Nonlinear optimization and applications

MATH 689 / CSI 747

Instructor: I. Griva

Tuesday, 7:20 – 10:00 pm

Markowitz portfolio problem

$i = 1,\dots,n$ assets

ξ_i random return from investing in asset *i*

We assume that the returns have a discrete joint distribution with realizations r_{ti}, $t = 1, \dots, T$, $i = 1, \dots, n$, attained with probabilities p_t, $t = 1, \dots, T$. For example, $p_t = 1/T$, t = 1, . . . , T.

Our aim is to invest our capital in these assets in order to obtain some desirable characteristics of the total return on the investment.

$r_i = E(\xi_i) = \sum_{t=1}^{T} p_t r_{ti}$ average return from investing in asset *i*

x_i proportion of the capital invested in asset *i*

$x = (x_1,\dots,x_n)$ selected portfolio

$R = x_1\xi_1 + \cdots + x_n\xi_n$ random return of portfolio x

Markowitz portfolio problem

Minimizing the risk
(the standard deviation):

$$\min E(R - E(R))^2$$

s.t. $E(R) = r$ desired return

$$\sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0$$

Markowitz portfolio problem

$x = (x_1, \dots, x_n)$ selected portfolio

$R = x_1\xi_1 + \cdots + x_n\xi_n$ random return of portfolio x

$$\begin{aligned} &\min E(R - E(R))^2 \\ \text{s.t.}\quad & E(R) = r \\ & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0 \end{aligned}$$

$$\begin{aligned} &\min x^T \sigma x \\ \text{s.t.}\quad & x_1 r_1 + \cdots + x_n r_n = r \\ & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0 \end{aligned}$$

$$E(R) = E(x_1\xi_1 + \cdots + x_n\xi_n) = x_1 E(\xi_1) + \cdots + x_n E(\xi_n) = x_1 r_1 + \cdots + x_n r_n = r$$

$$E(R - E(R))^2 = E(R^2) - (E(R))^2 = E(R^2) - r^2$$

$$E(R^2) = E(x_1\xi_1 + \cdots + x_n\xi_n)^2 = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i \xi_i \xi_j x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \sigma_{ij} x_j = x^T \sigma x$$

$$\sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix} \qquad \sigma_{ij} = E(\xi_i \xi_j)$$

Markowitz portfolio problem

$i = 1,\ldots,n$ assets

r_i average return from investing in asset *i*

σ covariance matrix

x_i proportion of the capital invested in asset *i*

Minimizing the risk

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \sigma_{ij} x_j$$

s.t.

$$\sum_{j=1}^{n} r_i x_i = r \quad \text{desired return}$$

$$\sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0$$

Markowitz portfolio problem

$i = 1, \ldots, n$ assets

r_i average return from investing in asset *i*

σ covariance matrix

x_i proportion of the capital invested in asset *i*

$\mu > 0$ risk tolerance parameter

Minimizing the risk $$\min \mu \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \sigma_{ij} x_j - \sum_{j=1}^{n} r_i x_i$$

s.t. $$\sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0$$

Markowitz portfolio problem

$x = (x_1, \ldots, x_n)$ selected portfolio

$R = x_1\xi_1 + \cdots + x_n\xi_n$ random return of portfolio x

$$\min E\,|R - E(R)|$$

s.t. $E(R) = r$

$$\sum_{i=1}^{n} x_i = 1$$

$$x_i \ge 0$$

$$\min \sum_{t=1}^{T} p_t u_t$$

s.t. $-u_t \le r_t - r \le u_t,\ t = 1,\ldots,T$

$$\sum_{i=1}^{n} x_i r_{ti} = r_t,\ \ t = 1,\ldots,T$$

$$\sum_{i=1}^{n} x_i r_i = r,\ \ \sum_{i=1}^{n} x_i = 1,\ \ x_i \ge 0$$

$|R - E(R)| = U$, another random variable with realization $u_t, t = 1,\ldots,T$

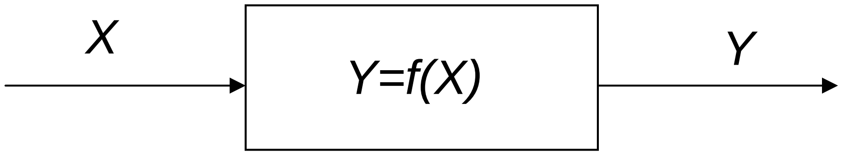
$$-U \le R - E(R) \le U$$

$$-u_t \le r_t - r \le u_t,\ t = 1,\ldots,T \qquad \sum_{i=1}^{n} x_i r_{ti} = r_t,\ \ t = 1,\ldots,T$$

