Hogyan tanítsunk gépeket tanulni?

Modern fizikai eredmények alkalmazása a gépi tanulás területén

Pósfay Péter

Wigner FK, Komputációs tudományok osztálya

A gépi tanulás ma

A gépi tanulás rövid története

- **1944**: A neurális hálók ötlete, McCullough – Pitts [1]
- **1957**: Az első tanítható neurális háló (**Perceptron**), Cornell University F. Rosenblatt
- **1959**: Minsky, Papert: Perceptronok gyakorlati célokra nem alkmasak
- **1969**-ig: az MIT-n aktívan kutatott terület a Neurális hálók, majd elhal
- **1980**-as évek: 2-3 rétegű hálók, ezekre már nem igazak Minsky és Papert kritikái, de nincsenek elég erős számítógépek, hogy növeljék a teljesítmény
- **1985**: tanítási módszerek fejlődése, Rumelhart, Hinton, and Williams újra felfedezik a "**backpropagation**"-t (Paul Werbos pszichologiai inspiráció alapján 1974) [2]

A modellek fejlesztését gyakran inspirálta a pszichológia és a neurológia

Első machine learning "tél" kezdete

Második machine learning "tél" kezdete

Machine learning tél: a neurális hálók fejlesztésének üteme visszaesik. Általában a technológia túlrekálmozása miatti túlzó elvárásokat nem tudják teljesíteni és a befektetők elfordulnak.

"Self Driving cars are less than a year away." – Elon Musk, 2015

A gépi tanulás rövid története

- 1989: LeCun, Bell Laboratóriumok, backpropagation demonstrációja, kézírott számjegyeket felismerő rendszer
- 1992: Cortes és Vapnik "**Support vector machines**", szebb matek jobban érthetőek, mint a neurális hálók [3]
- 1997: **LSTM** (rövid –hoszsú távú memória modellek) rekurrens neurális hálók fejlesztése Hochreiter, Schmidhuber [4] (nyelv felsimerő, fordító rendszerek alapjai)
- 1999: neurális hálók a GPU-knak köszönhetően kompetitívvé válnak a Support vector machine-okkal szemben
- 2012: DEEP Learning, ALexNet G. Hinton képfelismerő neurális háló, Sokkal több rétegű hálók megjelenése (10-50) [5]
- Ma: Big Data és mély neurális hálók korát éljük, ahol egyre komplexebb feladatok oldhatóak meg neurális hálókkal, akár társadalmi problémákat okozva:
  Esszé generátorok, AI festés

Házi feladat:
https://www.wombo.art/create

Prompt:
Cat reading a book
Comic style
10 másodperc alatt !

- GPU teljesítmény nő
- BIG DATA
- Cloud Computing

Modern gépi tanulás főbb elemei

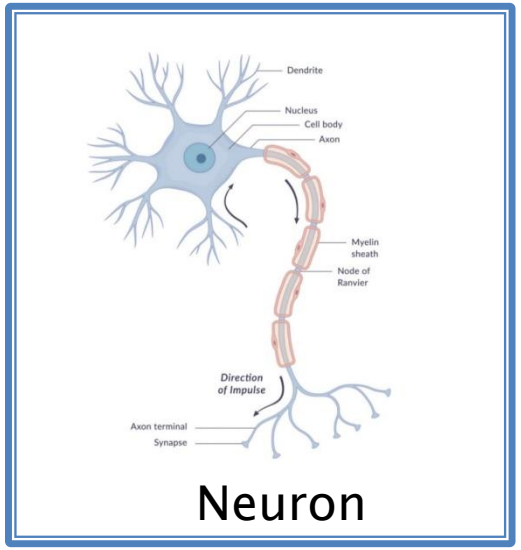
MÉLY
(DEEP LEARNING)

