Inteligență Artificială

Bogdan Alexe
bogdan.alexe@fmi.unibuc.ro

Secția Tehnologia Informației, anul III, 2022-2023
Cursul 5

Bonusul de la laborator

- puteți acumula maxim 1 punct bonus de la laborator;
- la examenul final, se adaugă acest bonus la notă;
- **pentru prima parte a semestrului aveți alocat un bonus de 0,5 puncte**
  - **Alexandra și Sergiu stabilesc care sunt aceste condiții (posibil să puncteze prezența la 2 ore de laborator cu 0,05 puncte);**
- pentru a doua parte a semestrului aveți alocat un bonus de 0,5 puncte
  - Irina stabilește care sunt aceste condiții;

Eroare în mulțimea de testare

test_data

1	**text**
2	Hr. formand, selv om vi i høj grad sympatiserer med de borgere, der befinder sig i denne uheldige situation, mener vi i UKIP, at problemet med fratagelse af ejendom i Valencia burde have været behandlet ve Byplanlægning er et område, der skal forblive på lokalt plan, for at lokalområdernes behov skal møde forståelse. En centraliseret politik gør ikke noget bedre, den forværrer tværtimod problemet. Det har vi set For det første er der den fælles fiskeripolitik med det særdeles skadelige kvotesystem. Hyldet som et miljøprojekt har den gjort næsten uoprettelig skade på fiskebestandene. En meget stor del af fiskeforarbej De problemer, som de ikke-spanske statsborgere, herunder mange britiske statsborgere, i Valencia står over for, bør løses mellem regeringerne. Jeg er ked af, at det ikke er lykkedes den britiske regering at op
3	Quiero dejar constancia de mi apoyo a este informe, como un gran paso para facilitar la aplicación del principio de «quien contamina paga». Cuando se producen daños medioambientales, la primera pregu El requisito de que la Comisión presente un informe, en el plazo de seis años, también nos dará la oportunidad de revisar esta importante cuestión y de evaluar convenientemente su impacto en la práctica. En un mundo ideal, yo esperaría que los ciudadanos estuviesen protegidos frente a cualquier riesgo de daño medioambiental, y afortunadamente estamos haciendo progresos con algunas medidas preventiv
7583	Mijnheer de fungerend voorzitter van de Raad, morgen zal dit Parlement stemmen over een amendement op uw nationaal waterplan, dus wij hopen dat u vandaag van uw glas water geniet. Als de Europese Unie het proces van Lissabon een schoolrapport zou geven, dan zou daarin vermoedelijk staan: "moet beter zijn best doen" . Wij hebben het afgelopen jaar enige vooruitgang gezien met betr Ik vrees echter dat uit dit schoolrapport eveneens zou blijken dat de Unie op teveel gebieden een onvoldoende haalt. In de Commissiemededeling aan de Europese Raad van Barcelona wordt verklaard dat me Het feit dat de richtlijn inzake het overnamebod verworpen is betekent een grote tegenslag voor het concurrentievermogen binnen de Unie en ik roep de andere politieke fracties en de Raad op samen met ons Een ander punt waarover de schoolmeester de Raad op de vingers zou tikken is het Europees octrooi. Als wij de kloof tussen Europa en de Verenigde Staten op het gebied van vernieuwing willen dichten, dan Tot slot zou ons rapport laten zien dat deze leerling te weinig ambitie heeft. Als wij willen dat de Europese Unie uitblinkt in de klas, moeten wij voortmaken met de liberalisering van de energie- en gasmarkten, Wanneer de Top van Barcelona wordt geëvalueerd, zullen wij u beoordelen naar uw schoolrapport. Zoals de zaken er nu voorstaan, moet de leerling beter zijn best doen.
7584	
7585	Señor Presidente, en primer lugar, acaso debemos precipitarnos a estrechar los lazos con Egipto, con la esperanza de que las cuestiones sin resolver y un fallo injusto se disiparán y se resolverán por sí solos ¿Aprueba el Parlamento las violaciones del derecho a un juicio justo, la libertad de expresión y la libertad de asociación, ya sea ésta social, política, religiosa o sexual? Porque eso es lo que estaremos diciend
7586	Hr. formand, jeg er blevet anmodet om at stille et mundtligt ændringsforslag. Som en mellemvej og for at få dette ændringsforslag vedtaget foreslår jeg, at vi i slutningen af punkt 37 tilføjer "og bidrage til bekæ (Formanden konstaterede, at ingen gjorde indsigelse, hvorfor det mundtlige ændringsforslag kunne tages i betragtning) (Forslaget til beslutning vedtoges) Stemmeforklaringer Budget 2002

Eroare în mulțimea de testare

7583	Mijnheer de fungerend voorzitter van de Raad, morgen zal dit Parlement stemmen over een amendement op uw nationaal waterplan, dus wij hopen dat u vandaag van uw glas water geniet. Als de Europese Unie het proces van Lissabon een schoolrapport zou geven, dan zou daarin vermoedelijk staan: "moet beter zijn best doen" . Wij hebben het afgelopen jaar enige vooruitgang gezien met betr Ik vrees echter dat uit dit schoolrapport eveneens zou blijken dat de Unie op teveel gebieden een onvoldoende haalt. In de Commissiemededeling aan de Europese Raad van Barcelona wordt verklaard dat me Het feit dat de richtlijn inzake het overnamebod verworpen is betekent een grote tegenslag voor het concurrentievermogen binnen de Unie en ik roep de andere politieke fracties en de Raad op samen met ons Een ander punt waarover de schoolmeester de Raad op de vingers zou tikken is het Europees octrooi. Als wij de kloof tussen Europa en de Verenigde Staten op het gebied van vernieuwing willen dichten, dan Tot slot zou ons rapport laten zien dat deze leerling te weinig ambitie heeft. Als wij willen dat de Europese Unie uitblinkt in de klas, moeten wij voortmaken met de liberalisering van de energie- en gasmarkten, Wanneer de Top van Barcelona wordt geëvalueerd, zullen wij u beoordelen naar uw schoolrapport. Zoals de zaken er nu voorstaan, moet de leerling beter zijn best doen.
7584	
7585	Señor Presidente, en primer lugar, acaso debemos precipitarnos a estrechar los lazos con Egipto, con la esperanza de que las cuestiones sin resolver y un fallo injusto se disiparán y se resolverán por sí solos ¿Aprueba el Parlamento las violaciones del derecho a un juicio justo, la libertad de expresión y la libertad de asociación, ya sea ésta social, política, religiosa o sexual? Porque eso es lo que estaremos diciend
7586	Hr. formand, jeg er blevet anmodet om at stille et mundtligt ændringsforslag. Som en mellemvej og for at få dette ændringsforslag vedtaget foreslår jeg, at vi i slutningen af punkt 37 tilføjer "og bidrage til bekæ (Formanden konstaterede, at ingen gjorde indsigelse, hvorfor det mundtlige ændringsforslag kunne tages i betragtning) (Forslaget til beslutning vedtoges) Stemmeforklaringer Budget 2002

Fișierul soluție:

7584	**7583**	Ireland

Recapitulare – cursul trecut

1. Clasificatorul naïve Bayes

2. Evaluarea performanței unui model

3. Strategii de împărțire a datelor

4. Proiect
   – concurs pe platforma Kaggle
   – demo
   – laborator săptămâna 5

Cuprinsul cursului de azi

1. Mașini cu vectori support (SVMs– support vector machines)

2. Perceptronul

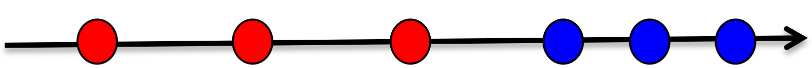
Mașini cu vectori suport
(SVMs – Support Vector Machines)

Separabilitate liniară

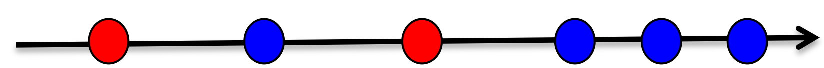
Două mulțimi de puncte într-un spațiu de dimensiune ***n*** sunt liniar separabile dacă există un hiperplan (subspațiu de dimensiune ***n -1***) care le separă perfect.

Exemple: n = 1

mulțimi liniar separabile



mulțimi liniar neseparabile

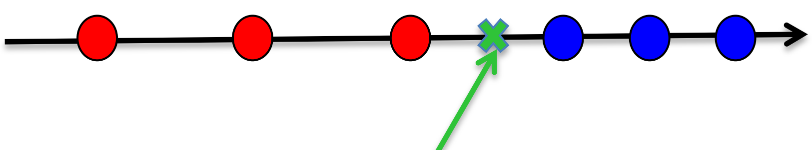


Separabilitate liniară

Două mulțimi de puncte într-un spațiu de dimensiune ***n*** sunt liniar separabile dacă există un hiperplan (subspațiu de dimensiune ***n -1***) care le separă perfect.

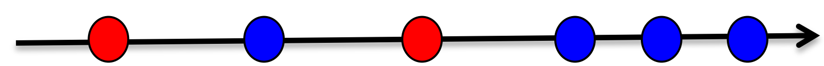
Exemple: n = 1

mulțimi liniar separabile



Hiperplan de dimensiune 0 = un punct.
Există o infinitate de hiperplane (puncte) care separă mulțimile perfect.

mulțimi liniar neseparabile



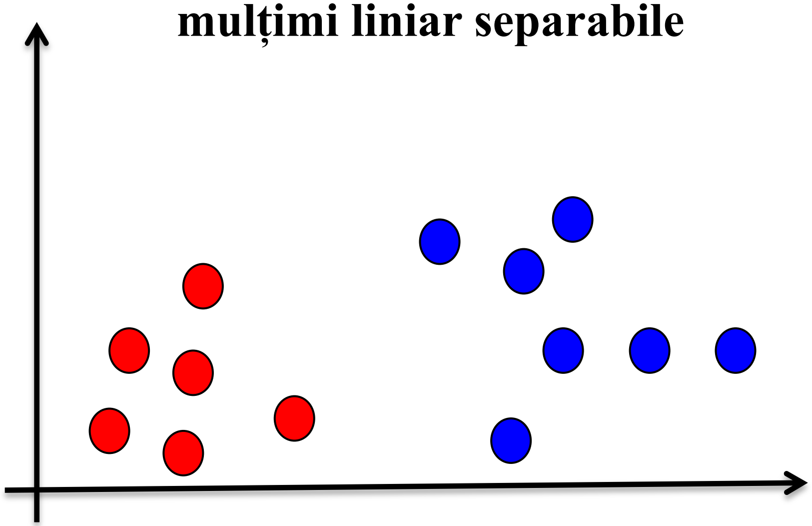
Nu există nici un hiperplan (punct) care separă mulțimile perfect.

Separabilitate liniară

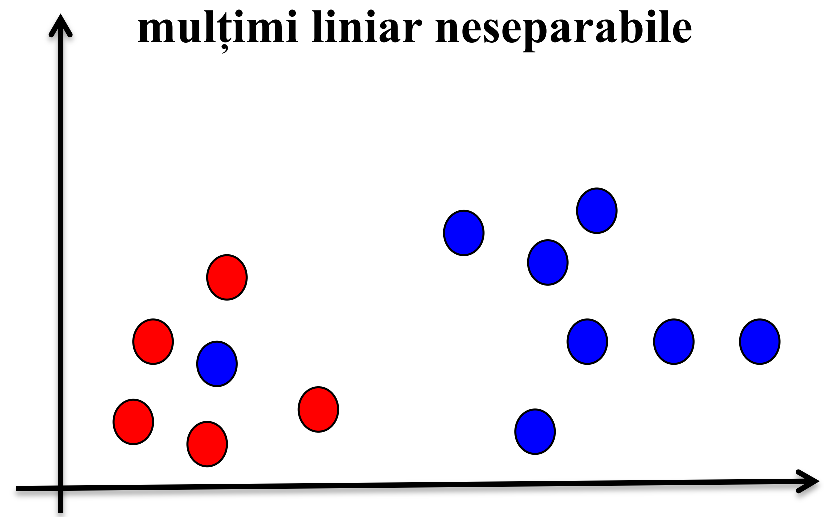
Două mulțimi de puncte într-un spațiu de dimensiune ***n*** sunt liniar separabile dacă există un hiperplan (subspațiu de dimensiune ***n -1***) care le separă perfect.

Exemple: n = 2

mulțimi liniar separabile



mulțimi liniar neseparabile

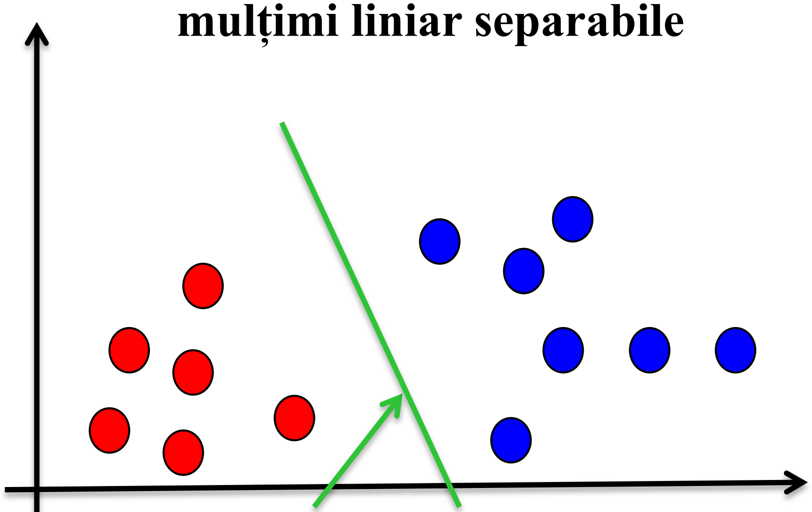


Separabilitate liniară

Două mulțimi de puncte într-un spațiu de dimensiune ***n*** sunt liniar separabile dacă există un hiperplan (subspațiu de dimensiune ***n -1***) care le separă perfect.

Exemple: n = 2

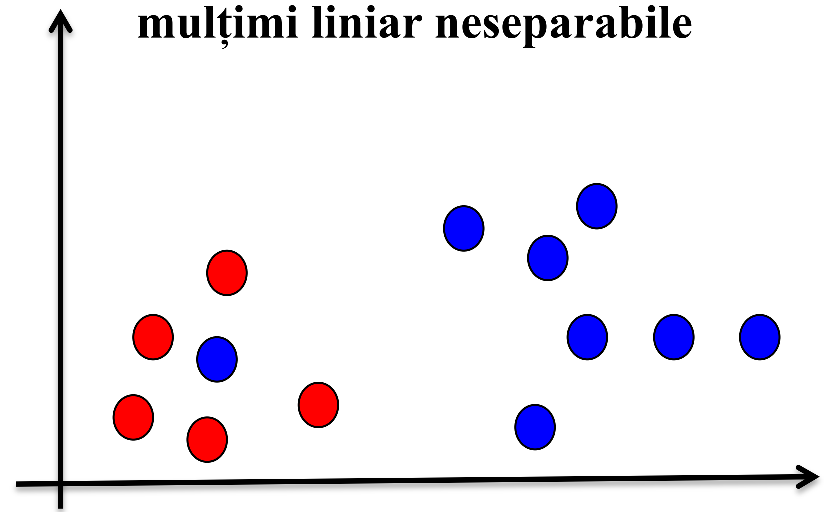
mulțimi liniar separabile



Hiperplan de dimensiune 1 = o dreaptă

Există o infinitate de hiperplane (drepte) care separă mulțimile perfect.

mulțimi liniar neseparabile



Nu există niciun hiperplan (dreaptă) care separă mulțimile perfect.

Separabilitate liniară

Două mulțimi de puncte într-un spațiu de dimensiune ***n*** sunt liniar separabile dacă există un hiperplan (subspațiu de dimensiune ***n -1***) care le separă perfect.