$$p_t = 1/T,\ \ t = 1,\ldots,T$$

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.*

Support vector machine

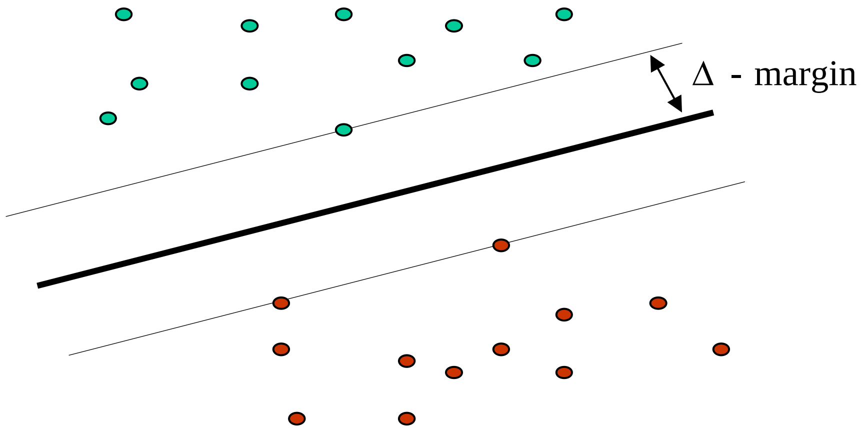
Given a set of training data

$$(x_1, y_1), \ldots, (x_l, y_l),\ x_i \in \Re^n,\ y_i \in \{+1, -1\}$$

find a function that can estimate

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

and minimize the frequency of the future error.