NEURÁLIS HÁLÓK

TANÍTÓ HALMAZOK

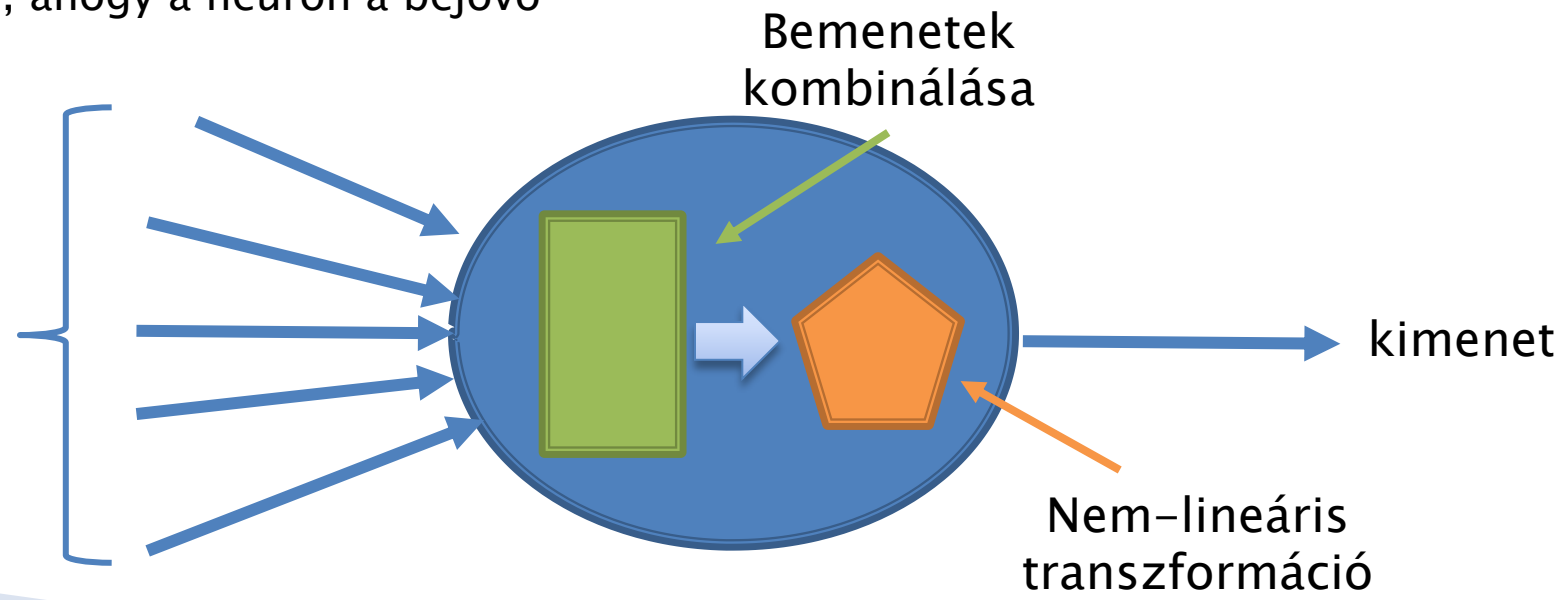
Neurális háló

- Az idegsejtek működése által inspirált komputációs hálózat
- “Neuronokból” áll, ami egy számítási egység ami kombinálja a bemenő információt és kiminetet képez belőle
  - A kombinációs lépés “egyszerű” : Lineáris kombinálás
  - Nem-lineáris transzformáció: Aktivációs függvény
- A neuronokat hálózatba rendezik általában rétegenként szervezve.
- A “tanulás” során az változik meg, ahogy a neuron a bejövő információt kombinálja



