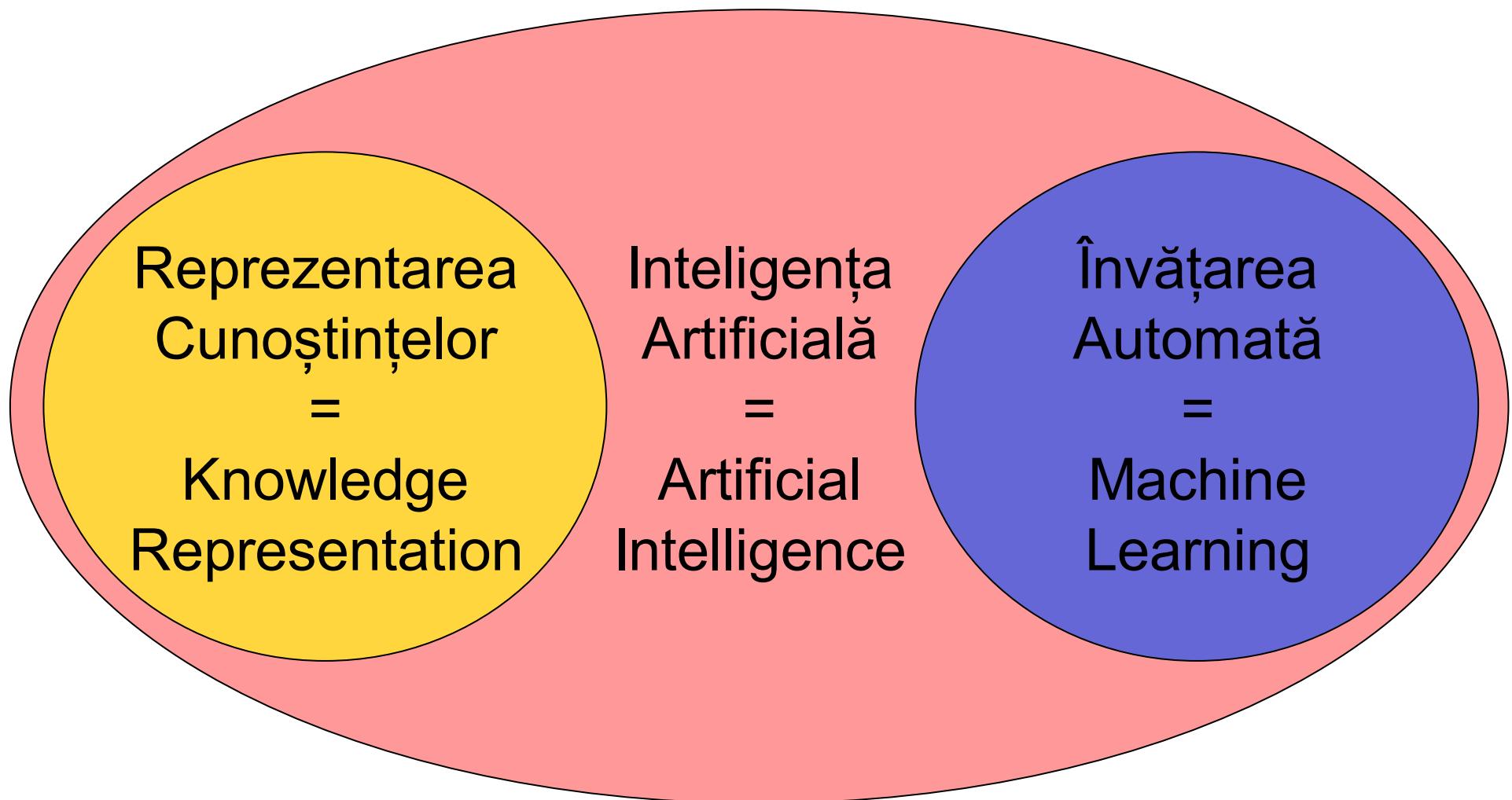


# Inteligenta artificială. Învățare automată. Concepte de bază.

Prof. Dr. Radu Ionescu  
[raducu.ionescu@gmail.com](mailto:raducu.ionescu@gmail.com)

Facultatea de Matematică și Informatică  
Universitatea din București

# Inteligenta artificială și învățarea automată



# Profesori - Învățarea automată

- Cursuri:
  - Radu Ionescu (raducu.ionescu[ML]@gmail.com)
- Laboratoare:
  - Mihail Burduja (warchildmd[ML]@gmail.com)
  - Iuliana Georgescu (georgescu\_lily[ML]@yahoo.com)
  - Andrei Manea (andecenu@gmail.com)
  - Cătălin Ristea (r.catalin196[ML]@yahoo.ro)
  - Petru Soviany (petru.soviany[ML]@yahoo.com)
  - Marius Todea (todeamarius2006[ML]@gmail.com)

# Sistem de notare

- Nota este formată din:
  - nota la examen 50%
  - nota la laborator 50%
- Vor fi două note la laborator, câte una pentru fiecare materie. Nota finală de la laborator este formată din media notelor de laborator
- Notele de la examen și laborator trebuie să fie ambele peste 5 (regula se aplică și la restanță, nu se reportează notele)
- La examen vor fi subiecte din ambele materii
- Pentru materia "Învățare automată", studenții vor prezenta un proiect individual în săptămâna a 8-a (respectiv a 11-a sau a 14-a)

# Sistem de notare

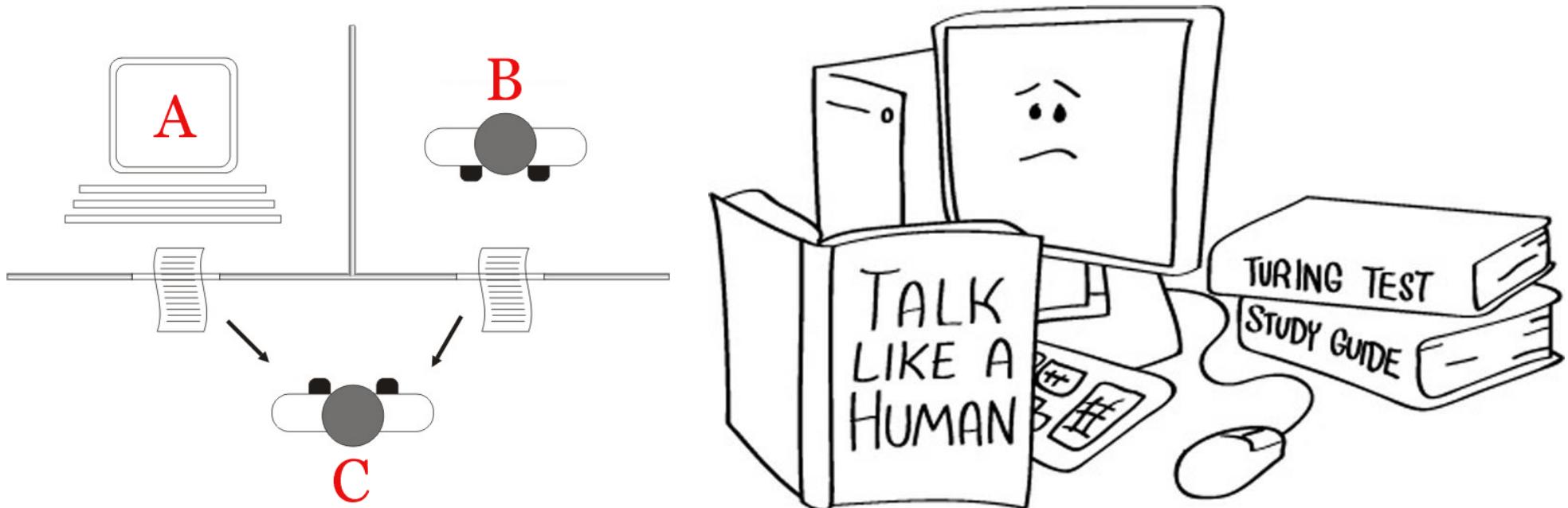
- Proiectul constă în dezvoltarea unei metode de clasificare și participarea la competiția (TBA) propusă pe platforma Kaggle
- Notele vor fi proporționale cu rata de acuratețe obținută:
  - Locurile 1-30 => nota maximă 10
  - Locurile 31-60 => nota maximă 9
  - Locurile 61-90 => nota maximă 8
  - Locurile 91-110 => nota maximă 7
  - Locurile 111-130 => nota maximă 6
  - Locul 131 sau mai jos => nota maximă 5
- Proiectul trebuie prezentat după ultimul laborator (se acordă 2 puncte pentru prezentare și documentație)
- Pentru nota  $\geq 5$ , trebuie depășită performanța baseline

# Sistem de notare

- Puncte extra în timpul cursurilor / laboratoarelor
  - Primul care răspunde la o întrebare / rezolvă un exercițiu primește 0.2 puncte
  - Maxim 0.4 puncte pe curs / laborator de persoană (se puntează doar primele două răspunsuri corecte)
  - Până la 1 punct în plus la nota din examen
  - Până la 1 punct în plus la nota din laborator
- 
- Codul va fi verificat cu soft anti-plagiat: NU este permisă preluare codului de pe web (sub nicio formă) sau de la colegi

# La ce se referă inteligența artificială?

- Scopul suprem al inteligenței artificiale este de a construi sisteme care să atingă nivelul de inteligență al omului
- Testul Turing: un computer prezintă un nivel de inteligență uman dacă un interlocutor uman nu reușește să distingă, în urma unei conversații în limbaj natural, că vorbește cu un om sau cu un calculator



# La ce se referă învățarea automată?

- O mare parte din cercetători consideră că acest scop poate fi atins prin imitarea modului în care o oamenii învăță
- **Învățarea automată** – domeniu care studiază modul în care calculatoarele pot fi înzestrate cu abilitatea de a învăța, fără ca aceasta să fie programată în mod explicit
- În acest context, **învățarea** se referă la:
  - recunoașterea unor tipare / structuri (patterns) complexe
  - luarea deciziilor inteligente bazate pe observațiile din **date**

# Problemă “bine pusă” de învățare automată

- Ce probleme pot fi rezolvate\* folosind învățarea automată?
- **Problemă “bine pusă” de învățare automată:**
- Spunem despre un program pe calculator că învață dintr-o experiență  $E$  în raport cu o clasă de task-uri  $T$  și o măsură de performanță  $P$ , dacă performanța sa în rezolvarea task-urilor  $T$ , măsurată prin  $P$ , se îmbunătățește odată cu experiența  $E$
- **(\*) rezolvate cu un anumit grad de acuratețe**

# Problemă “bine pusă” de învățare automată

- Arthur Samuel (1959) a scris un program pentru a juca dame (probabil primul program bazat pe conceptul de învățare)
- Programul a jucat împotriva lui însuși 10 mii de jocuri
- Programul a fost conceput să găsească ce poziții ale tablei de joc erau bune sau rele în funcție de probabilitatea de a câștiga sau pierde
- În acest caz:
  - E = 10000 de jocuri
  - T = joacă dame
  - P = dacă câștigă sau nu



# Strong AI versus Weak AI

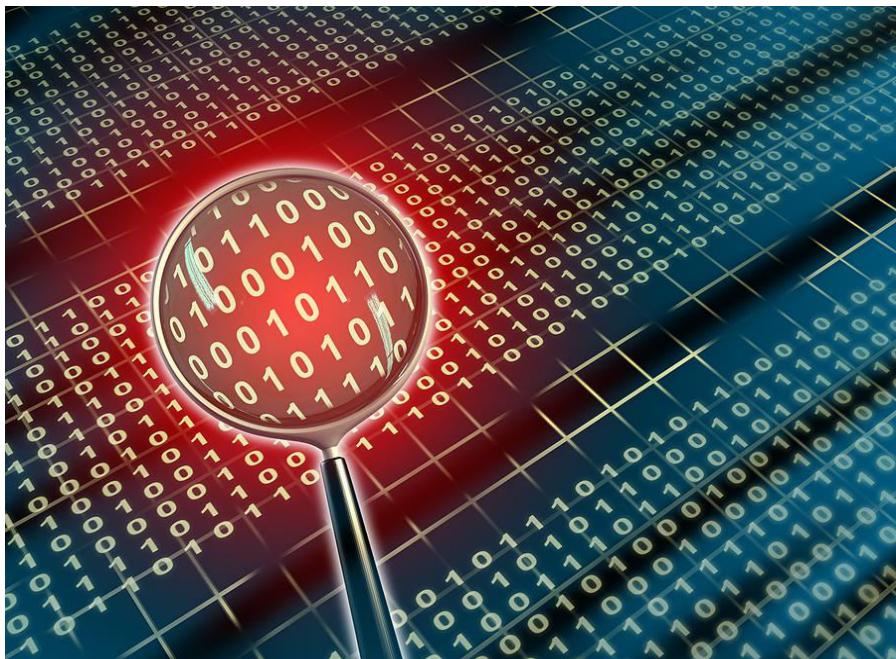
- strong / generic / true AI  
(vezi definiția lui Turing)
- weak / narrow AI  
(se focusează pe o anumită problemă)

