# Optimization Methods for Machine Learning (OMML)

**Laura Palagi**
Department of Computer, Control, and Management Engineering Antonio Ruberti

SAPIENZA
UNIVERSITÀ DI ROMA

## Support Vector Machines
## Hard SVM
## November 23, 2018

# Vapnik Chervonenkis bound on the risk

$$R(\alpha) \leq R_{emp}(\alpha) + C_{VC}(h, \ell, \eta)$$

VC dimension is the important parameter

The function to be minimized is

$$\min_{\alpha, f_\alpha} U(f_\alpha, \alpha) = R_{emp}(\alpha) + C_{VC}(h, \ell, \eta)$$

Penalization term on the complexity

Minimization with respect to both the class and the parameters

# Minimize the VC confidence

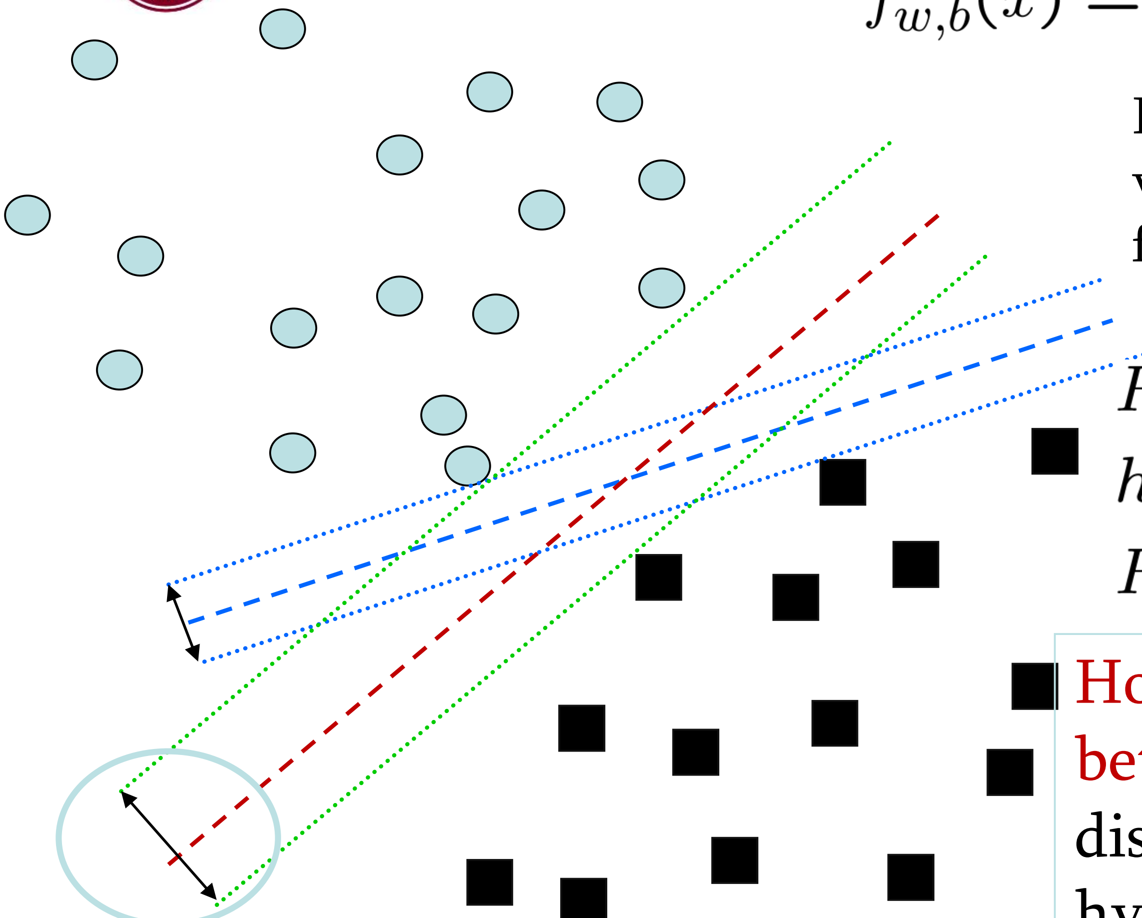
We need to calculate the VC dimension $h$ for a class of function.
Let's start with orienting hyperplanes

$$f_{w,b}(x) = sign(w^T x + b)$$

# Computation of the VC dimension

The VC dimension of class of function "hyperplanes" is n+1

$$f_{w,b}(x) = sign\{w^T x + b\}$$



In this case we get the same value of the bound for both the functions in the class

$$R_{emp}(f_\alpha) = 0$$

$$h = 3$$

$$R(f_\alpha) \leq 0 + C_{VC}(3, \ell, \eta)$$

However the red one seems better: it "maximizes" the distance among the hyperplane and points in the two sets

Which one of the two is the best ?