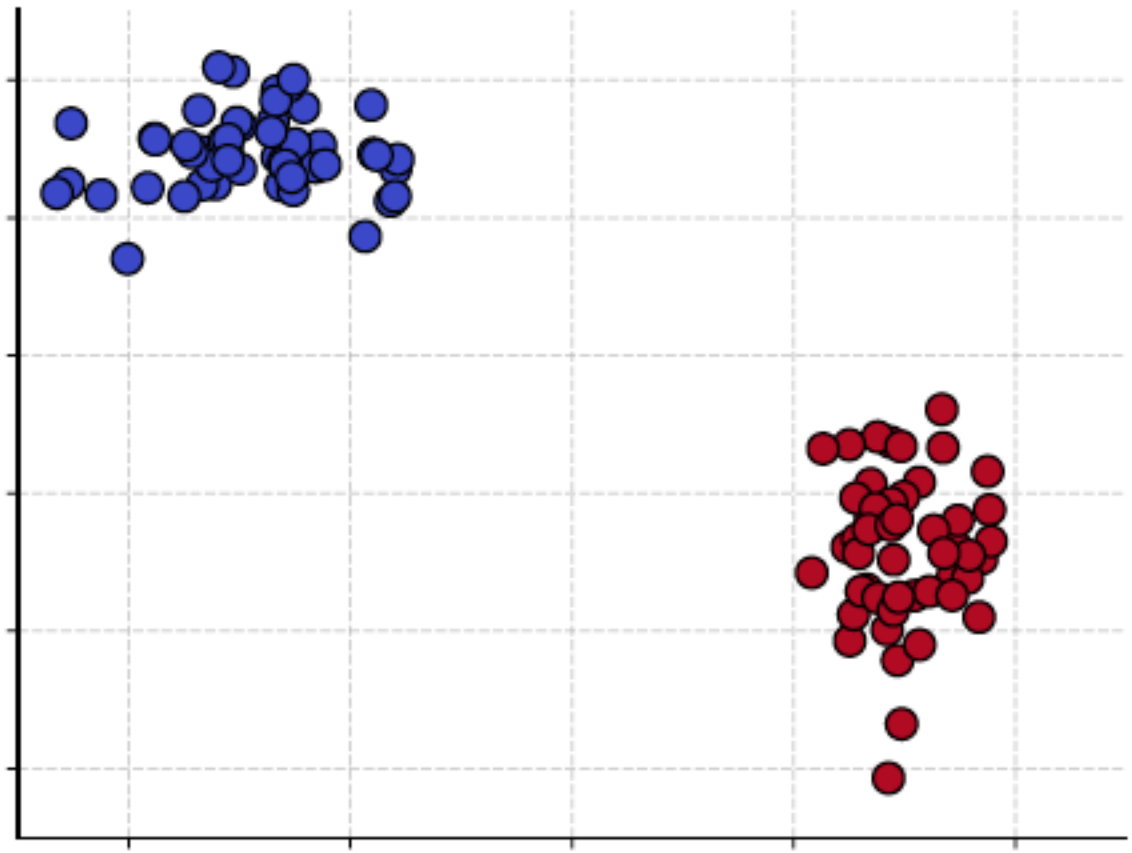
Un hiperplan în 1D este un **punct**.

Un hiperplan în 2D este o **dreaptă**.

Un hiperplan în 3D este un **plan**.

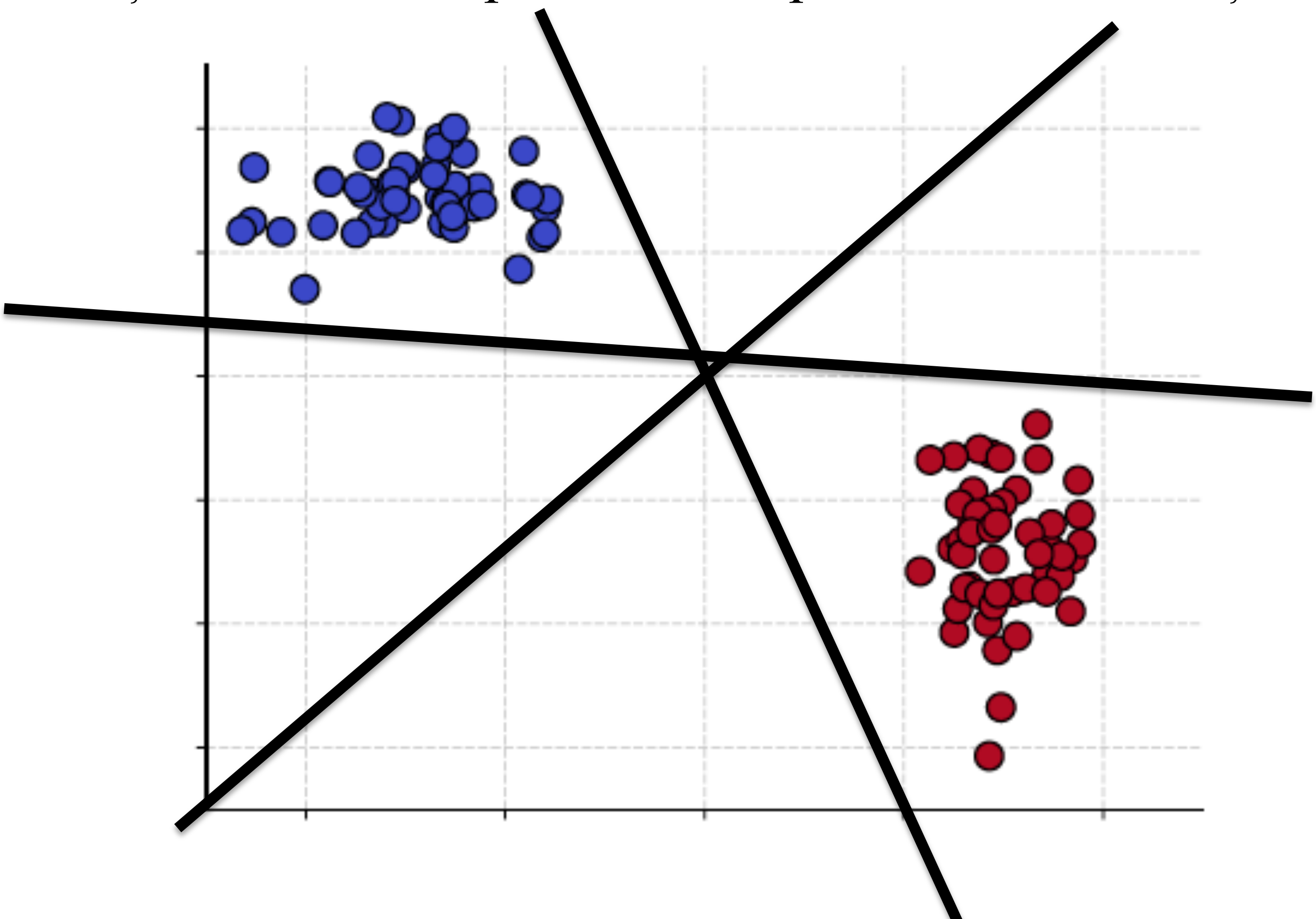
Frontiera de separare (decision boundary)

- cum ați desena o dreaptă care să separe cele două mulțimi de puncte?



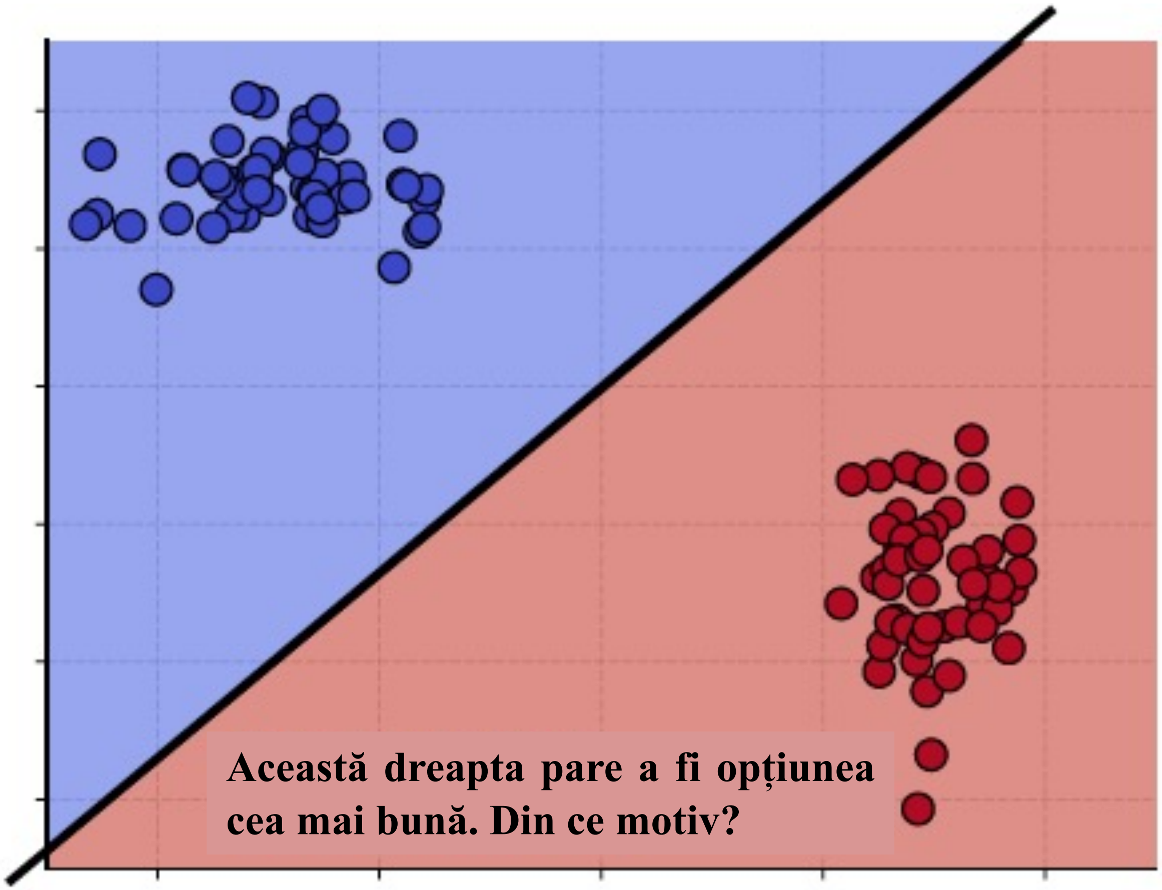
Frontiera de separare (decision boundary)

- cum ați desena o dreaptă care să separe cele două mulțimi de puncte?



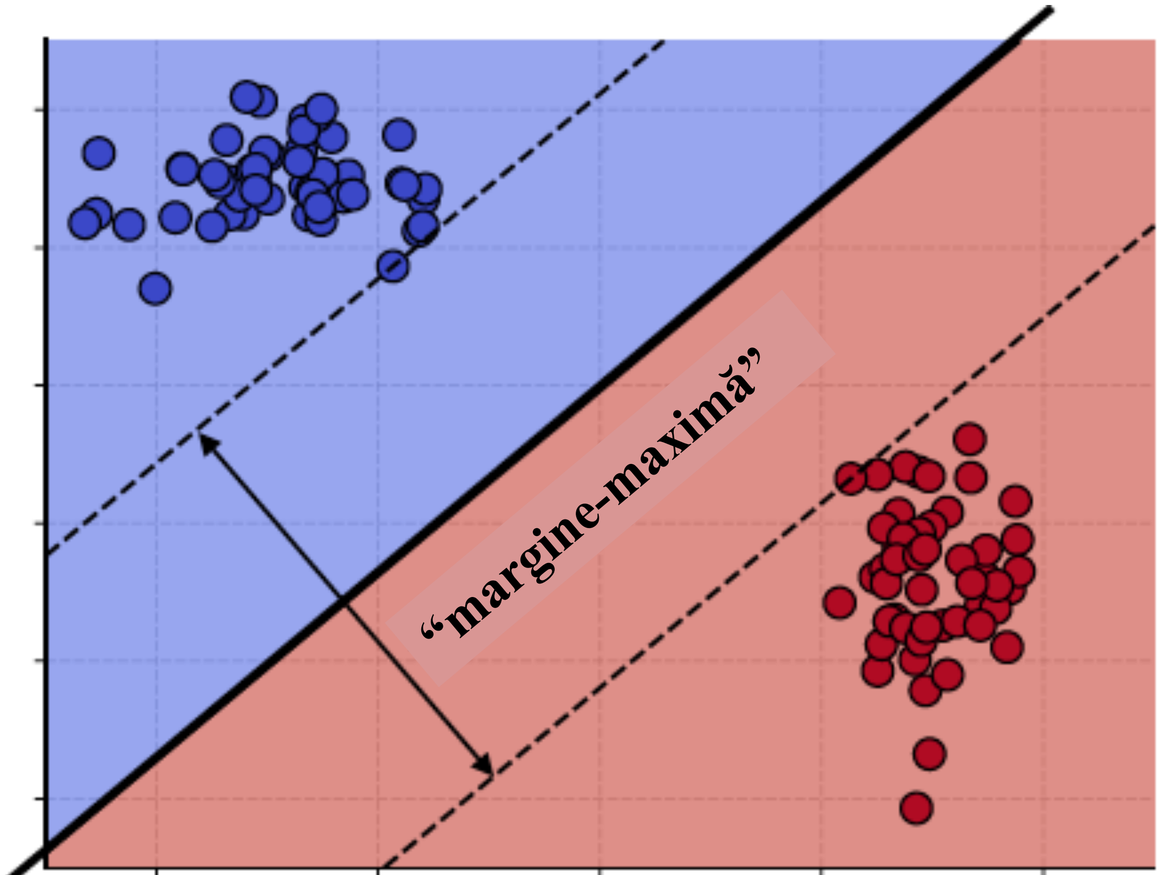
Frontiera de separare (decision boundary)

- cum ați desena o dreaptă care să separe cele două mulțimi de puncte?



Frontiera de separare (decision boundary)

- cum ați desena o dreaptă care să separe cele două mulțimi de puncte?



Definiția unui SVM

- Un **SVM** (support vector machine = mașină cu vectori suport) este un clasificator liniar binar nonprobabilist.
  - **clasificator**: metodă de învățare supervizată care are scop predicția de clase
  - **liniar**: frontiera de decizie este un hiperplan în n-dimensiuni
  - **binar**: învață să discrimineze între 2 clase (clasa + și clasa -)
  - **nonprobabilist**: rezultatul unui SVM nu este limitat, nu poate fi interpretat ca o probabilitate

Există extensii care fac ca SVM să poate fi folosit în probleme de regresie, să aibă o frontieră neliniară, cu mai multe clase și să fie probabilist

- Un SVM încearcă să găsească hiperplanul de separare care este cel mai departe posibil (are margine maximă) de punctele de antrenare
  - un punct este clasificat + sau -, în funcție de ce parte a hiperplanului se află

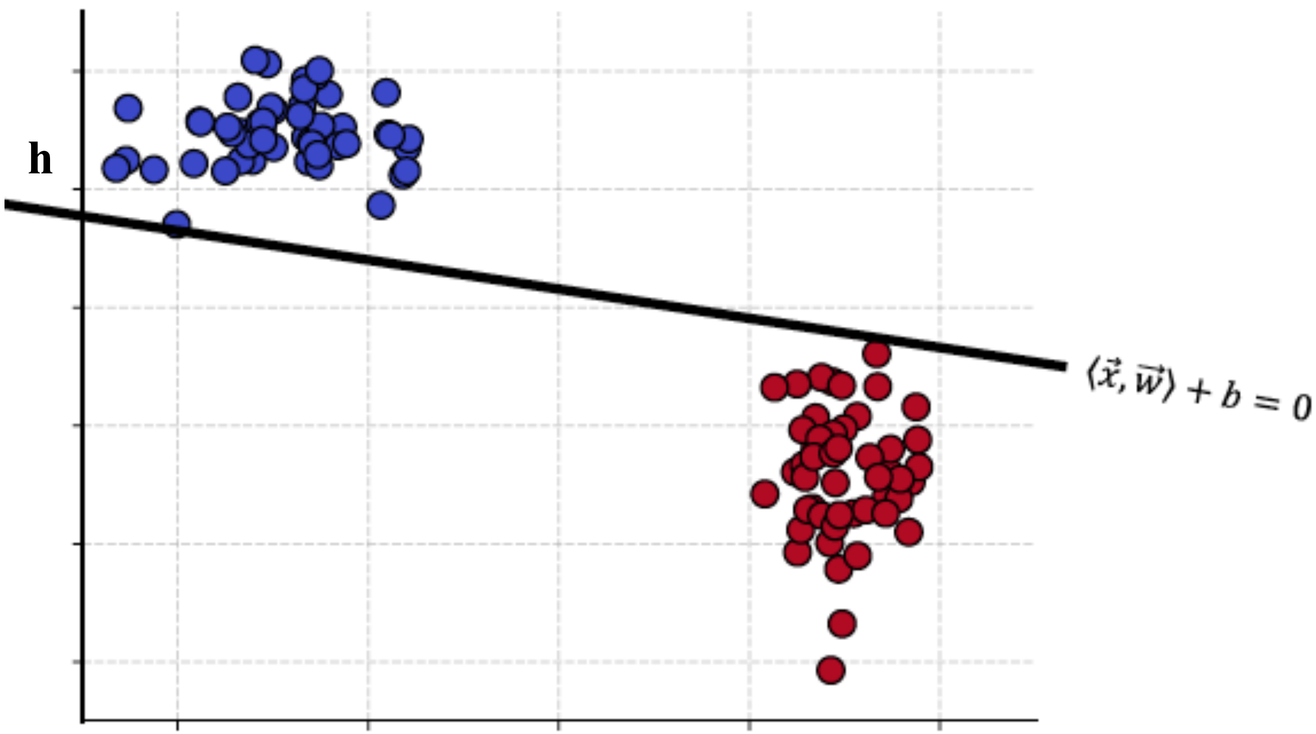
Învățarea unui hiperplan de separare “bun”

x – caracteristici (features)
y – etichete (labels)

$$\langle \overrightarrow{x_+}, \vec{w} \rangle + b \geq 0 \qquad \forall \overrightarrow{x_+} \in \{\vec{x}^{(i)} | y^{(i)} = +1\}$$
$$\langle \overrightarrow{x_-}, \vec{w} \rangle + b < 0 \qquad \forall \overrightarrow{x_-} \in \{\vec{x}^{(i)} | y^{(i)} = -1\}$$

Hiperplanul h satisface condițiile impuse dar nu are nicio margine.