# Când se aplică învățarea automată?

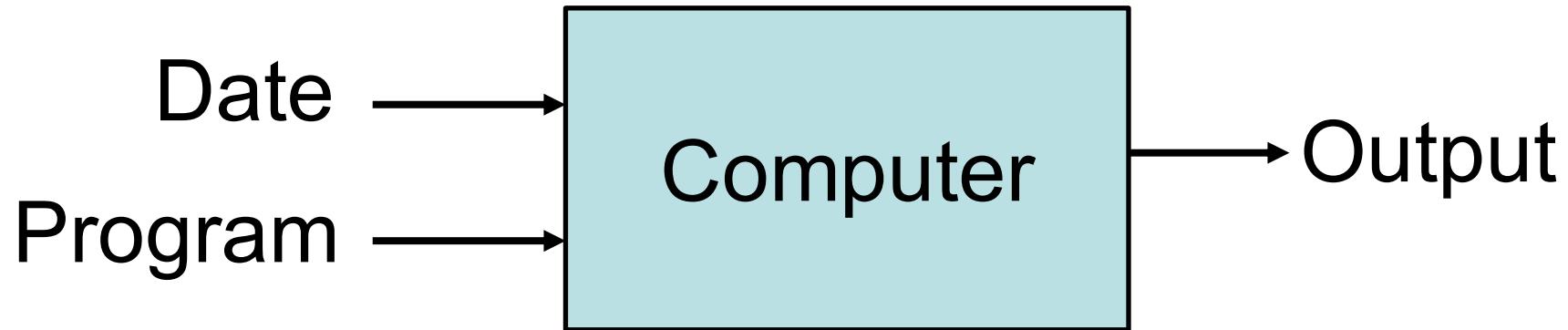
- Se aplică în situații în care este foarte greu (imposibil) să definim un set de reguli de mână / să scriem un program
- Exemple de probleme unde putem aplica învățarea automată:
  - Detectarea facială
  - Înțelegerea vorbirii
  - Prezicerea prețului acțiunilor
  - Recunoașterea obiectelor

# Esența învățării automate

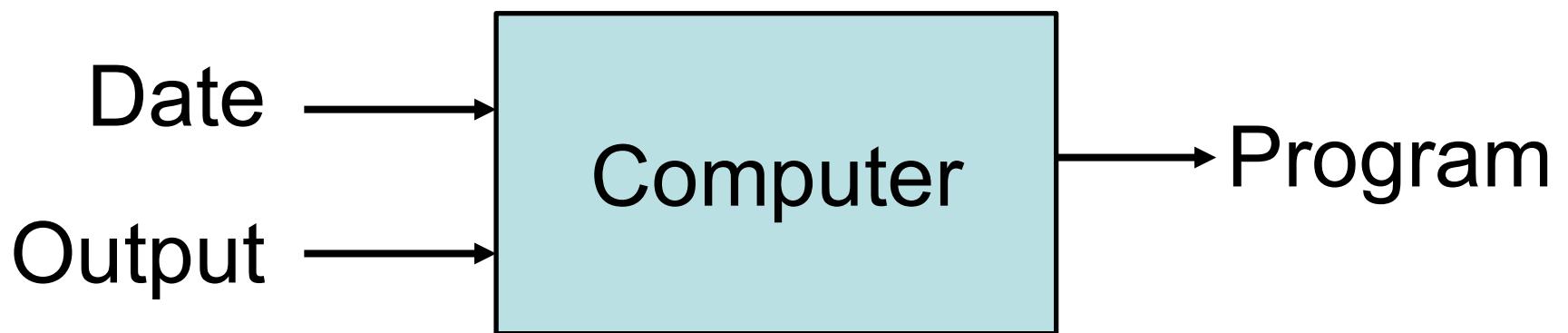
- Există un tipar
- Dar nu îl putem exprima programatic / matematic
- Avem date / exemple în care regăsim acest tipar



# Programare tradițională



# Învățare automată



# Ce este învățarea automată?

[Arthur Samuel, 1959] field of study that

- gives computers the ability to learn without being explicitly programmed

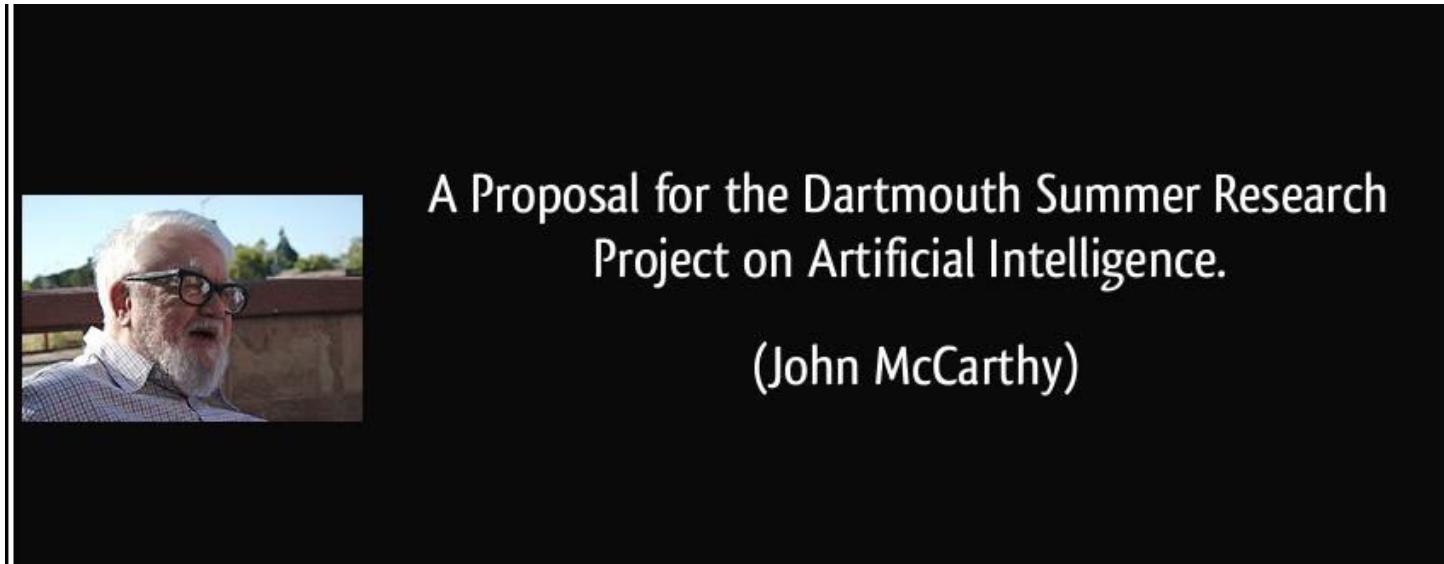
[Kevin Murphy] algorithms that

- automatically detect patterns in data
- use the uncovered patterns to predict future data or other outcomes of interest

[Tom Mitchell] algorithms that

- improve their performance (P)
- at some task (T)
- with experience (E)

# Scurt istoric al inteligenței artificiale



# Scurt istoric al inteligenței artificiale

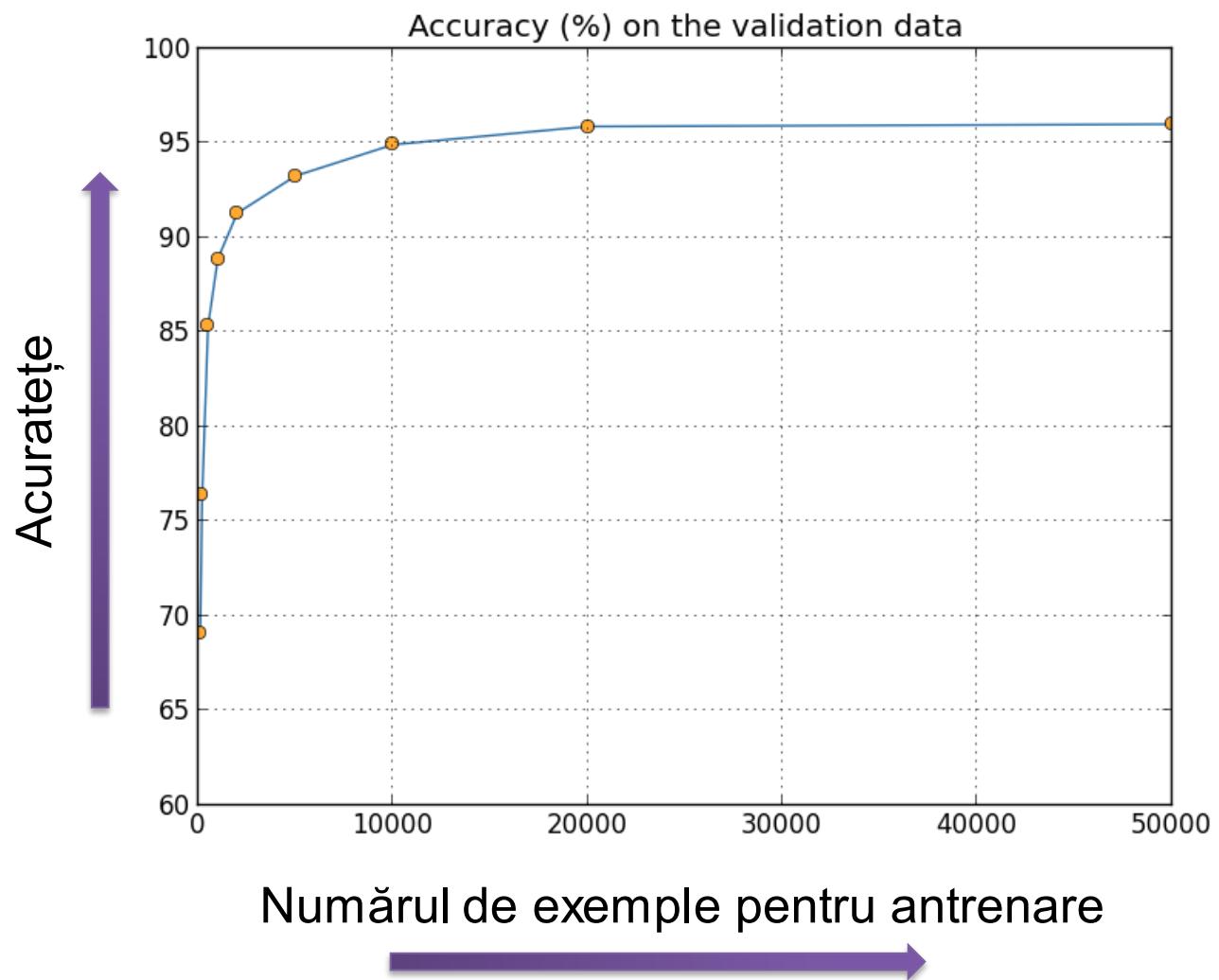
- “We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire.”
- The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.
- An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.
- We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

# Scurt istoric al inteligenței artificiale

- Anii 1960-1980: "AI Winter"
- Anii 1990: Rețelele neuronale domină, în principal datorită descoperirii algoritmului de propagare a erorii înapoi pentru rețele cu mai multe straturi
- Anii 2000: Metodele kernel domină, în principal din cauza instabilității rețelelor neuronale
- Anii 2010: Revenirea la rețelele neuronale, în principal datorită conceptului de învățare profundă (deep learning)

# De ce funcționează în prezent?

- Mai multă putere de calcul
- Mai multe date
- Modele mai bune



# Esența învățării automate

- Mii de algoritmi de învățare automată existenți
  - Cercetătorii publică sute de noi algoritmi în fiecare an
- Simplificând decenii de cercetare în domeniu, putem reduce învățarea automată la:
  - Învățarea unei funcții  $f$  care să mapeze un input  $X$  către un output  $Y$ , anume  $f: X \rightarrow Y$
  - Exemplu:  $X$ : email-uri,  $Y$ : {spam, non-spam}

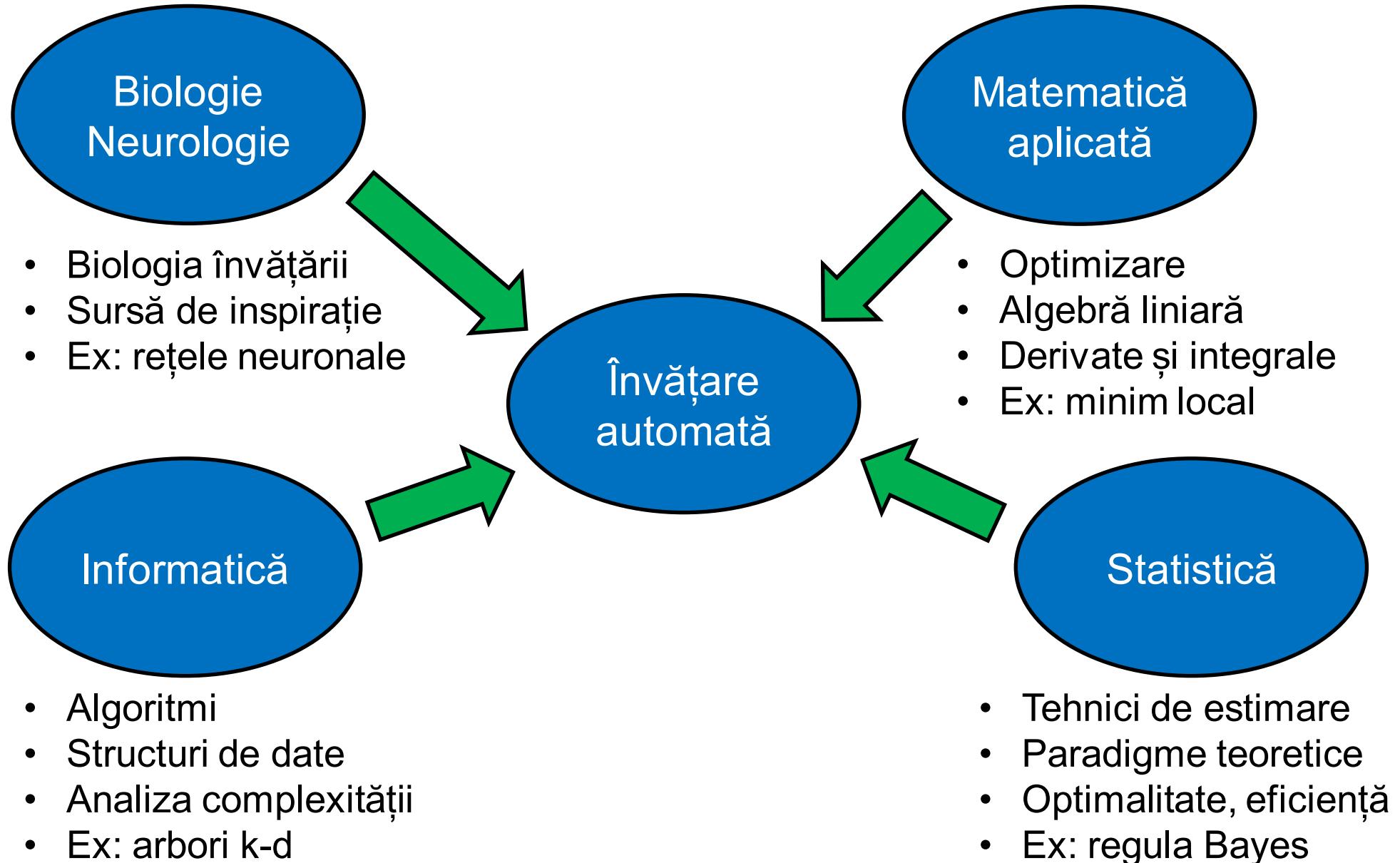
# Esența învățării automate

- Input: X (imagini, texte, email-uri...)
- Output: Y (spam sau non-spam...)
- Funcție Target (necunoscută)  
 $f: X \rightarrow Y$  (realitatea / "adevărata" mapare)
- Date  
 $(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)$
- Model  
 $g: X \rightarrow Y$   
 $y = g(x) = sign(w^T x)$

