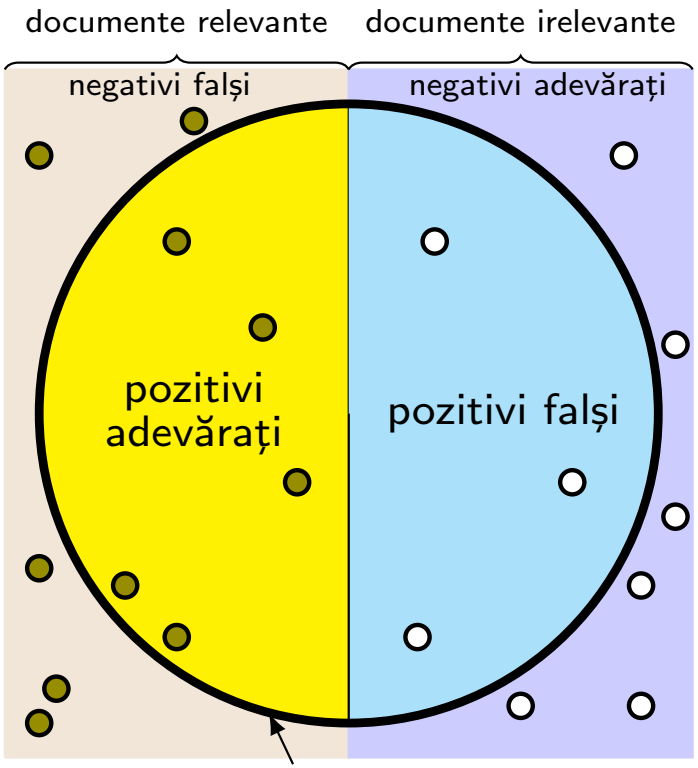
Rețele

Mașini Vector Suport.
Machine learning pentru clasificarea documentelor

Mircea Marin

Departamentul of Informatică
Universitatea de Vest din Timișoara
mircea.marin@e-uvt.ro

noiembrie 2018



● : document relevant
○ : document irelevant

$$\text{precizie} = \frac{\text{pozitivi adevăraţi}}{\text{nr.răspunsuri}} = \frac{5}{8} = 0.625$$

$$\text{reamintire} = \frac{\text{pozitivi adevăraţi}}{\text{nr.doc.relevante}} = \frac{5}{10} = 0.5$$

Precizia măsoară **exactitatea** sau **calitatea** răspunsurilor.

Reamintirea măsoară **completitudinea** sau **cantitatea** răspunsurilor.

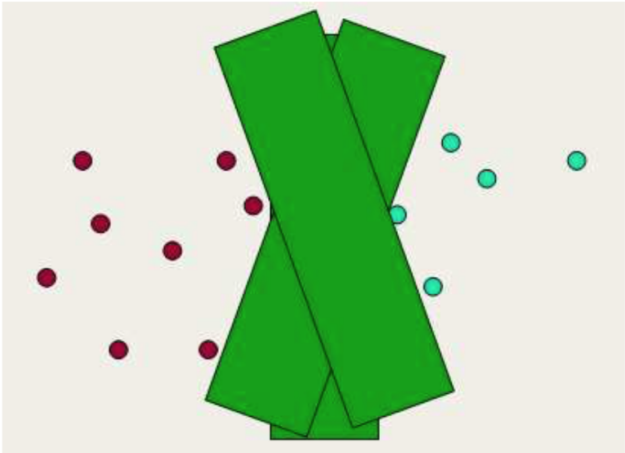
$$\text{acurateţe} = \frac{\text{pozitivi adevăraţi} + \text{falşi adevăraţi}}{\text{nr.total documente}} = \frac{11}{19} = 0.579$$

Acurateţea este o măsură a efectivităţii metodelor de clasificare bazate pe machine learning.

Eforturile de creştere a acurateţii din ultimele decenii au produs metode performante de clasificare:

- maşini vector suport (SVM): metodă de clasificare efectivă pt. documente reprezentate în spaţiu vectorial.
- arbori amplificaţi de decizie (engl. boosted decision trees)
- regresie logistică regularizată
- reţele neuronale
- păduri aleatoare (engl. random forests)

-
- SVM separă clasele calculând o suprafață de decizie aflată la distanță maximă de punctele clasificate.
- Exemplu ilustrat de suprafețe (benzi) de decizie posibile, colorate verde:



Presupunem dată o mulțime de antrenare $\mathbb{D} = \{(\vec{d_i}, y_i) \mid 1 \leq i \leq N\}$ cu

- documente $\vec{d_i}$ reprezentare în un spațiu vectorial $\mathbb{R}^M$.
- 2 clase: -1 și 1. Deci $y_i \in \{-1, 1\}$ pentru $1 \leq i \leq N$.