Un hiperplan h este cu atât mai bun cu cât are o margine mai mare.



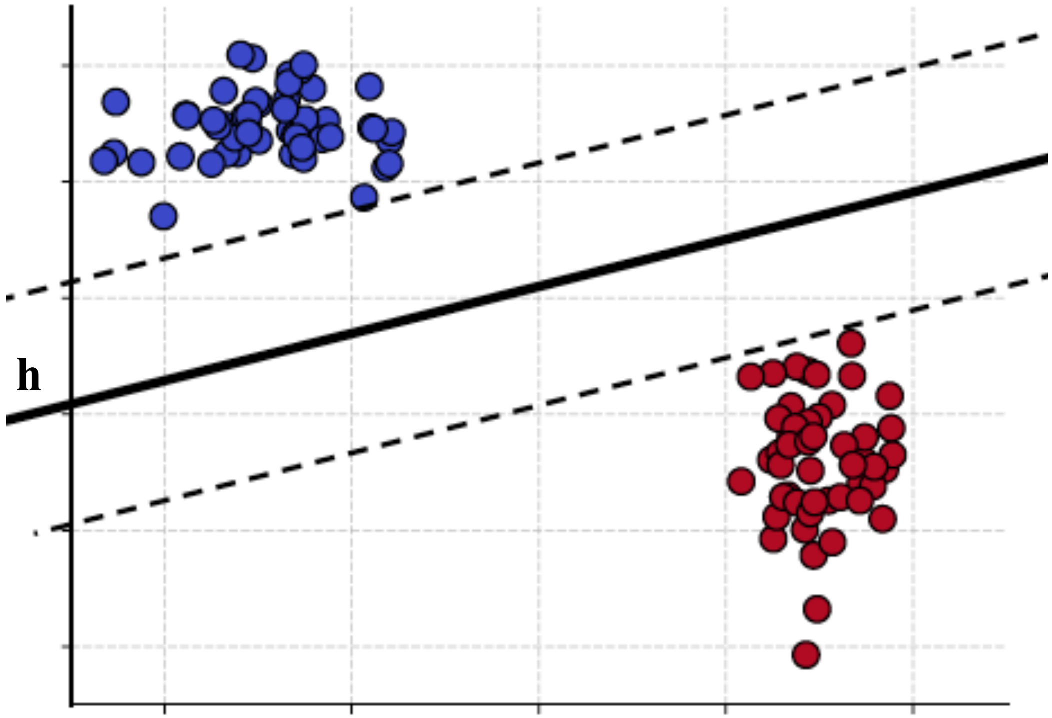
Învățarea unui hiperplan de separare “bun”

x – caracteristici (features)
y – etichete (labels)

$$\langle \overrightarrow{x_+}, \vec{w} \rangle + b \geq 0 \qquad \forall \overrightarrow{x_+} \in \{\vec{x}^{(i)} | y^{(i)} = +1\}$$
$$\langle \overrightarrow{x_-}, \vec{w} \rangle + b < 0 \qquad \forall \overrightarrow{x_-} \in \{\vec{x}^{(i)} | y^{(i)} = -1\}$$

Hiperplanul h satisface condițiile impuse dar nu are nicio margine.

Un hiperplan h este cu atât mai bun cu cât are o margine mai mare.



Regula de decizie + problema de optimizare

- Regula de decizie este:

$\vec{x}$ este exemplu + dacă $\langle \vec{x}, \vec{w} \rangle + b \geq 0$

$\vec{x}$ este exemplu - dacă $\langle \vec{x}, \vec{w} \rangle + b < 0$

- Pentru a obține hiperplanul de margine maximă trebuie să rezolvăm problema de optimizare (nu intrăm în detalii):

$$\text{minimize} \quad \frac{||\vec{w}||^2}{2}$$

cu constrângerile $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq 0$

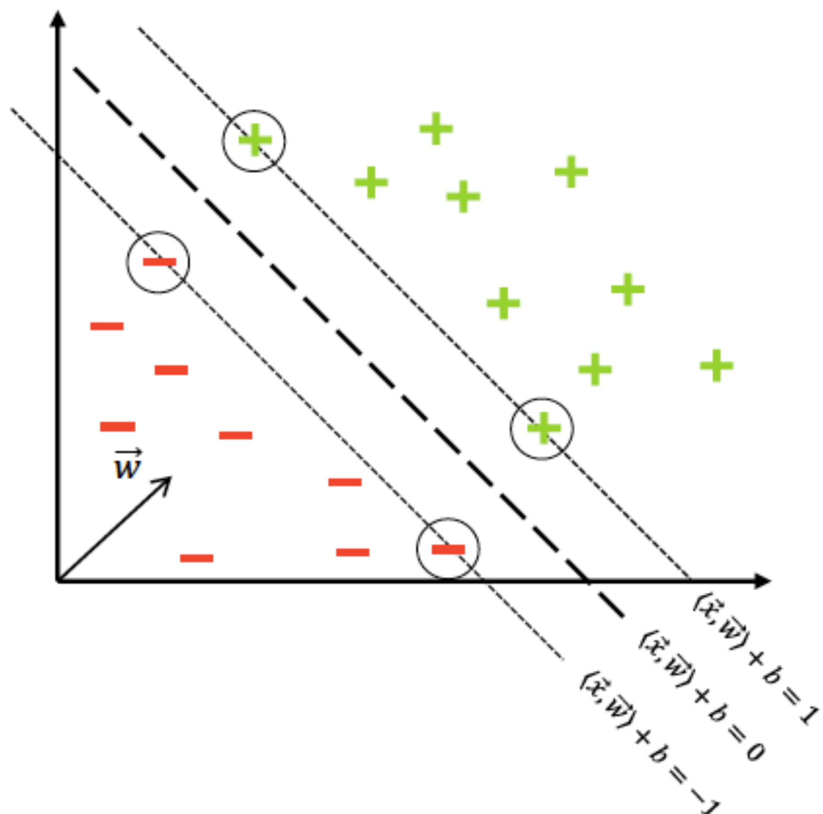
Forma primală pentru SVM

Vectori suport

- Exemplele de antrenare $(\vec{x}^{(i)}, y^{(i)})$ de la margine satisfac:

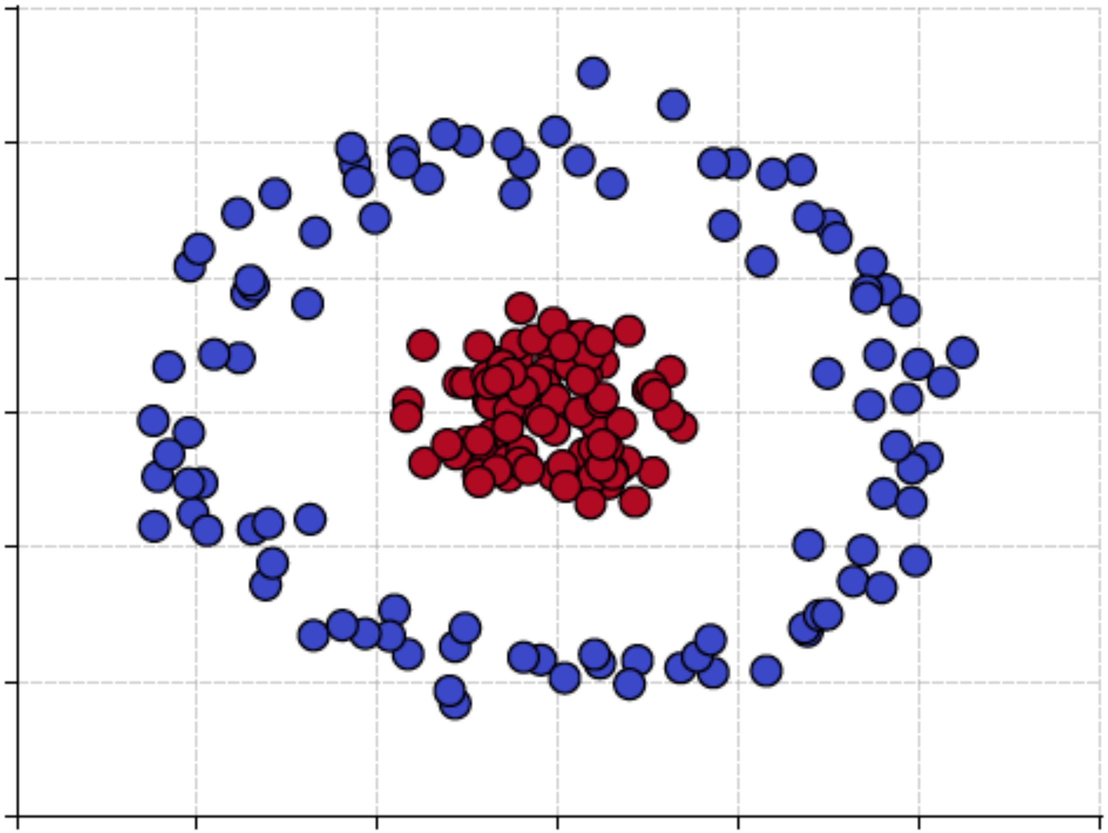
$$\boxed{y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 = 0}$$

- Aceste exemple de antrenare se numesc **vectori suport** (**Support Vectors**)



Liniar separabilitate?

- În practică, exemplele de antrenare pozitive și negative nu sunt liniar separabile



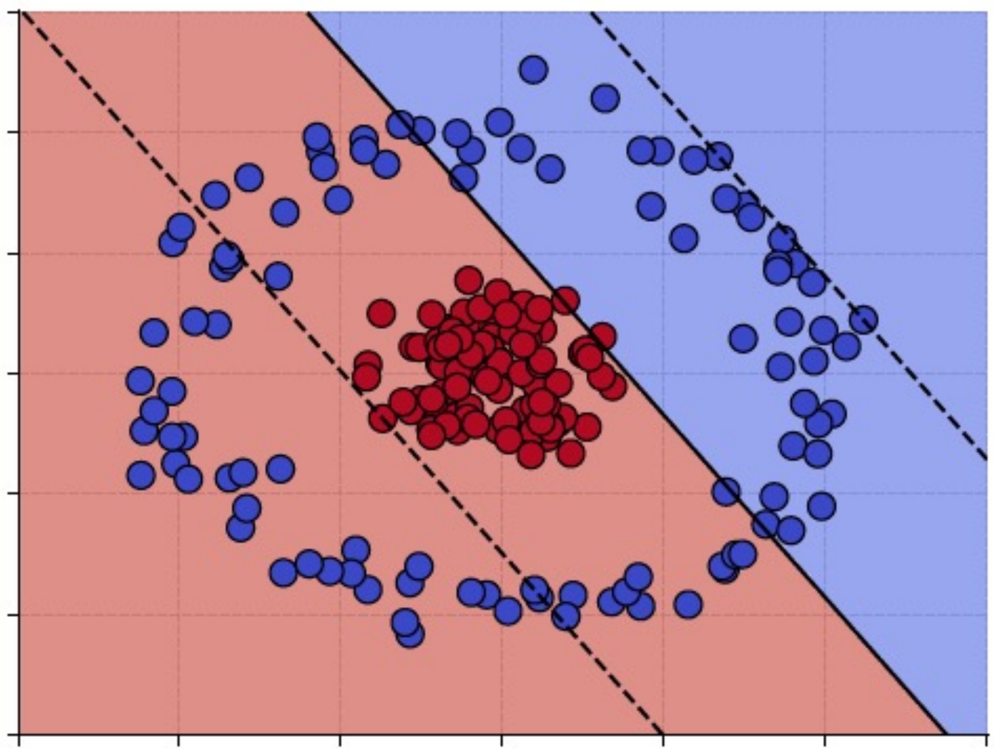
Liniar separabilitate?

- În practică, exemplele de antrenare pozitive și negative nu sunt liniar separabile
- Toate calculele precedente s-au bazat pe liniar separabilitatea exemplelor din mulțimea de antrenare
  - există un hiperplan h care separă + de – (**“hard-margin” linear SVM**)
- Cum putem trata cazul de date care nu sunt liniar separabile folosind SVM?
- Două soluții:

1. **“Kernel trick”** – maparea datelor într-un spațiu cu dimensionalitate mult mai mare în care acestea sunt separabile
2. **“Soft-margin” SVM** – permitem SVM-ului să facă câteva greșeli la antrenare (folosim o funcție cost) în schimbul obținerii unei margini bune

Date care nu sunt liniar separabile

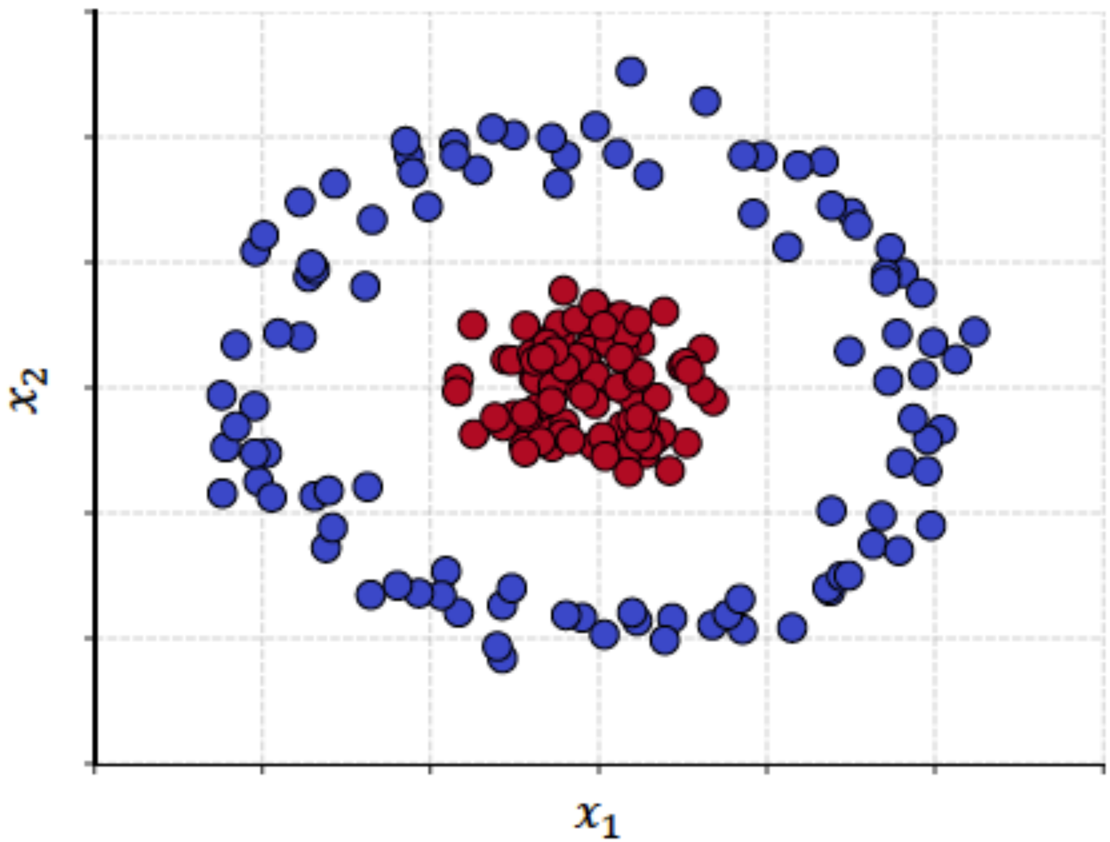
- În practică, exemplele de antrenare pozitive și negative nu sunt liniar separabile
- Nu putem găsi un hiperplan soluție folosind **"hard-margin" linear SVM**