# Esența învățării automate

- Orice algoritm de învățare automată are 3 componente:
  - Reprezentare / Modelare
  - Evaluare / Funcție obiectiv
  - Optimizare

# Ce cunoștințe sunt necesare?



# Paradigme ale învățării

- Învățare supervizată (supervised learning)
- Învățare nesupervizată (unsupervised learning)
- Învățare semi-supervizată (semi-supervised learning)
- Învățare ranforsată (reinforcement learning)
- Paradigme non-standard:
  - Învățarea activă (active learning)
  - Învățare prin transfer (transfer learning)

# Învățare supervizată

- Avem la dispoziție exemple de obiecte etichetate
- Exemplu 1: recunoașterea obiectelor din imagini cu eticheta obiectelor conținute



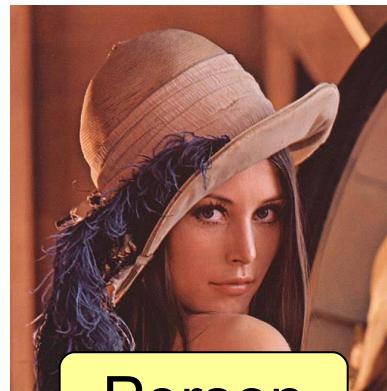
Car



Car



Person



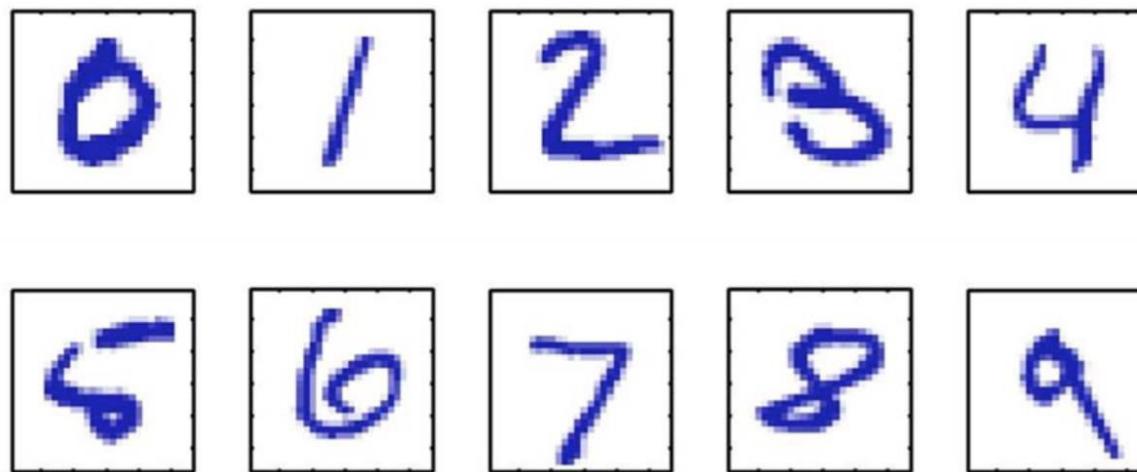
Person



Dog

# Învățare supervizată

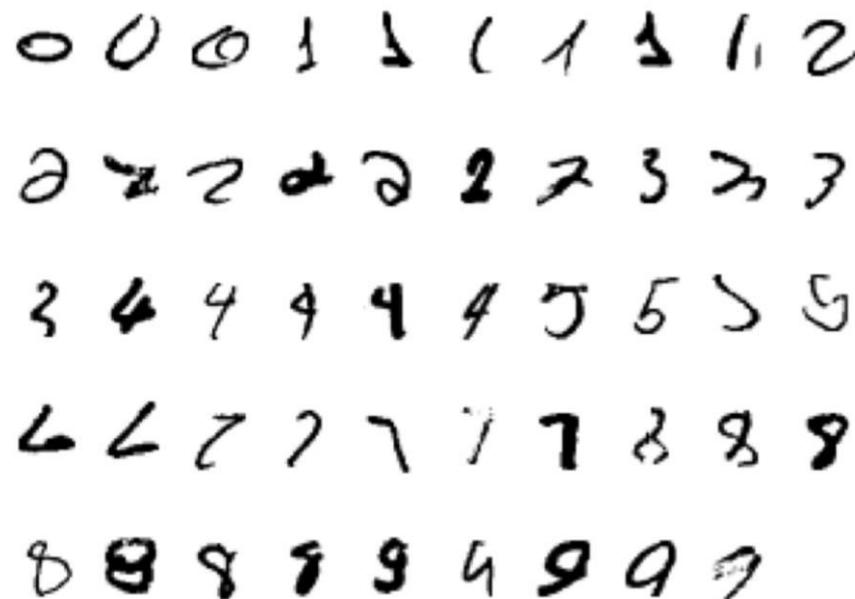
- Exemplu 2: recunoașterea caracterelor scrise de mână (setul de date MNIST)



- Imagini de 28 x 28 de pixeli
- Reprezentăm imagine ca un vector  $x$  cu 784 de componente
- Antrenăm un clasificator  $f(x)$  astfel încât:
- $f : x \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

# Învățare supervizată

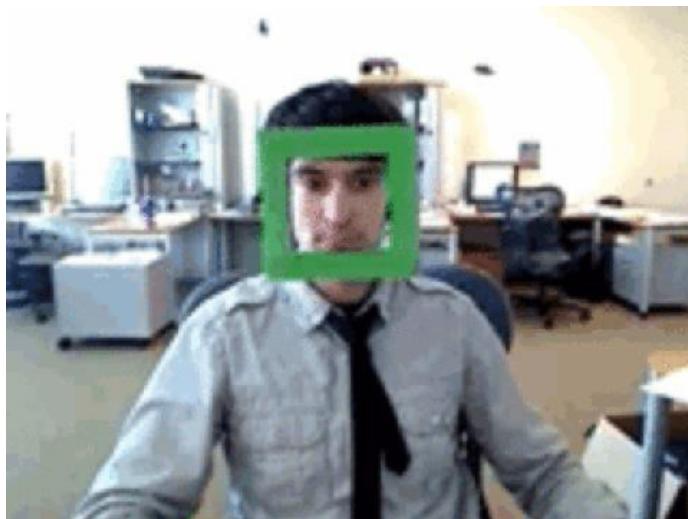
- Exemplu 2: recunoașterea caracterelor scrise de mână (setul de date MNIST)



- Pornind de la un set de antrenare, de exemplu 6000 de imagini per clasă
- Rata de eroare poate ajunge la 0.23% (cu rețele neuronale convolutionale)
- Printre primele sisteme (bazate pe învățare) comerciale utilizate pe scară largă pentru procesare de coduri poștale și cecuri bancare

# Învățare supervizată

- Exemplu 3: detectare facială



- O abordare constă în plimbarea unei ferestre peste imagine
- Scopul este să clasificăm fereastra într-una din cele două clase posibile: față sau non-față (transformarea problemei într-una de clasificare)

# Învățare supervizată

- Exemplu 3: detectare facială



- Pornim de la un set cu imagini cu fețe cu diverse variații de vârstă, gen, condiții de iluminare, dar nu translație.
- Și un set mult mai mare cu imagini care nu conțin fețe

# Învățare supervizată

- Exemplu 4: detectare de spam

The screenshot shows an email interface with a recipient icon and the text "rama rama ramaumar002@hotmail.com via yahoo.com to". Below this, the "From:" field is listed as "Mrs. Rama Umar" with the full name "Groupe Bank of Africa (Annexe) Burkina Faso Foreign Department Operation." The body of the email contains a message from the sender, Mrs. Rama Umar, stating she is working with Bank of Africa in Burkina Faso and has discovered a late foreign customer. She claims the account has neither deposits nor withdrawals and that none of the family members or relations of the late person are aware of it. The amount mentioned is "Five Million USA Dollars".

rama rama ramaumar002@hotmail.com via yahoo.com  
to

From: Mrs. Rama Umar  
Groupe Bank of Africa (Annexe) Burkina Faso  
Foreign Department Operation.

My name is Mrs.Rama Umar. I am working with Bank of Africa here in Burkina Faso. I have discovered a late foreign customer.  
When I discovered that there had been neither deposits nor withdrawals from this account, I found that none of the family member or relations of the late person are aware of this account, especially the amount involved (Five Million USA Dollars).

- Problema este de a clasifica un e-mail în spam și non-spam
- Apariția cuvântului “Dollars” este un indicator de spam
- Un exemplu de reprezentare este un vector cu frecvența cuvintelor

# Numărăm cuvintele

Obținem X

rama rama ramaumar002@hotmail.com via yahoo.com  
to ▾

From: Mrs. Rama Umar

Groupe Bank of Africa (Annexe) Burkina Faso

Foreign Department Operation.

My name is Mrs.Rama Umar. I am working with Bank of Africa here in Burkina Faso. I have been working here for 10 years. I work with the late foreign customer.

When I discovered that there had been neither deposits nor withdrawals from this account, I checked with all the family members and none of the family member or relations of the late person are aware of this account, which means that it is a dormant account. The amount in the account is (Five Million USA Dollars).

free	100
money	2
:	:
account	2
:	:



**Yoshua Bengio** <yoshua.bengio@gmail.com>

to Dong-Hyun, Ian, Dumitru, Pierre, Aaron, Mehdi, Ben, Will, Charlie,

Nice slides!

See you next week,

—Yoshua

free	1
money	1
:	:
account	2
:	:

# Algoritm de detectare a spam-ului



$$\begin{pmatrix} \text{free} & 100 \\ \text{money} & 2 \\ \vdots & \vdots \\ \text{account} & 2 \\ \vdots & \vdots \end{pmatrix}$$

De ce aceste cuvinte?

$$\begin{pmatrix} 100 \times 0.2 \\ 2 \times 0.3 \\ \vdots \\ 2 \times 0.3 \\ \vdots \end{pmatrix}$$

$$= 3.2$$



Confidență /  
garanția  
performanței?

$$\begin{pmatrix} 100 \times 0.01 \\ 2 \times 0.02 \\ \vdots \\ 2 \times 0.01 \\ \vdots \end{pmatrix}$$

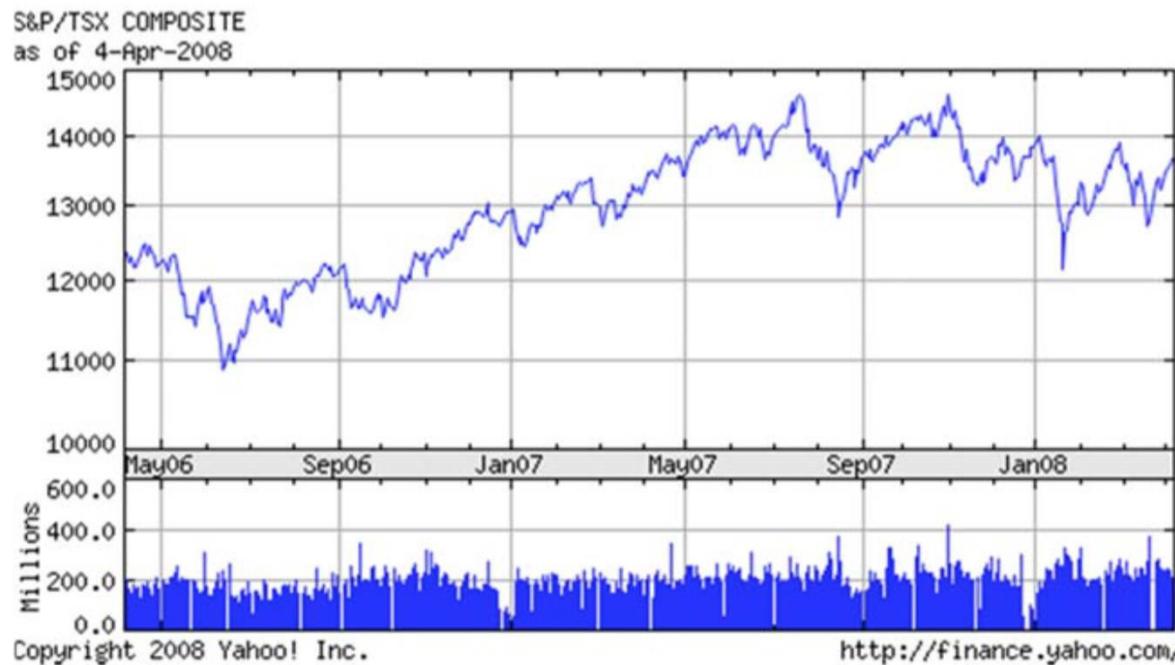
$$= 1.03$$

De ce combinație liniară?