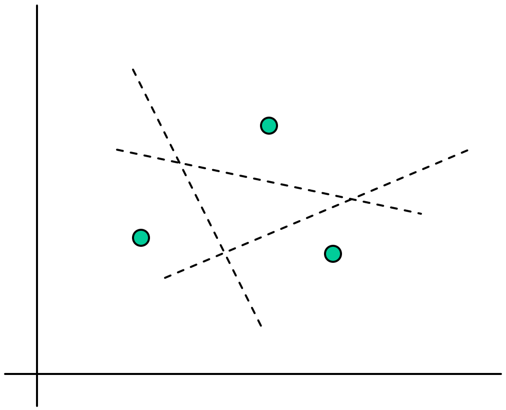


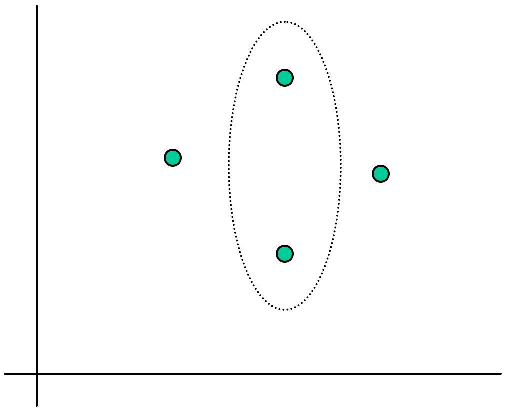
Examples

The VC dimension of linear indicator functions

$$I(x) = \operatorname{sgn}(x^T w + b), \quad x \in \Re^n, w \in \Re^n$$

is equal $n+1$



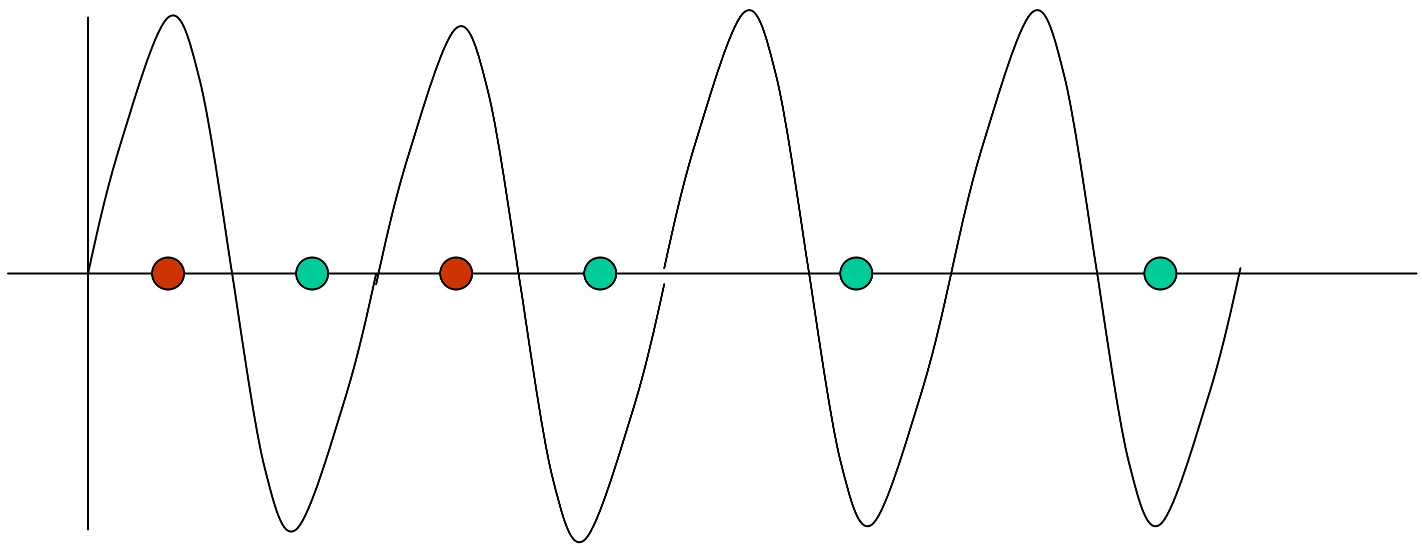


Examples

The VC dimension of the set of functions