The kernel trick

$\vec{x} \in \mathbb{R}, \vec{x} = (x_1, x_2)$

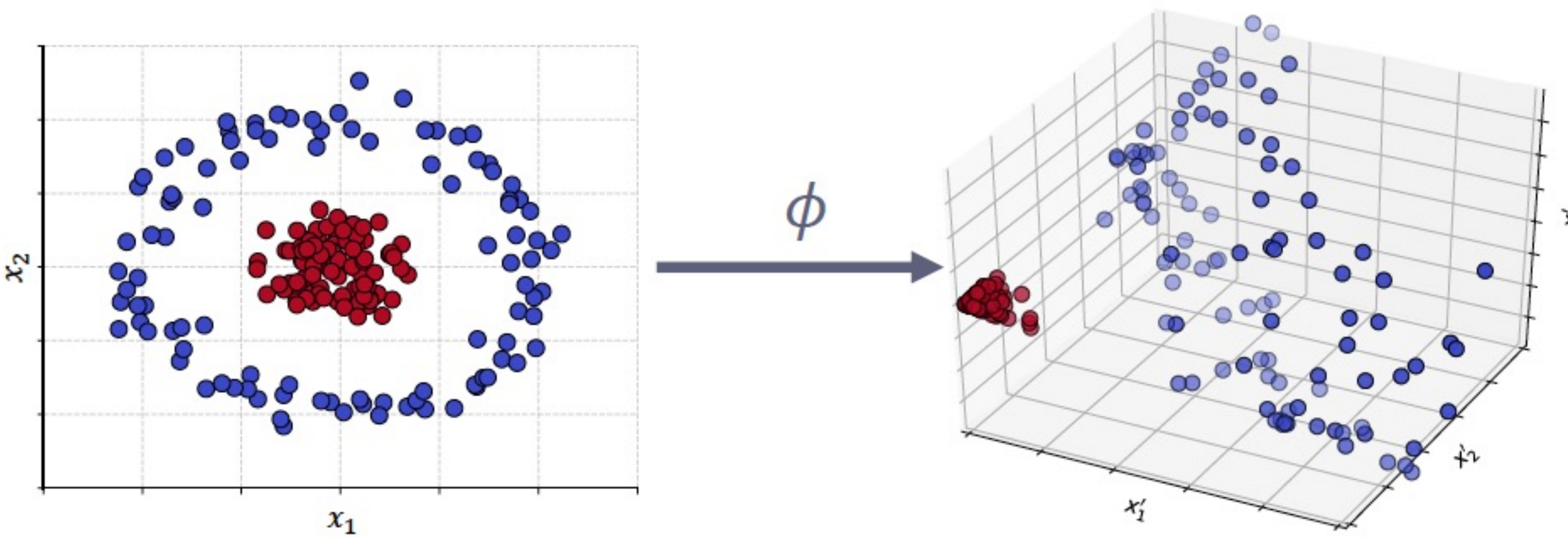
$\phi: \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \phi\big((x_1, x_2)\big) = (x_1', x_2', x_3') = \big(x_1^2, x_2^2, x_1 x_2 \sqrt{2}\big)$



The kernel trick

$\vec{x} \in \mathbb{R}, \vec{x} = (x_1, x_2)$

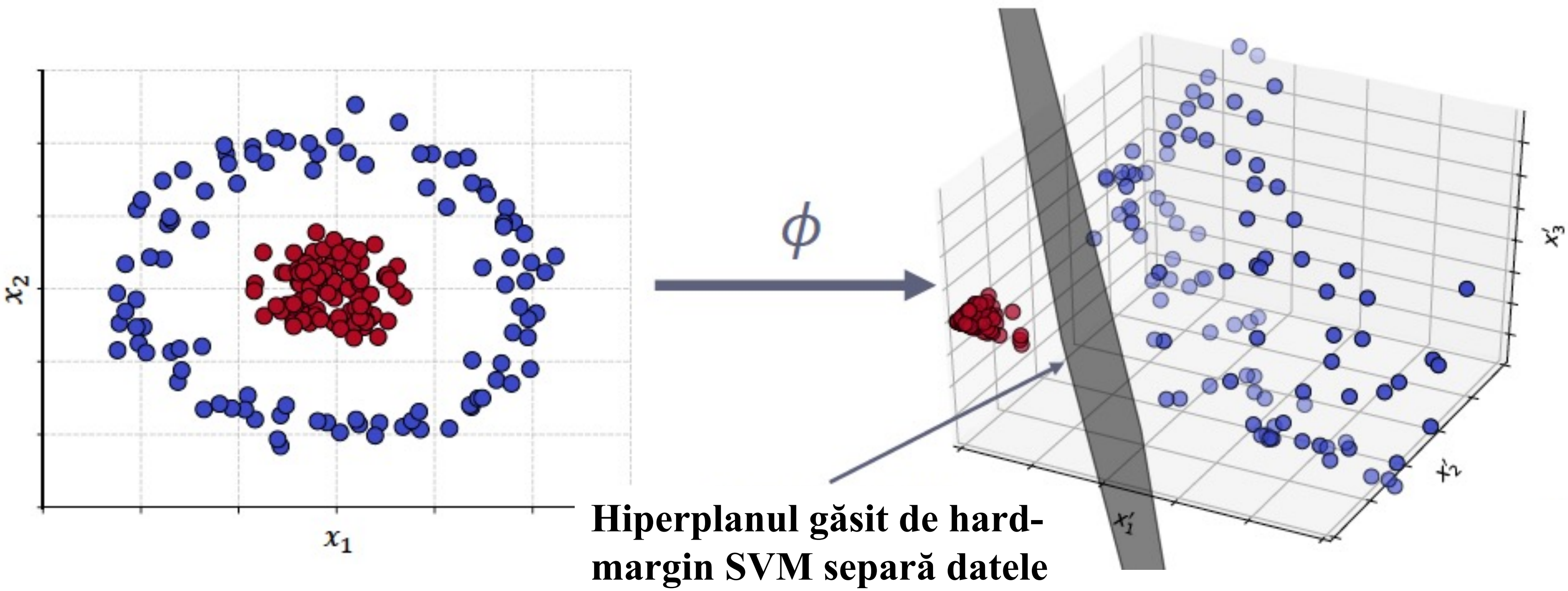
$\phi: \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \phi\big((x_1, x_2)\big) = (x_1', x_2', x_3') = \big(x_1^2, x_2^2, x_1 x_2 \sqrt{2}\big)$



The kernel trick

$$\vec{x} \in \mathbb{R}, \vec{x} = (x_1, x_2)$$

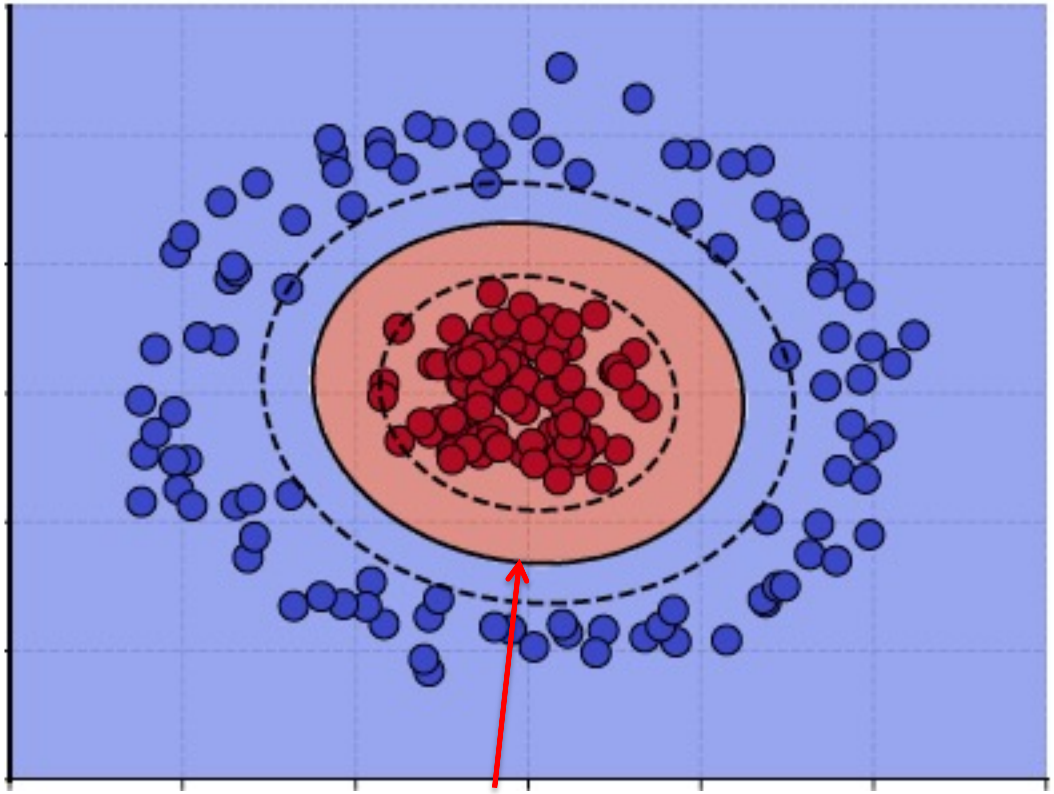
$$\phi: \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \phi\big((x_1, x_2)\big) = (x_1', x_2', x_3') = \big(x_1^2, x_2^2, x_1 x_2 \sqrt{2}\big)$$



The kernel trick

$\vec{x} \in \mathbb{R}, \vec{x} = (x_1, x_2)$

$\phi: \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \phi\big((x_1, x_2)\big) = (x'_1, x'_2, x'_3\,) = \big(x_1^2, x_2^2, x_1 x_2 \sqrt{2}\big)$



Hiperplanul găsit de hard-margin SVM se translatează în spațiul original într-o frontieră de decizie neliniară

The kernel trick

- Constă în folosirea unei funcții kernel (nucleu) ϕ care mapează datele într-un alt spațiu (de obicei de dimensionalitate mult mai mare) în care datele sunt liniar separabile

- Orice algoritm care folosește o asemenea abordare se numește metodă kernel

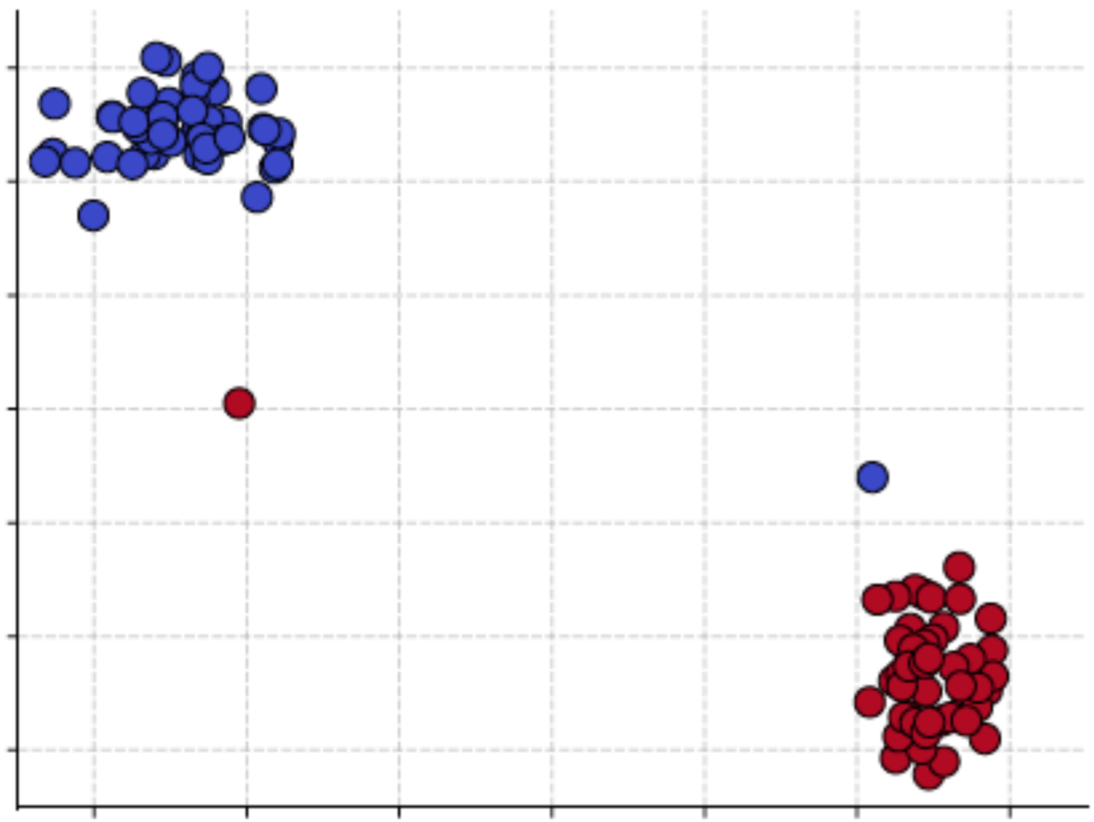
- Funcția kernel poate fi văzută ca o funcție de similarite între date.

Cele mai cunoscute funcții kernel

Linear kernel:	$K(\vec{x},\vec{y}) = \langle \vec{x},\vec{y} \rangle$
Polynomial kernel:	$K(\vec{x},\vec{y}) = (\langle \vec{x},\vec{y} \rangle + c)^d$
Radial basis function (RBF) kernel:	$K(\vec{x},\vec{y}) = e^{-\frac{\lVert \vec{x}-\vec{y} \rVert^2}{2\sigma^2}}$
Sigmoid kernel:	$K(\vec{x},\vec{y}) = \tanh(\gamma \langle \vec{x},\vec{y} \rangle + c)$

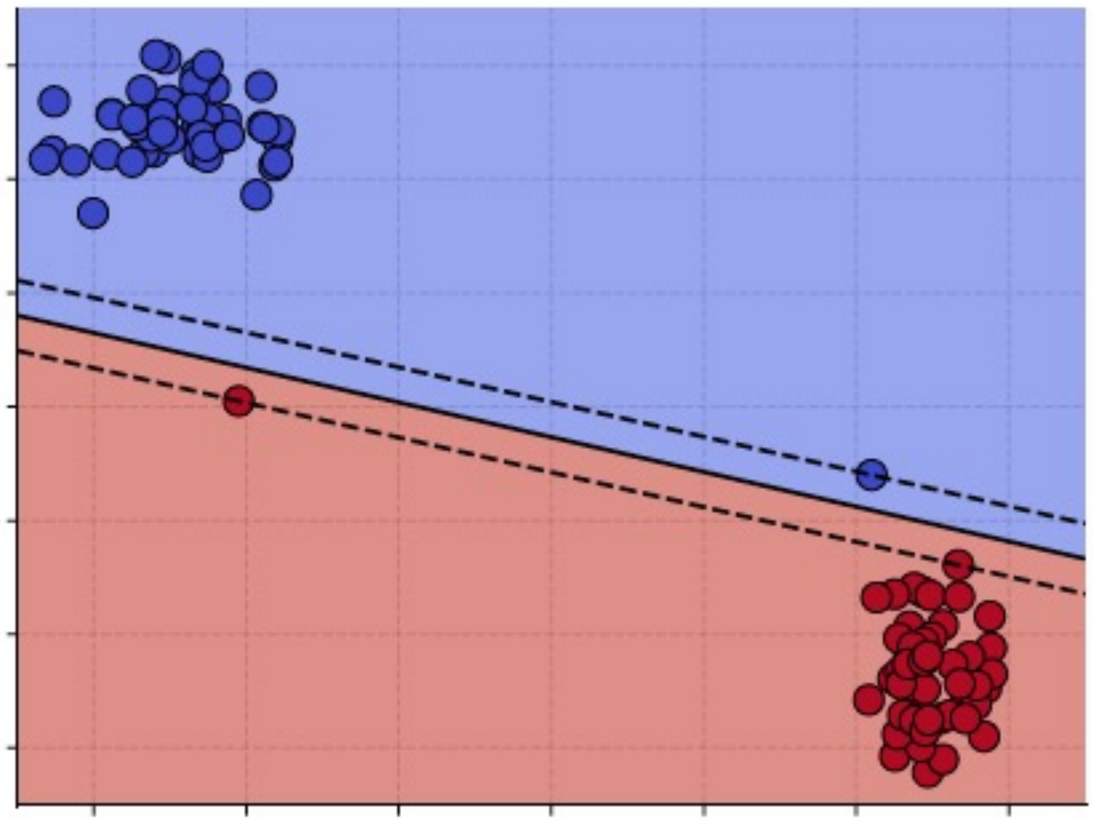
Soft-margin SVM

- Ce se întâmplă dacă antrenăm un hard-margin SVM pe setul de date de mai jos?



