Université Batna 2
Faculté de Mathématiques et Informatique
Département de Mathématique
Année universitaire 2019-2020

Support Vector Machine

SAADNA Yassmina
M2 SAD

Plan

1- *Induction*

2- *Les SVMs*

3- *Les méthodes à noyau*

4- *Mise en œuvre*

5- *Applications*

6- *Bilan*

Apprentissage inductif supervisé

À partir de l'*échantillon d'apprentissage* $S = \{(x_i, u_i)\}_{1,m}$

on cherche à identifier une loi de dépendance sous-jacente

- **Par exemple une fonction *h* aussi proche possible de *f* (fonction cible) tq : $u_i = f(x_i)$**
- **Ou bien de la distribution de probabilités $P(x_i, u_i)$**

afin de prédire l'avenir

Apprentissage inductif supervisé

$$\boxed{\begin{array}{c} P(\boldsymbol{x}, u) \\ u_i \;=\; f(\boldsymbol{x}_i) \end{array}}$$

$$S_m \;=\; \{(\boldsymbol{x}_i, u_i)\}_{1 \leq i \leq m}$$

Échantillon d'apprentissage

$$\boxed{u \;=\; h(\boldsymbol{x})}$$

- *Identification* : h « proche de » f
- *Prédiction* : h « bonne règle de décision »

Hyperplans séparateurs

- Tâche de classification
  - **Cas de la** séparation linéaire

- On cherche h sous forme d'une fonction linéaire : $h(x) = w.x + b$

- La *surface de séparation* est donc l'hyperplan :

$$w.\boldsymbol{x} + b = 0$$

- Elle est valide si

- L'hyperplan est dit sous forme canonique lorsque

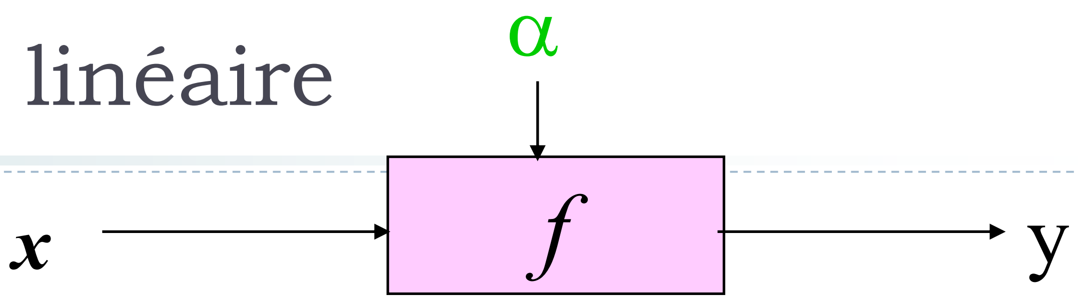
$$\forall i \quad u_i\, h(\boldsymbol{x}_i) \geq 0$$

$$\min_i |w.\boldsymbol{x} + b| = 1$$

ou encore

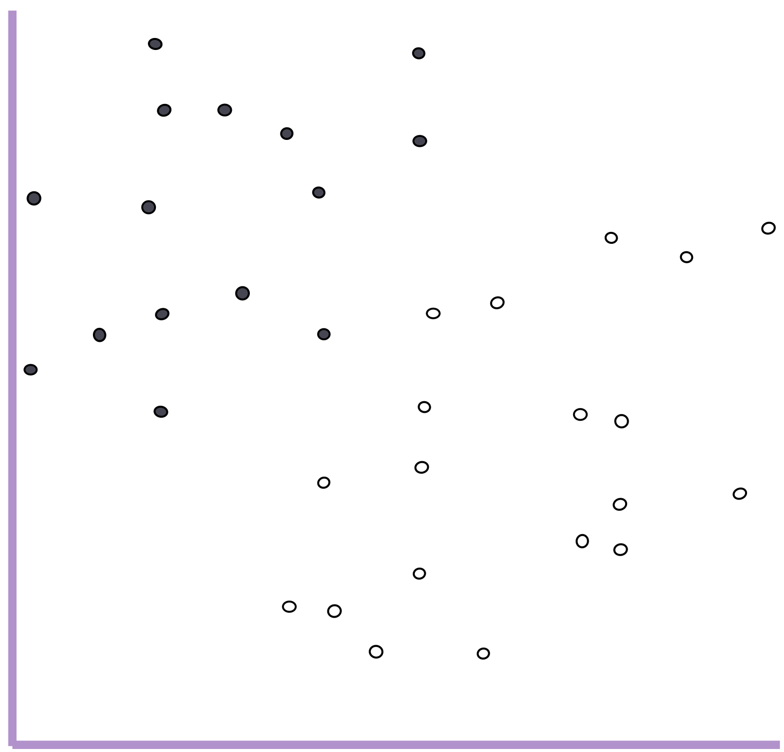
$$\forall i \quad u_i\,(w.\boldsymbol{x}_i + b) \geq 1$$

