

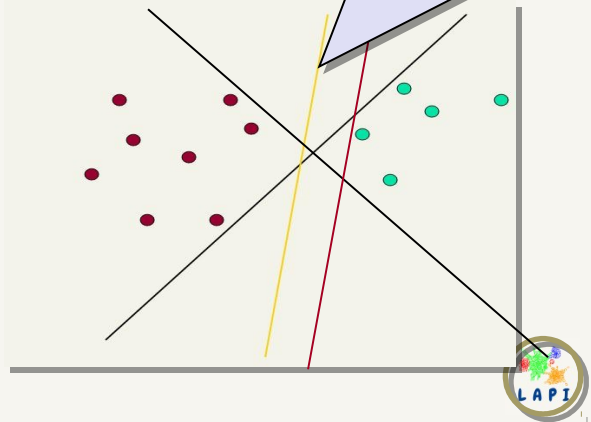
# **Support Vector Machine**

**Mașină cu Vectori Support**

# Clasificatori liniari: Ce Hyperplan folosim?

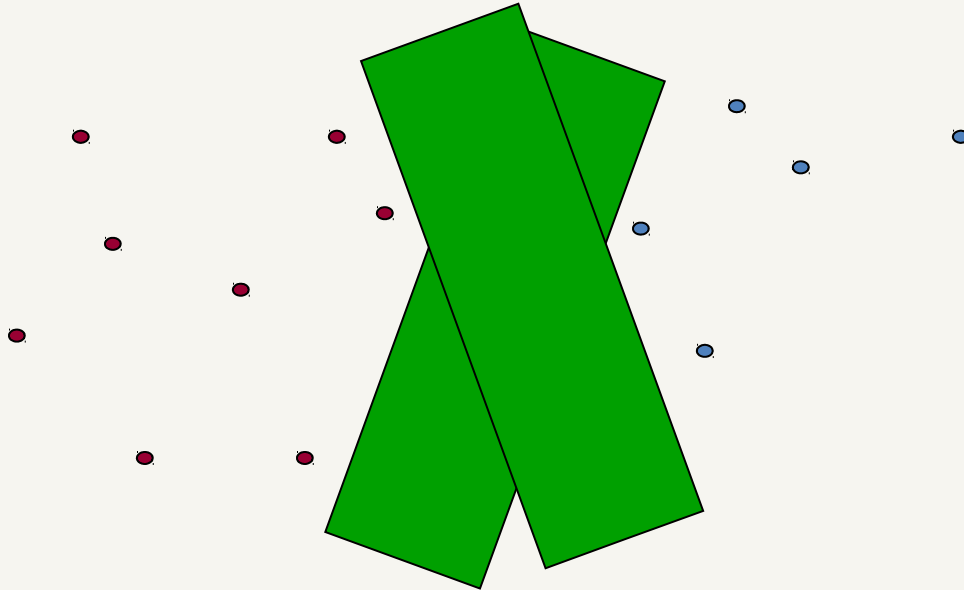
- O mulțime de soluții posibile pentru  $a$ ,  $b$ ,  $c$ .
- Unele metode găsesc un hiperplan de separație, dar nu cel optim **[conform unor criterii de potrivire așteptate]**
  - E.g., perceptron
- Support Vector Machine (SVM) găsește o soluție optimală .
  - Maximizează distanța dintre hiperplan și „punctele dificile” aproape de limita deciziei
  - O intuiție: dacă nu există puncte în apropierea suprafeței de decizie, atunci nu există decizii de clasificare foarte incerte

Această linie reprezintă limita de decizie (hiperplanul de separație):  
 $ax + by - c = 0$



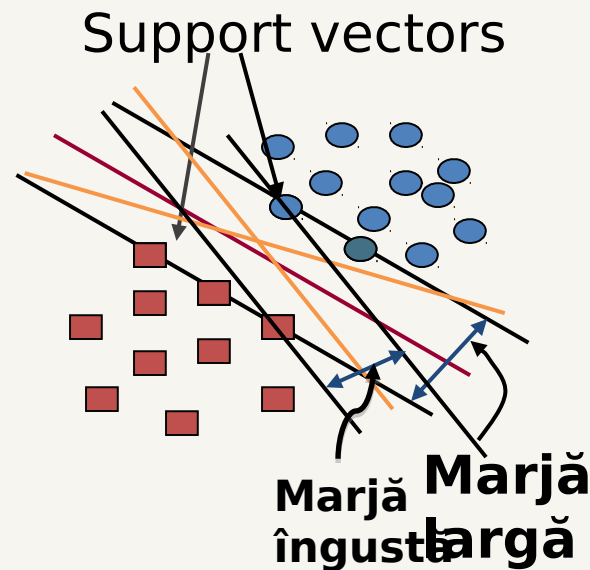
# Intuitiv

Dacă trebuie să așezați un separator lat între clase, aveți mai puține opțiuni și astfel capacitatea modelului este redusă