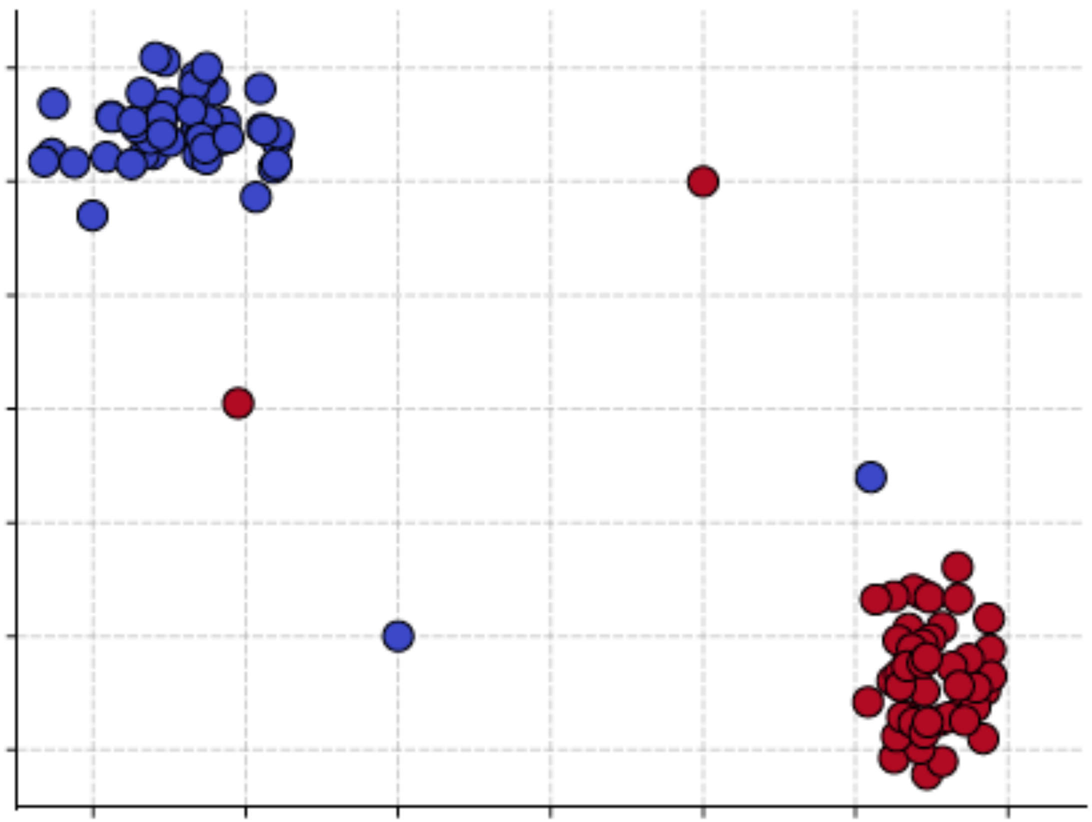
Soft-margin SVM

- Ce se întâmplă dacă antrenăm un hard-margin SVM pe setul de date de mai jos?
- SVM-ul găsește un hiperplan care separă datele perfect însă acesta are o margine foarte mică



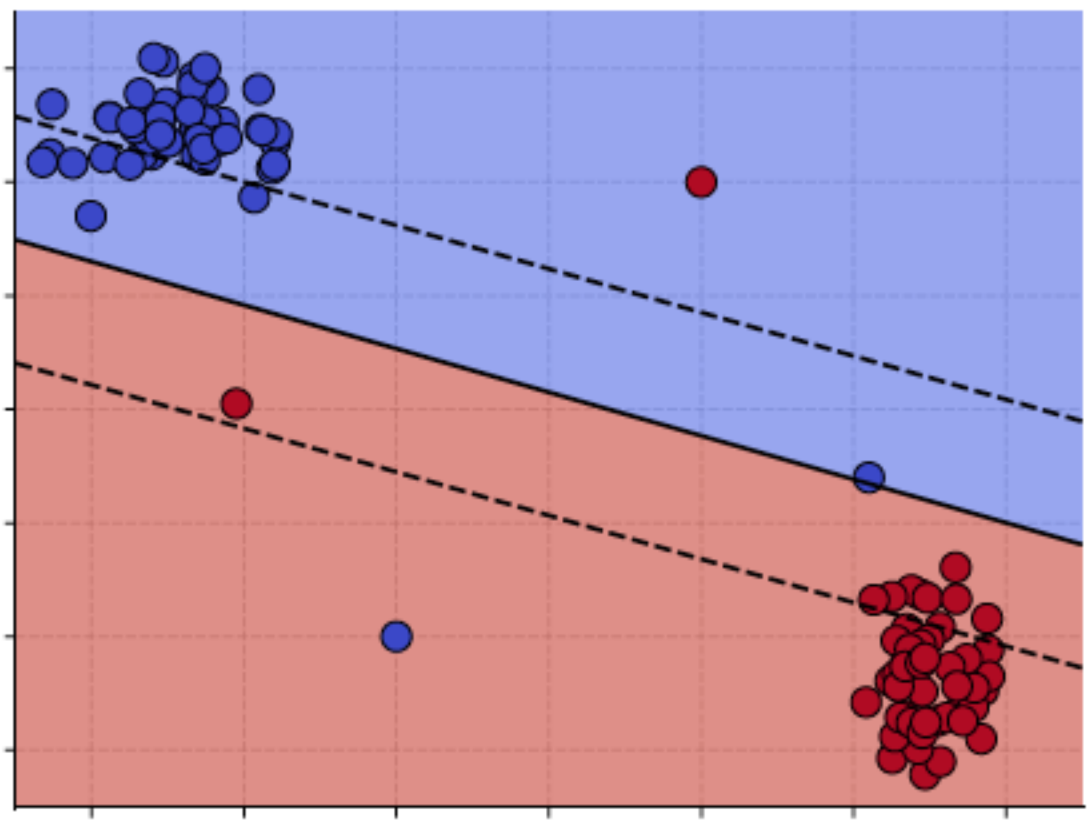
Soft-margin SVM

- Ce se întâmplă dacă antrenăm un hard-margin SVM pe setul de date de mai jos?



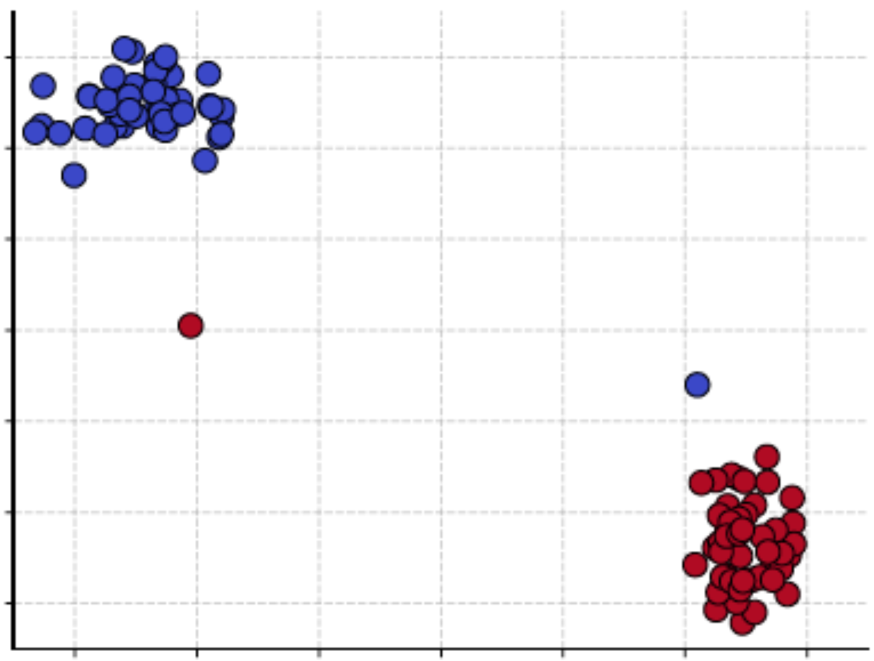
Soft-margin SVM

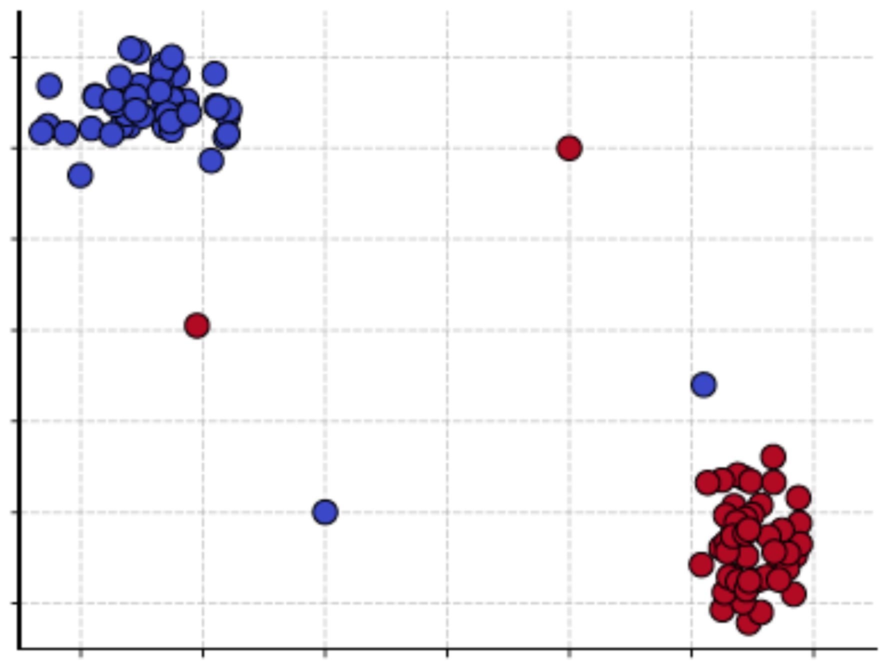
- SVM-ul nu reușește să găsească un hiperplan care separă datele perfect pentru că ele sunt liniar neseparabile
- Găsește totuși un hiperplan care face cele mai puține greșeli
  - dacă nu are un număr limită de iterații ciclează…



Soft-margin SVM

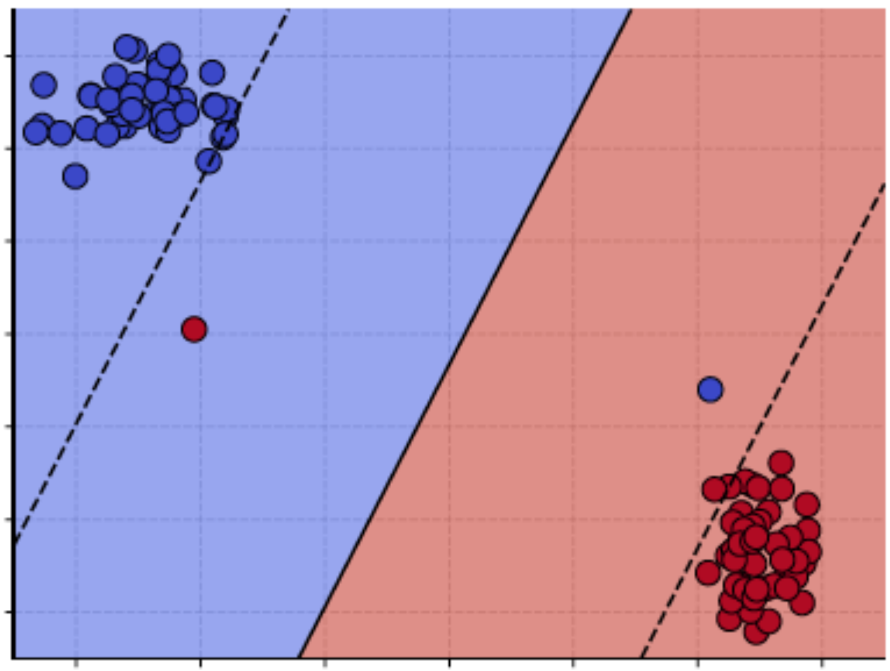
- Trade-off între găsirea unui hiperplan cu margine mică care separă perfect datele și găsirea unui hiperplan cu margine mare care face câteva greșeli (exemple de antrenare misclasificate)

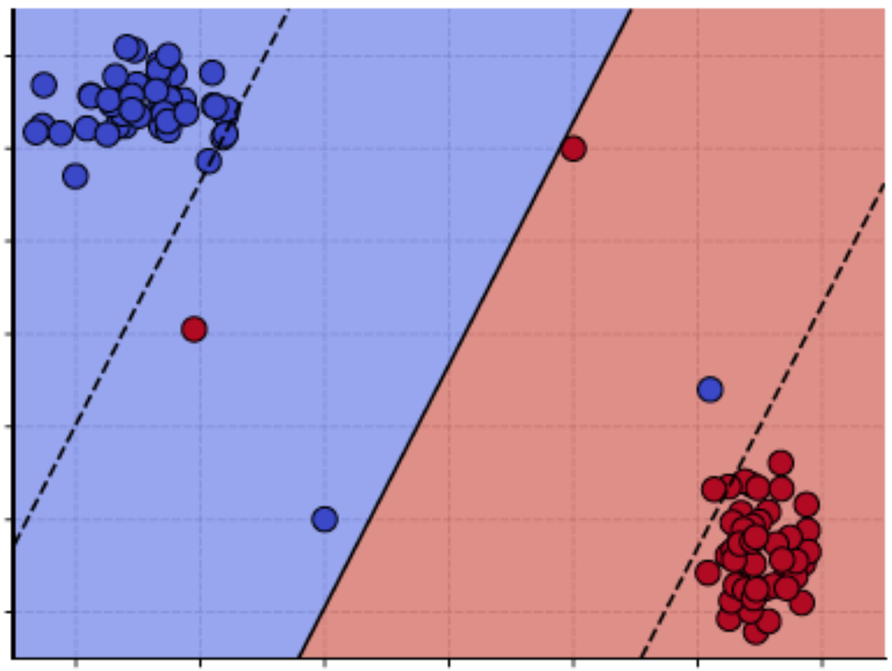




Soft-margin SVM

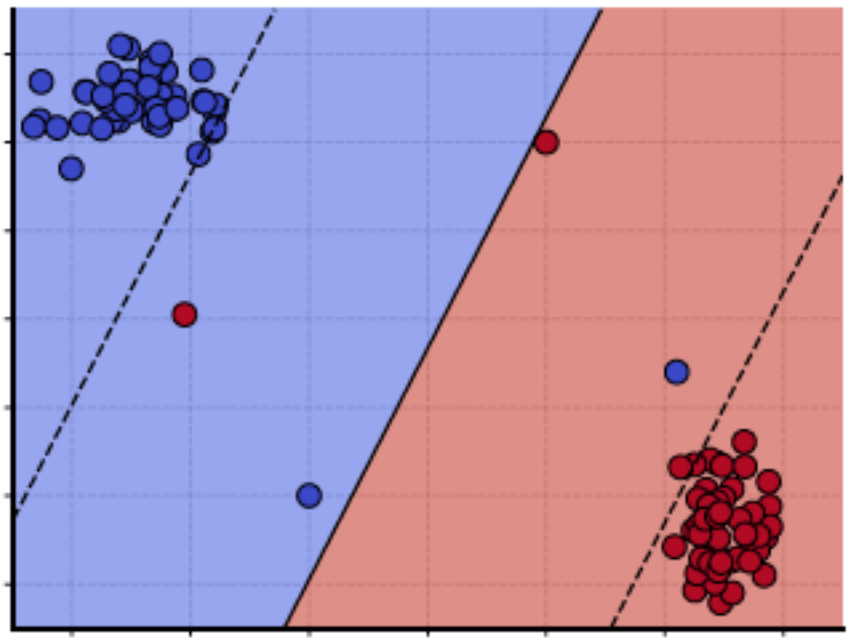
- Trade-off între găsirea unui hiperplan cu margine mică care separă perfect datele și găsirea unui hiperplan cu margine mare care face câteva greșeli (exemple de antrenare misclasificate)