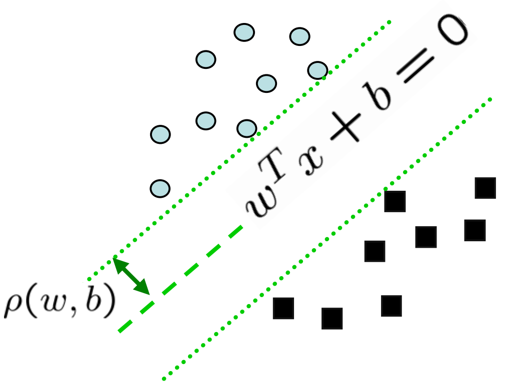
# Intuition

- A hyperplane that passes too close to the training examples will be sensitive to noise and less likely to generalize well for new data
- Instead, it seems reasonable to expect that a hyperplane that is farthest from all training examples will have better generalization capabilities

# Margin

Given two linearly separable sets A e B
and $w^T x + b = 0$ a separating hyperplane

$$w^T x^i + b > 0 \qquad y^i = 1$$
$$w^T x^i + b < 0 \qquad y^i = -1$$
$$y^i (w^T x^i + b) > 0 \qquad i = 1 \dots, \ell$$




Let's consider the distance of the points from the hyperplane $d(x^i; w, b)$. The margin is

$$\rho(w, b) = \min_{x^i \in A \cup B} \{d(x^i; w, b)\}$$

# Hyperplane with margin

The separating function gives

$$y^i = \begin{cases} 1 & \text{se } d(x^i; w, b) \geq \rho \\ -1 & \text{se } d(x^i; w, b) \leq -\rho \end{cases}$$

The highest the margin the lowest the VC dimension $h$

$$h \leq \min\left\{\left\lceil \frac{D^2}{\rho^2} \right\rceil, n\right\} + 1 \leq n + 1$$

# Maximize the margin

Given two sets of separable points, find the separating hyperplane with maximum margin

$$
\begin{aligned}
x^i \in A \;&\leftrightarrow\; y^i = 1 \\
x^i \in B \;&\leftrightarrow\; y^i = -1
\end{aligned}
$$

$$
\begin{aligned}
&\max_{w,b} \rho(w,b) \\
&w^T x^i + b \geq 1 \;\forall\; y^i = 1 \\
&w^T x^i + b \leq -1 \;\forall\; y^i = -1
\end{aligned}
$$

# Evaluate the margin

We need to find the Euclidean distance between a point $x^i$ and the hyperplane (convex set)

$$\mathcal{H} = \{y \in R^n : \; w^T y + b = 0\}$$

$$\min_{y \in \mathcal{H}} \quad \|x^i - y\|$$

It is a nonlinear optimization problem known as projection problem

$$\min_{y \in \mathcal{H}} \quad \frac{1}{2}\|x^i - y\|^2$$

It is a quadratic strictly convex pb. A unique solution exists

# Projection over a convex set

$$min_y \quad \frac{1}{2}\|x^i - y\|^2$$
$$w^T y + b = 0$$

$$\boxed{d(x^i, \mathcal{H}) = \frac{|w^T x^i + b|}{\|w\|}}$$

Proof
on the blackboard

# The margin

$$\min_{x^i \in A \cup B} \frac{|w^T x^i + b|}{\|w\|}$$

With the separating conditions

$$y^i(w^T x^i + b) \geq 1 \ \forall \ y^i$$

# Maximize the margin

$$\max_{w \in R^n, b \in R} \; \min_{x^i \in A \cup B} \frac{|w^T x^i + b|}{||w||}$$

With the separating constraints

$$y^i (w^T x^i + b) \geq 1 \;\forall\; y^i$$

The optimal separating hyperplane exists and it is unique

# Some observations

$$w^T x^i + b \geq 1 \ \forall \ y^i = 1$$
$$w^T x^i + b \leq -1 \ \forall \ y^i = -1$$