Classifieur linéaire



● denotes +1

○ denotes -1



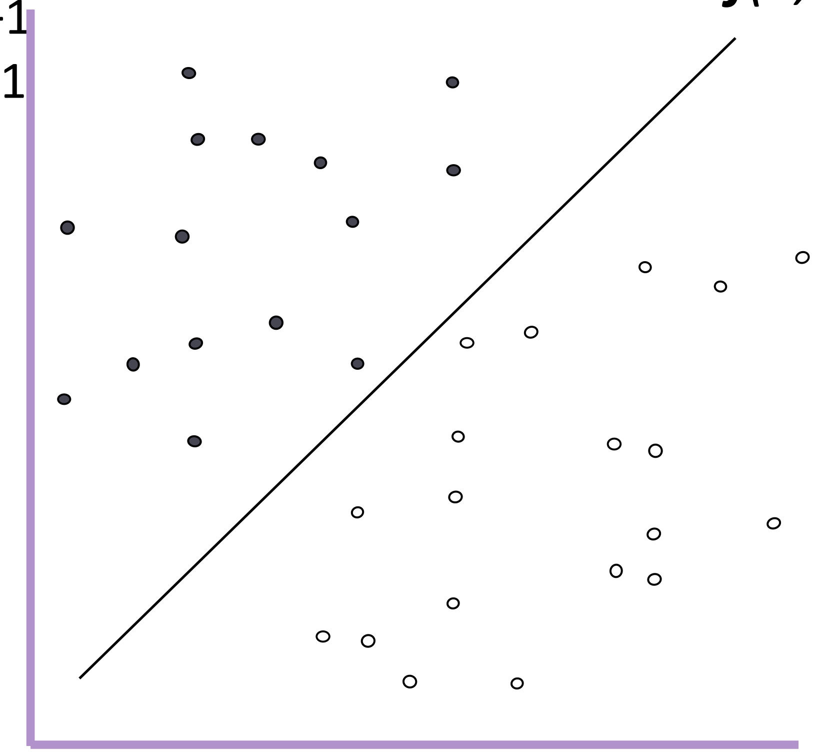
$$f(x,w,b) = sign(w \cdot x - b)$$

Comment classifier ces data

Cassifieur linéaire

$$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$$

- denotes +1
- denotes -1

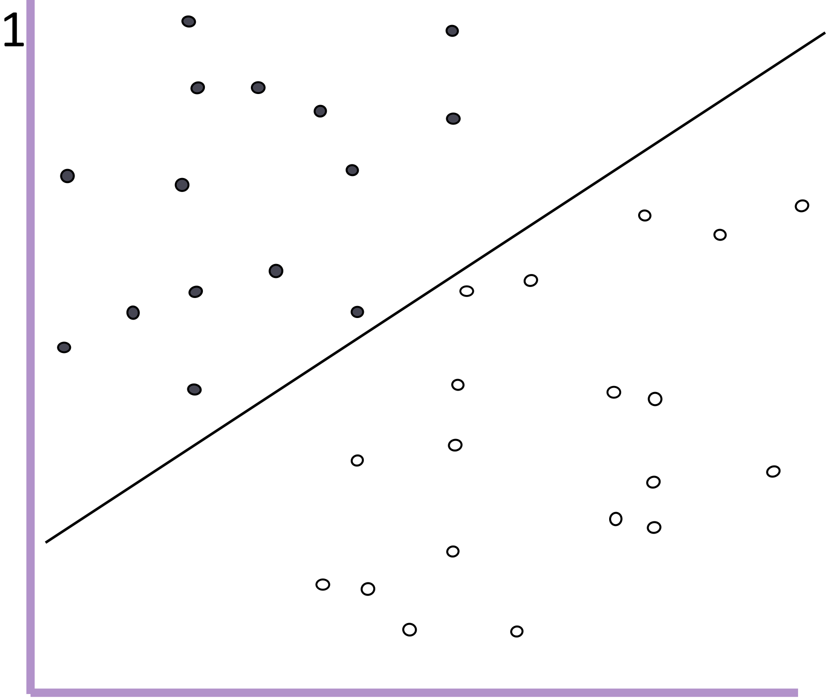


Comment classifier ces data

Cassifieur linéaire

$$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$$

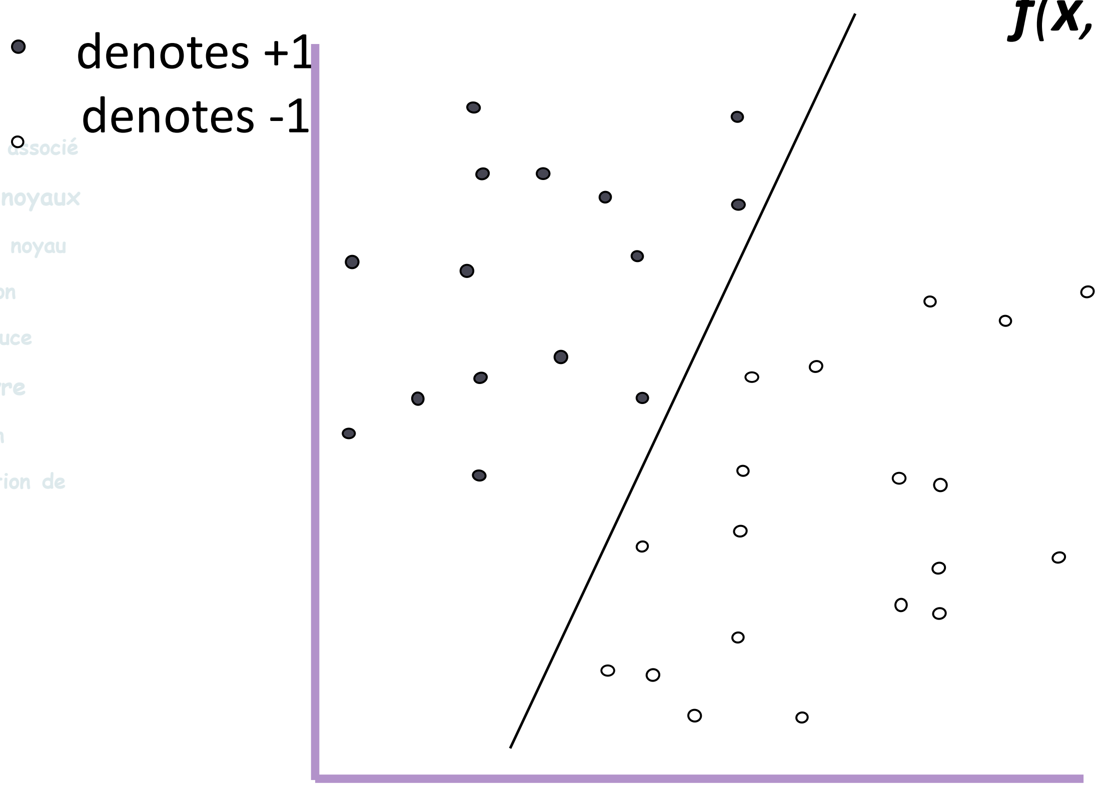
- ● denotes +1