Soft-margin SVM

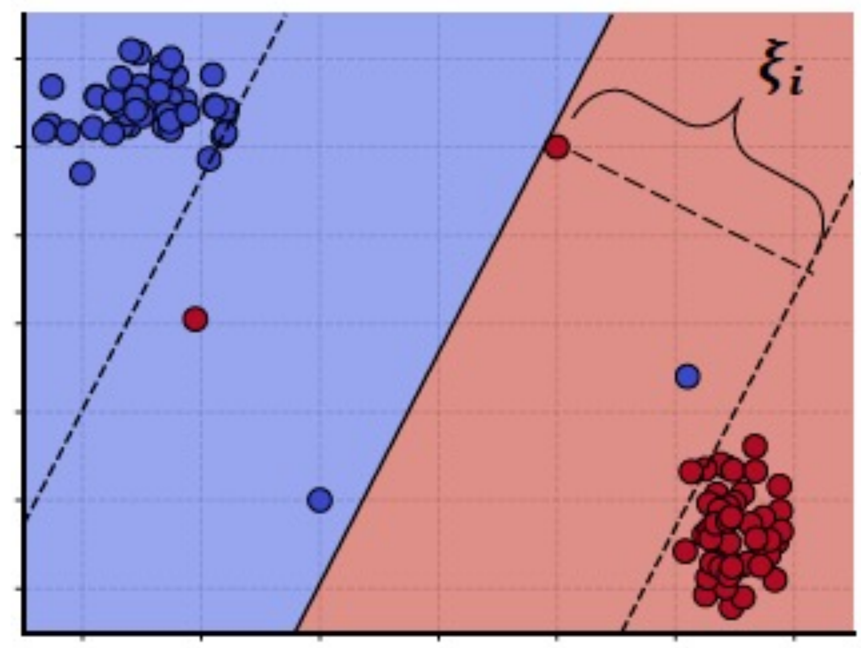
- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)



Soft-margin SVM

- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)
- Introducem variabilele $\xi_i \geq 0$
- Cuantifică cât de mult exemplul

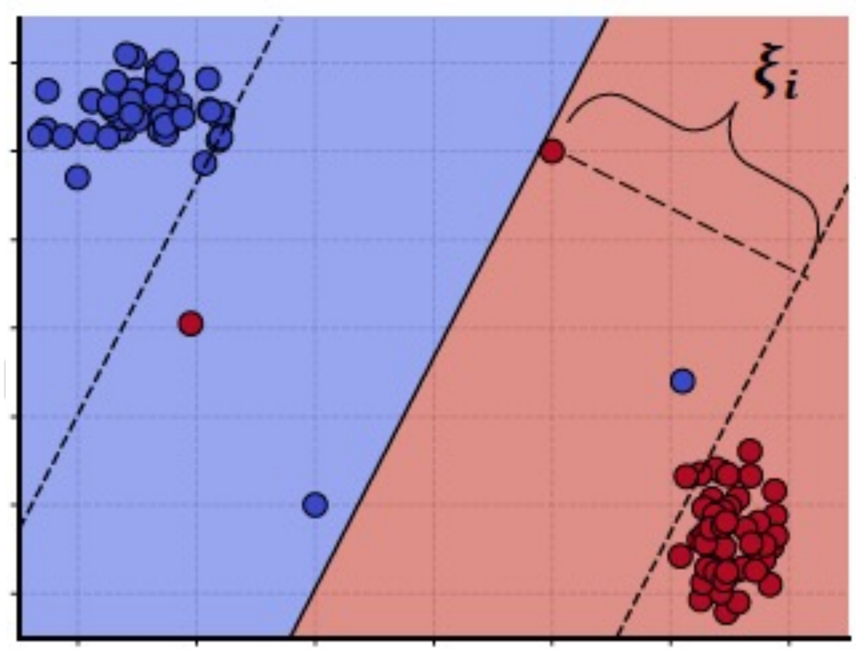
de antrenare $\vec{x}^{(i)}$ este de partea cealaltă a marginii



Soft-margin SVM

- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)
- Introducem variabilele $\xi_i \geq 0$
- Cuantifică cât de mult exemplul

de antrenare $\vec{x}^{(i)}$ este de partea cealaltă a marginii



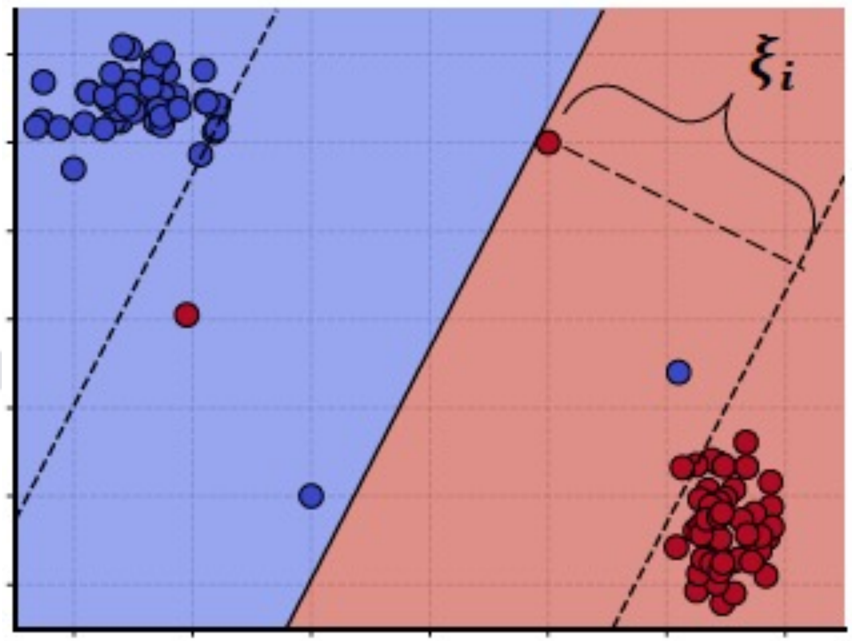
Constrângerile $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq 0$

devin: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq -\xi_i$

Soft-margin SVM

- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)
- Introducem variabilele $\xi_i \geq 0$
- Cuantifică cât de mult exemplul

de antrenare $\vec{x}^{(i)}$ este de partea cealaltă a marginii



Constrângerile $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq 0$

devin: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq -\xi_i$

Problema de optimizare nouă:

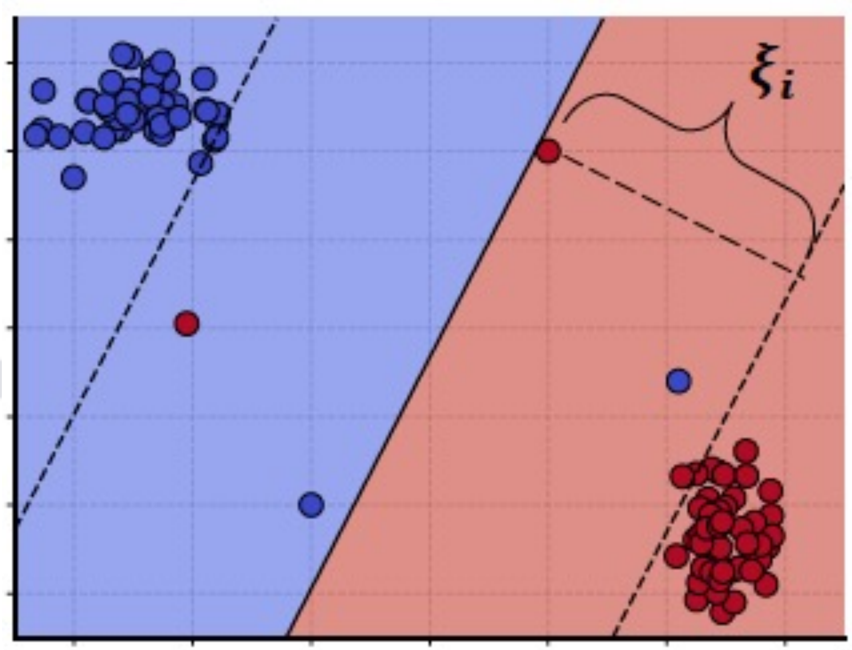
$$\text{minimize} \frac{\|\vec{w}\|^2}{2} + C \sum_i \xi_i$$

constrângeri: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) \geq 1 - \xi_i$

Soft-margin SVM

- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)
- Introducem variabilele $\xi_i \geq 0$
- Cuantifică cât de mult exemplul

de antrenare $\vec{x}^{(i)}$ este de partea cealaltă a marginii



Constrângerile $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq 0$

devin: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq -\xi_i$

Problema de optimizare nouă:

$$\text{minimize} \frac{\|\vec{w}\|^2}{2} + C \sum_i \xi_i$$

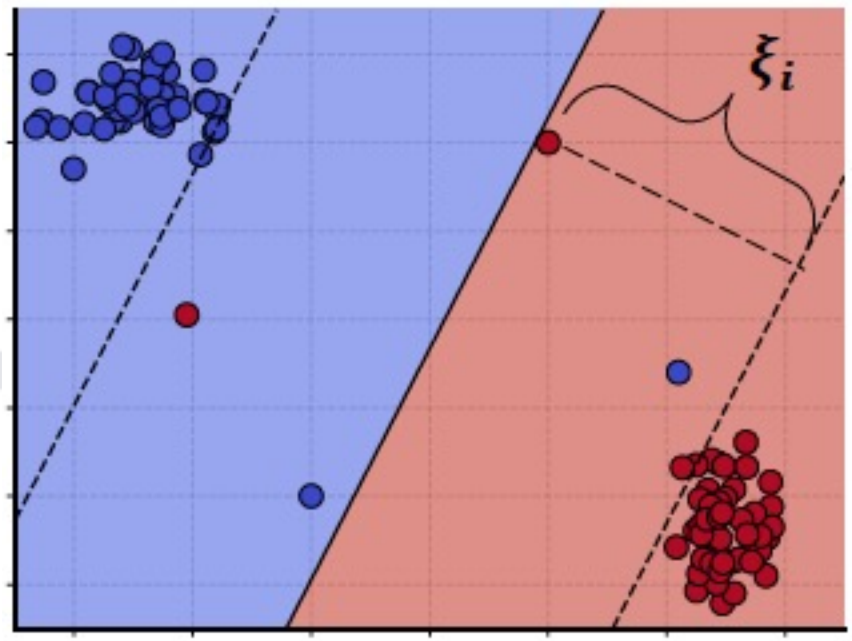
constrângeri: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) \geq 1 - \xi_i$

Hiperparametrul C controlează tradeoff-ul dintre maximizare marginii și acuratețea modelului

Soft-margin SVM

- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)
- Introducem variabilele $\xi_i \geq 0$
- Cuantifică cât de mult exemplul

de antrenare $\vec{x}^{(i)}$ este de partea cealaltă a marginii



Constrângerile $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq 0$

devin: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq -\xi_i$

Problema de optimizare nouă:

$$\text{minimize} \frac{\|\vec{w}\|^2}{2} + C \sum_i \xi_i$$

constrângeri: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) \geq 1 - \xi_i$

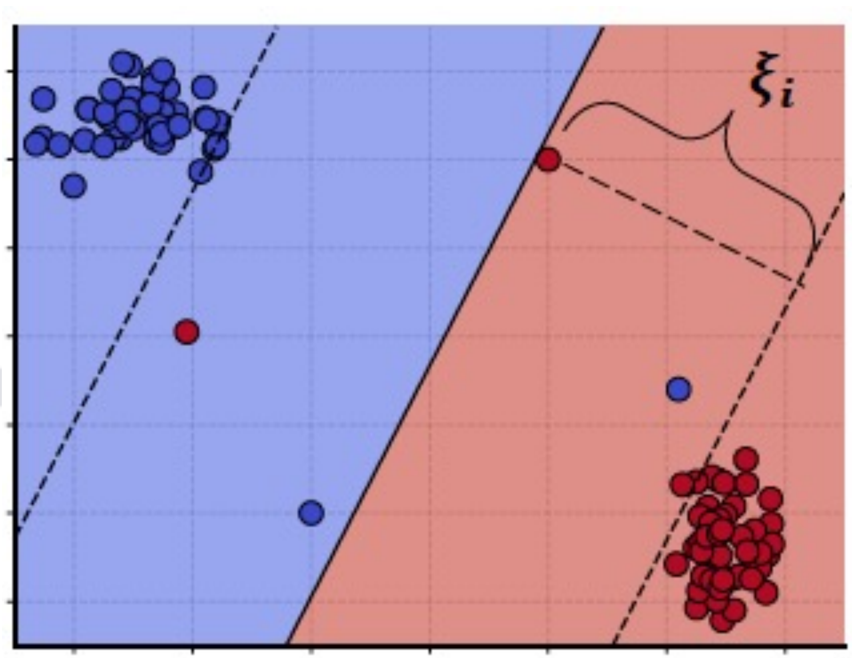
Valoare mică a lui C (aproape de 0) – maximizează marginea, marginea poate fi încalcată mai mult, penalitate mică

Soft-margin SVM

- Soft-margin SVM: permite ca anumite exemple de antrenare să fie misclasificate (clasificate greșit), cu un cost (penalitate)
- Introducem variabilele $\xi_i \geq 0$
- Cuantifică cât de mult exemplul

de antrenare $\vec{x}^{(i)}$ este de partea

cealaltă a marginii



Constrângerile $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq 0$

devin: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) - 1 \geq -\xi_i$

Problema de optimizare nouă:

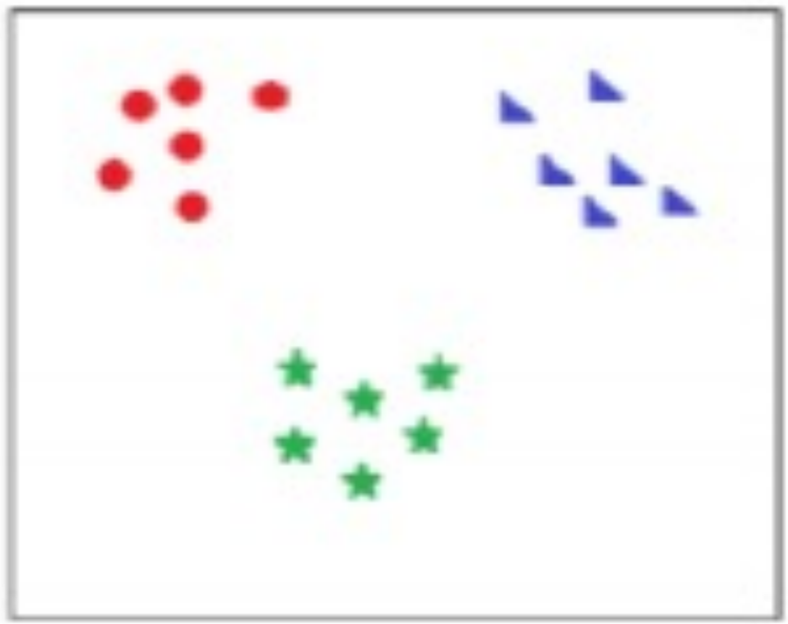
$$\text{minimize} \frac{\|\vec{w}\|^2}{2} + C \sum_i \xi_i$$

constrângeri: $y^{(i)}(\langle \vec{x}^{(i)}, \vec{w} \rangle + b) \geq 1 - \xi_i$

Valoare mare a lui C –margine nu prea mare, marginea nu poate fi încalcată mult, penalitate mare

SVM pentru clasificarea cu mai multe clase

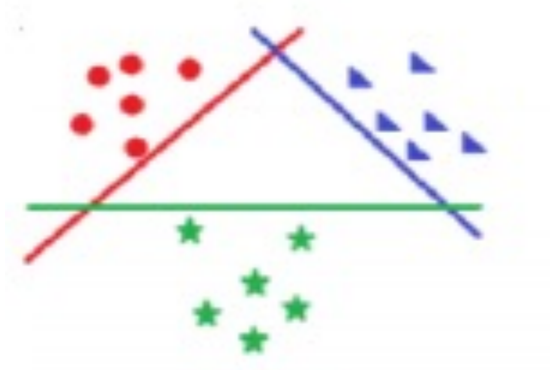
- Un SVM este un clasificator binar (prin construcție). Putem să folosim un SVM în problemele de clasificare cu mai multe clase?



SVM pentru clasificarea cu mai multe clase

- Un SVM este un clasificator binar (prin construcție). Putem să folosim un SVM în problemele de clasificare cu mai multe clase?

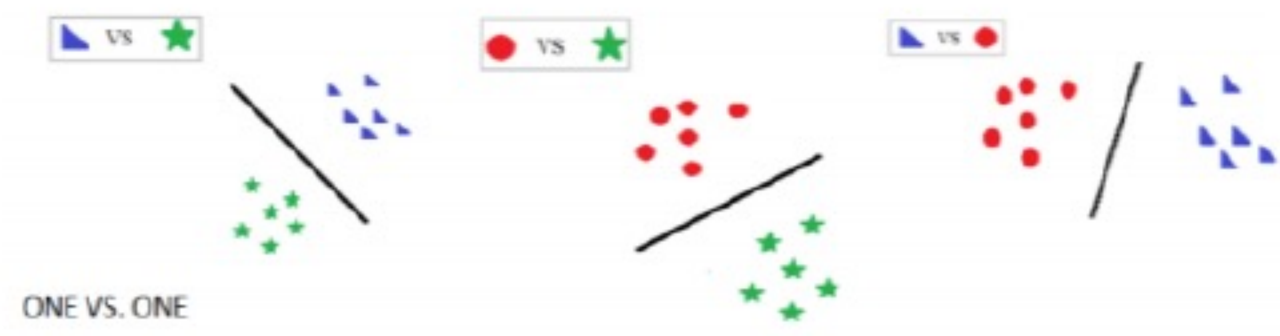
- One-versus-rest (OVR) = One-versus-all (OVA)
  - antrenează *n* clasificatori, câte unul pentru fiecare clasă
  - pentru antrenarea fiecărui clasificator: exemplele + sunt cele inițiale, exemplele – sunt date de reuniunea exemplelor celor *n-1* clase rămase
  - la inferență (testare): rulăm toți clasificatorii, alegem clasa pe baza clasificatorului cu marginea cea mai mare (cel mai încrezător că exemplul de testare este clasificat + în clasa respectivă)



SVM pentru clasificarea cu mai multe clase

- Un SVM este un clasificator binar (prin construcție). Putem să folosim un SVM în problemele de clasificare cu mai multe clase?