De unde vin aceste ponderi?

# Învățare supervizată

- Exemplu 5: prezicerea prețului acțiunilor la bursă



- Scopul este de a prezice prețul la o dată din viitor, de exemplu peste câteva zile
- Aceasta este un task de regresie, deoarece output-ul este unul continuu

# Învățare supervizată

- Exemplu 6: prezicerea dificultății unei imagini



2.78



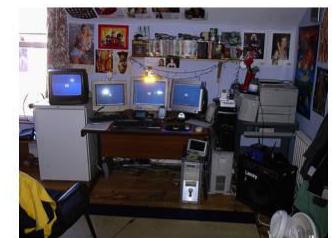
2.82



3.30



3.62



3.80

easy

image difficulty score

hard

2.81



3.15



3.45



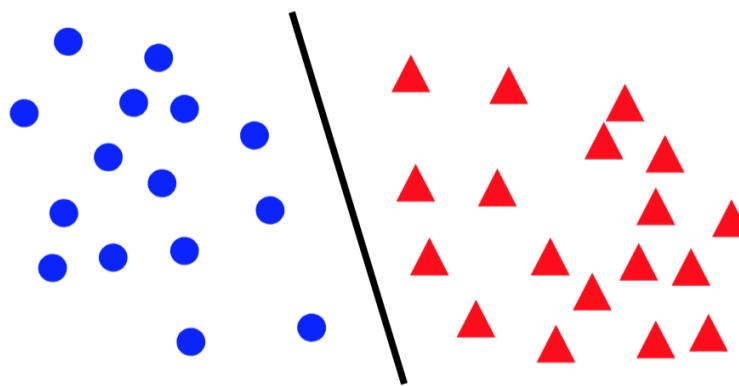
3.64



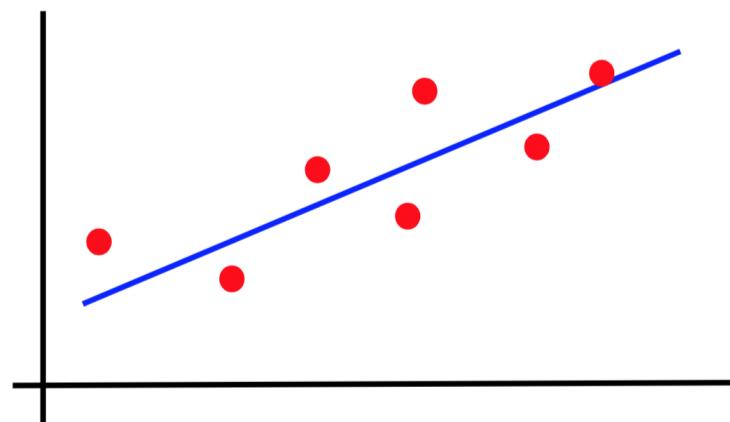
- Scopul este de a prezice cât de dificil ar fi pentru un om să recunoască obiectele din imagine
- Aceasta este un task de regresie, deoarece output-ul este unul continuu

# Formele canonice ale problemelor de învățare supervizată

- Clasificare

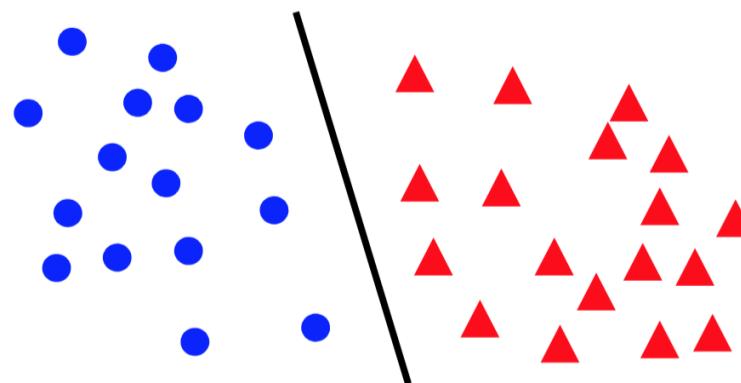


- Regresie

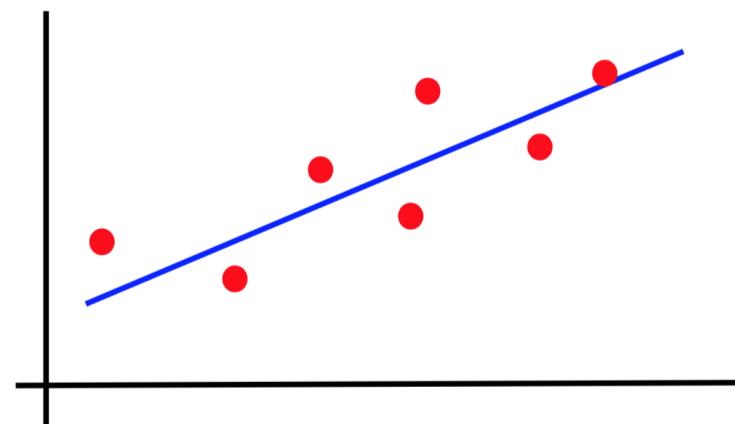


# Estimarea vârstei unei persoane din imagini

- Clasificare?



- Regresie?



Ce vârstă?

# Paradigma de învățare supervizată

Functions  $\mathcal{F}$

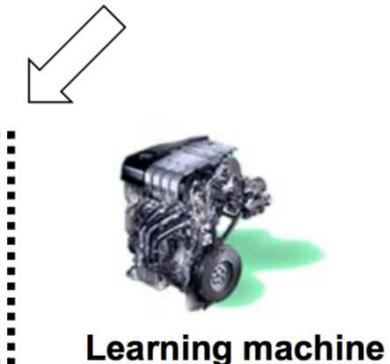
$$\textcolor{red}{f} : \mathcal{X} \rightarrow \mathcal{Y}$$

Training data

$$\{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}\}$$

LEARNING

find  $\hat{f} \in \mathcal{F}$   
s.t.  $y_i \approx \hat{f}(x_i)$



PREDICTION

$$\textcolor{red}{y} = \hat{f}(x)$$

New data

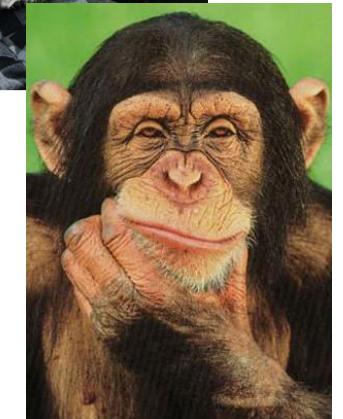
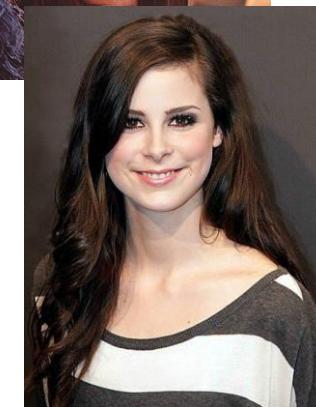
$$x$$

# Modele de învățare supervizată

- Clasificatorul Bayes naiv (cursul 2)
- Metoda celor mai apropiati vecini (cursul 3)
- Mașini cu vectori suport (cursul 4)
- Metode kernel (cursul 4)
- Rețele neuronale și învățare “deep” (cursurile 5, 6, 7)
- Arboi de decizie și random forests (la master)
- Altele

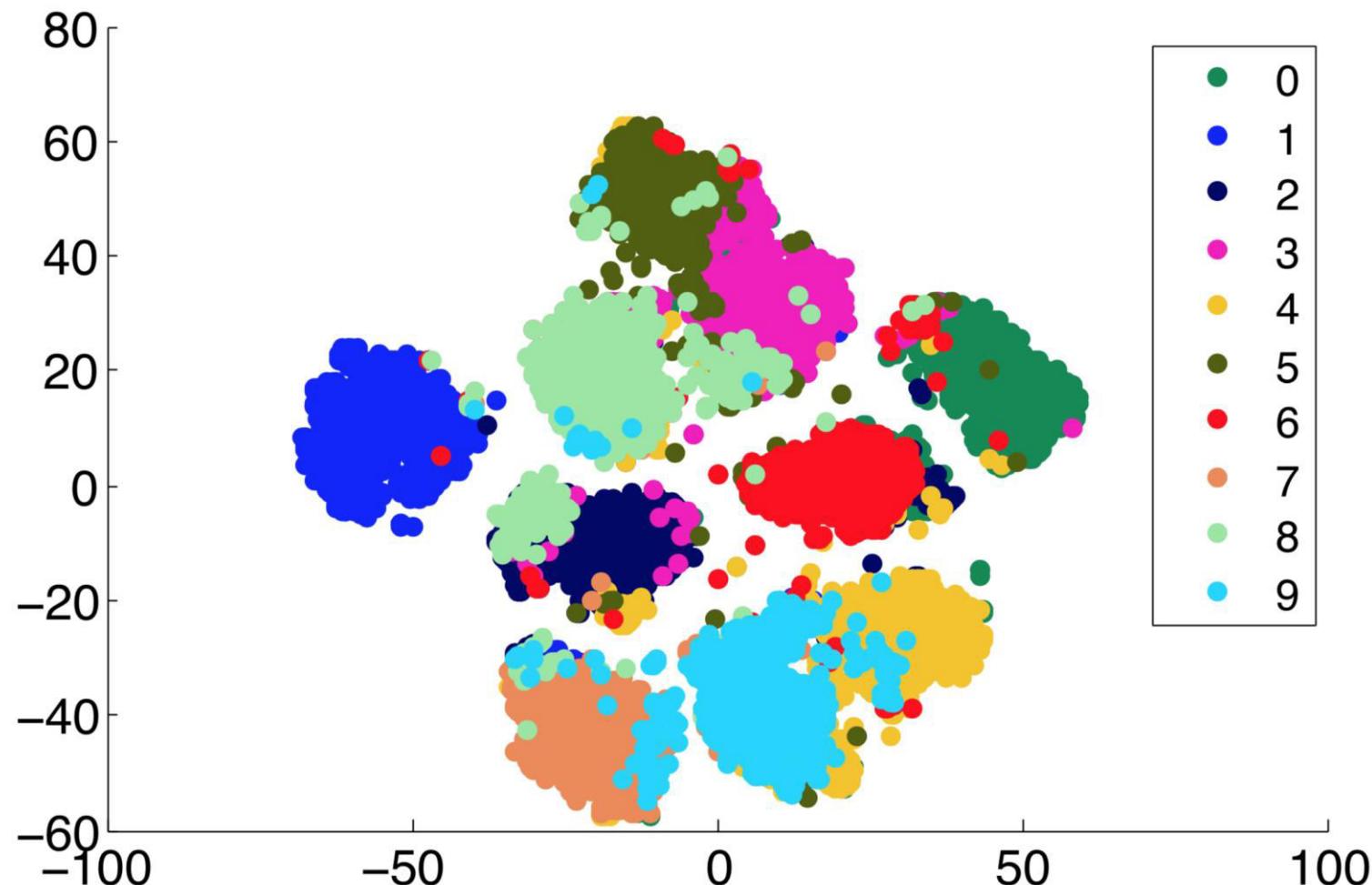
# Învățare nesupervizată

- Avem la dispoziție exemple de obiecte fără etichete
- Exemplu 1: gruparea imaginilor după similaritate



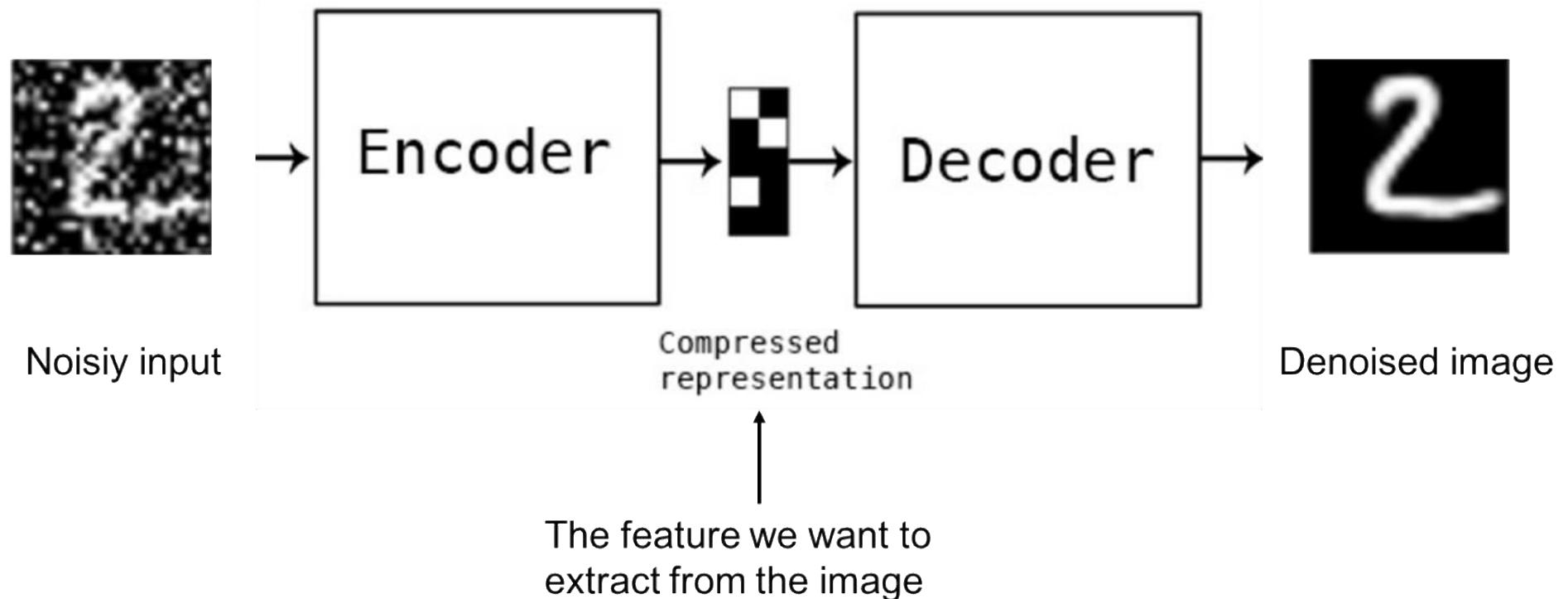
# Învățare nesupervizată

- Exemplu 1: clusterizarea aglomerativă a imaginilor MNIST [Georgescu et al. ICIP2019]



