

# Uczenie maszynowe wprowadzenie

---

Bartłomiej Śnieżyński  
AGH, Katedra Informatyki

bartlomiej.sniezynski@agh.edu.pl

20.03.2018

## Plan

- Wprowadzenie
- Definicje, formy wiedzy
- Obszary zastosowań, przykłady
- Systematyka, metody uczenia
- Weka, Knime
- Historia
- Nasze projekty
- Konsekwencje społeczne

# Wstęp

- Ludzie **gromadzą dane** od co najmniej **5300** lat, kiedy to Sumerowie użyli glinianych tabliczek do przechowywania danych podatkowych.
- Od skonstruowania **komputera**, ilość danych **rośnie eksponencjalnie** i zaczyna przekraczać możliwości przetworzenia.
- Obecnie tworzone jest ok. **2,5 EB** (exabyte) danych dziennie.
  - $1 \text{ EB} = 10^{18} \text{ B} = 1000000000000000000 \text{ B} = 1000 \text{ PB} = 1 \text{ mln TB} = 1 \text{ mld GB}$
  - $1 \text{ EB} = 100.000 \times \text{teksty Biblioteki Kongresu}$
  - $5 \text{ EB} \approx \text{wszystkie słowa wypowiedziane przez ludzi}$

3/16

# Problem

- Ilość danych szybko rośnie.
- Możliwości przetworzenia przez ludzi są mniej więcej stałe.
- Jednocześnie: **pragnienie wiedzy** z danych.
- Zagadnienie stworzenia efektywnych narzędzi do przetwarzania danych w wiedzę – jedno z wyzwań 21 wieku.

4/16

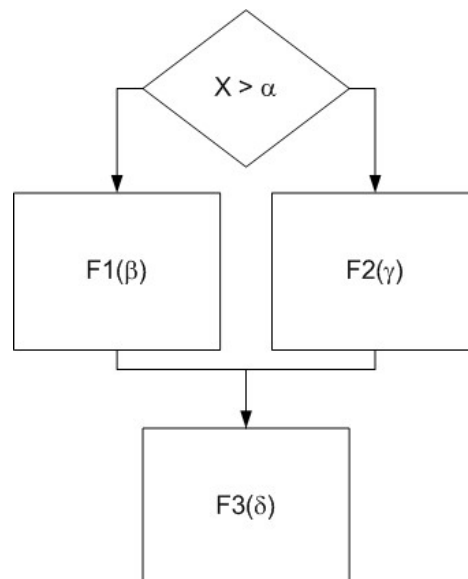
# Definicja uczenia maszynowego

- Uczenie to ***autonomiczna zmiana*** w systemie, zachodząca na podstawie doświadczeń, która prowadzi do ***poprawy*** jakości działania (Cichosz).
- Program się uczy z doświadczenia  $E$  dla zadań  $T$  i miary jakości  $P$  jeśli jego efektywność w zadaniach z  $T$  mierzona  $P$  wzrasta z doświadczeniem  $E$  (Mitchell).

5/16

## Formy wiedzy

- Formy:
  - wiedza (opis obiektu, sytuacji),
  - umiejętność (procedura, strategia).
- Uczenie  $\approx$  dobieranie parametrów.
- Parametry – hipoteza (element przestrzeni hipotez).