- One-versus-one (OVO)
  - antrenează n(n-1)/2 clasificatori, câte unul pentru fiecare pereche de clase
  - la inferență (testare): rulăm toți clasificatorii și alegem clasa care a fost selectată de cele mai multe ori



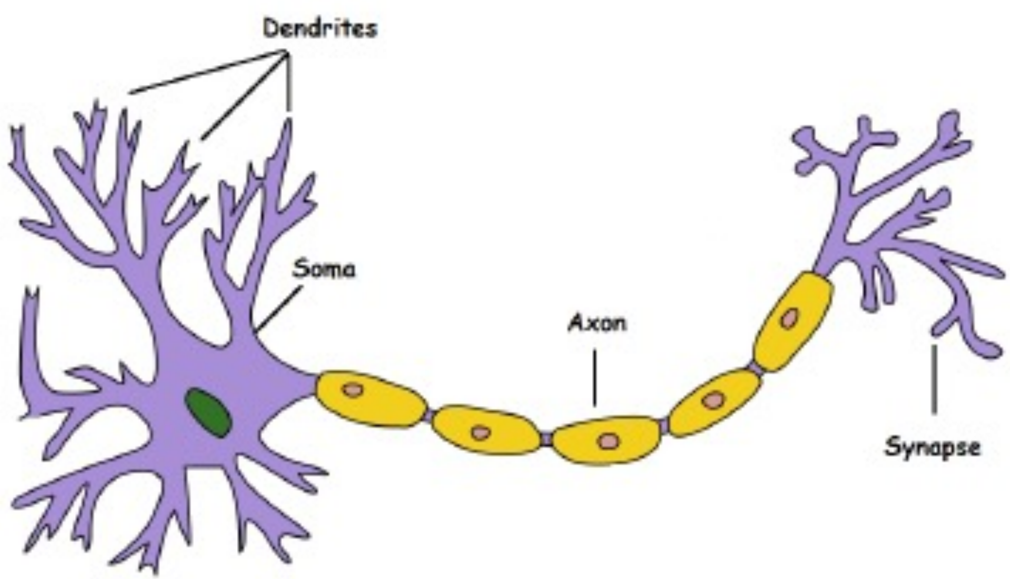
Perceptronul

Neuronul biologic

Perceptronul este un clasificator liniar inspirat de neuronul uman.

Cum funcționează un neuron:

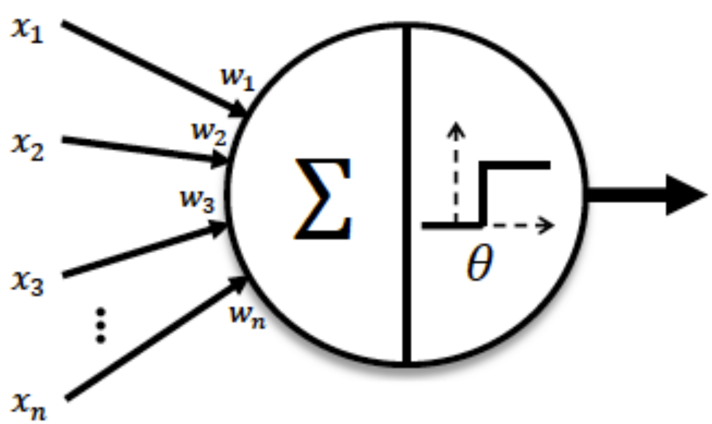
- semnale electrice (intrări) sunt transmise prin ***dendrite***
- aceste semnale duc la acumularea unui potențial electric în ***corpul neuronului (soma)***
- când potențialul electric acumulat depășește un anumit prag (threshold) un semnal electric (ieșirea neuronului) este transmis prin intermediul unui ***axon***
- conectarea axonului cu dendritele altor neuroni se realizează prin **sinapse**



Neuronul artificial

- Primul model al unui neuron artificial a fost propus de Warren McCulloch și Walter Pitts în 1943.
- Conform acestui model semnalele electrice de intrare erau valori booleene (0 - absent, 1 - prezent) de tip *excitator* sau *inhibitor.*
  - dacă numărul de intrări prezente (valoare 1) de tip excitator (pondere 1) era mai mare decât numărul de intrări prezente de tip inhibitor (pondere -1) atunci neuronul se activa
- Matematic, ieșirea neuronului este 1 dacă suma ponderată a intrărilor depășește un prag, altfel este 0.

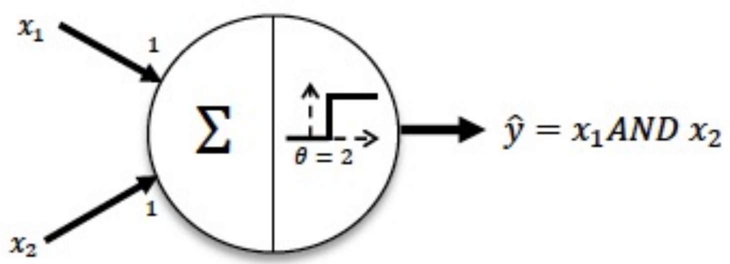
$$x_1, x_2, \dots, x_n \in \{0,1\}, \qquad w_1, w_2, \dots, w_n \in \{-1,1\}, \qquad \theta \in \mathbb{R}, \qquad \hat{y} \in \{0,1\}$$

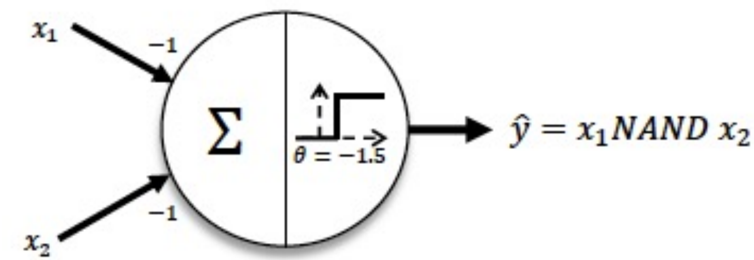


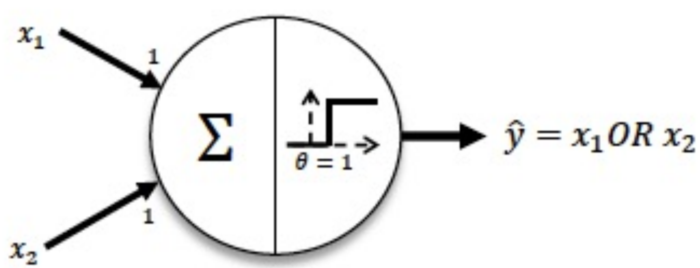
$$\hat{y} = \begin{cases} 1 & \text{if} \sum_{i=1}^{n} w_i x_i \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

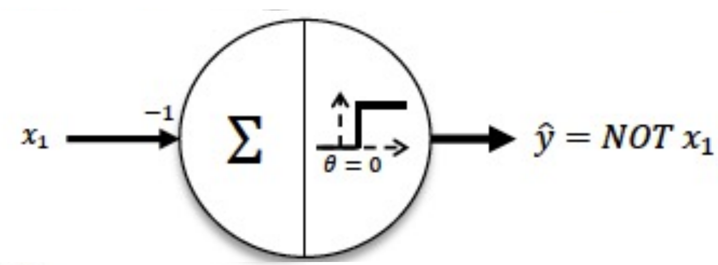
Neuronul artificial

$$\hat{y} = \begin{cases} 1 & \text{if} \sum_{i=1}^{n} w_i x_i \geq \theta \\ 0 & \text{otherwise} \end{cases}$$









Regula de învățare a lui Hebb

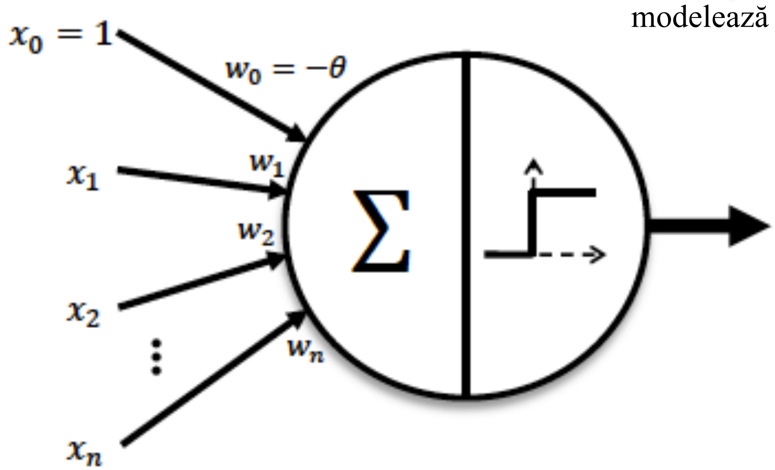
- Regula de învățare a lui Donald Hebb (1949) spune că legătura dintre doi neuroni devine mai puternică pe măsura ce un neuron trimite semnale electrice care activează celălalt neuron:
  - conform acestei reguli, intrările nu sunt numai de tip excitator sau inhibitor dar pot fi ponderate de sinapse (unele conexiuni dintre neuroni sunt mai puternice decât celelalte)
  - regula de învățare actualiza ponderile sinaptice (creștea ponderea sinaptică a unui neuron asupra altui neuron dacă acesta este activat)

Perceptronul

- Plecând de la modelul neuronului artificial McCulloch –Pitts și cu ideea lui Hebb în minte, Frank Rosenblatt inventează în 1957 o mașină și un algoritm de învățare asociat, pe care le va numi Perceptron, creat pentru recunoașterea cifrelor din imagini.
  – răspunsul celulelor fotosensibile – intrări, potențiometre – ponderi sinaptice, motoare electrice se ocupau de actualizarea de ponderi

$x_0 = 1, \quad x_1, x_2, \dots, x_n \in \{0,1\}, \quad w_0, w_1, w_2, \dots, w_n \in \mathbb{R}, \quad \hat{y} \in \{0,1\}$

modelează pragul θ



$$\hat{y} = \begin{cases} 1 & \text{if} \sum_{i=0}^{n} w_i x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{i=1}^{n} w_i x_i = 0$$ Ecuația unui hiperplan

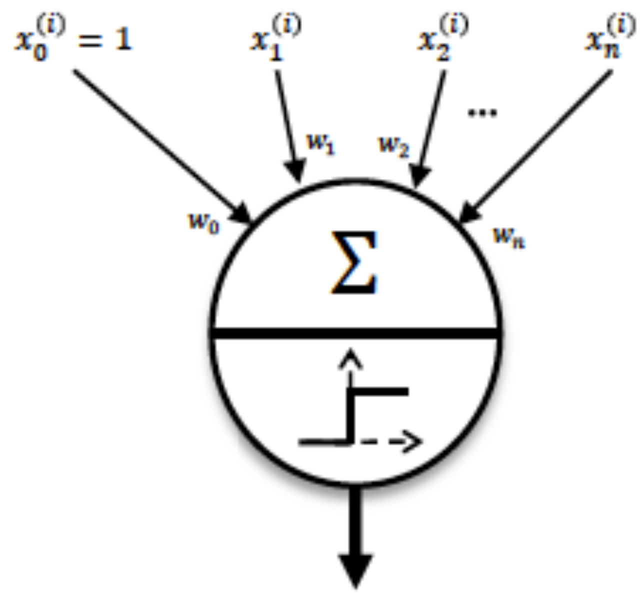
Algoritmul de învățare al perceptronului

- Mulțimea de exemple de antrenare:

$$E = \{(\vec{x}^{(1)}, y^{(1)}), \ldots, (\vec{x}^{(m)}, y^{(m)})\},\ \vec{x}^{(i)} \in \{0,1\}^{n+1} \left(x_0^{(i)} = 1, \forall i\right),\ y^{(i)} \in \{0,1\}$$

- Ponderile w_j sunt inițializate cu 0

- ***Scopul învățării este de a învăța un set de ponderi w_0, w_1, ..., w_n care minimizează eroare reală. Aceste ponderi se învață pe mulțimea de antrenare E.***



$$\hat{y}^{(i)} = \begin{cases} 1 & \text{if } \sum_{j=0}^{n} w_j x_j^{(i)} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Algoritmul de învățare al perceptronului

- Mulțimea de exemple de antrenare:

  $$E = \{(\vec{x}^{(1)}, y^{(1)}), \ldots, (\vec{x}^{(m)}, y^{(m)})\},\ \vec{x}^{(i)} \in \{0,1\}^{n+1} \left(x_0^{(i)} = 1, \forall i\right),\ y^{(i)} \in \{0,1\}$$

- Ponderile w_j sunt inițializate cu 0
- Pentru fiecare exemplu de antrenare $(\vec{x}^{(i)}, y^{(i)})$ din mulțimea E:
  - dacă eticheta prezisă = eticheta reală $\hat{y}^{(i)} == y^{(i)}$ nu facem actualizare
  - dacă $\hat{y}^{(i)} == 0$ și $y^{(i)} == 1$ creștem ponderile tuturor intrărilor active
  - dacă $\hat{y}^{(i)} == 1$ și $y^{(i)} == 0$ scădem ponderile tuturor intrărilor active

$$\hat{y}^{(i)} = \begin{cases} 1 & \text{if } \sum_{j=0}^{n} w_j x_j^{(i)} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Algoritmul de învățare al perceptronului

- Mulțimea de exemple de antrenare:

$$E = \{(\vec{x}^{(1)}, y^{(1)}), \dots, (\vec{x}^{(m)}, y^{(m)})\},\ \vec{x}^{(i)} \in \{0,1\}^{n+1} \left(x_0^{(i)} = 1, \forall i\right),\ y^{(i)} \in \{0,1\}$$

- Ponderile w_j sunt inițializate cu 0
- Pentru fiecare exemplu de antrenare $(\vec{x}^{(i)}, y^{(i)})$ din mulțimea E:
  - dacă eticheta prezisă = eticheta reală $\hat{y}^{(i)} == y^{(i)}$ nu facem actualizare
  - dacă $\hat{y}^{(i)} == 0$ și $y^{(i)} == 1$ creștem ponderile tuturor intrărilor active
  - dacă $\hat{y}^{(i)} == 1$ și $y^{(i)} == 0$ scădem ponderile tuturor intrărilor active
- Regula de actualizare a perceptronului

$w_j = w_j + \Delta w_j$, unde

$$\Delta w_j = \eta\left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

$$\hat{y}^{(i)} = \begin{cases} 1 & \text{if } \sum_{j=0}^{n} w_j x_j^{(i)} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Cantitatea prin care schimbăm ponderea (rata de învățare)

Setează semnul actualizării

Selectează intrările active

Algoritmul de învățare al perceptronului

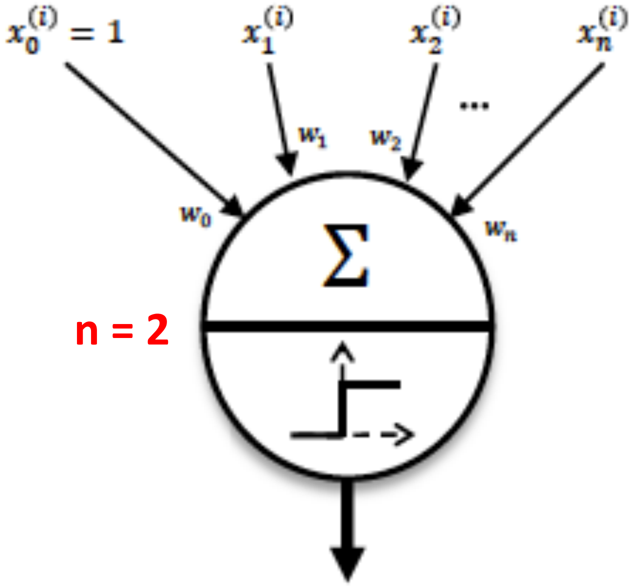
```
def perceptron(X, y, n_epochs, η):
    m, n = X.shape # number of samples, number of inputs
    for j in range(n):
        w_j = 0
    for epoch in range(n_epochs): # an "epoch" is a run through all training data.
        for i in range(m):          # a "training step" is one    update of the weights.
            ŷ^(i) = unit_step_function(Σ_{j=0}^{n} w_j x_j^(i))
            for j in range(n):
                w_j += η(y^(i) − ŷ^(i)) · x_j^(i)
```

$$\hat{y}^{(i)} = \begin{cases} 1 & \text{if } \sum_{j=0}^{n} w_j x_j^{(i)} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Exemplu – învațarea funcției OR