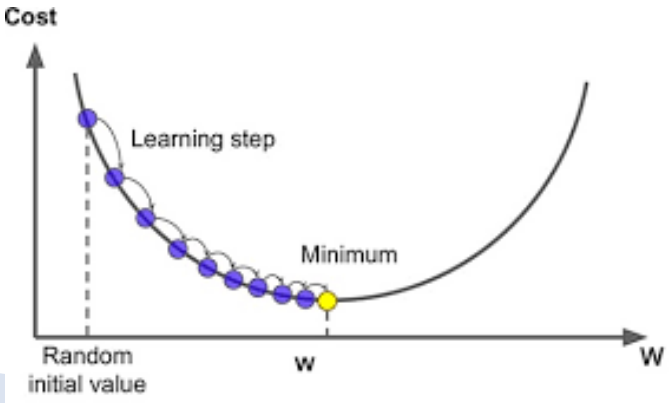
Neuron

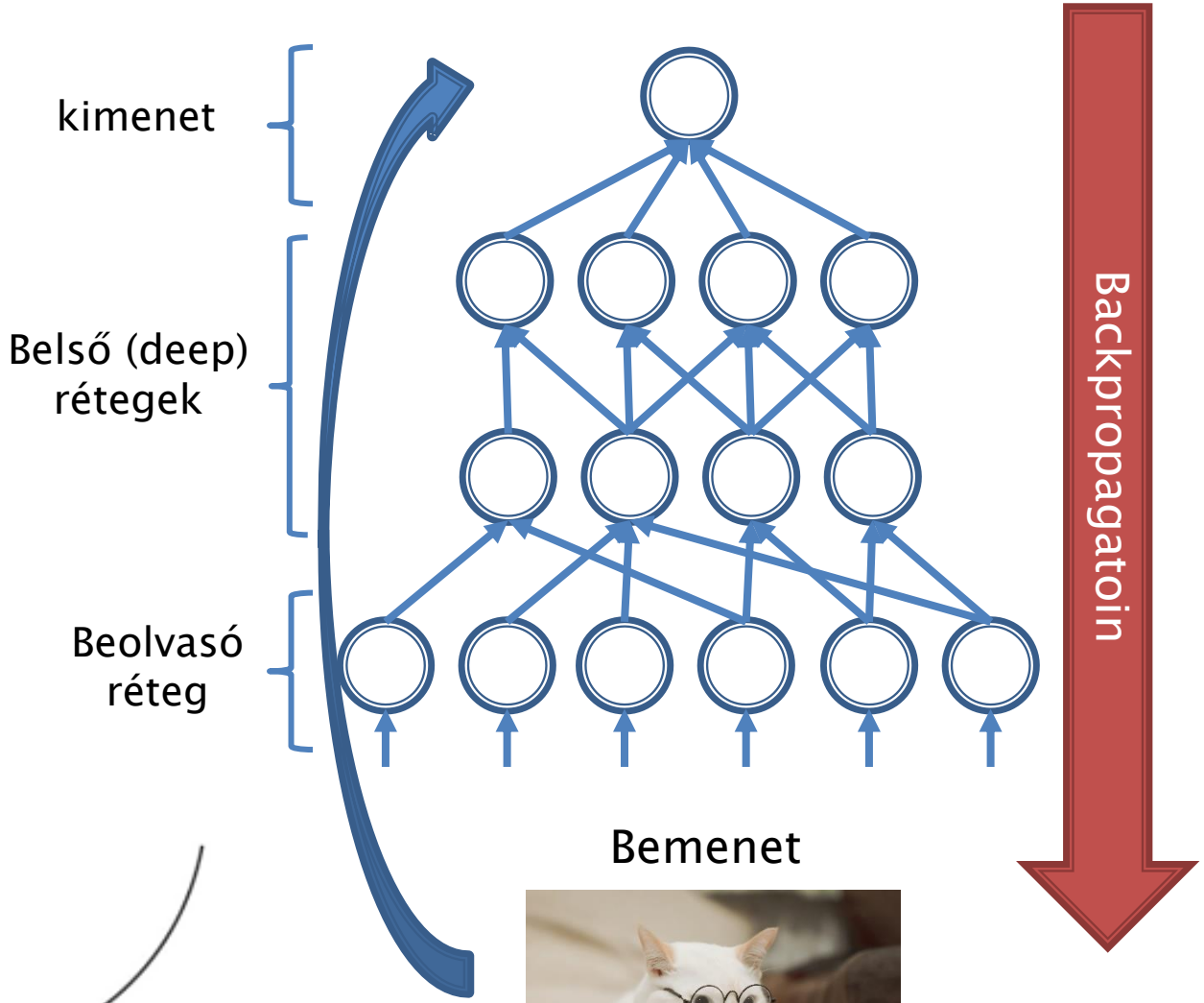
Itt történik a tanulás !!



Hogyan “tanul” a neurális háló?

- A beolvasó réteg kap egy ismert mintát amihez előre tudjuk milyen kimenet tartozik
- A háló feldolgozza az adatokat és ad egy kimenetet, amit összehasonlítunk az elvárt értékkel
  - Összehasonlítás: **COST** függvény. Számszerűen megmondja mennyire hasonlóak.
- A kettő különbsége alapján módosítjuk a háló belső paramétereit úgy, hogy közelebb legyen a kimenethez.
  - **Backpropagation**: megmutatja, hogy a háló paramétereinek kis változtatás milyen hatással lenne a kimenetre
  - **Gradient descent**: Felhasználva a backpropagation eredményét megmutatja, hogyan kell változtatni a paramétereket, hogy az elvárt kimenethez közelebbi eredményt kapjunk

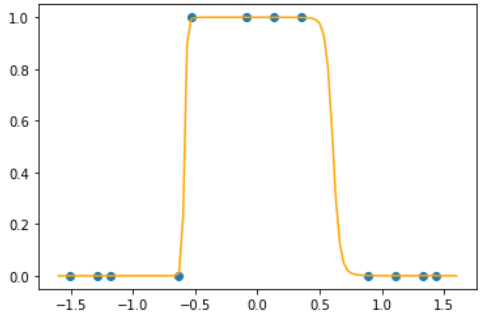




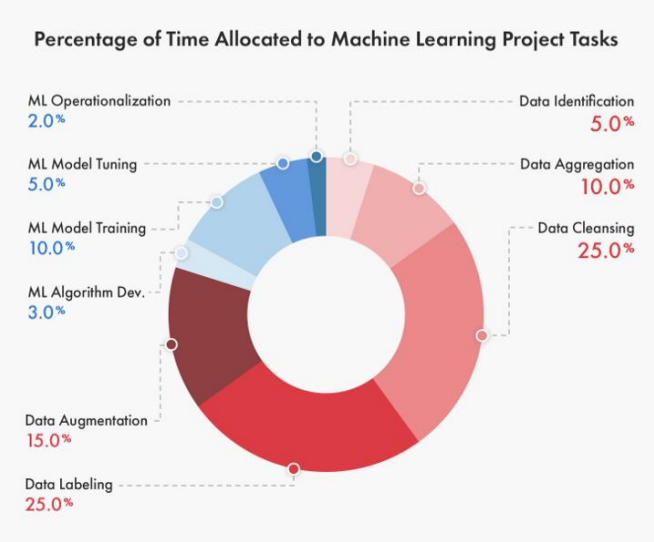
Miért ilyen sikeresek a neurális hálók?