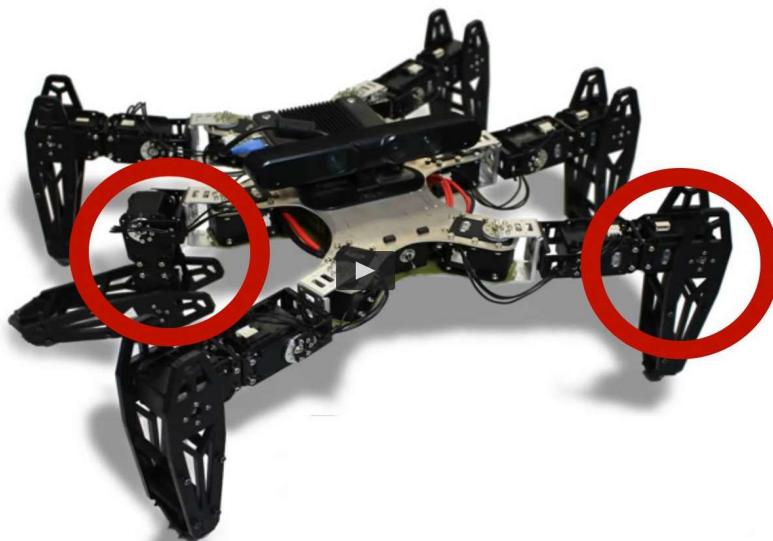
6/16

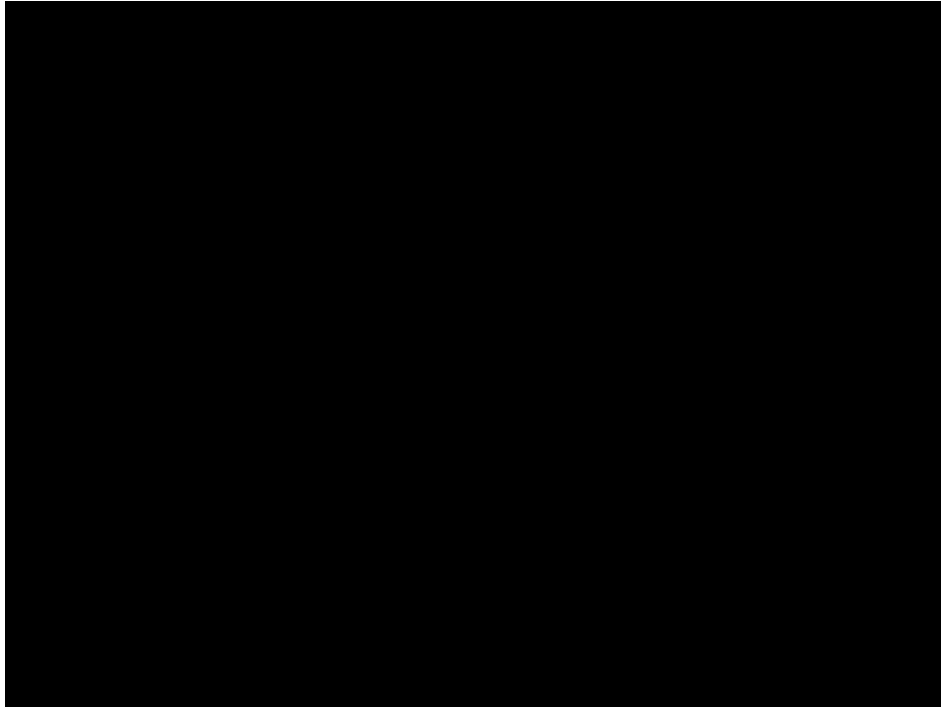
# Motywacja

- Dla naprawdę **złożonych zadań** trudno jest sformułować wprost gotowe programy.
- Często **zbiory** dostępnych **danych** są zbyt **duże** i skomplikowane.
- Złożone środowiska są trudne do opisu, często nie posiadają wystarczających **modeli teoretycznych** albo ich uzyskanie jest bardzo kosztowne.
- „**Ręcznie**” zakodowane programy dla takich środowisk byłyby **mało wiarygodne**.
- Inteligentne systemy powinny być w maksymalnym stopniu **autonomiczne**, co nie jest możliwe bez adaptacyjności.

