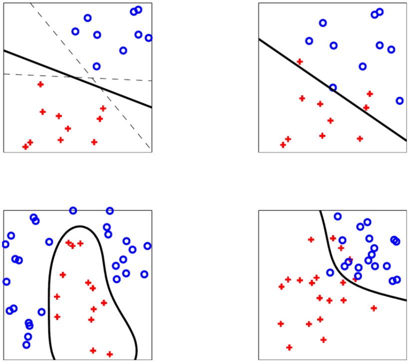
SVM

SVMs

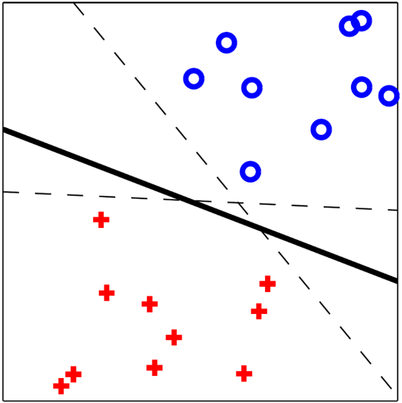
Plan du cours

- The linearly separable case
- The linearly separable case
- The kernel trick
- Non linear SVMs

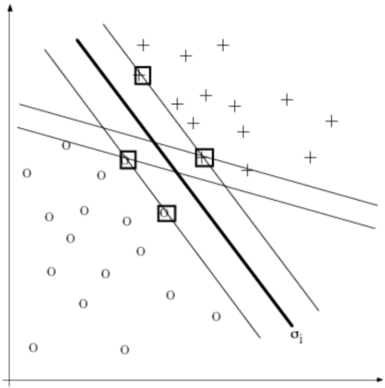
Some examples of classification problems



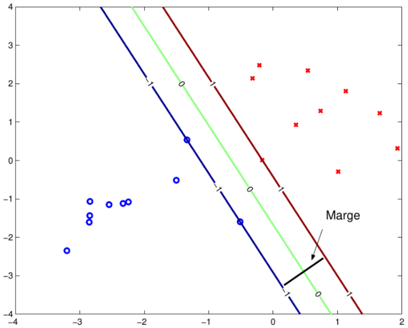
Some examples of classification problems



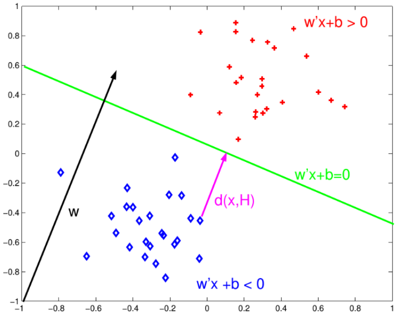
Notion of margin



Notion of margin



Notion of margin



Definitions

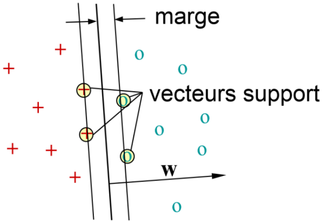
- Margin : Distance between the nearest example of the training set and the decision boundary.
- Dataset : $\{(x_i, y_i), i = 1, \cdots, n\}$, $x_i \in \mathbb{R}^d$, $y_i \in \{1, -1\}$
- Decision function : $f(x) = w^T x + b = 0$
  - $f(x) = 0$: separating hyperplane
  - $f(x) > 0$: class 1 ($y_i = 1$)
  - $f(x) < 0$: class 2 ($y_i = 1$)

Formalization of the problem

- Decision function : $h_w(x) = w^T x + b$. Separating hyperplane $h_w(x) = 0$
- Parameters :
  - w is normal to the hyperplane
  - b is the offset
- The parameters w and b are non unique. Indeed, kw and kb yield the same decision boundary

Formalization of the problem

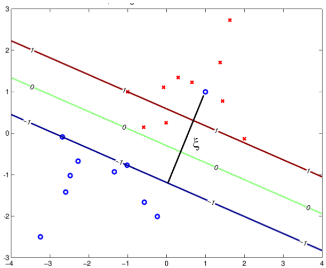
How should we choose $h_w(x) = w^T x + b = 0$?



Solution : we maximize the margin!

Non linearly separable case

What can we do if the data are not linearly separable?

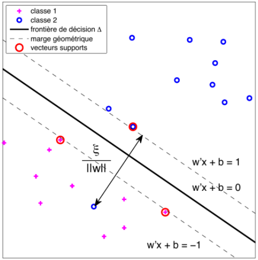


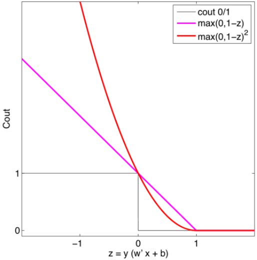
Non linearly separable case

- Idea : model potential errors using slack positive variables ξ_i associated to the observations (x_i, y_i)
- We have two situations :
  - No error : $y_i(w^T x_i + b) \geq 1 \Rightarrow \xi_i = 0$
  - Error : $y_i(w^T x_i + b) < 1 \Rightarrow \xi_i = 1 - y_i(w^T x_i + b) > 0$

Non linearly separable case

We associate to this situation a cost function :

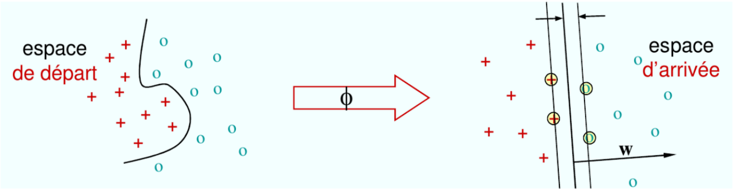




In this example one single point is misclassified

Non linear SVMs

- How can we extend this algorithm to the non linear case ?
- Principle : we transport the data into another space where the data are linearly separable using a transformation $\varphi : x \rightarrow \varphi(x)$

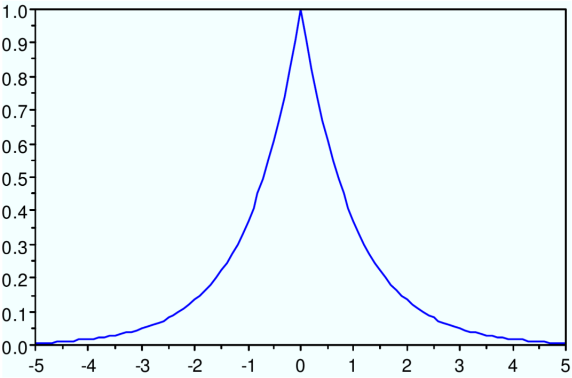


Non linear SVMs

The Kernel trick

- We observe that in the SVM algorithm, the only quantities involved are $K(x, y) =< \phi(x), \phi(y) >$
- We are given an explicit expression of K and forget ϕ

Non linear SVMs
Examples of kernels



Non linear SVMs
Examples of kernels