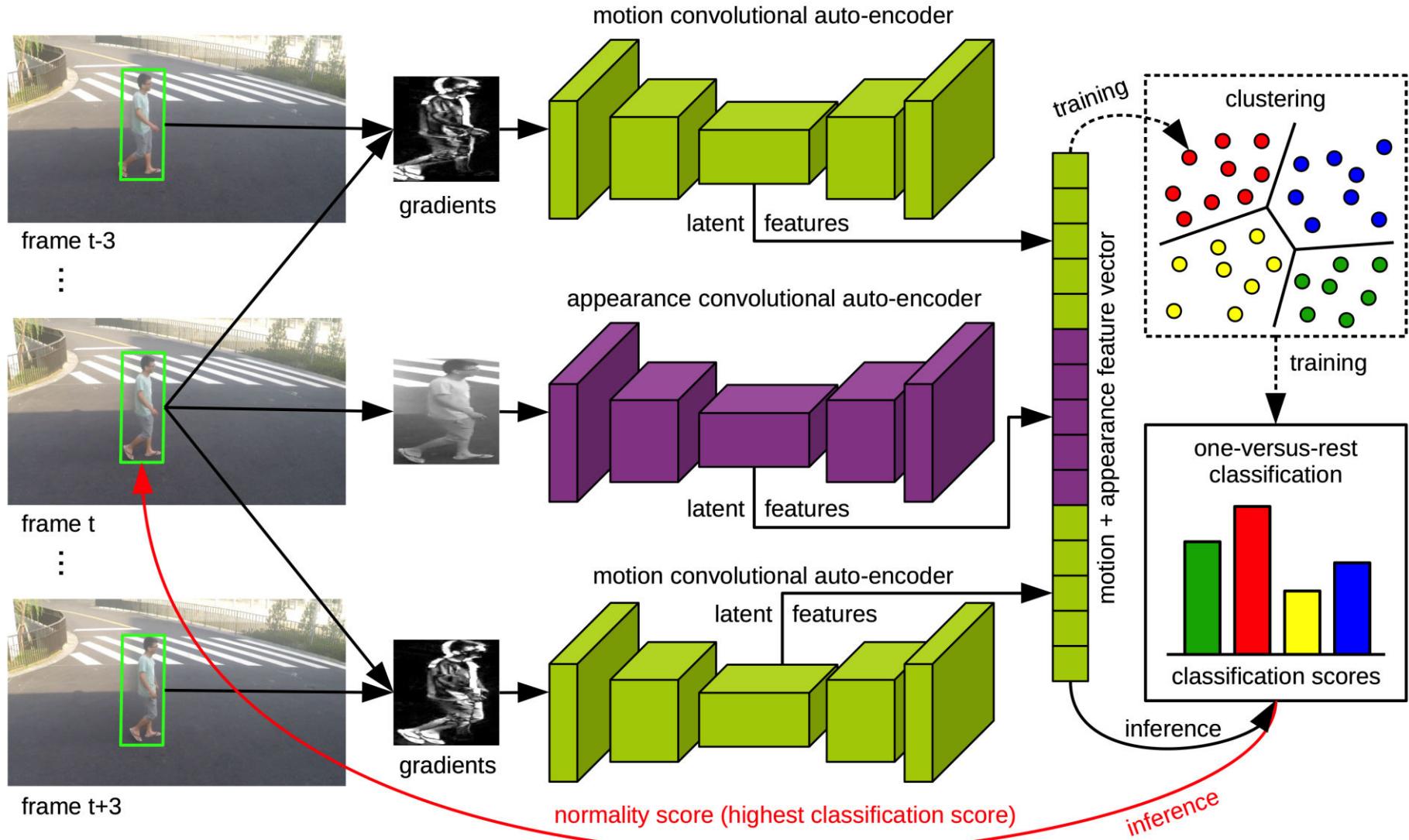
# Învățare nesupervizată

- Exemplu 2: Învățarea de trăsături folosind principiul “bottleneck”



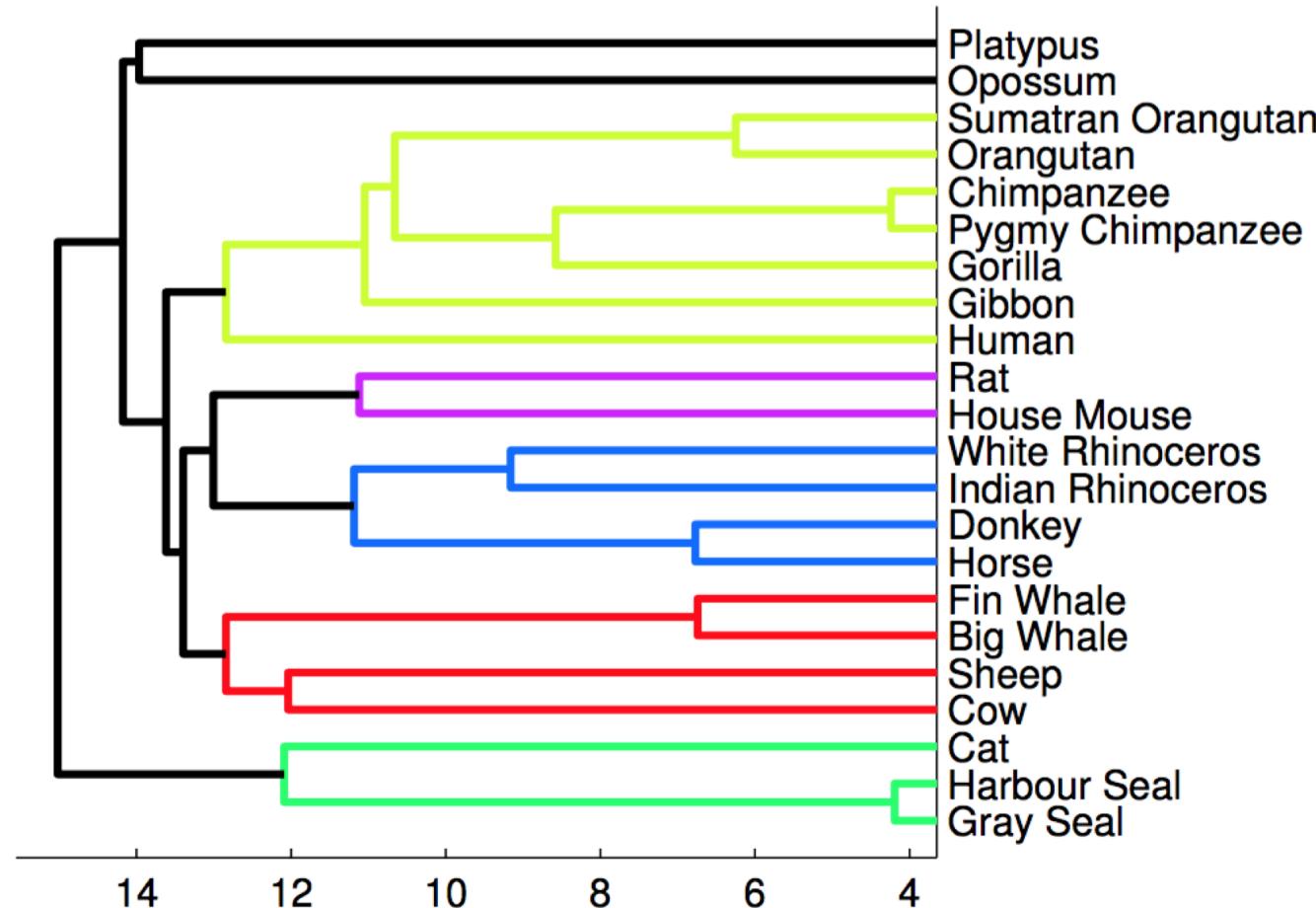
# Învățare nesupervizată

- Exemplu 2: Învățarea de trăsături pentru detectarea evenimentelor anormale [Ionescu et al. CVPR2019]



# Învățare nesupervizată

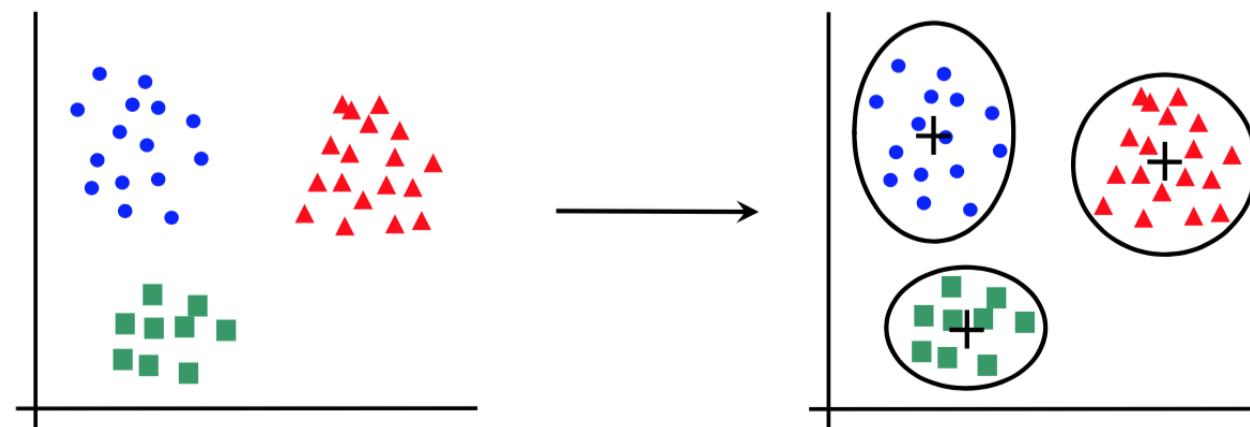
- Exemplu 2: gruparea mamiferelor pe familii, specii, etc.



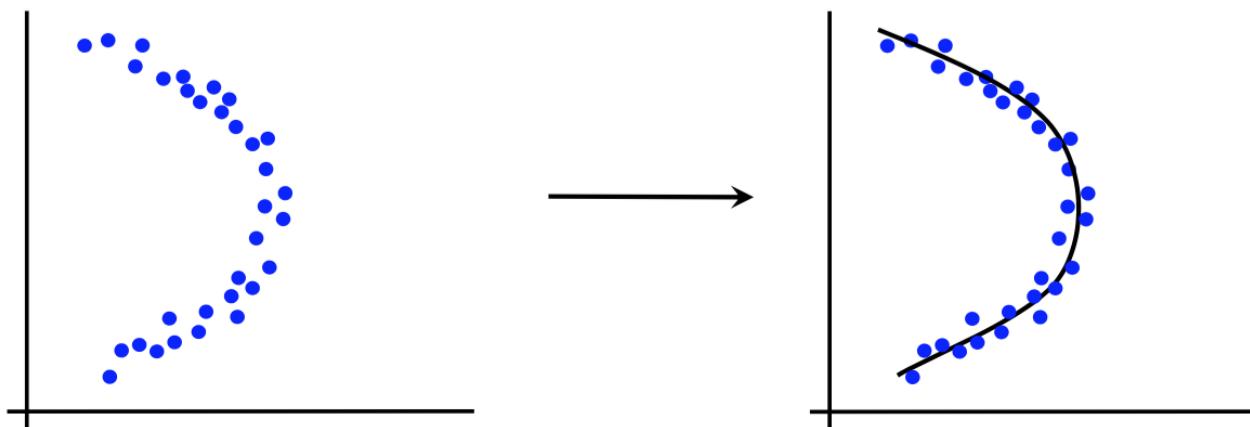
- Generarea arborelui filogenetic pe baza secvențelor ADN

# Formele canonice ale problemelor de învățare nesupervizată

- Grupare (clustering)



- Reducerea dimensiunii



# Modele de învățare nesupervizată

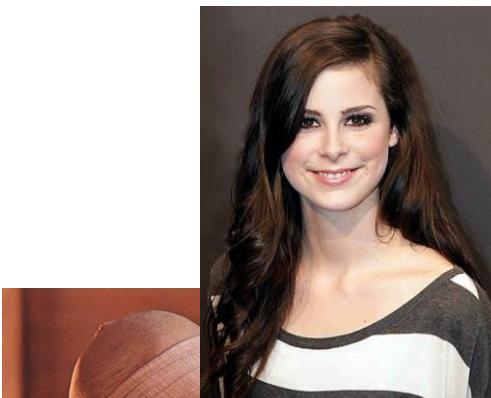
- K-means clustering
- Clustering ierarhic
- Analiza în componente principale
- Altele

# Învățare semi-supervizată

- Avem la dispoziție exemple de obiecte etichetate și exemple de obiecte netichetate
- Exemplu 1: recunoașterea obiectelor din imagini, unele cu eticheta obiectelor conținute