7/16

## Potrzeba autonomii





## Przykłady

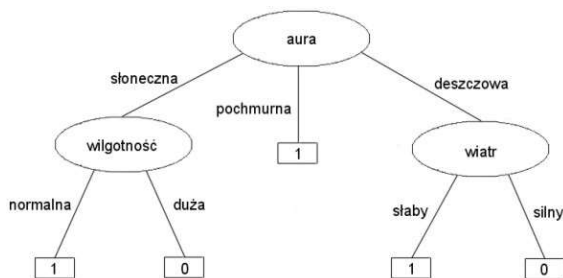
- Uczenie się gry w warcaby ( $T$ : granie w warcaby,  $P$ : procent wygranych partii,  $E$ : praktyka, np. granie przeciw sobie).
  - A. L. Samuel, 1959 – program grał sam z sobą.
- Uczenie się rozpoznawania chorób na podstawie symptomów ( $T$ : diagnozowanie,  $P$ : procent poprawnych diagnoz,  $E$ : baza przypadków z diagnozami).
  - IBM Watson Health – np. wspomaganie pracy onkologów.
- Uczenie się kierowania samochodem na podstawie obserwacji i naśladowania instruktora ( $T$ : kierowanie pojazdem,  $P$ : średni dystans bez błędu,  $E$ : sekwencja obserwacji i reakcji kierowcy).
  - nVidia – Deep learning Self-Driving Car.
- Inne:
  - Wykrywanie spamu.
  - Uczenie się klasyfikowania tekstów do grup tematycznych.
  - Uczenie się odnajdywania drogi w nieznanym środowisku.
  - Uczenie się zależności funkcyjnych pomiędzy danymi obserwacyjnymi.

# Systematyka

- Reprezentacja wiedzy
  - Symboliczna
    - Drzewa decyzyjne
    - Reguły
    - Sieci Bayesowskie
  - Niesymboliczna
    - Sieci neuronowe
    - Maszyny wektorów nośnych
- Zadanie
  - Klasyfikacja
  - Aproksymacja funkcji
  - Grupowanie
  - Rekomendacja
- Dane uczące
  - Uczenie pod nadzorem
    - Przykłady
    - Odpowiedzi na pytania
    - Eksperymentowanie
  - Uczenie ze wzmocnieniem
  - Uczenie bez nadzoru
- Mechanizmy uczenia
  - Indukcja
    - sekwencyjne pokrywanie
  - Przypisywanie zasług
    - uczenie ze wzmocnieniem
  - Metody statystyczne
    - sieci bayesowskie, lasy losowe, boosting
  - Optymalizacja
    - metoda gradientu prostego
  - Analogia
    - Metoda najbliższych sąsiadów

11/16

## Drzewa decyzyjne, reguły

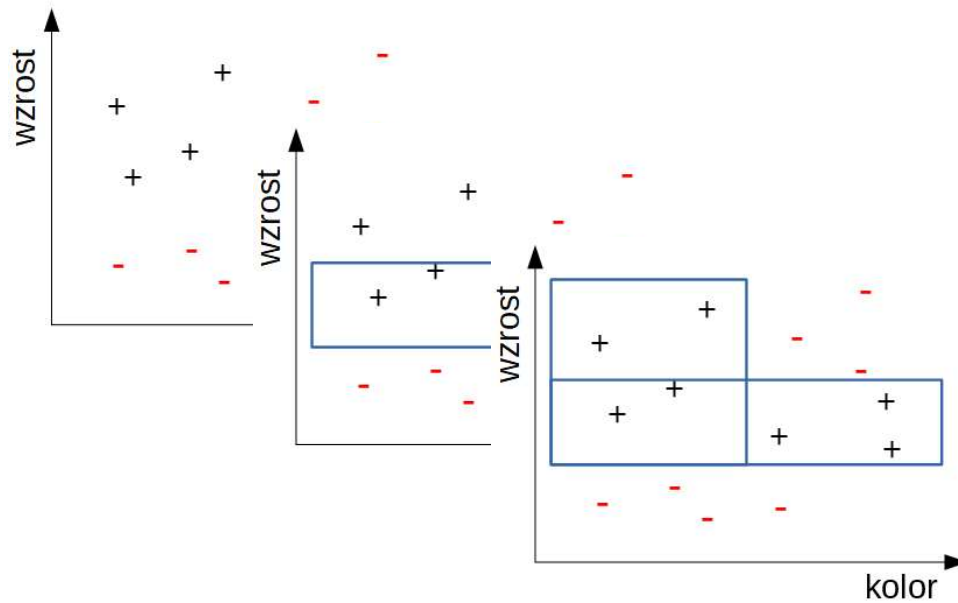


- **Jeśli** aura = pochmurna  
**To tak**
- **Jeśli** aura = słoneczna i wilgotność = normalna  
**To tak**
- **Jeśli** aura = słoneczna i wilgotność = duża  
**To nie**