Un hiperplan de decizie $\langle \vec{w}, b \rangle$ este dat de ecuația

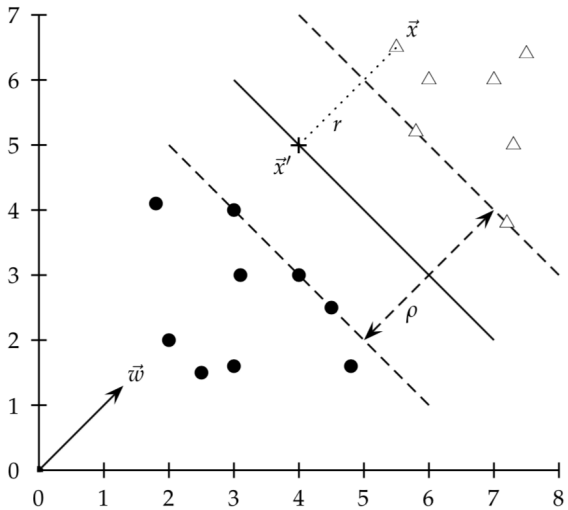
$$\vec{w} \cdot \vec{x} = -b$$

unde $\vec{w} \in \mathbb{R}^M$ și $b \in \mathbb{R}$. Hiperplanul de decizie definește clasificatorul liniar

- $f : \mathbb{R}^M \to \mathbb{R}$, $f(\vec{d}) := \text{sign}(\vec{w} \cdot \vec{d} + b)$

Distanța de la un punct $\vec{d}$ într-o clasă $y \in \{-1, 1\}$ la un hiperplan $\langle \vec{w}, b \rangle$ dat de ecuația

$$\vec{w} \cdot \vec{x} = -b \quad \text{este} \quad r = y \cdot \frac{\vec{w} \cdot \vec{d} + b}{|\vec{w}|}$$



$y \cdot (\vec{w} \cdot \vec{d} + b)$ se numește marginea funcțională a lui $\langle \vec{d}, y \rangle$ în raport cu hiperplanul $\langle \vec{w}, b \rangle$.

Problema de învățare a valorilor $\vec{w}, b, \rho$ a mașinilor vector suport (SVM) pentru 2 clase liniar separabile este:

(1) Se impune condiția ca marginile funcționale ale tuturor documentelor $\vec{d_i}$ din setul de antrenare să fie ≥ 1, adică

$$y_i \cdot (\vec{w} \cdot \vec{d_i} + b) \geq 1$$

și 1 pentru cel puțin un document de antrenare $\langle \vec{d_i}, y_i \rangle \in \mathbb{D}$.

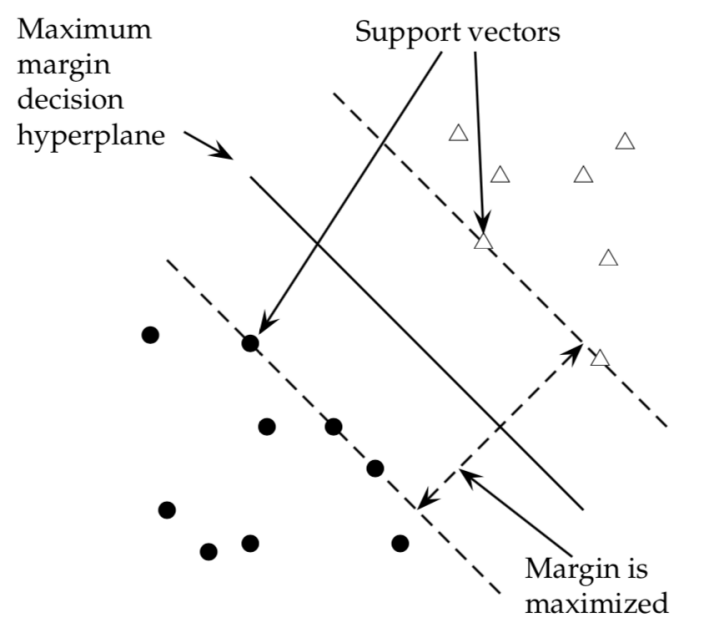
- vectorii $\vec{d_i}$ ai documentelor de antrenare pt. care $y_i \cdot (\vec{w} \cdot \vec{d_i} + b) = 1$ se numesc vectori suport

(2) Se maximizează valoarea lui $\rho = 2/|\vec{w}|$
$\Leftrightarrow$ minimizarea valorii lui $\frac{1}{2}|\vec{w}|^2$
când au loc constrângerile impuse la (1).