- ○ denotes -1



Comment classifier ces data

Cassifieur linéaire

$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$

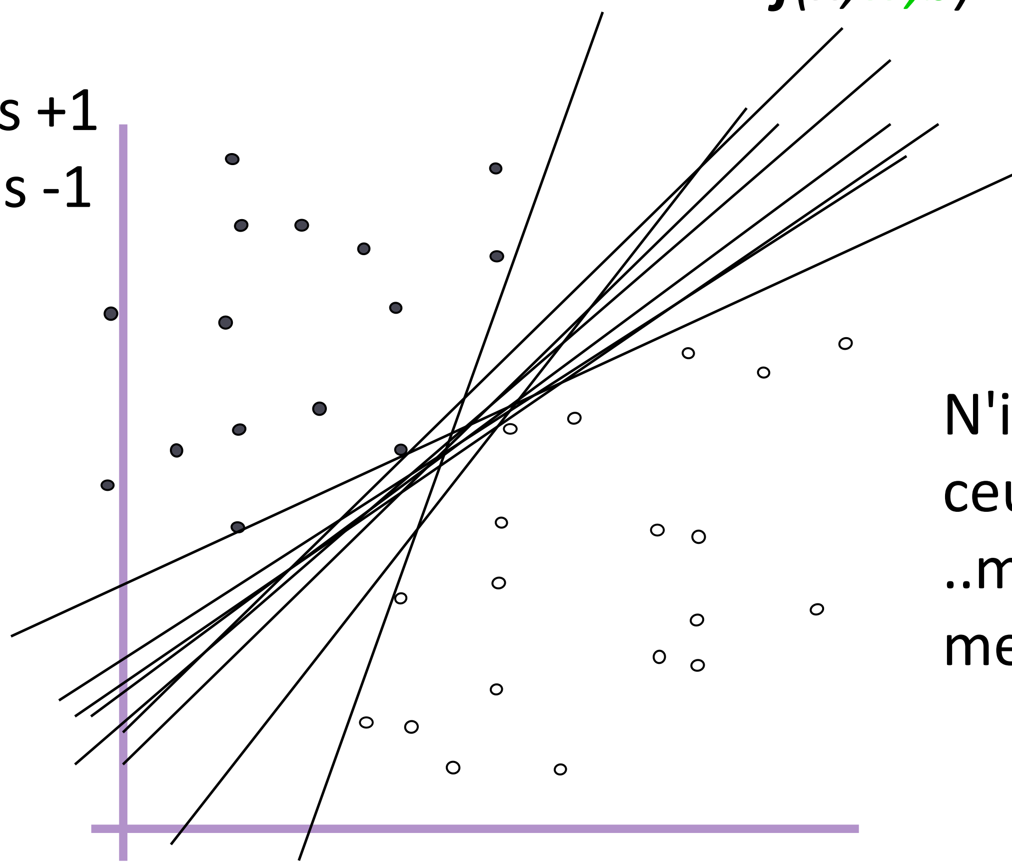


Comment classifier ces data

Cassifieur linéaire

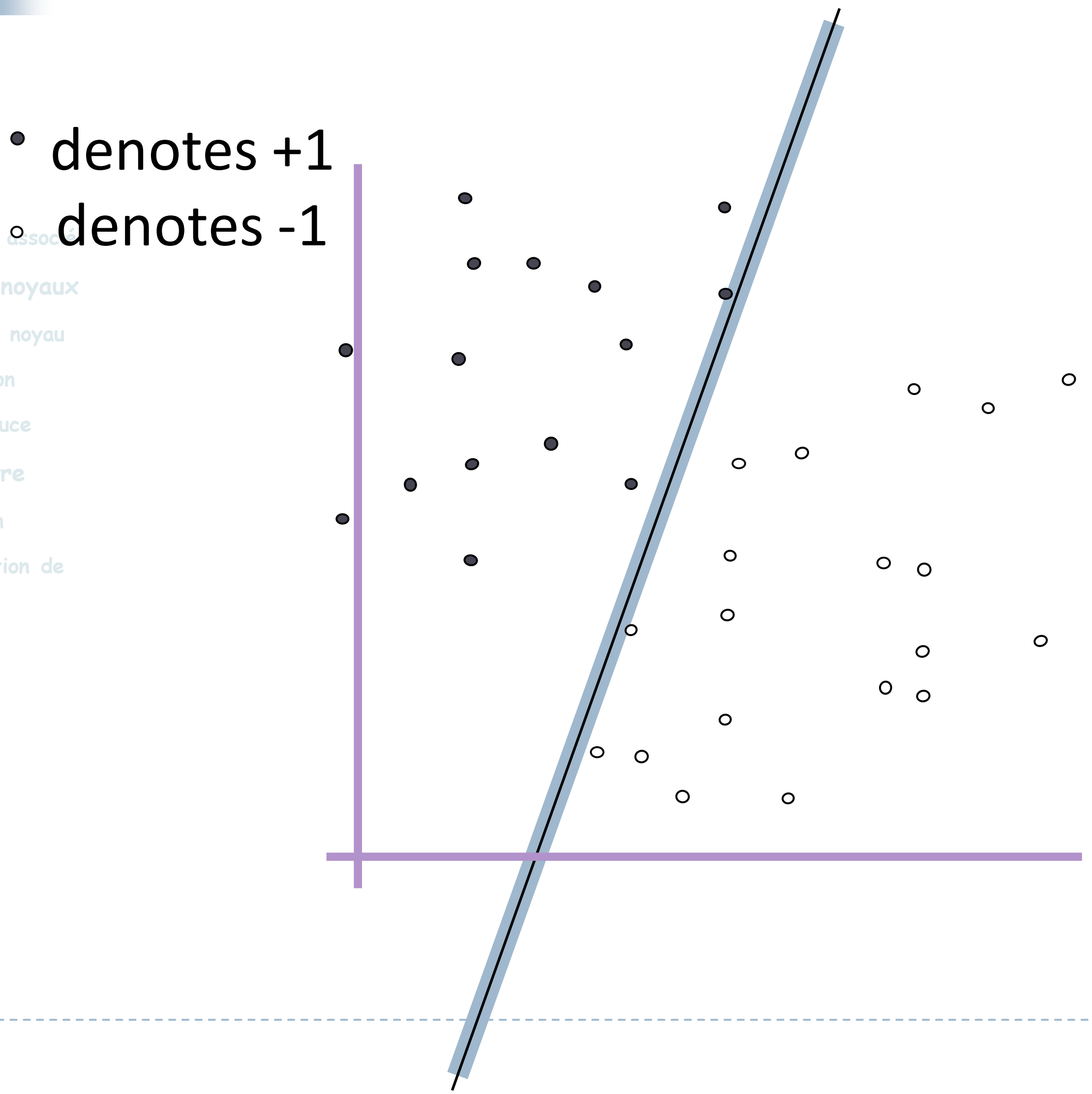
$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}.\ \mathbf{x} - b)$

- ● denotes +1
- ○ denotes -1



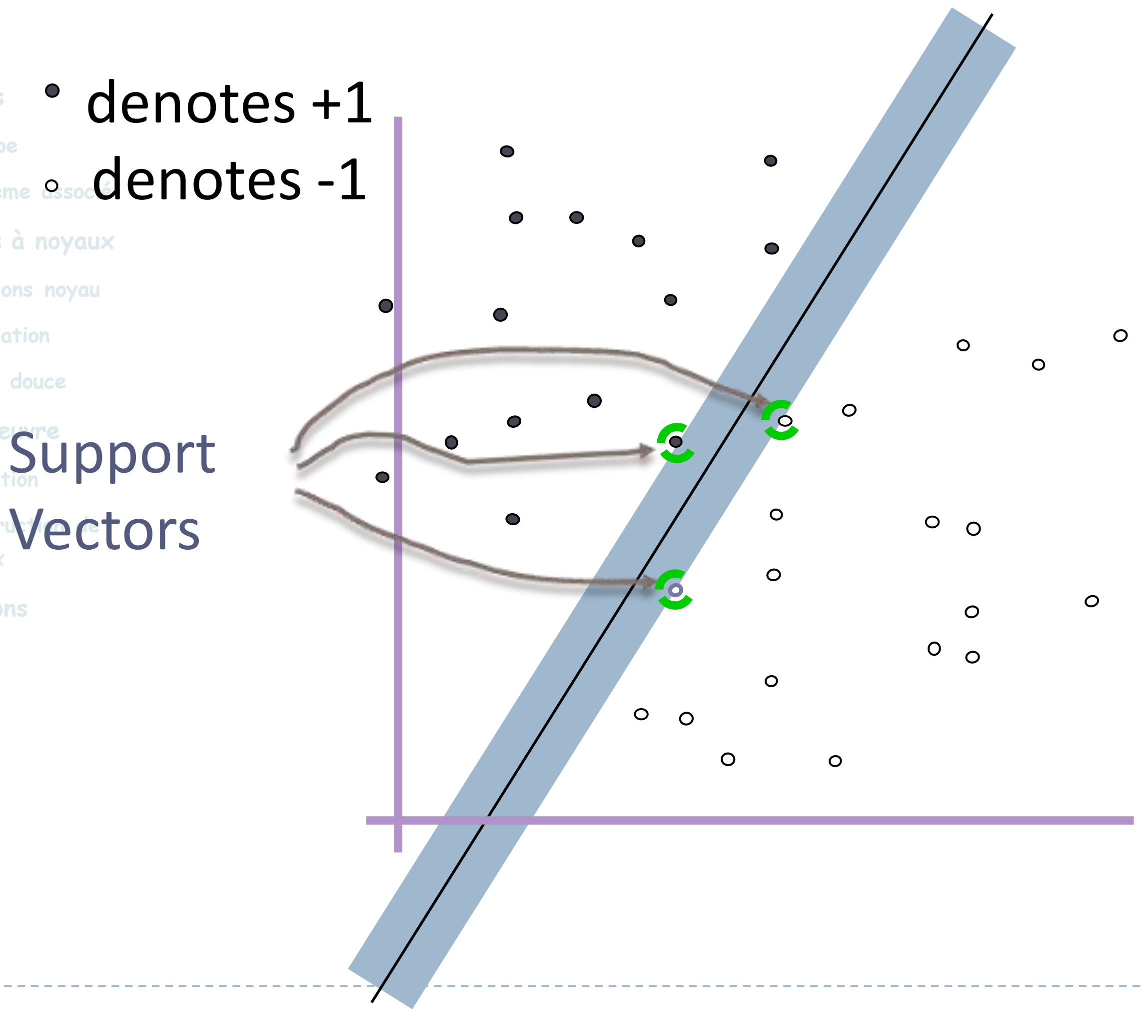
N'importe lequel de ceux-ci serait bien.
..mais quel est le meilleur?