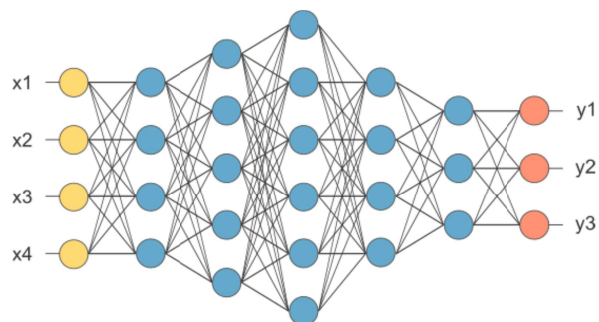
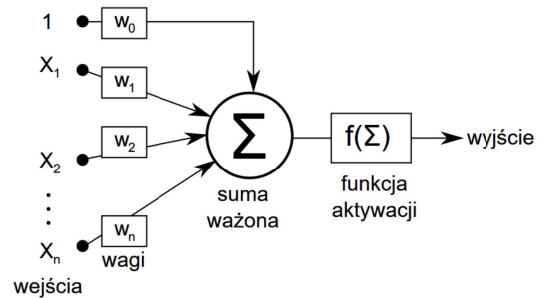
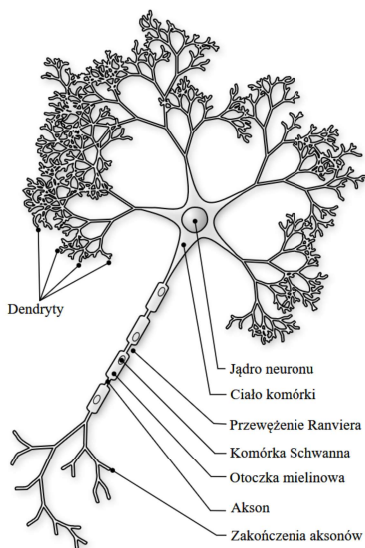
iAQ

<https://www.mli.gmu.edu>

# Sekwencyjne pokrywanie



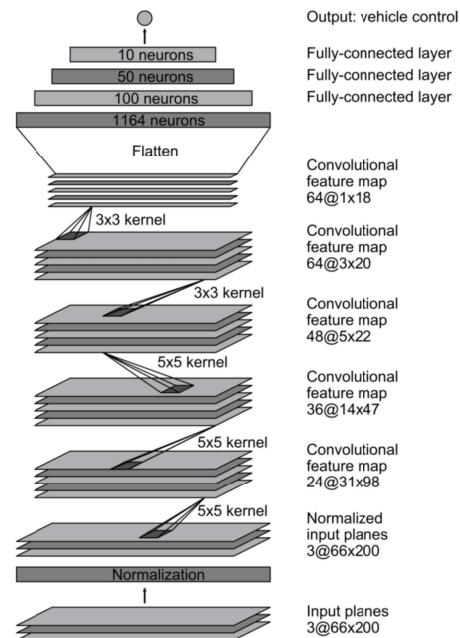
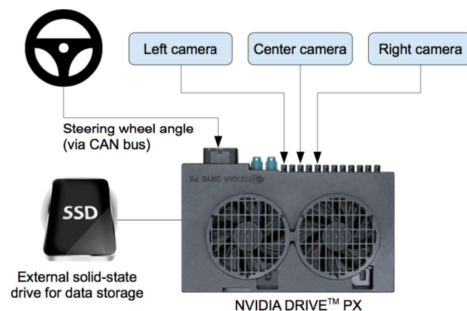
# Sieci neuronowe



<https://playground.tensorflow.org>

# Deep learning, big data

- Architektura głębokiej sieci dla autonomicznych samochodów (nVidia)
- 27 mln połączeń
- 72 godziny nagrań, 10FPS