$\Rightarrow$ **problemă de minimizare pătratică** (vezi slide-ul următor)

- ρ este lățimea benzii de separare, și se numește margine geometrică a SVM.
- $\rho/2$ coincide cu distanța de la vectorii suport la hiperplanul de decizie.

Exemplu ilustrat cu 5 vectori suport

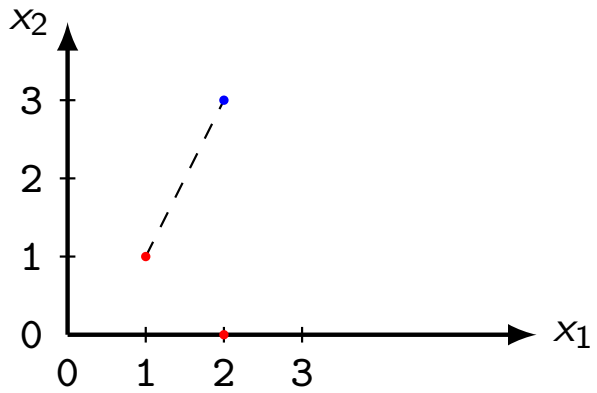


Să se determine $\vec{w} \in \mathbb{R}^M$ și $b \in \mathbb{R}$ a.î.

- $|\vec{w}|^2/2 = \frac{1}{2}\sum_{i=1}^{M} w_i^2$ are valoare minimă, și
- $y_i \cdot (\vec{w} \cdot \vec{d_i} + b) \geq 1$ pentru toți $\langle \vec{d_i}, y_i \rangle \in \mathbb{D}$

Metode de rezolvare a problemei de învățare pentru SVM:

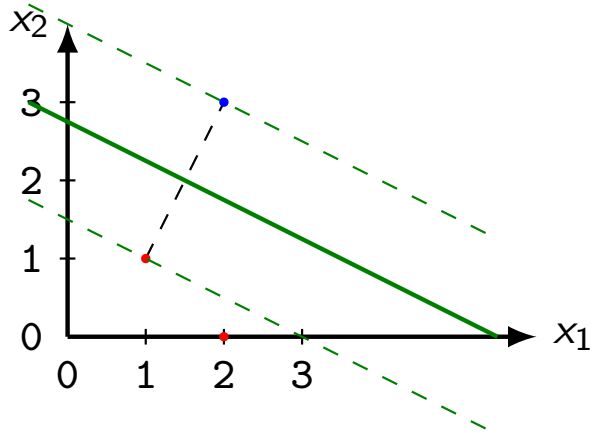
1. folosind biblioteci/algoritmi standard de rezolvare a problemelor de optimizae pătratică
2. folosind algoritmi optimizați (mai rapizi și mai scalabili) pt. problema de optimizare pătratică specifică mașinilor vector suport.



$\mathbb{D} = \{\langle(1,1),-1\rangle, \langle(2,0),-1\rangle, \langle(2,3),1\rangle\}$:

1. În acest exemplu, **hiperplanul optim de decizie** $\langle\vec{w}, b\rangle$ este paralel cu segmentul cel mai scurt dintre vectori din clase diferite, care este $(1,1)$–$(2,3)$:
   - $(1,1)$ și $(2,3)$ devin vectori suport.
   - $\vec{w} = \langle 1, 2\rangle$ fiindcă $\vec{w}$ este paralel cu vectorul de la (1,1) la (2,3),
   - $b = 5.5$ fiindcă hiperplanul trece prin mijlocul segmentului $(1,1)$–$(2,3)$, adică prin (1.5,2).

   $\Rightarrow$ hiperplanul de decizie $a \cdot (x_1 + 2\,x_2 - 5.5) = 0$ cu $a > 0$.
2. **Marginile funcționale** ale vectorilor suport $(1,1)$ și $(2,3)$ trebuie să fie 1, deci $-a \cdot (1 + 2 - 5.5) = 1$ și $a \cdot (2 + 6 - 5.5) = 1 \Rightarrow a = \frac{2}{5}$, deci vom considera $\vec{w} = (2/5, 4/5)$ și $b = 11/5$.