La marge d'un classifieur



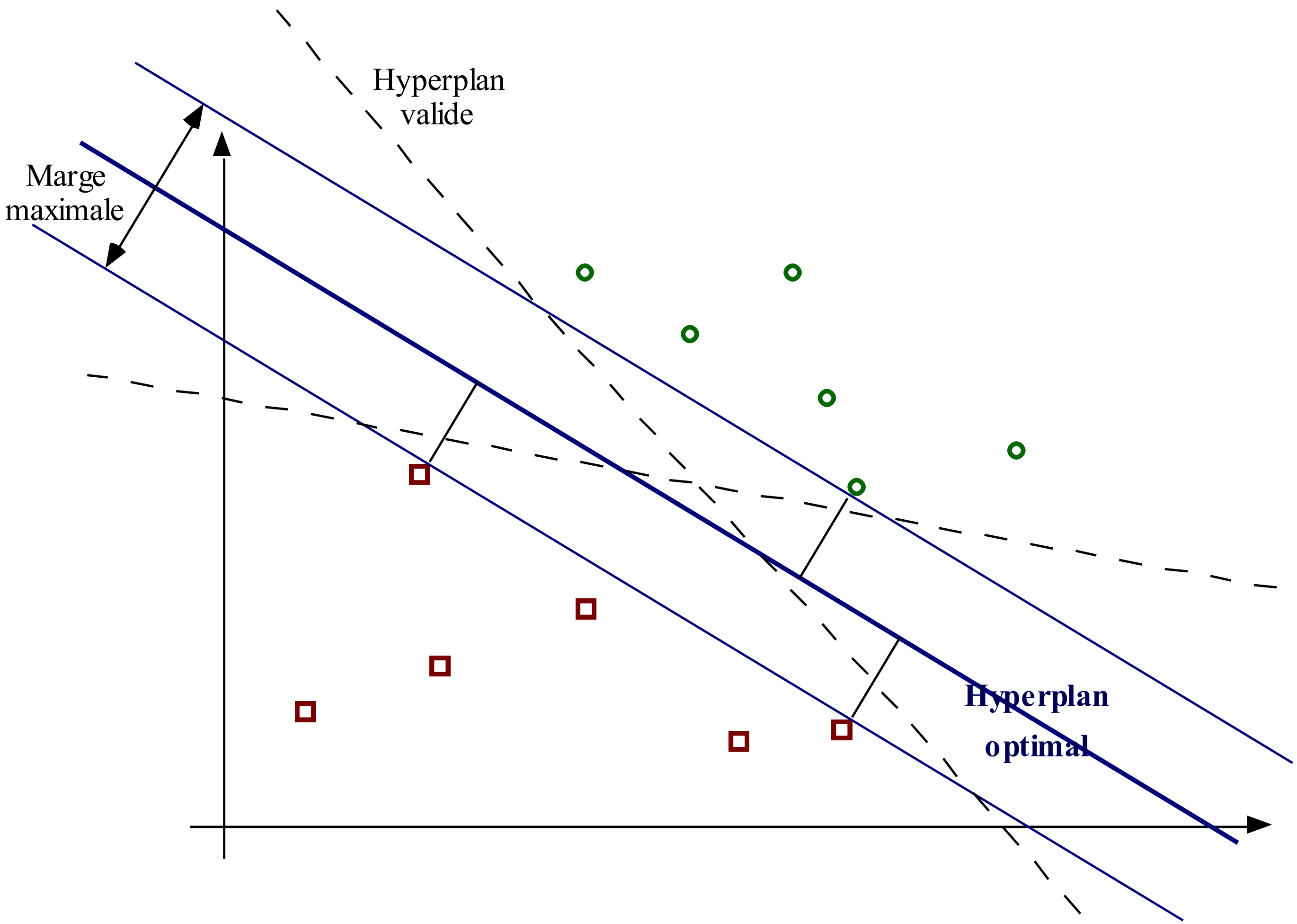
Définissez la marge d'un classificateur linéaire comme la largeur dont la limite peut être augmentée avant d'atteindre un point de données.