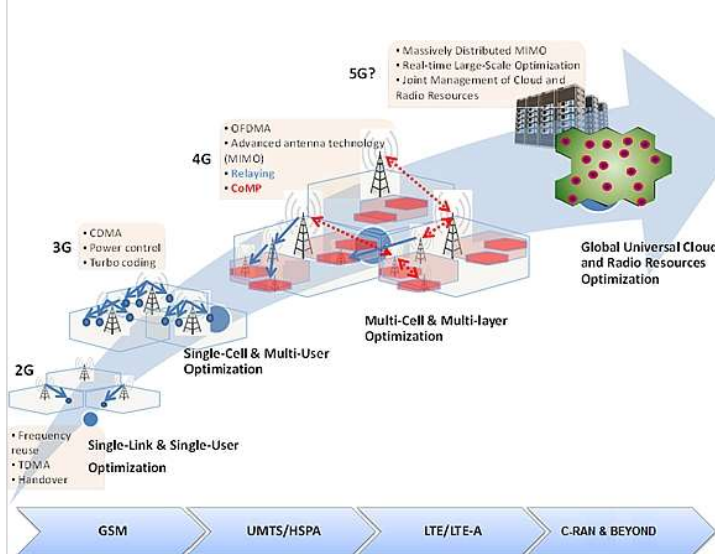


## Historia

- Początki (50, 60) – wczesne sieci neuronowe i metody oparte na prawdopodobieństwie.
- “Living in a Dungeon” (60, 70) – uczenie zbyt skąplikowane, zajmowało się nim mało osób.
- Rozkwit (80). Punkt zwrotny – konferencja w CMU, 1980 oraz książki Machine Learning: An Artificial Intelligence Approach
- Zastosowania (późne 80, 90) – Nacisk na eksperymentalne aplikacje i ulepszanie algorytmów, kilka nowych kierunków, np. uczenie ze wzmocnieniem, maszyny wektorów nośnych, indukcyjne programowanie logiczne.
- Stopniowy postęp (90 do teraz) – Stopniowe ulepszanie algorytmów.
- Sieci głębokie (10) – Rozwój sprzętu pozwala na przetwarzanie dużych ilości danych (Big Data)
- 2016: AlphaGo – program wygrywa 4:1 z profesjonalnym graczem w GO (monte carlo tree search + 2 x deep NN kierujące przeszukiwaniem)

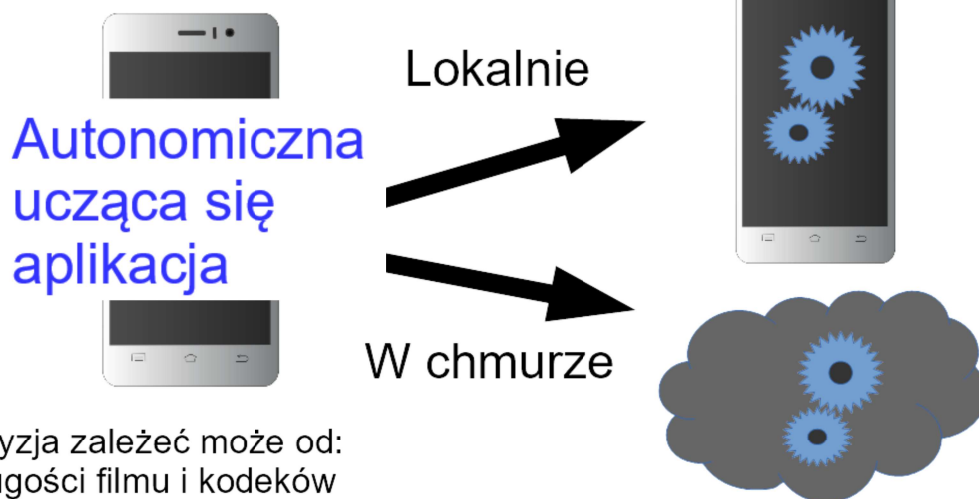


## Adaptacyjne planowanie zasobów w chmurach obliczeniowych (Samsung)



- Telefonia **5G** opiera się na chmurach, które mają **ograniczone zasoby**
- Potrzebne są **plany rezerwacji**
- **Uczenie maszynowe** stosowane do **adaptacji planów**

## Wybór miejsca wykonywania zadań



Decyzja zależy może od:

- długości filmu i kodeków
- rodzaju połączenia z siecią
- wydajności urządzenia
- stanu baterii