- A neurális hálók lényegében nagyon sok paraméteres függvények
- A tanulás a függvény “fittelése”, illesztése az adatokra.
- Bizonyítható, hogy **univerzális approximátorok**, azaz nagyon bonyolult összefüggéseket is le tudnak írni [6,7].
  - A bizonyítás nem konstruktív, azaz nincs egy általános módszer tetszőleges adatok leírására.
- Nagyon jó tanító halmazokkal rendelkezünk.
  - Előállításuk drága és emberi munkaerőt igényel.
  - Nem elhanyagolható tényező, nagy cégek sikere mögött sokszor az “adatvásárlás áll”.
  - Az adattisztítás és címkézés teszi ki a legtöbb project időt kb (65%) még a tényleges fejlesztés kb. 20%. [8]
- A neurális hálók achilles-sarka pont az tanulási halmaz:
  - Bonyolult dolgok fitteléséhez egyre nagyobb tanító halmazokra van szükség (**kombinatorikai robbanás**) és ezt már nem tudjuk előállítani.
  - Nem tudunk minden furcsa esetre elég sok példát hozni. (a szemüveget viselő macskát eltérő esetekben elétérő módon kellene kategorizálni.

$$f(x) = \frac{1}{1 + e^{-\left(\frac{w_{21}}{1+e^{-(w_{11}x+b_{11})}} + \frac{w_{22}}{1+e^{-(w_{12}x+b_{12})}} + b2\right)}}$$



Két neuronos “hálózat”, mint függvény



A neurális hálók problémái

- A bonyolultabb feladatokhoz szükséges tanító halmazok mérete túl nagyra nő
- Nem tudjuk, hogy pontosan hogyan valósítják meg a feladatot, mit tanulnak meg.
  - Nagyon más módon tanulnak mint az emberi agy: a bemenet-kimenet függvény nagyon nem folytonos [9]
  - Pontos hibabecslés szinte lehetetlen
  - Nem alkalmazhatóak biztonsági előírásokat alkalmazó rendszerekben
- Nincs visszajelzés arra vonatkozóan, hogy egy döntés mennyire biztos
  - Az ember számára detektálhatatlan különbségek katasztrófális hibákat okoznak [10]
  - Az ember számára nyilvánvaló különbségeket nem detektálja
- Ha a tanítási folyamat befejeződött már nehéz új információt hozzáadni a rendszerhez (Catastrophic forgetting)

A módosított képeket mind “strucc”-ként ismeri fel a NN

Eredeti kép különbség Módosított kép

Fizika a gépi tanulásban

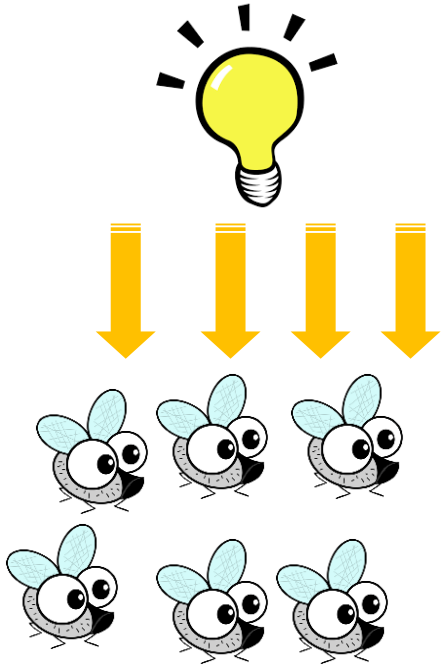
Miben segíthet a fizika?

- Neurális hálók működésének megértése
  - Tanítási folyamat optimalizálása
  - Hálók szerkezetének optimalizálása
  - Jobban megérthetjük miért hibáznak a hálók bizonyos esetekben
  - Megérthetjük milyen problémákra optimális neurális hálókat használni
- Új algoritmusok kifejlesztése
  - Kiszóró minták detektálása (a ML rendszer jelezzen ha valami nagyon ismeretlent lát) -> biztonságtechnikai alkalmazások
  - Egyszerű kategorizálás helyett valamilyen becslést adni arra, hogy mennyire biztos a rendszer az eredményben.
- A gépi tanulási feladatok pontos matematikai definiálása
  - Mit jelent a “tanulás”
  - Mit jelent a megértés?