Maximum Margin

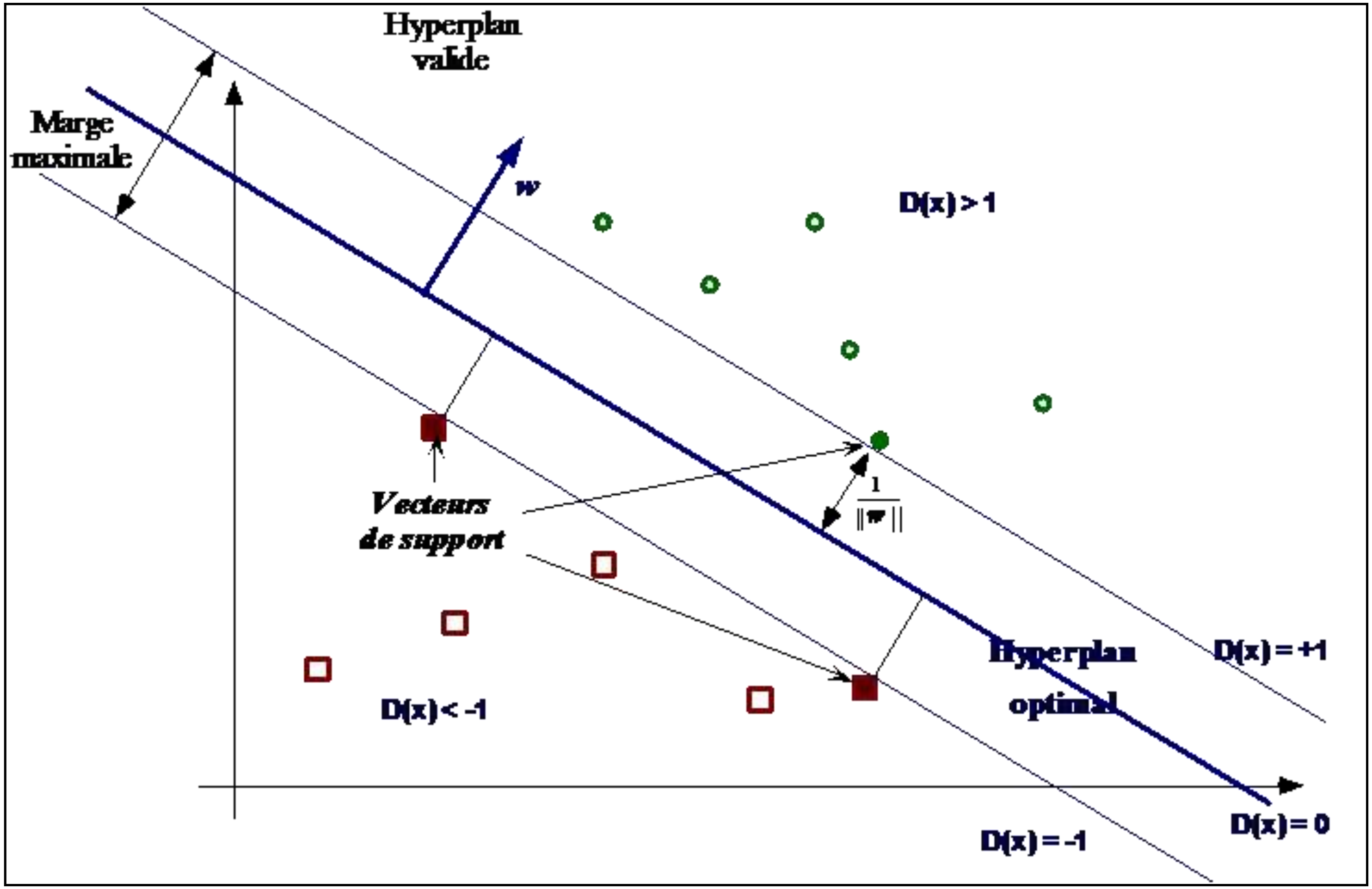


Le type le plus simple de SVM est **Linear** SVM avec une marge linéaire maximale.