# Support Vector Machine (SVM)

- SVM maximizează marja (**margin**) din jurul hiperplanului de separație.
  - A.k.a. large margin classifiers
- Funcția de decizie este complet specificată de un subset de exemple de antrenare – vectorii suport (**support vectors**).
- Este o metodă de clasificare foarte eficientă



# Marja maximă: formalizare

- $\mathbf{w}$ : vectorul normal pe hiperplanul de decizie
- $\mathbf{x}_i$ : data  $i$
- $y_i$ : clasa datei  $i$  (+1 or -1) NB: Not 1/0
- Clasisifcatorul este :  $f(\mathbf{x}_i) = \text{sign}(\mathbf{w}^T \mathbf{x}_i + b)$

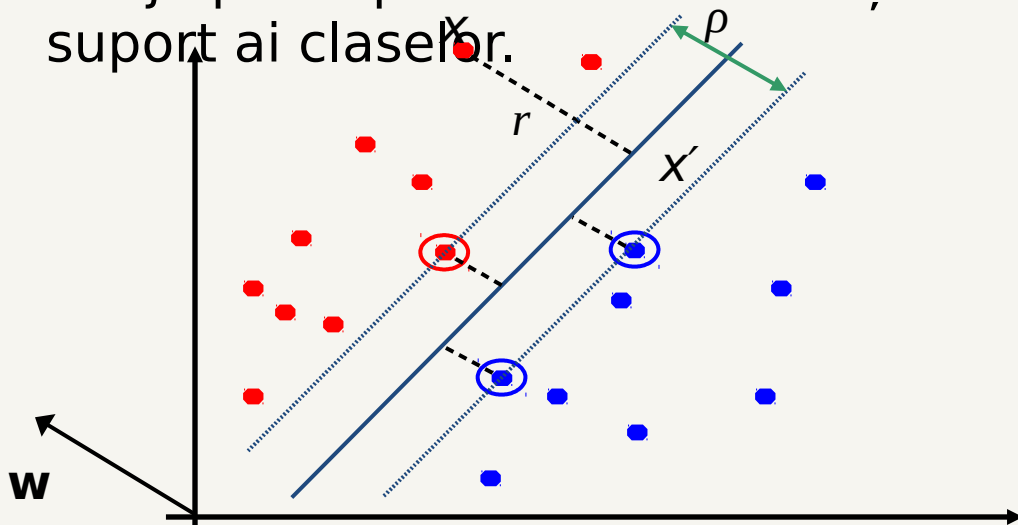
*Marja Funcțională a lui  $\mathbf{x}_i$  este:*  $y_i (\mathbf{w}^T \mathbf{x}_i + b)$

– But note that we can increase this margin simply by scaling  $\mathbf{w}$ ,  $\mathbf{b}$ ....

- Marja funcțională a setului de date este de două ori mai mare decât marja funcțională minimă pentru orice punct