Renormálás

- A récsecskefizika és statisztikus fizika területén fejlődött ki
- Fázisátalakulások leírása
- Annak leírása hogyan alakulnak ki a mérhető mennyiségek a mikroszkópikus törvényekből
- A mikroszópikus és makroszkópikus mennyiségek közötti kapcsolat vizsgálata

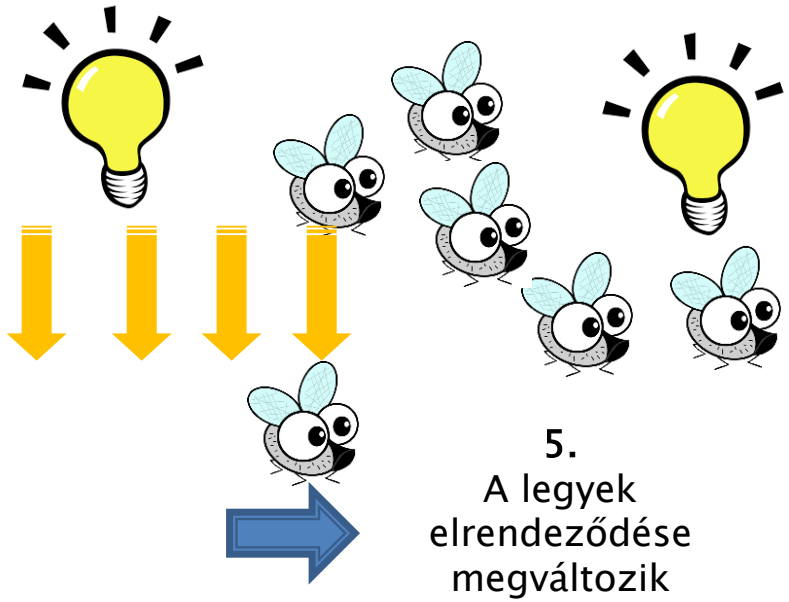
Renormálás= Árnyékolás



1.
Legyek reagálnak a fényre és köré gyűlnek

2.
A fényt árnyékolja az aktív “közeg”

3.
A mért fény kisebb erősségű, mintha nem lennének legyek.



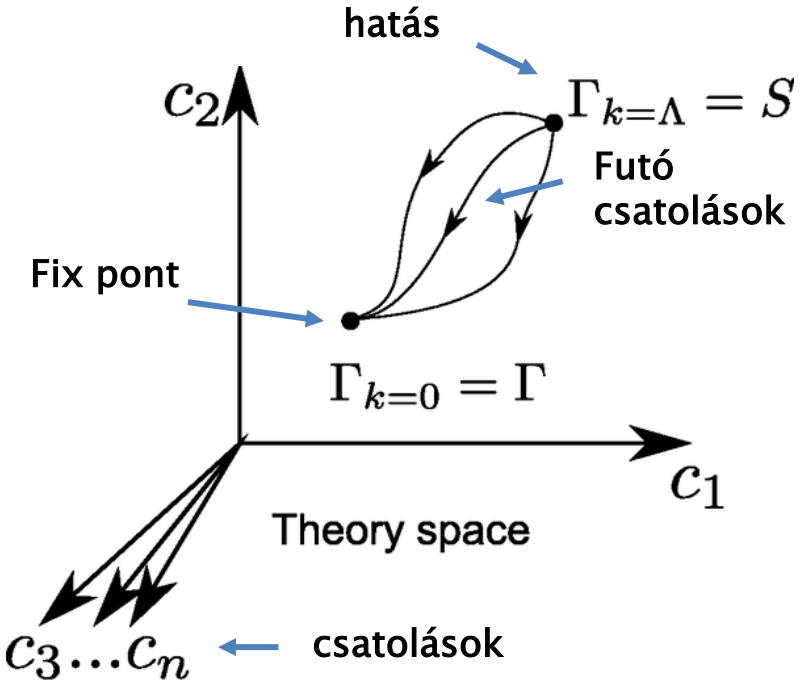
4.
Bekapcsol egy másik lámpa

5.
A legyek elrendeződése megváltozik

6.
A fényerősség amit mérünk szintén megváltozik.

- Legyek: a mikroszkópikus mennyiségek: amiket közvetlenül nem látunk de a kölcsönhatást írják le
- Fényerősség, izzók helyzete: makroszkópikus mennyiség amit mérünk

Modern formalizmus – Elméletek tere



Analógia
Csatolások – neurális hálók paraméterei
Hatás – COST függvény
Fix pontok: a háló tanult állapota

- **Skálafüggő effektív hatás (Γ_k)**:
  - A rendszert jellemző kölcsönhatásokat írja le
  - A skálafüggő módon változik
- **Scale (k)**:
  - Energia, impulzus, felbontás
  - Az a mennyiség ami befolyásolja a rendszer viselkedését
- **Fixed point**:
  - A hatás futása megáll
  - **Méréseket** a fix pontok körül teszünk
  - Releváns operátorok: Azok a kölcsönhatások amik meghatározzák a rendszer viselkedését a fix pont körül
  - Irreleváns operátorok: Az adott fix pontban nincsenek hatással a rendszer működésére
  - **NOTE**: más fixpontban más hatások lesznek relevánsak vagy irrelevánsak

A gépi megértés renormálás alapú megközelítése

- X : Véges halmaz aminek az elemeit vizsgáljuk
  - PL: Minden 1 MP -es kép
- C: Részhalmaz amit meg akarunk érteni
  - PL: 1 MP-es macskás képek
- Koordinátázás: ξ: X $\rightarrow B^N$ (bijekció)
  - Minden X-beli elem egy B halmaz beli vektornak felel meg
  - B : R részhalmaza, pl B = {0,1} – monokróm képek
- Példa:

Megértés:
A **releváns** és **irreleváns** koordináták megkülönböztetése a C halmaz azonosításának céljából.