Car



Person



Dog

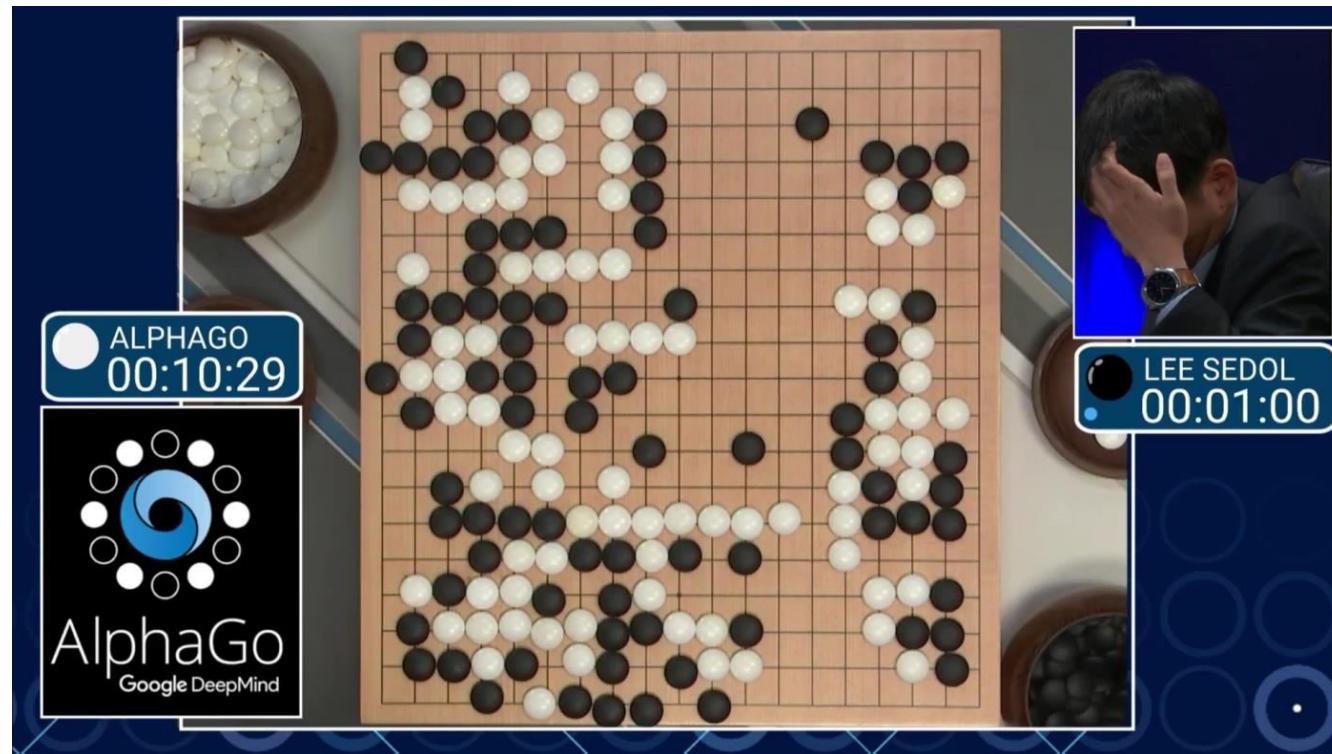


# Învățare ranforsată

- Cu ce diferă această paradigmă de învățare?
- Sistemul învăță comportamentul intelligent pe baza unei recompense (reinforcement signal)
- Recompensa este primită după mai multe acțiuni (nu vine instant)
- Timpul contează (datele sunt secvențiale, nu i.i.d.)
- Acțiunea sistemului influențeză datele

# Învățare ranforsată

- Exemplu 1: Învățarea jocului Go
- recompensă +/- pentru câștigarea/pierderea unui joc



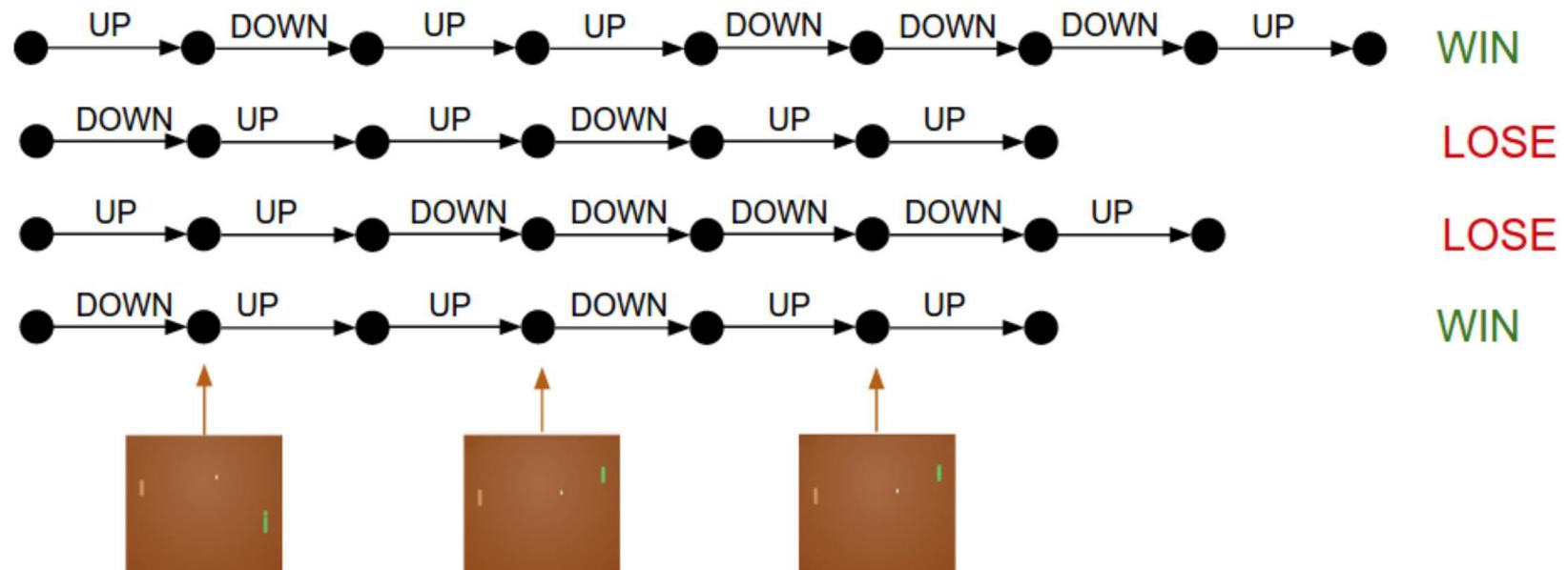
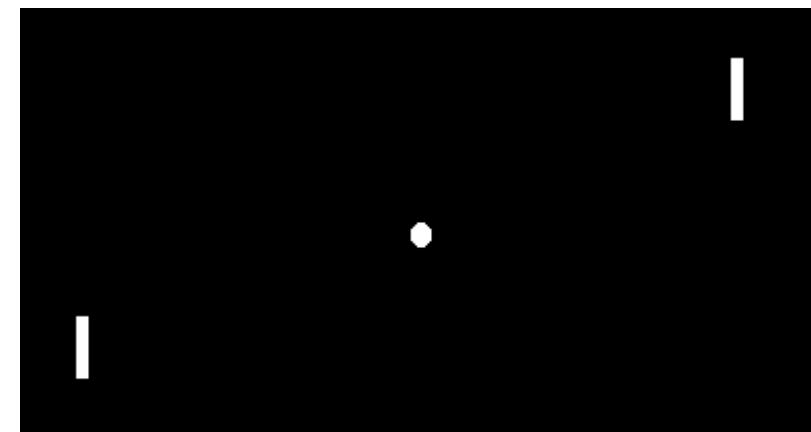
# Învățare ranforsată

- Exemplu 2: Învățarea unui robot să meargă pe bicicletă
- recompensă +/- pentru mișcare înainte/cădere

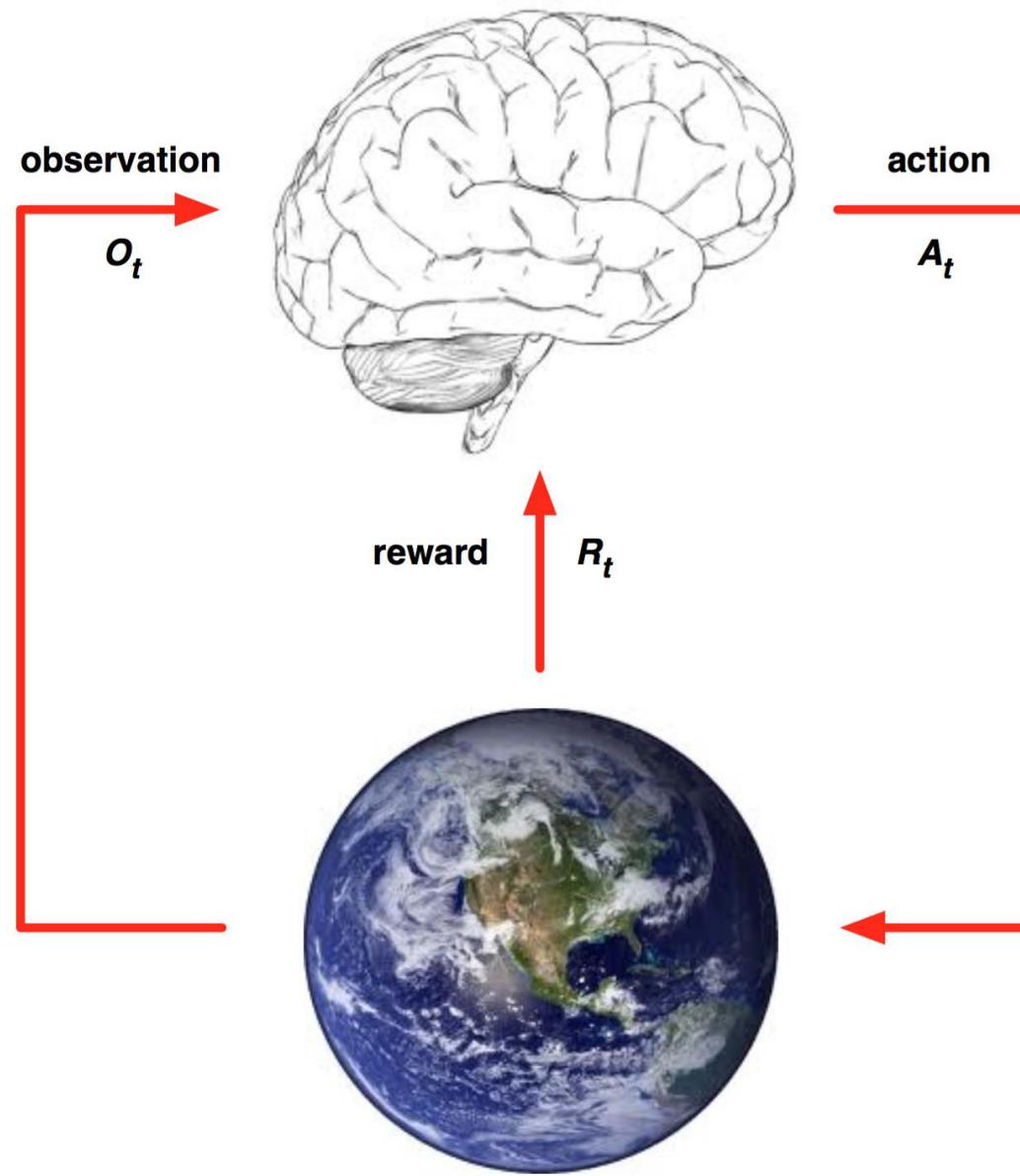


# Învățare ranforsată

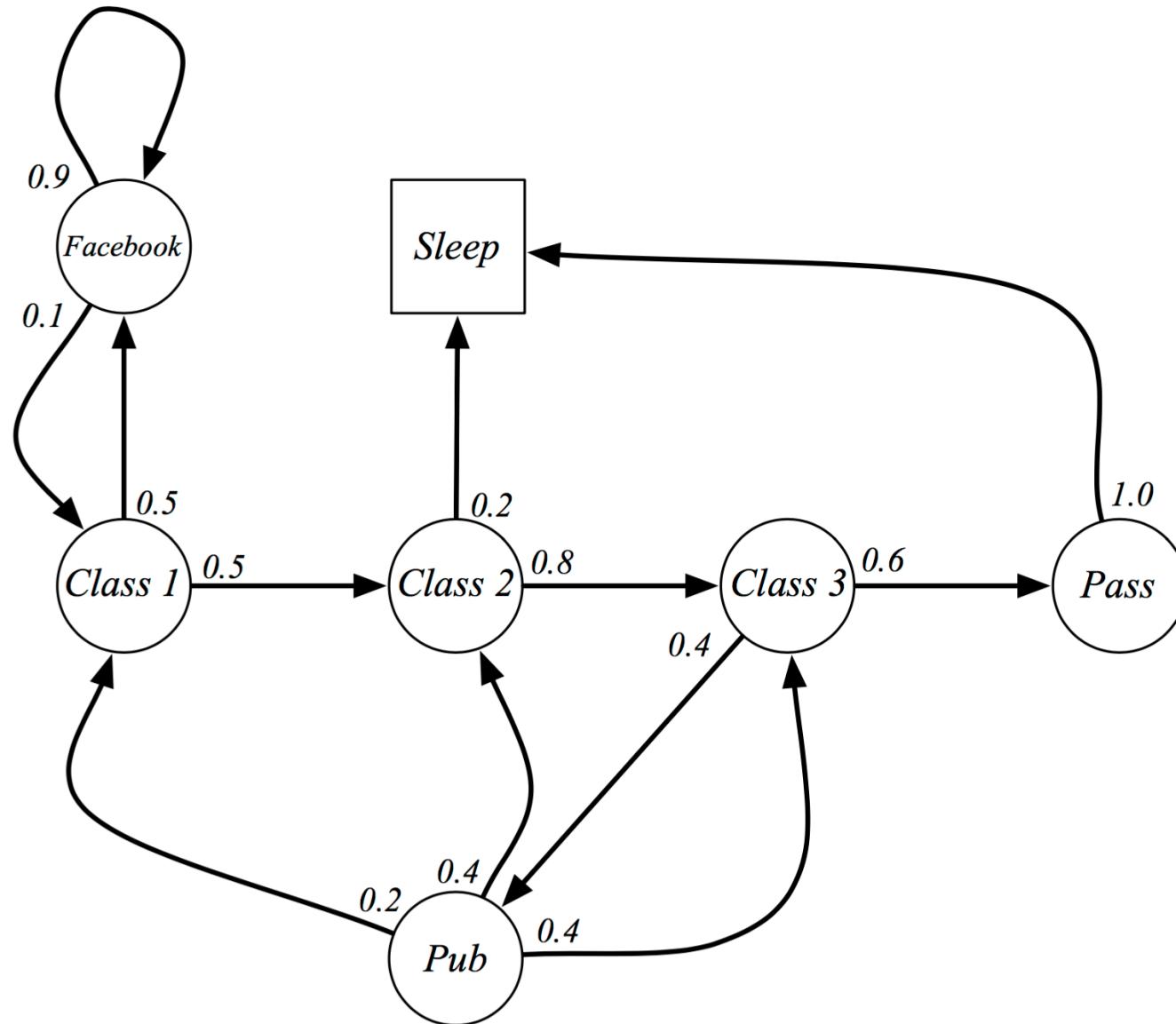
- Exemplu 3: învățarea jocului Pong din pixeli
- recompensă +/- pentru creșterea scorului personal/al adversarului