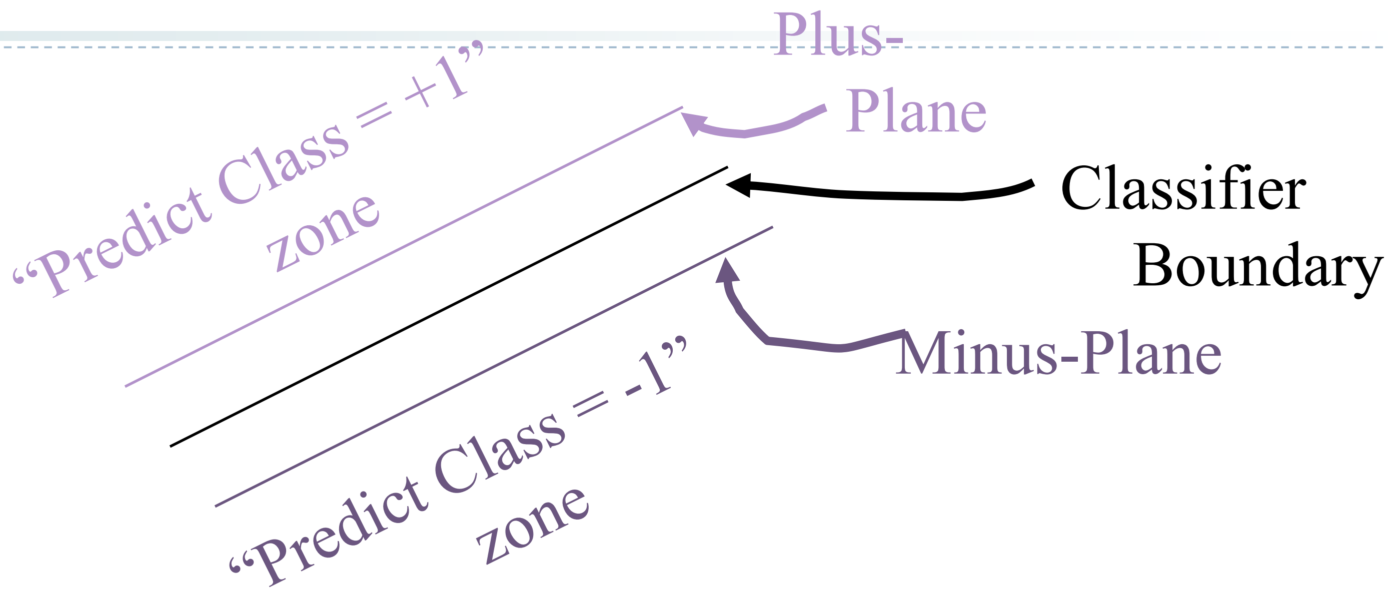
Hyperplan de plus vaste marge



Optimisation de la marge

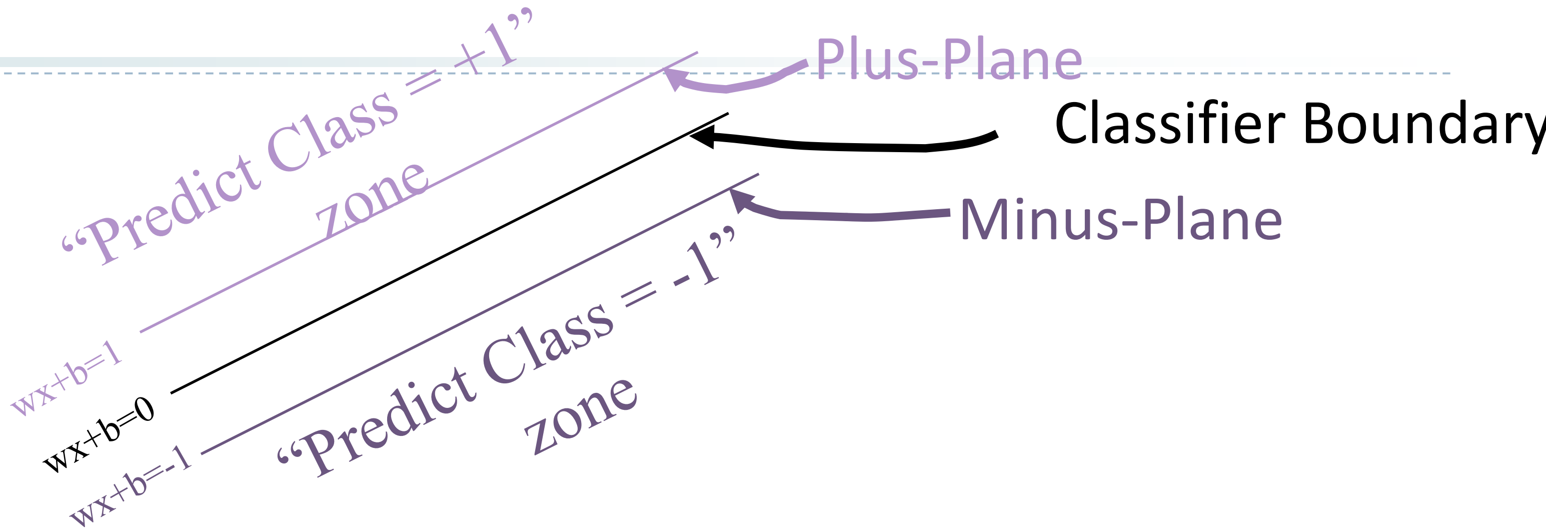


Spécifier une ligne et une marge



- Comment représenter cela mathématiquement?
- …dans *m dimensions d'entrée?*

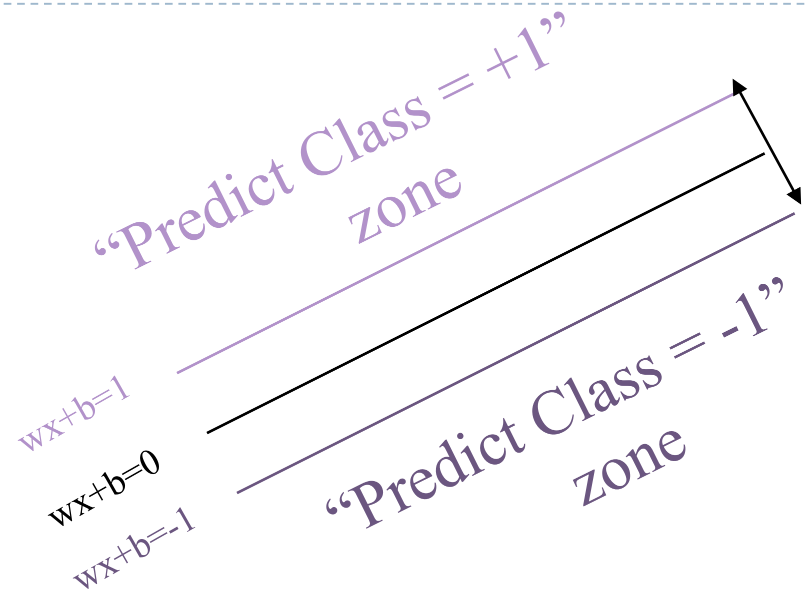
Spécifier une ligne et une marge



- Plus-plane = $\{ \mathbf{x} : \mathbf{w} \,.\, \mathbf{x} + b = +1 \}$
- Minus-plane = $\{ \mathbf{x} : \mathbf{w} \,.\, \mathbf{x} + b = -1 \}$

Classify as..	+1	if	$\mathbf{w} \,.\, \mathbf{x} + b >= 1$
	-1	if	$\mathbf{w} \,.\, \mathbf{x} + b <= -1$
	Universe explodes	if	$-1 < \mathbf{w} \,.\, \mathbf{x} + b < 1$

Calculer la largeur de la marge



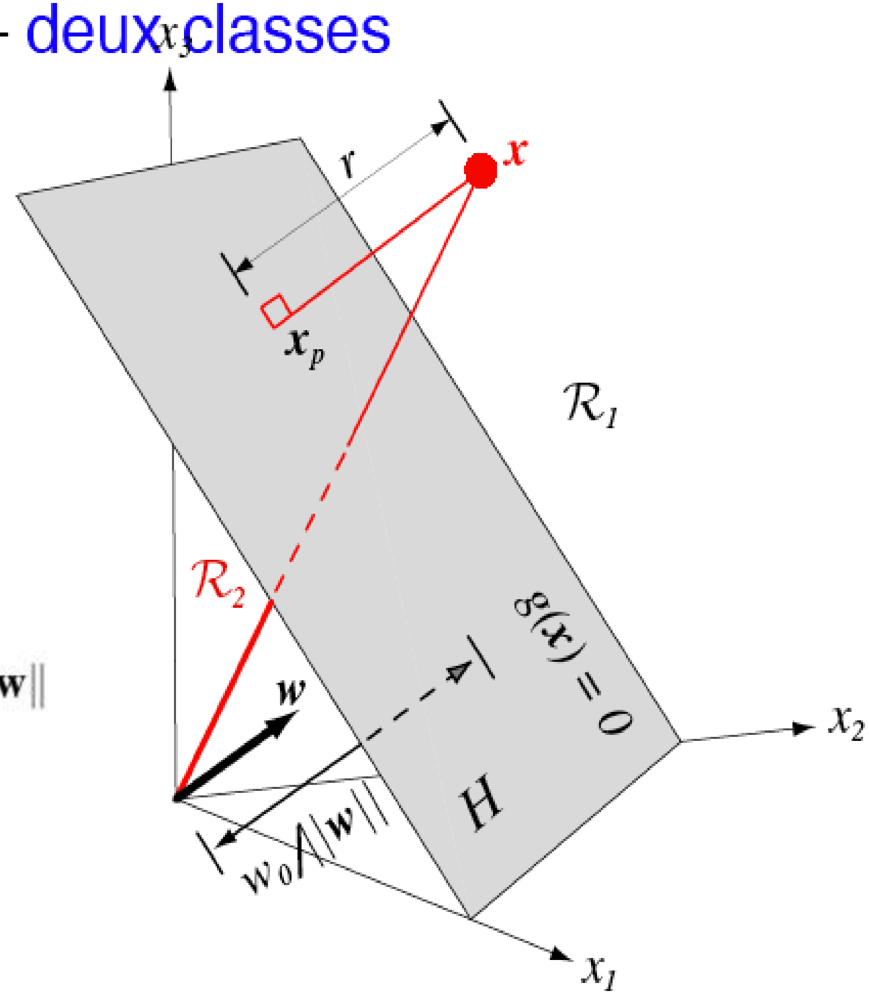
M = Largeur de la marge

Calculer M en terme de $\boldsymbol{w}$ et b?