[0,1,1,0]

2x2
Monokróm
képek

Koordinátázás:
B = {0,1}
N = 4

Megértés – a koordinátázás tulajdonságai

- **A koordináták feltételes valószínűségi eloszlása:**
  - **C**: megérteni kívánt halmaz
  - **I**: indexek listája
  - **σ**: fix koordináta vektor
  - Vegyük az elemeket C-ből és nézzük meg, hogy I-ben szereplő koordinátáknak milyen a hisztogramja.

$$p_I^{(C)}(\xi=\sigma)$$

- **faktorizáció**: adott I szet eloszlása független a többitől
- **Determinisztikus koordináta**: C minden elemére ugyanaz az értéke – RELEVÁNS KOORDINÁTÁK
- **Egyenletes eloszlás**: C-n a koordináta minden értéke felvehet – IRRELEVÁNS KOORDINÁTÁK

Megértés

C olyan koordinátázása ahol a koordináták függetlenek és a $p_i^{(C)}(\xi=\sigma)$ eloszlások vagy determinisztikusak vagy egyenletesek C-felett

Részletes bizonyítások [11]

Szokásos AI feladatok-átfogalmazva

- **Klasszifikáció** $C = \bigcup_a C_a$
  - Minden C_a halmazhoz meg kell találni a releváns (determinisztikus) koordinátákat, melyek más halmazokra nézve nem azok.
  - Kiszóró minták detektálás: a koordináták alapján még akkor is lehetséges ha nincs a tanító halmazban
- **Regresszió**
  - Speciális klasszifikáció: C_a -halmazok függvény értékek amiket az illesztendő paraméterek koordinátáznak.
- **Dekódolás**
  - Egy C halmaz eleme egyszerűen generálható: a koordináták teljesen megadják az elemet. A determinisztikusak adottak, az egyenletes eloszlásúakból pedig véletlenszerűen választunk.
- **Tömörítés**
  - Mivel a releváns koordináták állandóak, azokat nem kell külön tárolni csak az irrelevánsakat.

Szemléltetés

Értsük meg a macskás képeket

Releváns koordináták: Ezek minden macskás képre ugyanazok függetlenül a részletektől. A "macskaságot" ragadják meg.

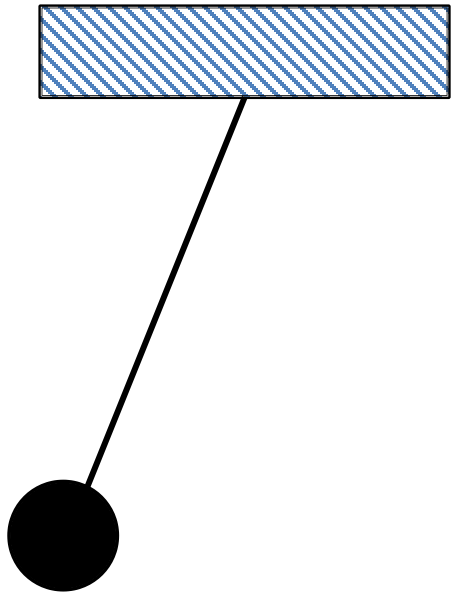
Irreleváns koordináták: A macskás kép részletei, hol van a macska milyen szögben, milyenek a fények stb…

Alkalmazások

Mechanikai mozgások rekonstruálása [12]

A feladat

- Adott egy mechanikai mozgás
- **r**(t), **v**(t), **a**(t) adatok elérhetőek
- Ezek alapján folytassuk a mozgást.
- Azaz az idősorok alapján keressük meg a mozgástörvényt
- Határozzunk meg megmaradó mennyiséget.



Gravitációs inga

Megoldási stratégia

- Tipikus LSTM feladat, de ehelyett a fentebb bemutatott módszert követjük
- Konstans mennyiségeket keresünk, amik a tanító halmaz minden elemére igazak. Ezek a determinisztikus koordináták.
- A mozgás folytatásához felhasználjuk az így megtalált összefüggéseket:
  - A jósolt mozgás koordinátáinak teljesíteniük kell ezeket az összefüggéseket – analogia a megmaradási törvényekkel. Ezeket a megmaradó mennyiségeket használjuk a rekurzió numerikus pontatlanságának javítására.