# Podsumowanie

- Można **tworzyć programy automatycznie**, na podstawie danych, przykładów, nie trzeba ich kodować ręcznie.
- **Proces rozwoju przyspiesza.**
  - Dekada obecnie = cały XVIII wiek = 0-XAD = 10000BC-0
- **Uczenie może być uruchamiane równolegle**
  - Eve – robot naukowiec w dziedzinie biologii molekularnej, sprawdza 10.000 składników dziennie. Pomógł odkryć lek na malarię.
- **Rośnie moc obliczeniowa.**
- **Komputery kwantowe!**
  - Analogia:  
komputer może przeczytać  
po kolei wszystkie książki,  
a komputer kwantowy  
przeczyta wszystkie na raz!



# Konsekwencje społeczne

- Foxconn zwalnia 60.000 tys. pracowników -> roboty (a ich zarobki < 5\$/dzień)
- Transport/logistyka
  - Roboty pracują w magazynach Amazona, obsługują transport kontenerów w porcie LA
  - Uber w Pittsburghu
  - Autonomiczna ciężarówka jeździ od 4 lat po USA
  - 4 00 mln ludzi może stracić pracę!
- Emily Howell – algorytm uczący się, który gra na pianinie
- Ross – AI prawnik – zatrudniony w Baker & Hostetler USA (zajmuje się upadłościami)
- Watson (IBM) w podwójnie ślepej próbie osiągnął zgodność 90% diagnoz z konsylium onkologicznym



## Ważna edukacja

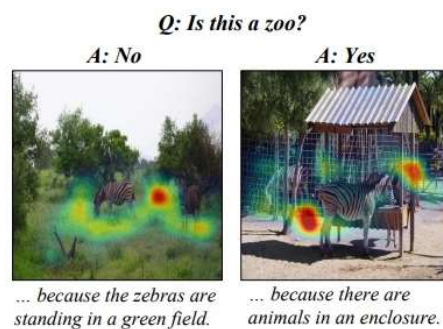
- Trzeba postawić na to co ludzkie, niepowtarzalne:
  - Kreatywność
  - Miłość
  - Empatia
  - Umiejętność krytycznego myślenia
  - Dialog
- W powtarzalnych czynnościach komputery mają przewagę.

## Społeczeństwo modeli (Pedro Domingos)

- Aktualnie firmy mają **fragmentaryczne modele** (Netflix, Amazon, Google, FB)
- Niedługo będą **kompletne modele**, które będą nam pomagać – od **zakupów**, przez **podróże**, **zawieranie znajomości** (czyją wiadomość pokazać na FB?) – trudno się będzie bez tego obejść.
- Społeczeństwo modeli:
  - modele będą wchodzić w **interakcje** aby ustalić co będzie w przyszłości,
  - po kliknięciu w LinkedIn model może przejść rozmowę kwalifikacyjną na wszystkie interesujące pozycje,
  - W tym samym czasie model może szukać samochodu i targować się ze sprzedawcą,
  - Model może iść na tysiąc randek aby wybrać najbardziej obiecującą osobę.

# Jak się do tego przygotować?

- **Uwaga! Model powinien być pod naszą kontrolą, a nie kogoś kto ma sprzeczne interesy**
- Google – chce być „trzecią połową mózgu”, ale czy chcemy by ona zajmowała się dobieraniem ogłoszeń???
- Może trzeba stworzyć **banki modeli**, analogicznie jak zwykłe banki?
- Może potrzebujemy odpowiednika związków zawodowych aby reprezentowały **nasze interesy** przed firmami?
- Potrzebna możliwość interakcji z modelem:
  - definiowanie celi, sprawdzanie uzasadnienia sugestii
  - wskazywanie gdzie popełniono błąd i dlaczego (**nie czarne skrzynki!**)
- **Nowy model społeczeństwa?**
  - co dozwolone, co nie, aby wszyscy odnosili korzyści



## Wizja przyszłości (Ray Kurzweil)

- Coraz szybsze odkrycia
- Rozwiązanie trudnych problemów
  - choroby, głód, energia
- Nieśmiertelność
- Porzucenie ciał, pozostaną jedynie umysły
- Jeden połączony umysł dla całej ludzkości?