$$I(x) = \mathrm{sgn}(\sin ax), \quad x \in \Re^1, w \in \Re^1$$

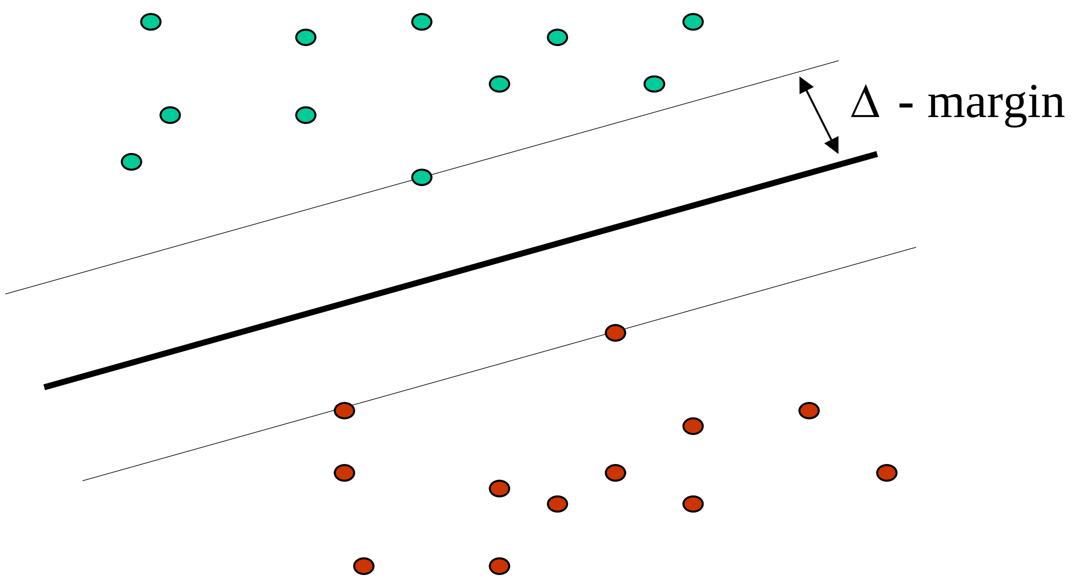
is infinity



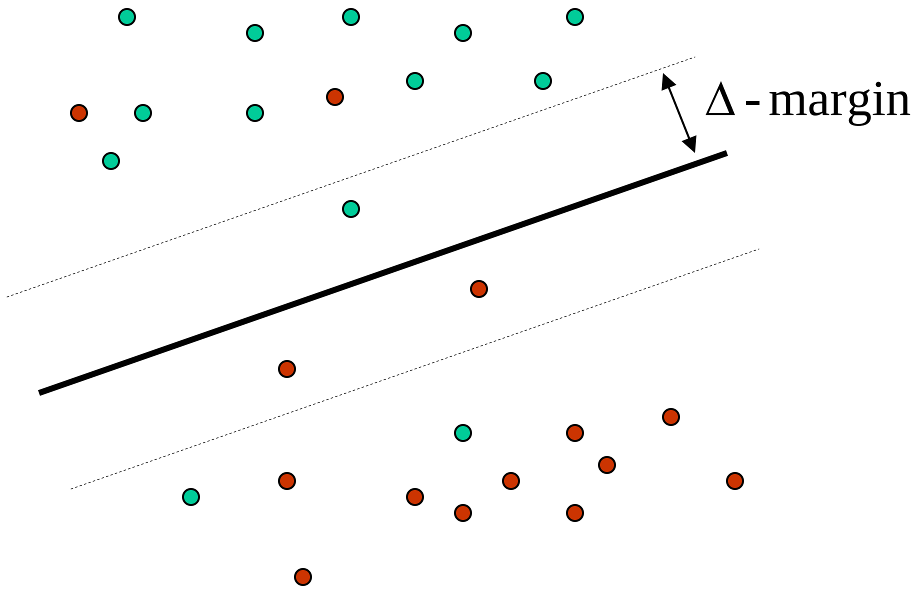
Support vector machine

Let the vector $x \in \Re^n$ belong to a sphere of radius R. Then the set of Δ - margin separating hyperplaines has a VC dimention bounded as follows