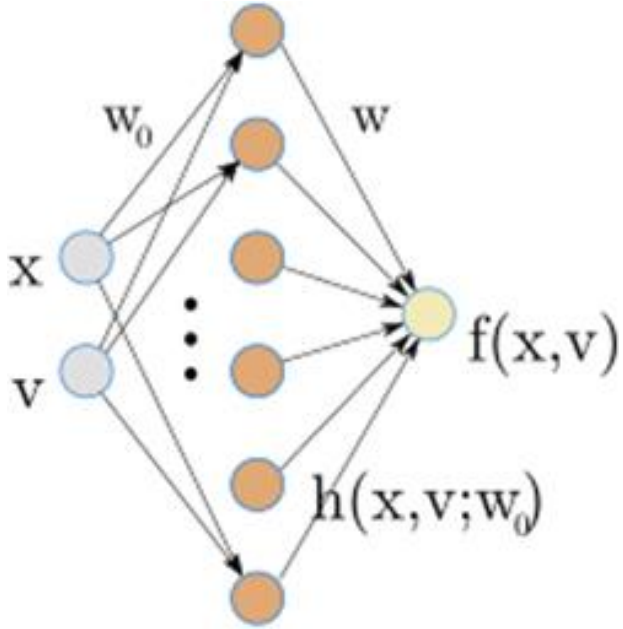
Mechanikai mozgások rekonstruálása II

Megoldás menete

- Newton 2. törvénye alapján a mozgás: $a_n = f_{\Delta t}(x_n, v_n)$
- Ha a diszkrét erőtörvény ismert a mozgás már folytatható
- **Extreme learning machine**: csak egy rejtett réteg van
- h: folytonos függvény (feature). A neurális háló megtanulja, hogy kell ezeket kombinálni, hogy megkapjuk az erőtörvényt.

Tanítás

- Az **Extreme learning machine** tanításához nem kell backpropagation, nagyon gyors.
- Numerikus esetben a kényszerek és az erő nem feltétlen következnek egymásból, ezért az erőtörvényt, és a kényszereket (megmaradó mennyiségeket) külön tanítjuk.
- A tanítás a **PCA technikával** analóg műveletekre vezet – **Sajátérték egyenletek**



Extreme learning machine

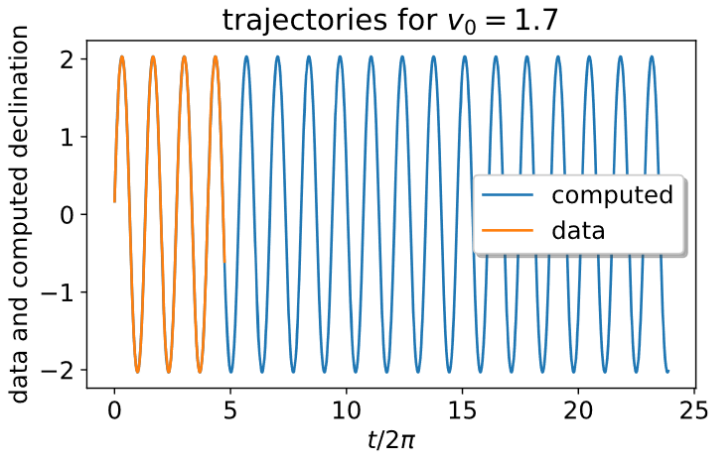
$$F^{\mathrm{T}} F\bar{w} = F^{\mathrm{T}} a.$$

Erőtörvény

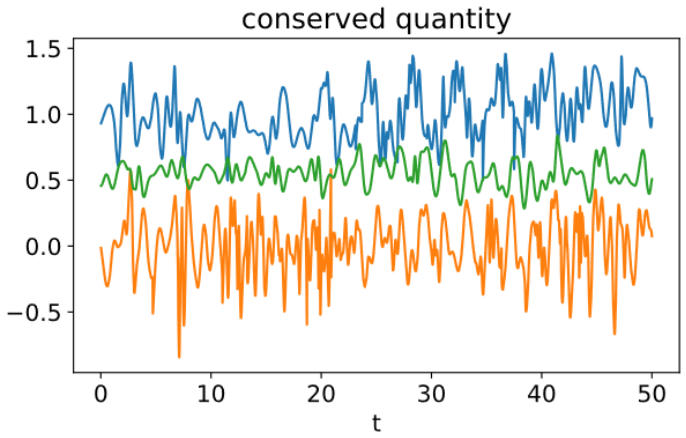
$$\mathrm{d}F^{\mathrm{T}}\,\mathrm{d}Fw = \lambda w$$