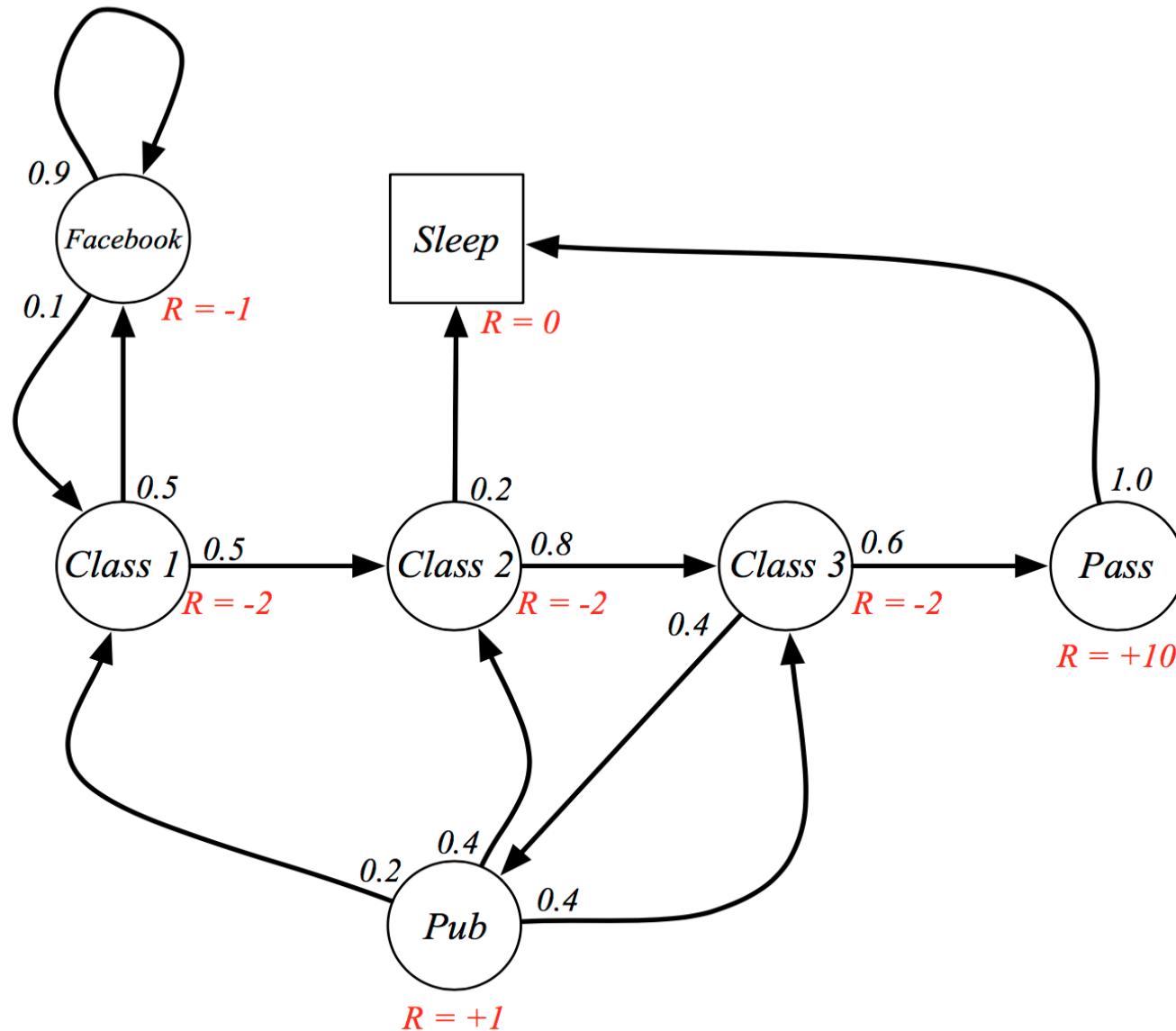
# Paradigma de învățare ranforsată



# Formalizarea cu Procese de Decizie Markov



# Formalizarea cu Procese de Decizie Markov



# Formalizarea cu Procese de Decizie Markov

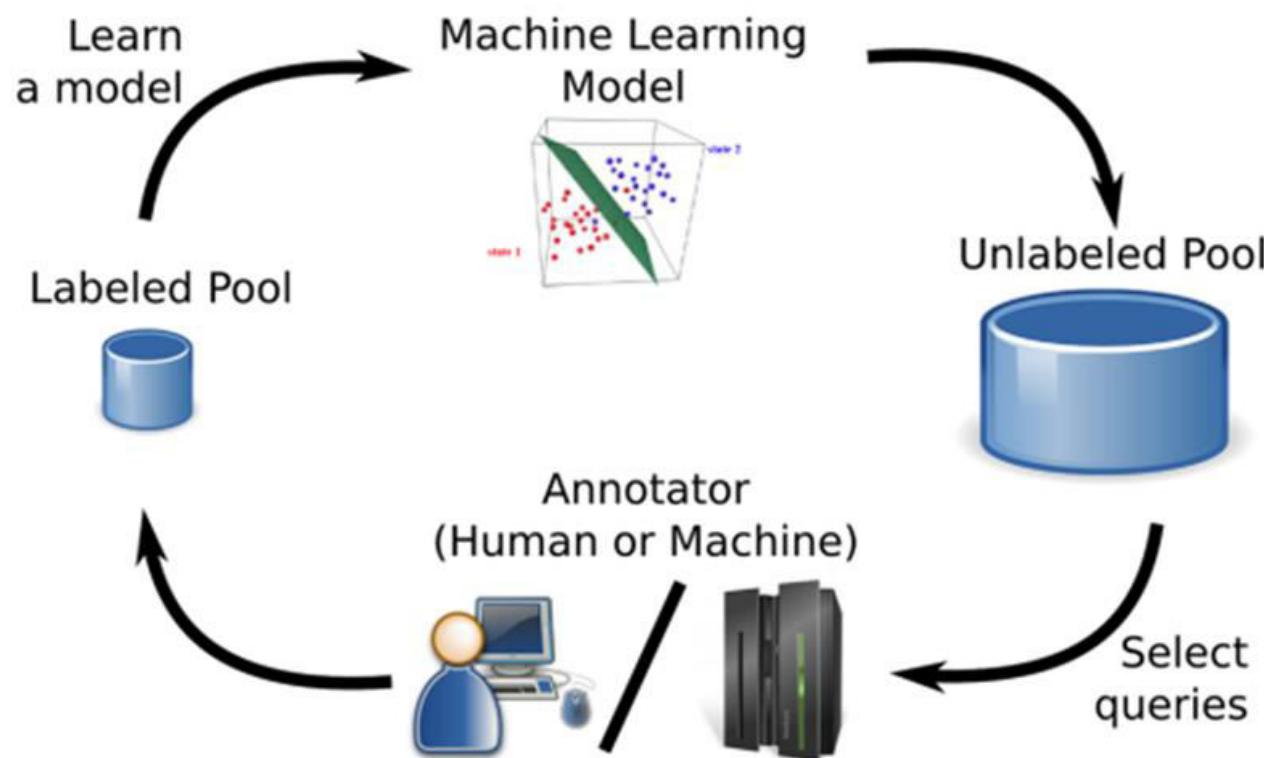
- Soluția bazată pe programare dinamică (grafuri mici) sau aproximare (grafuri mari)
- Scop: selectarea acțiunilor pentru a maximiza recompensa totală finală
- Acțiunile pot avea consecințe pe termen lung
- Sacrificarea unei recompense imediate poate conduce la câștiguri mai mari pe termen lung

# Formalizarea cu Procese de Decizie Markov

- Exemplu AlphaGo:
- Comentator 1: “That’s a very strange move”
- Comentator 2: “I thought it was a mistake”
- But actually, “the move turned the course of the game. AlphaGo went on to win Game Two, and at the post-game press conference, Lee Sedol was in shock.”
- <https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/>

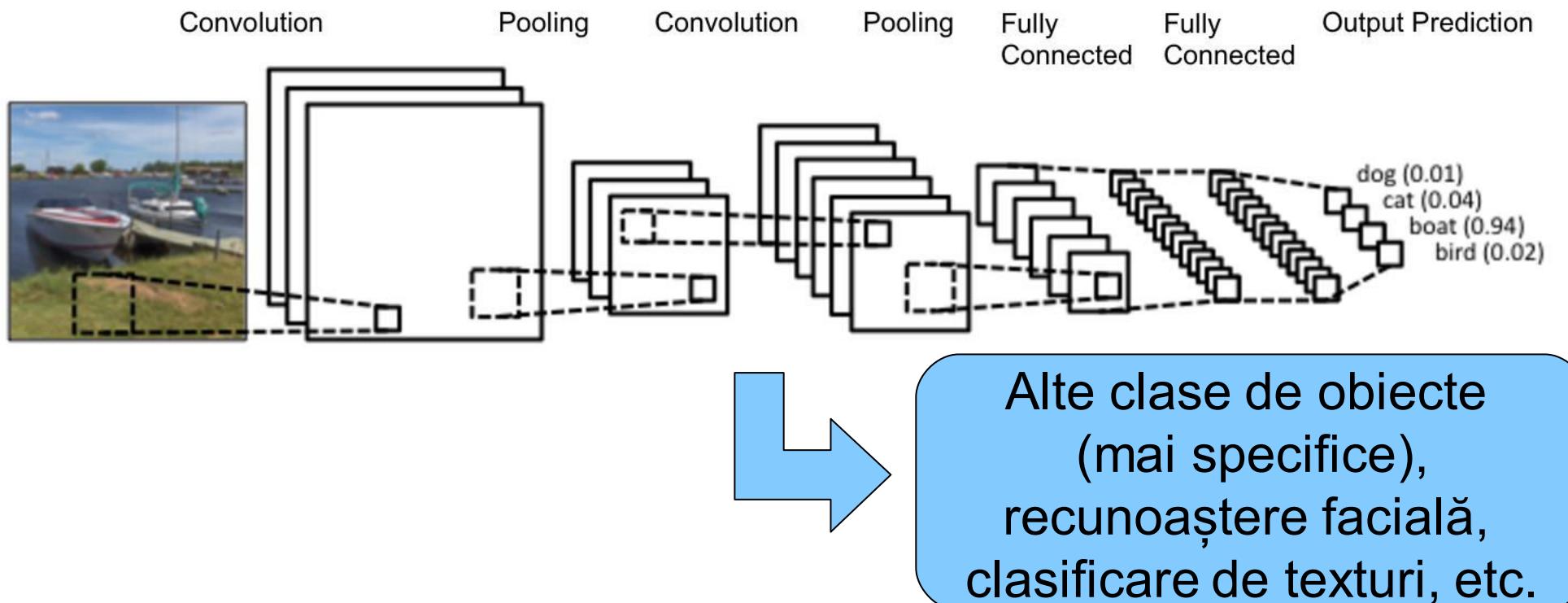
# Învățarea activă

- Având un set mare de exemple netichetate, trebuie să alegem un subset mult mai mic pe care să îl etichetăm pentru a obține un clasificator cât mai bun