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



Support vector machine



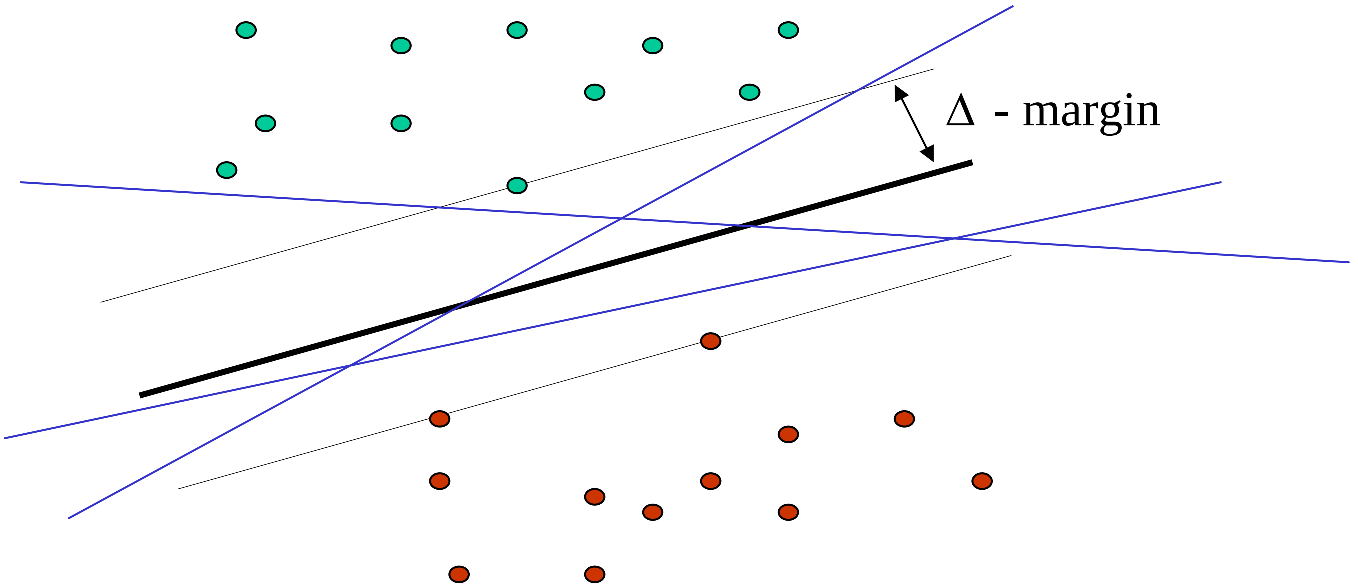
Support vector machine

Suppose that the data

$$(y_1, x_1), \ldots, (y_l, x_l) \qquad x \in \Re^n \qquad y \in \{+1; -1\}$$

can be separated by a hyperplane

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



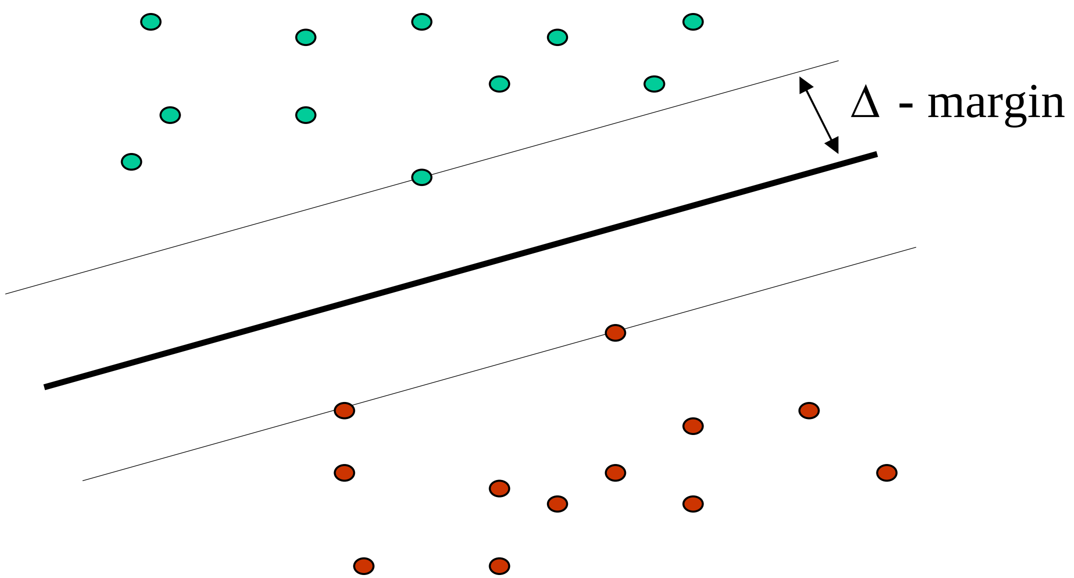
Support vector machine

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

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

Combined: $y_i\left[(w \cdot x_i) - b\right] \geq 0, \quad \forall i$

Variables: w and b



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

$$\|w\| = 1$$

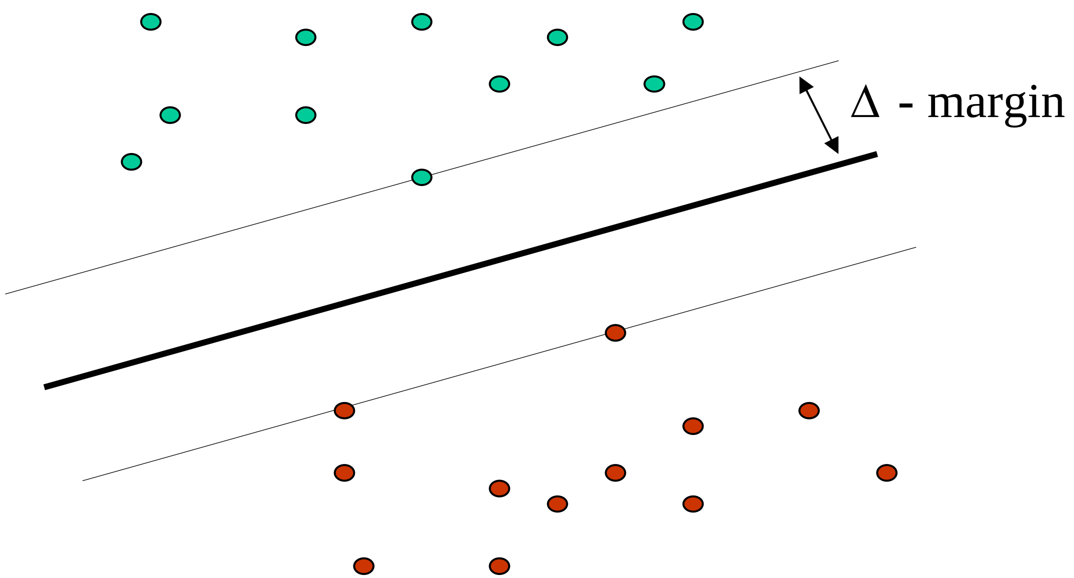
Support vector machine

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

$$\Delta \geq 0$$

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

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



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

$$\|w\| = 1$$

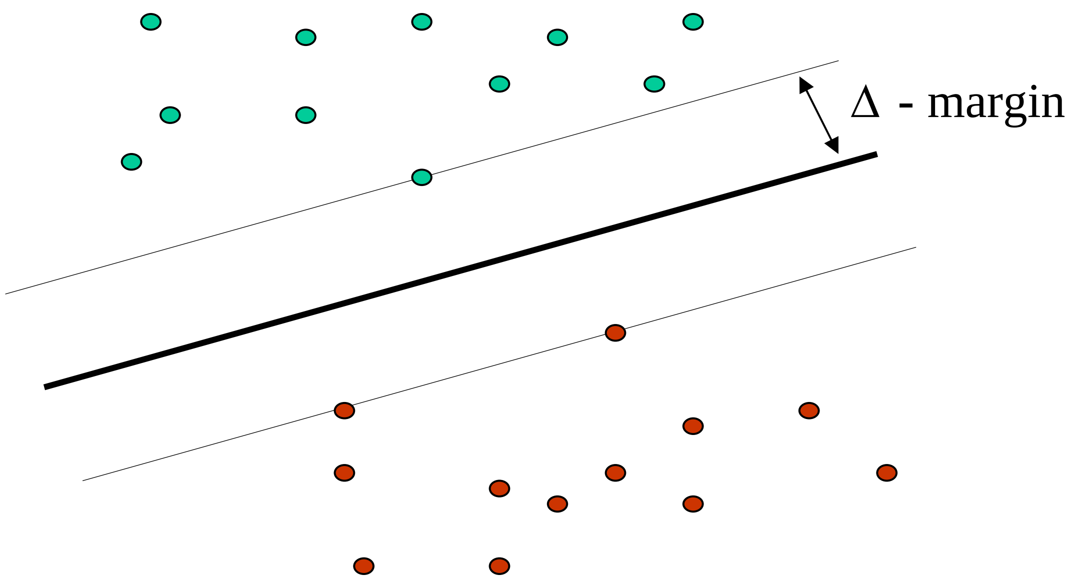
Support vector machine

Maximize the margin: $\max \Delta$

s.t. $y_i\left[(w \cdot x_i) - b\right] \geq \Delta, \quad \forall i$

$$w_1^2 + \cdots + w_2^2 = 1, \quad \Delta \geq 0$$

Variables: w, Δ and b



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

$$\|w\| = 1$$

Support vector machine

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

Maximize the margin: $\min \left\|\overline{w}\right\|$

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

Variables: $\overline{w}$ and $\overline{b}$

Support vector machine

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

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

Variables: w and b

Support vector machine

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

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

Variables: w and b

Support vector machine

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

