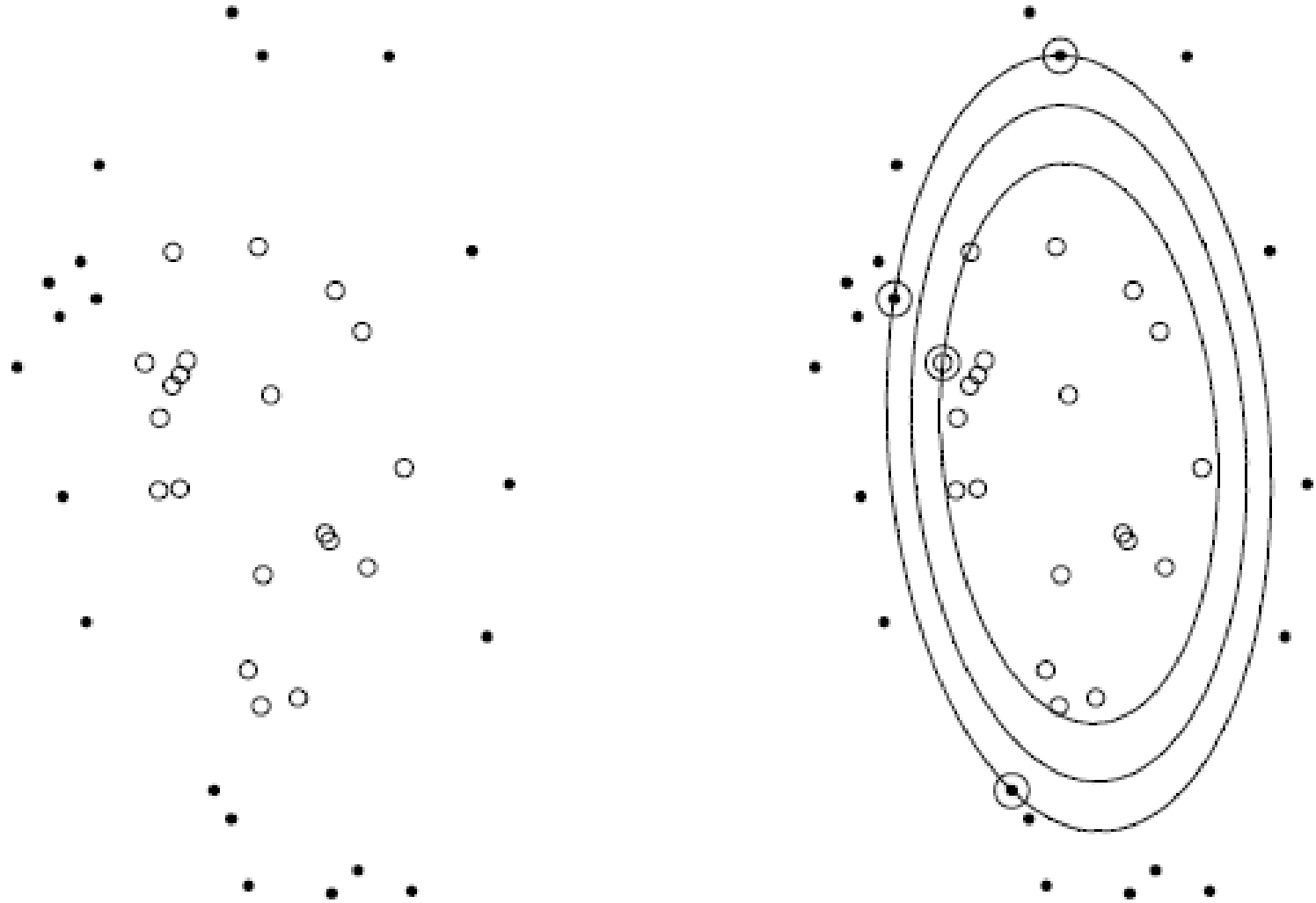
Polynomial machine:

$$K(x, x_i) = [\alpha(x \cdot x_i) + \beta]^d$$

A radial basis function machine:

$$K(x, x_i) = \exp\left\{-\gamma \|x - x_i\|^2\right\}$$

Support vector machine



Support vector machine

*Optimization
problem for finding
support vectors*

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^l \alpha_i - 0.5 \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & \\ & \sum_{i=1}^l \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \end{aligned}$$

*Decision rules with a kernel using found α_i^**

if $\left(\sum_{i=1}^l y_i \alpha_i^* K(x_i, x) \right) - b \geq 0$ then x is *blue*

if $\left(\sum_{i=1}^l y_i \alpha_i^* K(x_i, x) \right) - b < 0$ then x is *red*

$$b = \sum_{i=1}^l y_i \alpha_i^* K(x_i, x_{i_0}) - y_{i_0} \text{ for some } \alpha_{i_0}^* : 0 < \alpha_{i_0}^* < C, (\alpha_{i_0}^* \neq 0, \alpha_{i_0}^* \neq C)$$

if there is no such $\alpha_{i_0}^*$, increase C and train again

Support vector machine

x_i that correspond to positive α_i^* are called the support vectors!!!

Only the support vectors carry important information!!!

They correspond to the active constraints of the primal problem!!!

Let $I^* = \{i : \alpha_i > 0\}$ be the set of support vectors

Decision rules using only the support vectors

if $\left(\sum_{i \in I^*} y_i \alpha_i^* K(x_i, x) \right) - b \geq 0$ then x is *blue*

if $\left(\sum_{i \in I^*} y_i \alpha_i^* K(x_i, x) \right) - b < 0$ then x is *red*

$$b = \sum_{i \in I^*} y_i \alpha_i^* K(x_i, x_{i_0}) - y_{i_0} \text{ for some } \alpha_{i_0}^* : 0 < \alpha_{i_0}^* < C, (\alpha_{i_0}^* \neq 0, \alpha_{i_0}^* \neq C)$$

if there is no such $\alpha_{i_0}^*$, increase C and train again