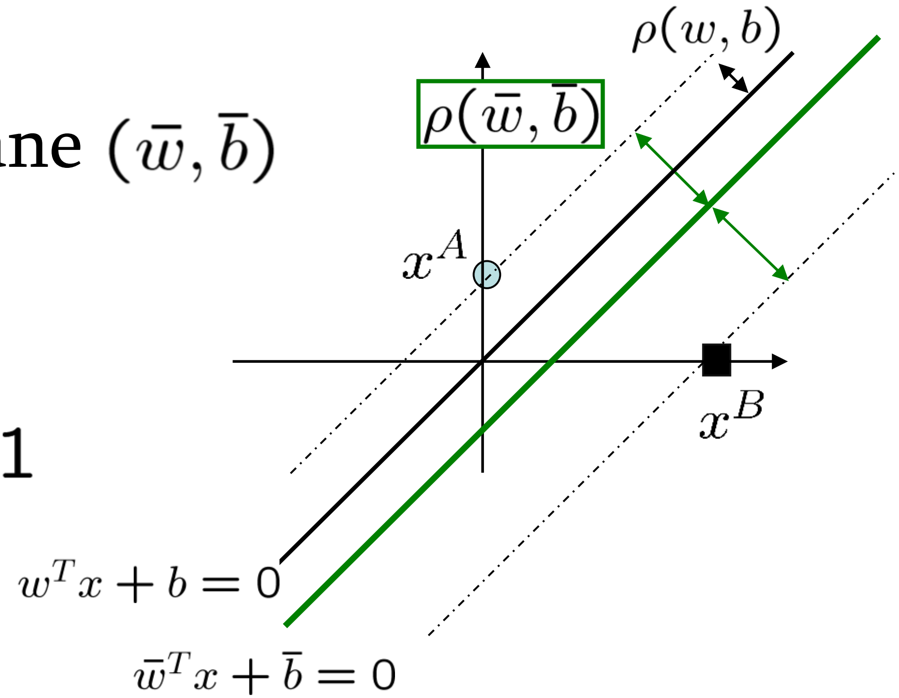
$$|w^T x^i + b| \geq 1 \ \forall \ y^i$$

$$\boxed{\rho(w,b)} = \min_{x^i \in A \cup B} \frac{|w^T x^i + b|}{\|w\|} \boxed{\geq \frac{1}{\|w\|}}$$

# A "standard" separating hyperplane

Given a separating hyperplane with margin $\rho(w,b)$,

there exists another hyperplane $(\bar{w},\bar{b})$




$$\exists\ x^A \in A \quad \bar{w}^T x^A + \bar{b} = 1$$
$$\exists\ x^B \in B \quad \bar{w}^T x^B + \bar{b} = -1$$

with margin

$$\rho(\bar{w},\bar{b}) = \left\{\frac{|\bar{w}^T x^A + \bar{b}|}{\|\bar{w}\|}\right\} = \left\{\frac{|\bar{w}^T x^B + \bar{b}|}{\|w\|}\right\} = \frac{1}{\|\bar{w}\|}$$

$$\rho(w,b) \le \rho(\bar{w},\bar{b}) = \frac{1}{\|\bar{w}\|}.$$

# The max margin problem

$$\max_{w,b} \frac{1}{\|w\|}$$

$$w^T x^i + b \geq 1 \ \forall \ i : \ \ y^i = 1$$
$$w^T x^i + b \leq -1 \ \forall \ i : \ \ y^i = -1$$

$\exists(\bar{w}, \bar{b})$ t.c.

$$\rho(w, b) \leq \rho(\bar{w}, \bar{b}) = \frac{1}{\|\bar{w}\|}.$$

$(w^*, b^*)$
soluzione ottima

$$\frac{1}{\|w\|} \leq \rho(w, b) \leq \frac{1}{\|w^*\|} \qquad \frac{1}{\|w^*\|} \leq \rho(w, b) \leq \frac{1}{\|w^*\|}$$

# The max margin problem

$$\max_{w,b} \rho(w,b)$$

$$y^i(w^T x^i + b) \geq 1 \quad \forall\ i$$

$$\min_{w,b} \|w\|^2$$

$$y^i\left(w^T x^i + b\right) \geq 1\ \forall\ i$$