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

Variables: w and b

Non separable case:

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

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

$\xi_i \geq 0$

Variables: w, b and ξ

Support vector machine

$(y_1, x_1), \ldots, (y_l, x_l) \qquad x \in \Re^n$

Primal problem

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

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

Variables: w, b and ξ

$$y_i\left[(x_i \cdot w) - b\right] \geq 1 - \xi_i, \; i = 1, \ldots, l$$

Dual problem

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

s.t.

Variables: α

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \; i = 1, 2, \ldots, l$$

Support vector machine

Optimization problem for finding support vectors

$$\max_{\alpha} \sum_{i=1}^{l} \alpha_i - 0.5 \sum_{i,j}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

s.t.

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \; i = 1,2,\ldots,l$$

Kernels

Polynomial machine:

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

A radial basis function machine:

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

Support vector machine

Optimization problem for finding support vectors

$$\max_{\alpha} \sum_{i=1}^{l} \alpha_i - 0.5 \sum_{i,j}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

s.t.

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \; i = 1,2,\ldots,l$$

Decision rules with a kernel using found α_i^*

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

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

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

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

Support vector machine

Positive $\alpha^*_{i_0}$ correspond to the support vectors!!!

Only the support vectors carry inportant 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 $\sum_{i \in I^*} y_i \alpha_i^* K(x_i, x) - b \geq 0$ then x is *blue*

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