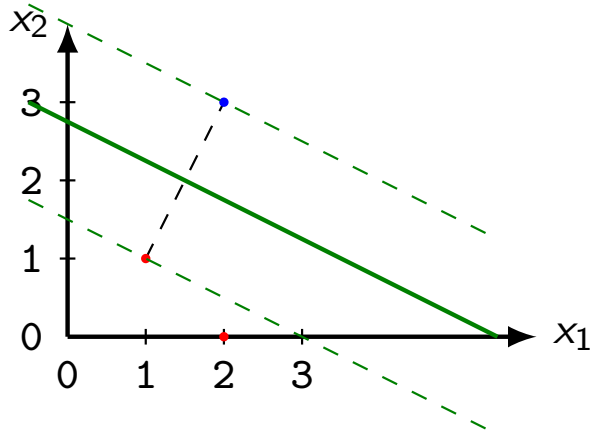


$\mathbb{D} = \{\langle(1,1),-1\rangle, \langle(2,0),-1\rangle, \langle(2,3),1\rangle\}$:

1. În acest exemplu, **hiperplanul optim de decizie** $\langle \vec{w}, b\rangle$ este paralel cu segmentul cel mai scurt dintre vectori din clase diferite, care este $(1,1)$–$(2,3)$:
   - $(1,1)$ și $(2,3)$ devin vectori suport.
   - $\vec{w} = \langle 1, 2\rangle$ fiindcă $\vec{w}$ este paralel cu vectorul de la (1,1) la (2,3),
   - $b = 5.5$ fiindcă hiperplanul trece prin mijlocul segmentului $(1,1)$–$(2,3)$, adică prin (1.5,2).

   $\Rightarrow$ hiperplanul de decizie $a \cdot (x_1 + 2\,x_2 - 5.5) = 0$ cu $a > 0$.
2. **Marginile funcționale** ale vectorilor suport $(1,1)$ și $(2,3)$ trebuie să fie 1, deci $-a \cdot (1 + 2 - 5.5) = 1$ și $a \cdot (2 + 6 - 5.5) = 1 \Rightarrow a = \dfrac{2}{5}$, deci vom considera $\vec{w} = (2/5, 4/5)$ și $b = 11/5$.

Mașini vector suport
Rezolvare geometrică a unui exemplu concret



$\mathbb{D} = \{\langle(1,1),-1\rangle, \langle(2,0),-1\rangle, \langle(2,3),1\rangle\}$:

1. În acest exemplu, **hiperplanul optim de decizie** $\langle \vec{w}, b\rangle$ este paralel cu segmentul cel mai scurt dintre vectori din clase diferite, care este $(1,1)$–$(2,3)$:
   - $(1,1)$ și $(2,3)$ devin vectori suport.
   - $\vec{w} = \langle 1, 2\rangle$ fiindcă $\vec{w}$ este paralel cu vectorul de la (1,1) la (2,3),
   - $b = 5.5$ fiindcă hiperplanul trece prin mijlocul segmentului $(1,1)$–$(2,3)$, adică prin (1.5,2).

   $\Rightarrow$ hiperplanul de decizie $a \cdot (x_1 + 2\,x_2 - 5.5) = 0$ cu $a > 0$.
2. **Marginile funcționale** ale vectorilor suport $(1,1)$ și $(2,3)$ trebuie să fie 1, deci $-a \cdot (1 + 2 - 5.5) = 1$ și $a \cdot (2 + 6 - 5.5) = 1 \Rightarrow a = \frac{2}{5}$, deci vom considera $\vec{w} = (2/5, 4/5)$ și $b = 11/5$.
3. **Marginea geometrică** a mașinii vector suport este $\rho = 2/|\vec{w}| = 2/\sqrt{20/25} = \sqrt{5}$.

Permite definirea unei mașini vector suport pt. colecții de date care nu sunt separabile liniar.

- Dacă mulțimea de antrenare $\mathbb{D}$ nu este liniar separabilă, SVM permite clasificarea eronată a unui nr. mic de documente, numite *zgomot* (engl. noise)
  - Fiecare clasificare greșită a unui exemplu $\vec{d}_i$ este penalizată cu un cost proporțional cu val. unei variabile ζ_i

Definiție formală a problemei de învățare SVM cu margini soft:
Să se determine $\vec{w}$, b și $\zeta_i \geq 0$ astfel încât:

- $\frac{1}{2}|\vec{w}|^2 + C\sum_i \zeta_i$ este minimizat și
- $y_i \cdot (\vec{w} \cdot \vec{d}_i + b) \geq 1 - \zeta_i$ pentru toți $(\vec{d}_i, y_i) \in \mathbb{D}$.

Parametrul C este un termen de regularizare