$E = \{(\vec{x}^{(1)}, y^{(1)}), \dots, (\vec{x}^{(m)}, y^{(m)})\},\ \vec{x}^{(i)} \in \{0,1\}^{n+1} \left(x_0^{(i)} = 1, \forall i\right),\ y^{(i)} \in \{0,1\}$

	x_0	x_1	x_2	y
$\vec{x}^{(1)}$	1	0	0	0
$\vec{x}^{(2)}$	1	0	1	1
$\vec{x}^{(3)}$	1	1	0	1
$\vec{x}^{(4)}$	1	1	1	1



$$\hat{y}^{(i)} = \begin{cases} 1 & \text{if } \sum_{j=0}^{n} w_j x_j^{(i)} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	0	0

Accuracy: 25%

$\hat{y}$
0
0
0
0

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	0	0

Accuracy: 25%

$\hat{y}$
0
0
0
0

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	0	0	0
$\vec{x}^{(2)}$			
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 1

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	0	0

Accuracy: 25%

$\hat{y}$
0
0
0
0

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	0	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 1

$\eta = 1$

$$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	0	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	0	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 1

$\eta = 1$

$$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	0	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	0	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	0	0	0
$\vec{x}^{(4)}$			

Epoch 1

$\eta = 1$

$$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	0	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	0	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	0	0	0
$\vec{x}^{(4)}$	0	0	0

Epoch 1

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	0	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$			
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 2

$\eta = 1$

$$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	0	1

Accuracy: 75%

$\hat{y}$
0
1
0
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$			
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 2

$\eta = 1$

$$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	0	1

Accuracy: 75%

$\hat{y}$
0
1
0
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$	0	0	0
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 2

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	0	1

Accuracy: 75%

$\hat{y}$
0
1
0
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$	0	0	0
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$			

Epoch 2

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	1	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$	0	0	0
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$			

Epoch 2

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	1	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$	0	0	0
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	0	0	0

Epoch 2

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
1	1	1

Accuracy: 75%

$\hat{y}$
1
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$			
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 3

$\eta = 1$

$$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	1	1

Accuracy: 100%

$\hat{y}$
0
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$			
$\vec{x}^{(3)}$			
$\vec{x}^{(4)}$			

Epoch 3

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

y
0
1
1
1

w_0	w_1	w_2
0	1	1

Accuracy: 100%

$\hat{y}$
0
1
1
1

	Δw_0	Δw_1	Δw_2
$\vec{x}^{(1)}$	-1	0	0
$\vec{x}^{(2)}$	0	0	0
$\vec{x}^{(3)}$	0	0	0
$\vec{x}^{(4)}$	0	0	0

Epoch 3

$\eta = 1$

$\Delta w_j = \left(y^{(i)} - \hat{y}^{(i)}\right) \cdot x_j^{(i)}$

Exemplu – învațarea funcției OR

	x_0	x_1	x_2
$\vec{x}^{(1)}$	1	0	0
$\vec{x}^{(2)}$	1	0	1
$\vec{x}^{(3)}$	1	1	0
$\vec{x}^{(4)}$	1	1	1

OR

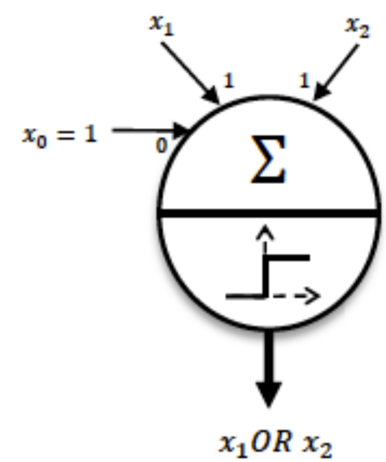
y
0
1
1
1

w_0	w_1	w_2
0	1	1

Accuracy: 100%

$\hat{y}$
0
1
1
1

Perceptronul a învațat funcția OR în 3 epoci



$$\hat{y}^{(i)} = \begin{cases} 1 & \text{if } \sum_{j=0}^{n} w_j x_j^{(i)} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Eticheta 1

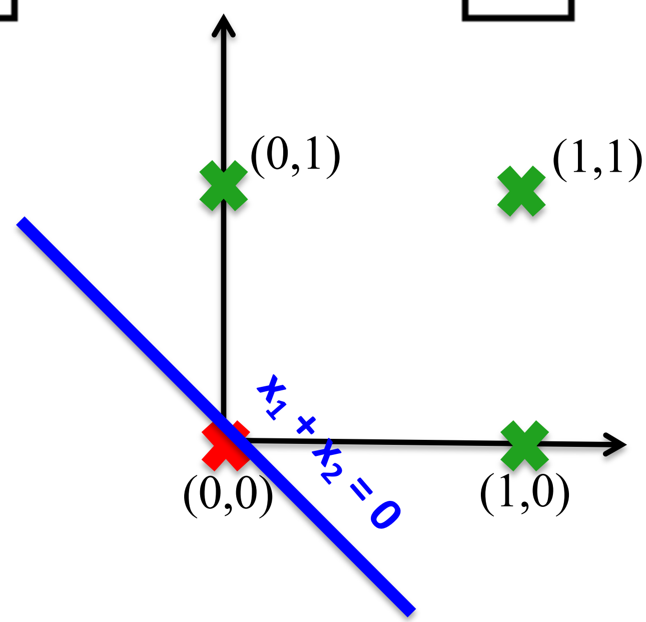
Eticheta 0

$w_0x_0 + w_1x_1 + w_2x_2 > 0$

$x_1 + x_2 > 0$

$w_0x_0 + w_1x_1 + w_2x_2 \leq 0$

$x_1 + x_2 \leq 0$



Limitările perceptronului

Limitările perceptronului

- Perceptronul învață funcțiile booleene AND/OR și reușește să rezolve probleme de recunoaștere a cifrelor. Probleme simple sau grele?

- Considerăm funcția XOR (exclusiv OR):

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Pentru ca un Perceptron să învețe funcția XOR e nevoie ca:

$$w_0 + 0 \cdot w_1 + 0 \cdot w_2 < 0$$
$$w_0 + 0 \cdot w_1 + 1 \cdot w_2 \geq 0$$
$$w_0 + 1 \cdot w_1 + 0 \cdot w_2 \geq 0$$
$$w_0 + 1 \cdot w_1 + 1 \cdot w_2 < 0$$

Limitările perceptronului

- Perceptronul învață funcțiile booleene AND/OR și reușește să rezolve probleme de recunoaștere a cifrelor. Probleme simple sau grele?

- Considerăm funcția XOR (exclusiv OR):

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Pentru ca un Perceptron să învețe funcția XOR e nevoie ca:

$$w_0 + 0 \cdot w_1 + 0 \cdot w_2 < 0 \Rightarrow \boldsymbol{w_0 < 0}$$
$$w_0 + 0 \cdot w_1 + 1 \cdot w_2 \geq 0 \Rightarrow w_0 \geq -w_2$$
$$w_0 + 1 \cdot w_1 + 0 \cdot w_2 \geq 0 \Rightarrow w_0 \geq -w_1$$
$$w_0 + 1 \cdot w_1 + 1 \cdot w_2 < 0 \Rightarrow w_0 < -w_1 - w_2$$

Limitările perceptronului

- Perceptronul învață funcțiile booleene AND/OR și reușește să rezolve probleme de recunoaștere a cifrelor. Probleme simple sau grele?

- Considerăm funcția XOR (exclusiv OR):

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Pentru ca un Perceptron să învețe funcția XOR e nevoie ca:

$$w_0 + 0 \cdot w_1 + 0 \cdot w_2 < 0 \Rightarrow \boldsymbol{w_0 < 0}$$

$$\left.\begin{array}{l} \left.\begin{array}{l} w_0 + 0 \cdot w_1 + 1 \cdot w_2 \geq 0 \Rightarrow w_0 \geq -w_2 \\ w_0 + 1 \cdot w_1 + 0 \cdot w_2 \geq 0 \Rightarrow w_0 \geq -w_1 \end{array}\right\} 2w_0 \geq -w_1 - w_2 \\ w_0 + 1 \cdot w_1 + 1 \cdot w_2 < 0 \Rightarrow w_0 < -w_1 - w_2 \end{array}\right\} 2w_0 > w_0 \Rightarrow \boldsymbol{w_0 > 0}$$

Limitările perceptronului

- Perceptronul învață funcțiile booleene AND/OR și reușește să rezolve probleme de recunoaștere a cifrelor. Probleme simple sau grele?

- Considerăm funcția XOR (exclusiv OR):

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Pentru ca un Perceptron să învețe funcția XOR e nevoie ca:

$$w_0 + 0 \cdot w_1 + 0 \cdot w_2 < 0 \Rightarrow \mathbf{w_0 < 0}$$

$$\left.\begin{array}{l} w_0 + 0 \cdot w_1 + 1 \cdot w_2 \geq 0 \Rightarrow w_0 \geq -w_2 \\ w_0 + 1 \cdot w_1 + 0 \cdot w_2 \geq 0 \Rightarrow w_0 \geq -w_1 \end{array}\right\} 2w_0 \geq -w_1 - w_2$$

$$w_0 + 1 \cdot w_1 + 1 \cdot w_2 < 0 \Rightarrow w_0 < -w_1 - w_2$$

$$\left.\begin{array}{l} 2w_0 \geq -w_1 - w_2 \\ w_0 < -w_1 - w_2 \end{array}\right\} 2w_0 > w_0 \Rightarrow \mathbf{w_0 > 0}$$

Contradicție

- Perceptronul nu poate învăța XOR oricât de mulți pași ar face
- Mai mult, Perceptronul poate învață numai clase care sunt liniar separabile

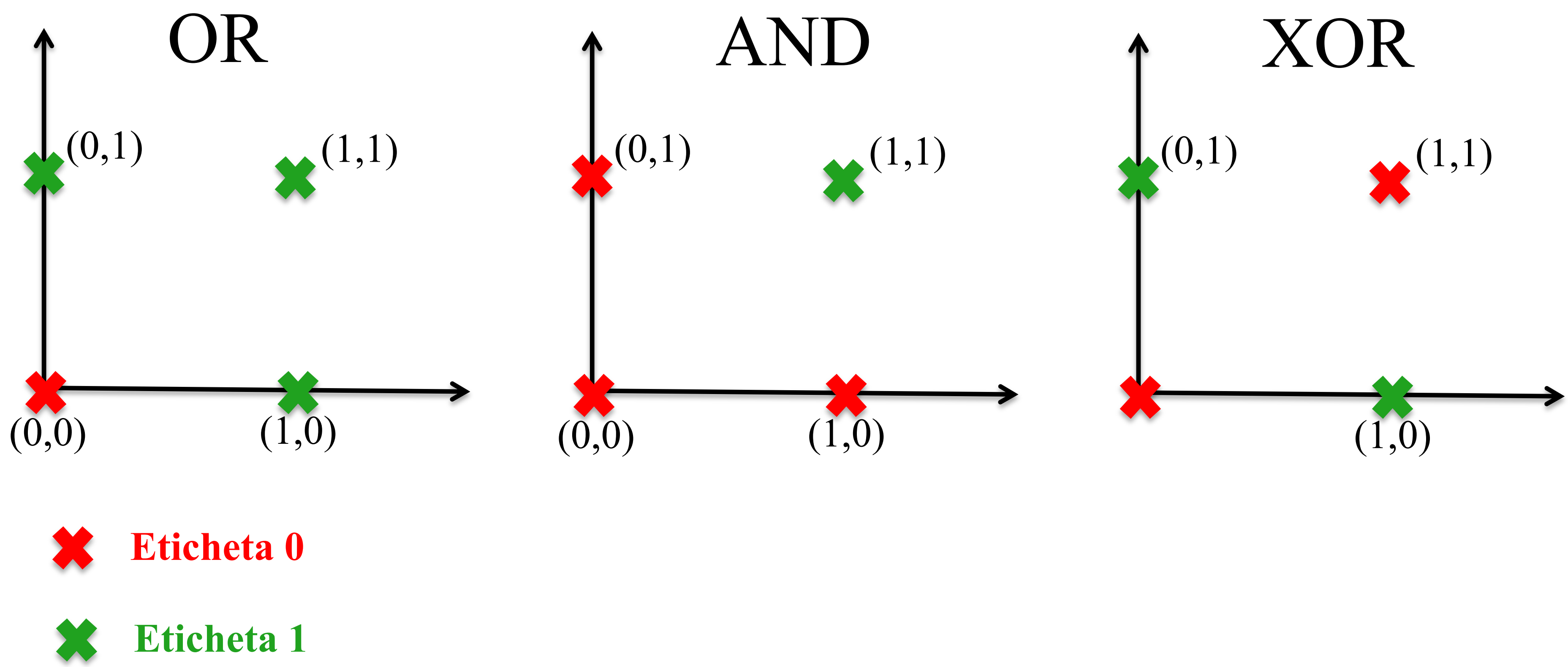
Teorema de convergență a perceptronului

- Dacă mulțimea de antrenare E este liniar separabilă cu margine γ, algoritmul de învățare a perceptronului este garantat că va converge într-un număr finit de pași către o soluție în care nu se fac greșeli pe mulțimea de antrenare

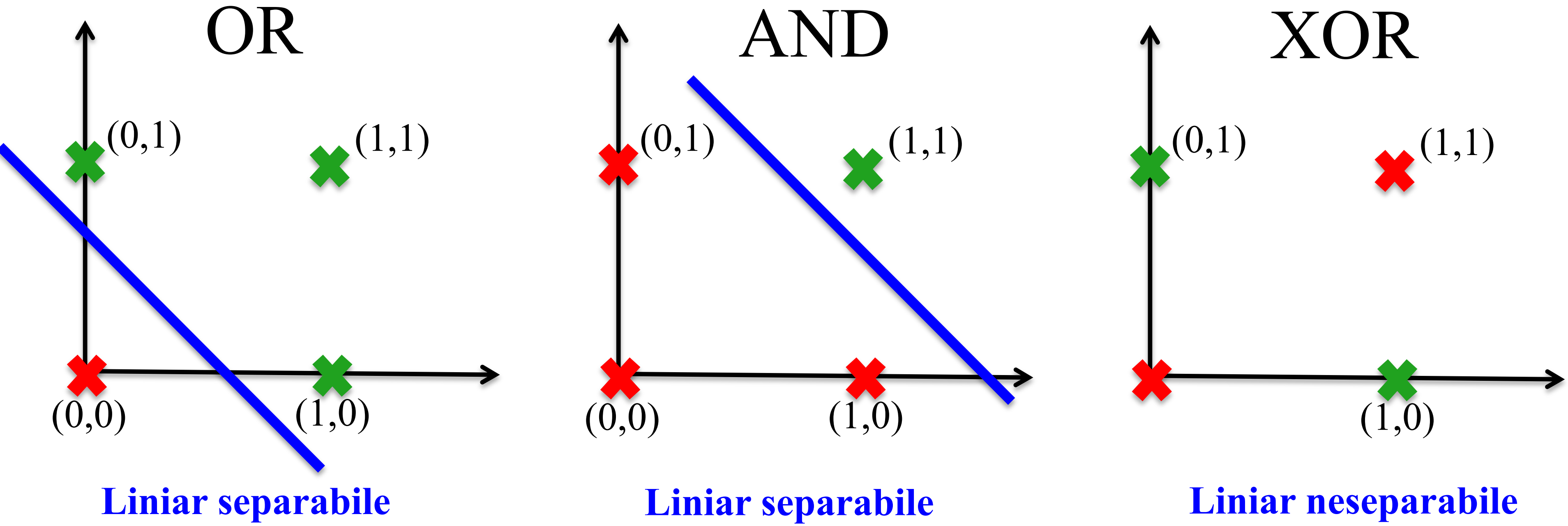
- Numărul de pași k satisface relația $k \leq \frac{R^2}{\gamma^2}$, unde R este raza sferei din spațiul caracteristicilor care cuprinde toate exemplele de antrenare

- Teorema spune că va converge către o soluție, nu este necesar să fie o soluție bună (SVM oferă soluția de margine maximă). Dacă mulțimea de antrenare E nu este liniar separabilă algoritmul nu va converge către o soluție.

Perspectiva geometrică



Perspectiva geometrică



Eticheta 0

Eticheta 1