- Plus-plane = $\{ \boldsymbol{x} : \boldsymbol{w} \, . \, \boldsymbol{x} + b = +1 \}$
- Minus-plane = $\{ \boldsymbol{x} : \boldsymbol{w} \, . \, \boldsymbol{x} + b = -1 \}$

Géométrie

- Géométrie – deux classes



$$
\begin{aligned}
\mathbf{x} &= \mathbf{x}_p + r\frac{\mathbf{w}}{\|\mathbf{w}\|} \\
g(\mathbf{x}) &= \mathbf{w}^t\mathbf{x} + w_0 = r\|\mathbf{w}\| \\
r &= \frac{g(\mathbf{x})}{\|\mathbf{w}\|}
\end{aligned}
$$

Calculer la marge

What we know:

- $\mathbf{w} \, . \, \mathbf{x}^{+} + b = +1$
- $\mathbf{w} \, . \, \mathbf{x}^{-} + b = -1$
- $\mathbf{x}^{+} = \mathbf{x}^{-} + \lambda \, \mathbf{w}$
- $|\mathbf{x}^{+} - \mathbf{x}^{-}| = M$

$\mathbf{w} \, . \, (\mathbf{x}^{-} + \lambda \, \mathbf{w}) + b = 1$

=>

$\mathbf{w} \, . \, \mathbf{x}^{-} + b + \lambda \, \mathbf{w} \, . \mathbf{w} = 1$

=>

$-1 + \lambda \, \mathbf{w} \, . \mathbf{w} = 1$

=>

$$\lambda = \frac{2}{\mathbf{w.w}}$$

$$= \frac{2}{\|w\|}$$

Optimisation de la marge

- La distance d'un point à l'hyperplan est : $$d(\boldsymbol{x}) = \frac{|\boldsymbol{w}.\boldsymbol{x} + w_0|}{\|\boldsymbol{w}\|}$$
- L'hyperplan optimal est celui pour lequel la distance aux points les plus proches (*marge*) est maximale. Cette distance vaut $\frac{2}{\|\boldsymbol{w}\|}$
- Maximiser la marge revient donc à minimiser $||w||$ sous contraintes:

$$\boxed{\begin{cases} \min \frac{1}{2}\|\boldsymbol{w}\|^2 \\ \forall \boldsymbol{i} \quad y_i(\boldsymbol{w}.\boldsymbol{x}_i + w_0) \geq 1 \end{cases}}$$

SVMs : un problème d'optimisation quadratique

- Il faut donc déterminer *w* et w_0 minimisant :

$$\eta(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2$$

(afin de maximiser le pouvoir de généralisation)

- sous les contraintes (hyperplan séparateur) :

$$u_i \left[(\boldsymbol{w} \,.\, \boldsymbol{x_i}) + w_0 \right] \geq 1 \,, \qquad i = 1, \ldots, n$$

SVMs : un problème d'optimisation quadratique

d : dimension de l'espace d'entrée

Il faut régler $d + 1$ paramètres

- Possible quand d est assez petit

  avec des méthodes d'optimisation quadratique

- Impossible quand d est grand (> qqs 10^3)

Transformation du problème d'optimisation

- Méthode des multiplicateurs de Lagrange

$$\begin{cases} L(\boldsymbol{w}, w_0, \alpha) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{l} \alpha_i \{(\boldsymbol{x}_i.\boldsymbol{w} + w_0)\boldsymbol{y}_i - 1\} \\ \forall i \;\; \alpha_i \geq 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial Q(w,b,\alpha)}{\partial w} = 0 & (a) \\ \dfrac{\partial Q(w,b,\alpha)}{\partial b} = 0 & (b) \\ \alpha_i\{y_i(w^T x_i + b) - 1\} = 0 & (c) \\ \alpha_i \geq 0 & (d) \end{cases}$$

$$\begin{cases} w = \sum_{i=1}^{n} \alpha_i y_i x_i \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases}$$

Propriétés de la forme duale

Problème dual

$$\begin{cases} \max\limits_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \boxed{(\boldsymbol{x}_i . \boldsymbol{x}_j)} \\ \forall i \;\; \alpha_i \geq 0 \\ \sum_{i=1}^{l} \alpha_i y_i = 0 \end{cases}$$

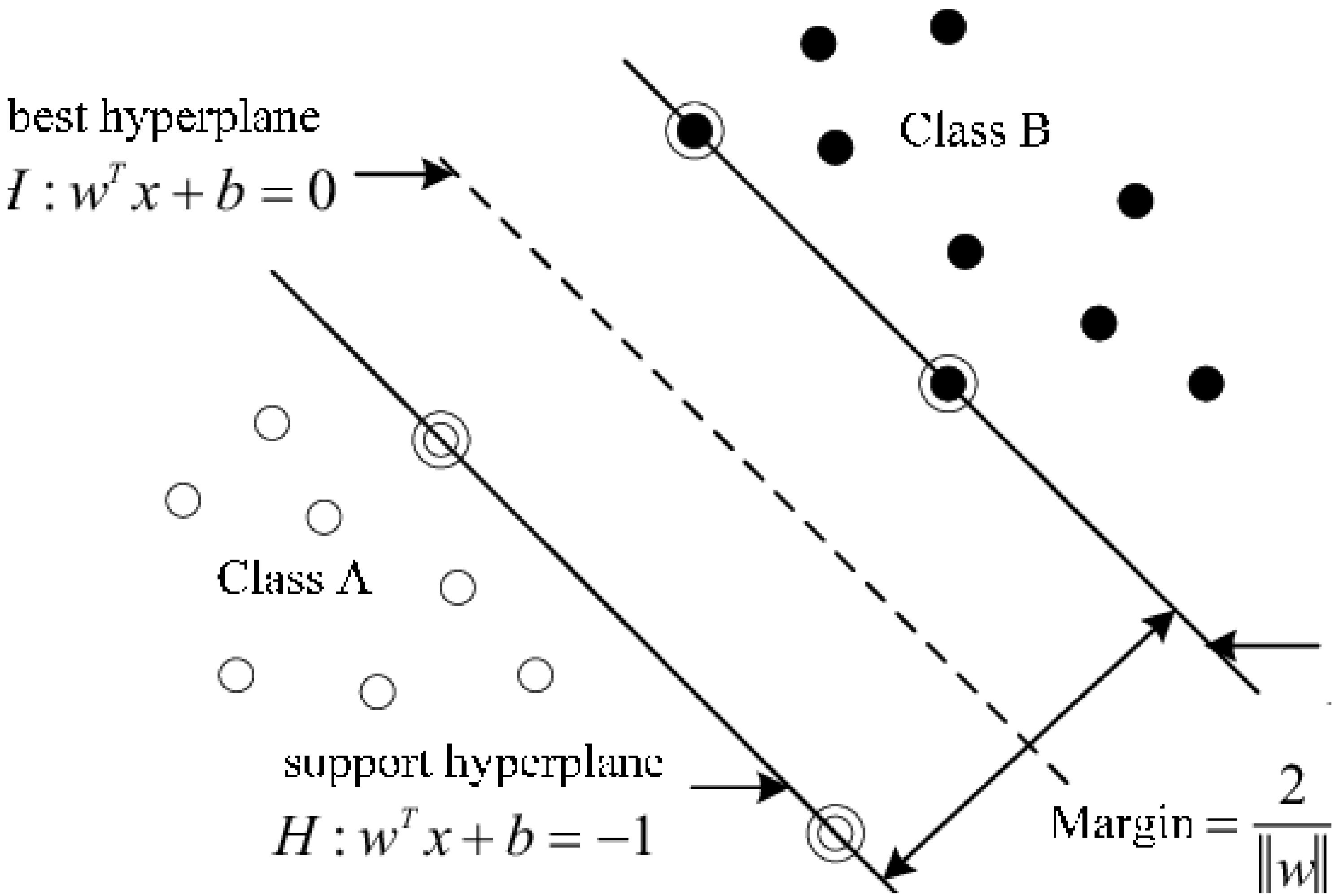
Fonction de décision

Fonction de décision

$$H(x) = \sum_{S} \alpha_i y_i x^T x_i + b$$

$$\begin{cases} x \in Classe+1 \quad si \;\; H(x) > 0 \\ x \in Classe-1 \quad si \;\; H(x) < 0 \\ x \; est \; inclassifiable \; si \; H(x) = 0 \end{cases}$$