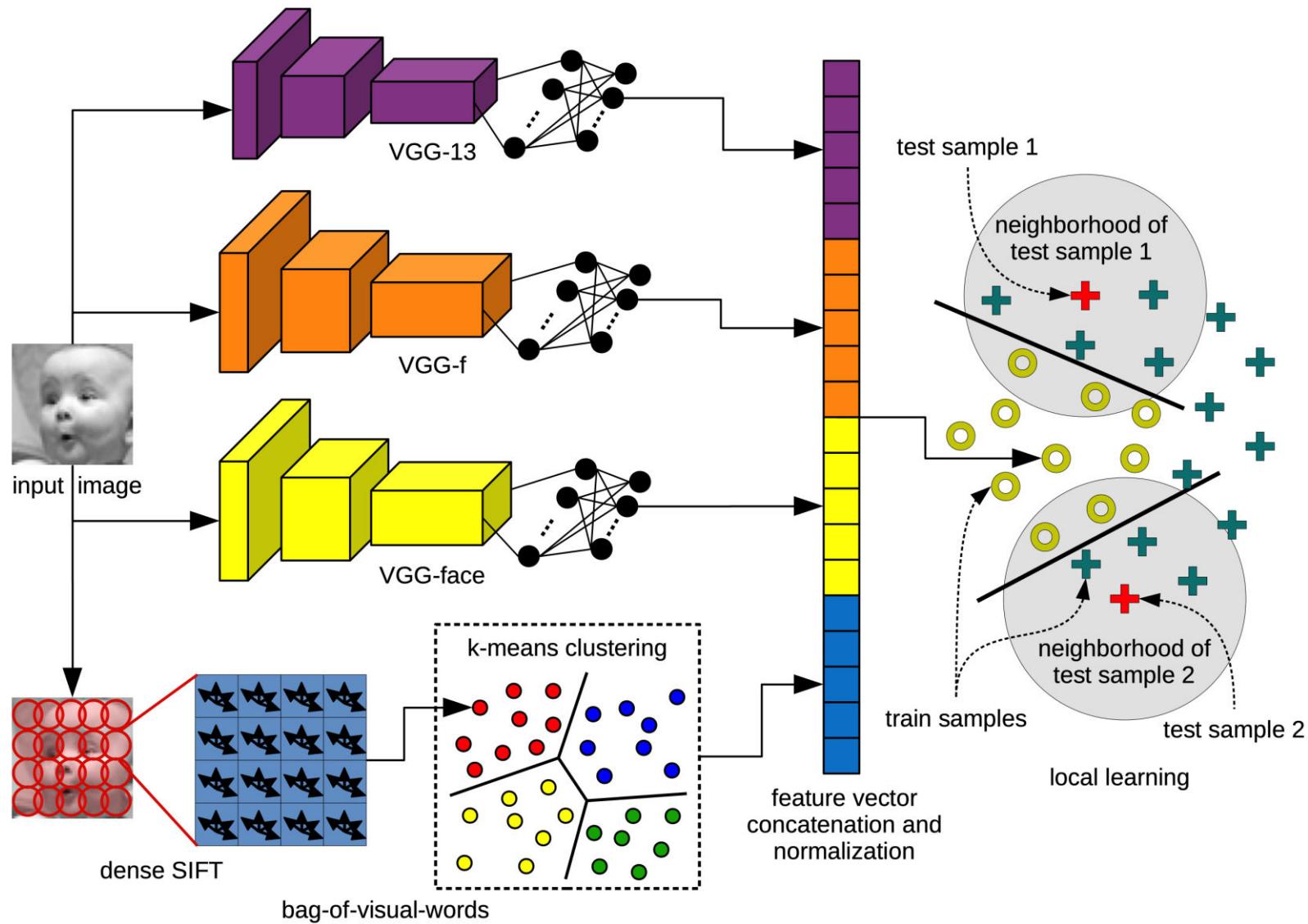
# Învățarea prin transfer

- Pornind la un model antrenat pe un domeniu / o problemă anume, doresc să îl folosesc pentru o altă problemă / alt domeniu
- Exemplu 1: rețele neuronale conoluționale



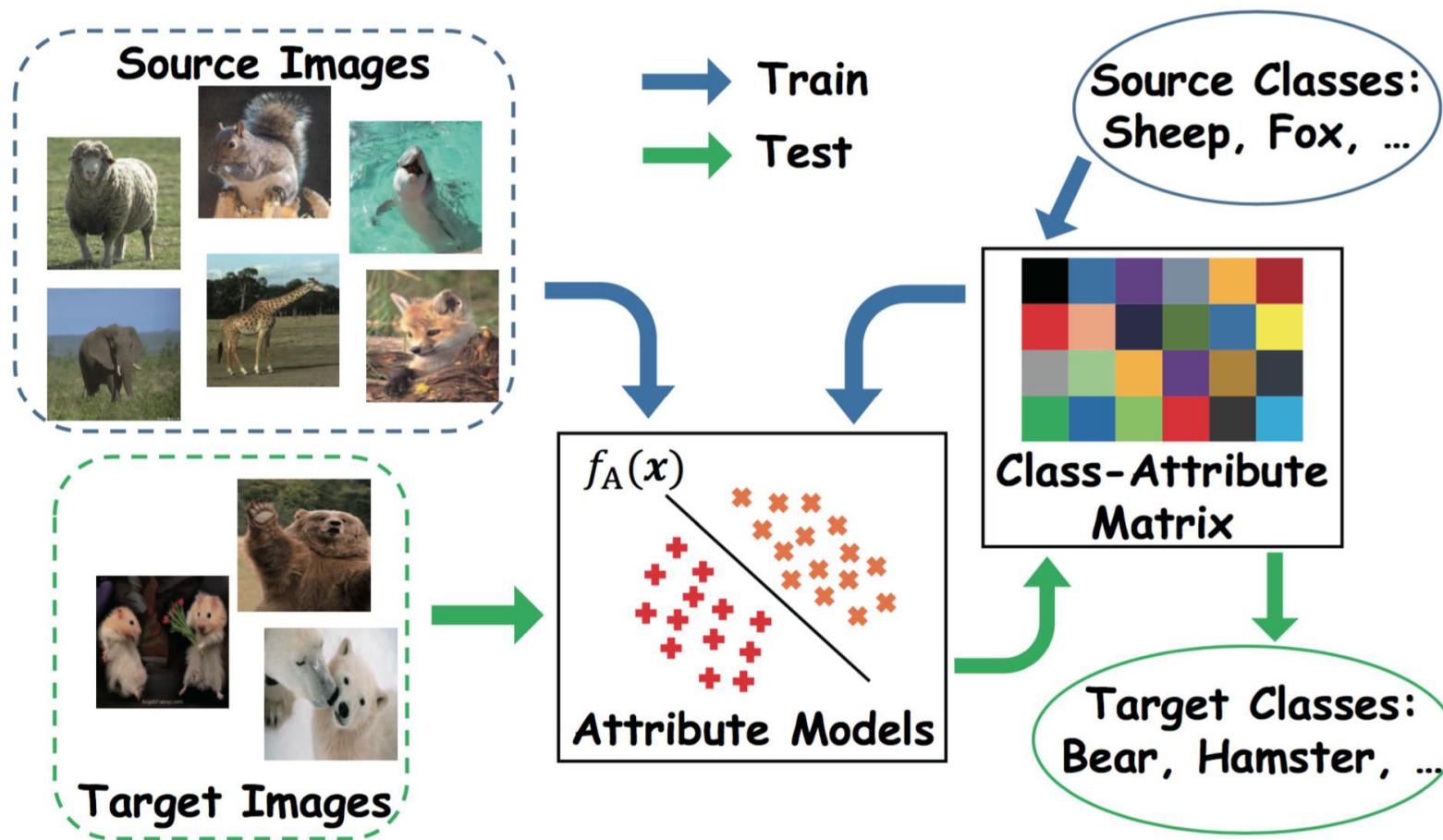
# Învățarea prin transfer

- Exemplu 1: recunoașterea expresiilor faciale [Georgescu et al. Access2019]



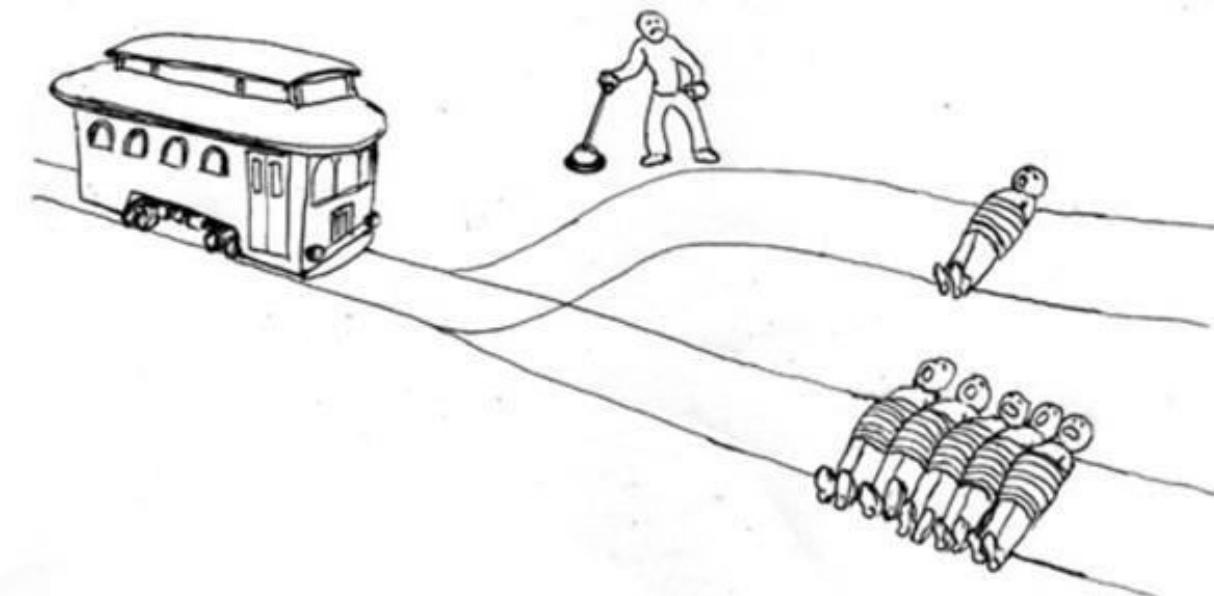
# Învățarea prin transfer

- Exemplu 2: Învățare cu zero-exemple (Zero-shot learning)



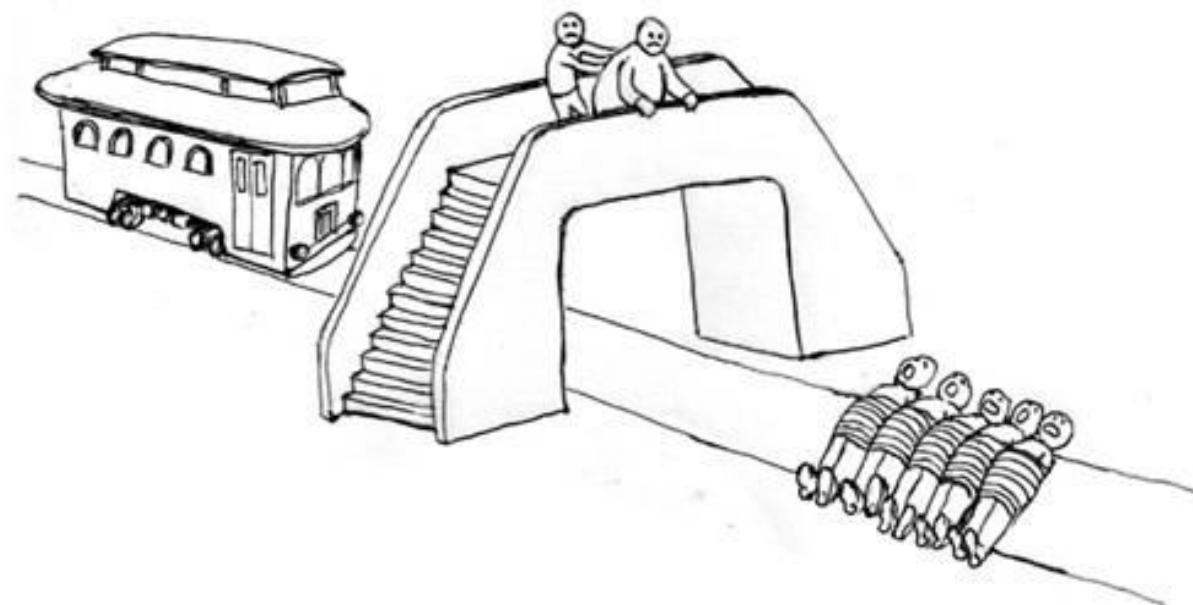
# Multe aplicații interesante, dar...

- Ce este etic și ce nu?
- Trolley paradox



# Multe aplicații interesante, dar...

- Ce este etic și ce nu?
- Trolley paradox



# Multe aplicații interesante, dar...

- Ce este etic și ce nu?
- Trolley paradox
- <http://moralmachine.mit.edu>

# Bibliografie

Springer Series in Statistics

Trevor Hastie  
Robert Tibshirani  
Jerome Friedman

# The Elements of Statistical Learning

Data Mining, Inference, and Prediction

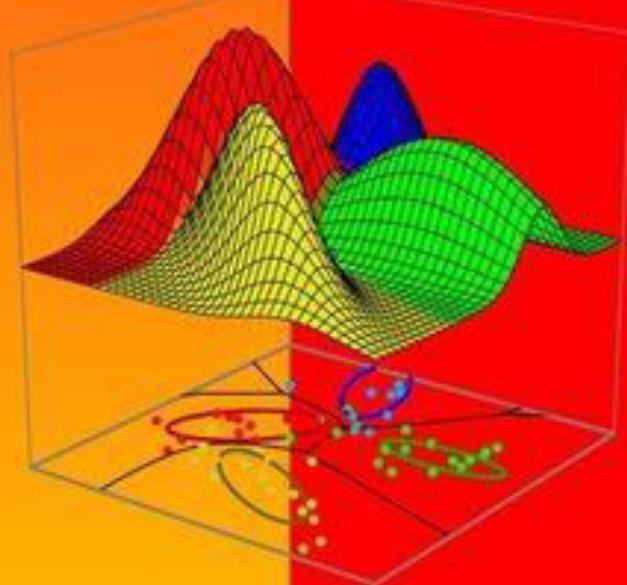
Second Edition



Springer

Richard O. Duda  
Peter E. Hart  
David G. Stork

# Pattern Classification



Second Edition

O'REILLY®

# Hands-On Machine Learning with Scikit-Learn & TensorFlow

CONCEPTS, TOOLS, AND TECHNIQUES  
TO BUILD INTELLIGENT SYSTEMS



powered by  
 jupyter

Aurélien Géron

Advances in Computer Vision and Pattern Recognition



Radu Tudor Ionescu  
Marius Popescu

# Knowledge Transfer between Computer Vision and Text Mining

Similarity-based Learning Approaches

 Springer