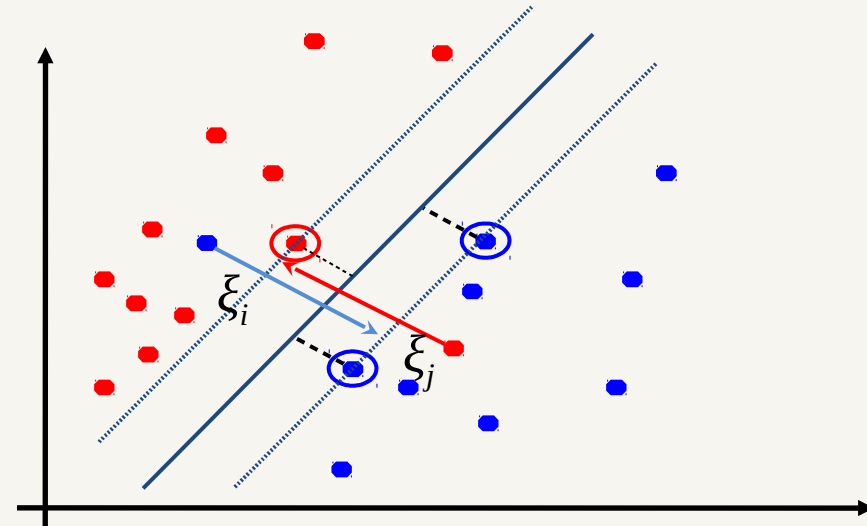
# Marja geometrică

- Distanța de la exemplu la separator este  $\epsilon = y \frac{\mathbf{w}^T \mathbf{x} + b}{\|\mathbf{w}\|}$
- Exemplele cele mai apropiate de hiperplan sunt **vectorii suport**.
- Marja  $\rho$  a separatorului este lățimea de separare între vectorii suport ai claselor.



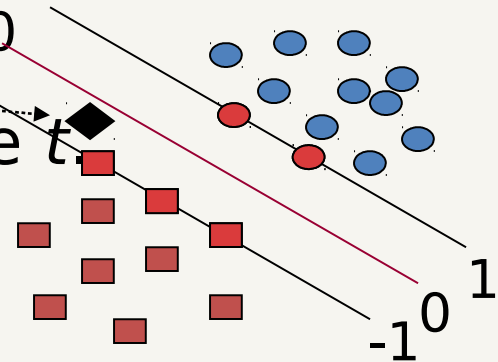
# Soft Margin Classification

- Dacă datele de instruire nu sunt liniar separabile, se pot adăuga variabile ajustabile  $\xi_i$  pentru a permite clasificarea greșită a unor exemple dificile sau zgomotoase.
- Permiteți unele erori
  - Lăsați anumite puncte să se mute în locul în care aparțin, contra cost
- Totuși, încercați să minimalizați erorile de formare și să plasați hiperplanul „departe” de fiecare clasă (mariă mare)



# Clasificare cu SVM-uri

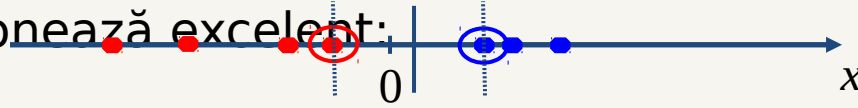
- Având în vedere un nou punct  $x$ , putem marca proiecția sa pe normala la hiperplan:
  - I.e., calculează:  $\mathbf{w}^T \mathbf{x} + b = \sum \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b$
  - Decide clasa pe baza dacă  $<$  or  $> 0$
- Putem stabili pragul de încredere  $t$ .
  - Scorul  $> t$ : yes
  - Scorul  $< -t$ : no
  - Altfel: nu decidem





# SVM-uri neliniare

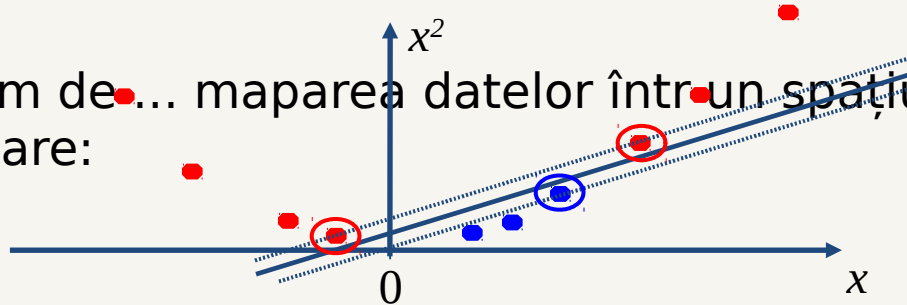
- Seturile de date care se pot separa liniar (cu ceva zgomot) funcționează excelent:



- Dar ce vom face dacă setul de date este prea greu?



- Ce zicem de... maparea datelor într-un spațiu de dimensiuni superioare:



# Kernels

- Why use kernels?
  - Make non-separable problem separable.
  - Map data into better representational space
- Common kernels
  - Linear
  - Polynomial  $K(x_i, x_j) = (1 + x_i x_j)^n$ 
    - Gives feature conjunctions
  - Radial basis function (infinite dimensional space)

$$K(x_i, x_j) = \frac{e^{-\|x_i - x_j\|^2}}{2\sigma^2}$$