Classification



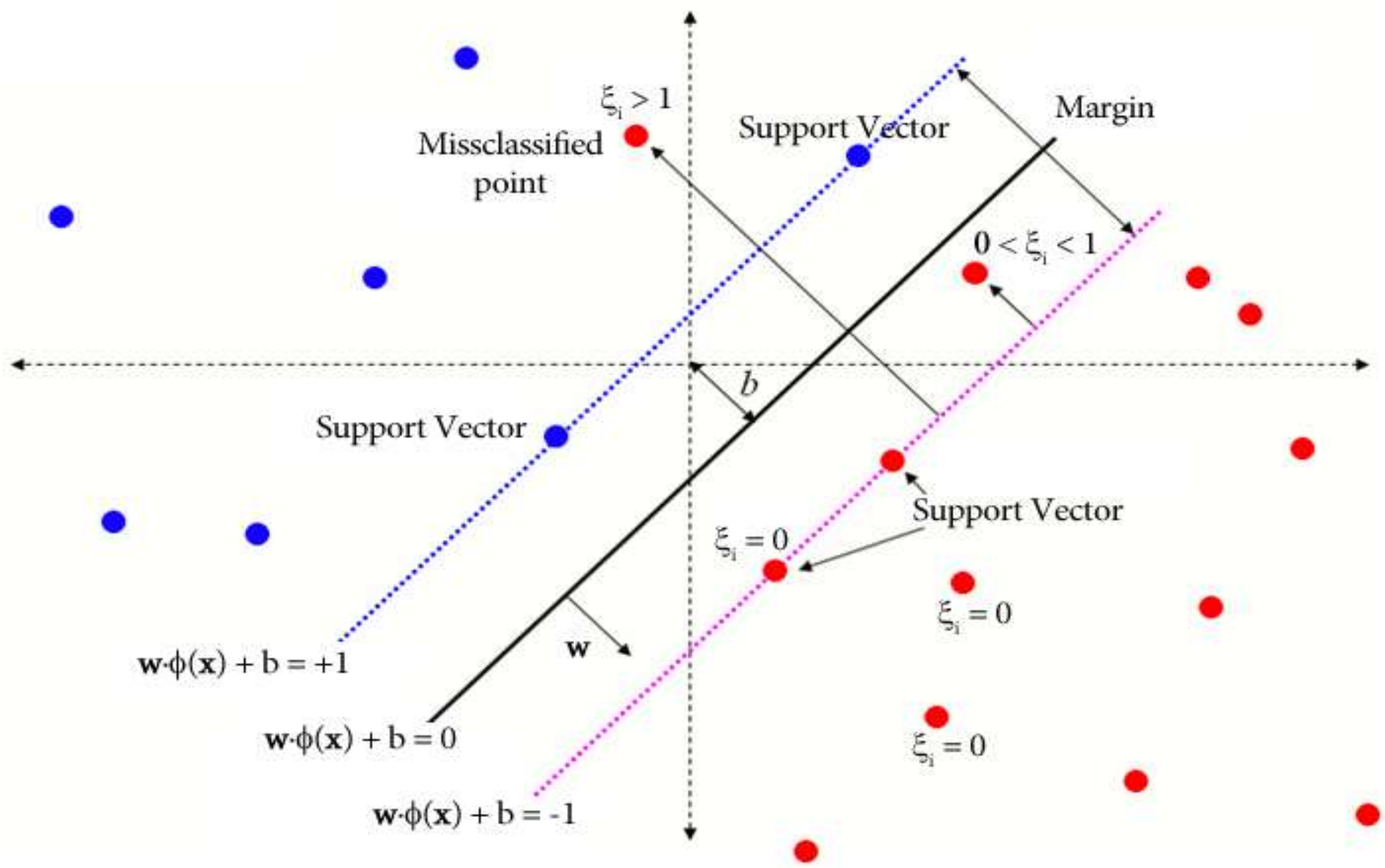
SVMs : un problème d'optimisation quadratique

- En réalité, un hyperplan séparateur n'existe pas toujours, et même s'il existe, il ne représente pas généralement la meilleure solution pour la classification.

- On relaxe un peut les contraintes en admettant une certaine erreur de classification des données ce qui est appelé "SVM à marge souple (Soft Margin).

- On introduit alors sur les contraintes des variables i dites de relaxation pour obtenir la contrainte de l'équation :

SVM binaire à marge souple.



SVM binaire à marge souple.

- introduire alors sur les contraintes des variables dites de relaxation pour obtenir la contrainte de l'équation :

$$y_i(w^T x_i + b) \leq 1 - \varepsilon_i \ , i = 1..n$$

- Le problème dual devient donc :

$$\begin{cases} Minimiser \quad \dfrac{1}{2}\|w\|^2 + C\sum_{i=0}^{n} \varepsilon_i \\ \text{Sous contraintes} \\ y_i(w^T x_i + b) \geq 1 - \varepsilon_i \ , i = 1..n \\ \varepsilon_i \geq 0 \end{cases}$$

SVM binaire à marge souple.

multiplicateurs de Lagrange

$$Q(w,b,\alpha,\varepsilon,\beta) = \frac{1}{2}w^T w + C\sum_{i=1}^{n}\varepsilon_i - \sum_{i=1}^{n}\alpha_i y_i(w^T x_i + b) - 1 + \varepsilon_i - \sum_{i=0}^{n}\beta_i\varepsilon_i$$

$$\begin{cases} \dfrac{\partial Q(w,b,\varepsilon,\alpha,\beta)}{\partial w} = 0 & (a) \\ \dfrac{\partial Q(w,b,\varepsilon,\alpha,\beta)}{\partial b} = 0 & (b) \\ \alpha_i\{y_i(w^T x_i + b) - 1 + \varepsilon_i\} = 0 & (c) \\ \beta_i\varepsilon_i = 0 & \\ \alpha_i \geq 0, \beta_i \geq 0\ , \varepsilon_i \geq 0 & (d) \end{cases}$$

SVM binaire à marge souple.

$$\begin{cases} w = \sum_{i=1}^{n} \alpha_i y_i x_i \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \\ \alpha_i + \beta_i = C \end{cases}$$

$$\begin{cases} \textit{Maximiser} \quad Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \textbf{\textit{sous contrainte}} \\ \sum_{i=1}^{n} \alpha_i \, y_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}$$