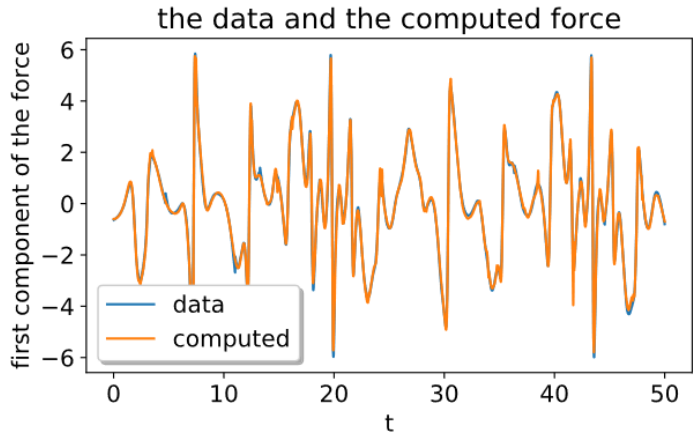
Megmaradó mennyiség

Mechanikai mozgások rekonstruálása – eredmények



Gravitációs ingára
vonalvastagságon belül
pontos eredmény





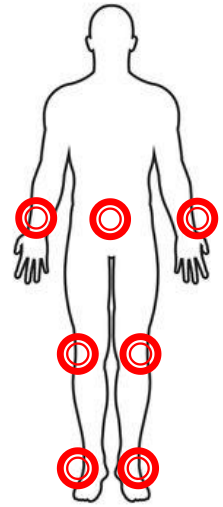
Kettős inga (kaotikus rendszer)

- Pontos rekonstrukció
- A mozgás véges marad nincs megszaladó megoldás
- A megmaradó mennyiségek stabilizálják az iterációt

Mozgástípusok osztályozása [12]

A feladat

- Adottak kísérleti alanyok mozgásszenzoros adatai.
- Formátum: Több féle szenzor jeleinek idősora.
- Ezt felhasználva készítsünk egy ML rendszert ami sikeresen kategorizálja az ilyen jeleket mozgástípusok szerint.



Megoldási stratégia

- Az előzőhöz hasonlóan idősoroként konstans mennyiségeket keresünk (Releváns koordináták) majd ezeket használjuk osztályozásra.
- Ezzel transzformáltuk a bejövő adatokat, azaz olyan koordináta rendszerre váltottunk amiben a feladat egyszerűbb.
- Akkor működik jól, ha az idősorokhoz rendelt megmaradó mennyiségek különbözőek, és ezért diszjunkt koordináták.
- Ezeken a transzformált releváns koordinátákon egyszerűbb osztályozó algoritmussal is nagy sikert lehet elérni.

	Applied feature space			
	Original		Transformed by LLT	
	Accuracy (%)	Time (sec)	Accuracy (%)	Time (sec)
Ensemble	75.7	238.4	99.7	313.4
KNN	75.6	40.5	100.0	59.2
DT	71.5	32.2	98.8	33.6
SVM	73.0	1379.2	100.0	1726.3

Köszönöm a figyelmet!

Referenciák

- [1] Mcculloch, Warren S. & Pitts, Walter (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. _Journal of Symbolic Logic_ 9 (2):49–50.
- [2] Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. Nature 323, 533–536 (1986). https://doi.org/10.1038/323533a0
- [3] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995). https://doi.org/10.1007/BF00994018
- [4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9, 8 (November 15, 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90. https://doi.org/10.1145/3065386
- [6] Cybenko, G. Approximation by superpositions of a sigmoidal function. Math. Control Signal Systems 2, 303–314 (1989). https://doi.org/10.1007/BF02551274
- [7] Zhou, D. X. (2020). Universality of deep convolutional neural networks. Applied and computational harmonic analysis, 48(2), 787-794.
- [8] https://towardsdatascience.com/data-labeling-is-chinas-secret-weapon-in-the-connected-car-battle-e8e395965380
- [9] Szegedy, Christian & Zaremba, Wojciech & Sutskever, Ilya & Bruna, Joan & Erhan, Dumitru & Goodfellow, Ian & Fergus, Rob. (2013). Intriguing properties of neural networks.

Referenciák

- [10] Biggio, B. et al. (2013). Evasion Attacks against Machine Learning at Test Time. In: Blockeel, H., Kersting, K., Nijssen, S., Železný, F. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2013. Lecture Notes in Computer Science(), vol 8190. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40994-3_25
- [11] Understanding understanding: a renormalization group inspired model of (artificial) intelligence A. Jakovac (Wigner RCP, Budapest), D. Berenyi (Wigner RCP, Budapest), P. Posfay (Wigner RCP, Budapest) e-Print: 2010.13482 [cs.AI]
- [12] Reconstruction of observed mechanical motions with artificial intelligence tools, Antal Jakovác,1, Marcell T Kurbucz and Péter Pósfay, 2022 New J. Phys. 24 073021
- [13] Kurbucz, Marcell Tamás and Pósfay, Péter and Jakovác, Antal, Facilitating Time Series Classification by Linear Law-Based Feature Space Transformations (July 7, 2022). Available at SSRN: https://ssrn.com/abstract=4161139 or http://dx.doi.org/10.2139/ssrn.4161139

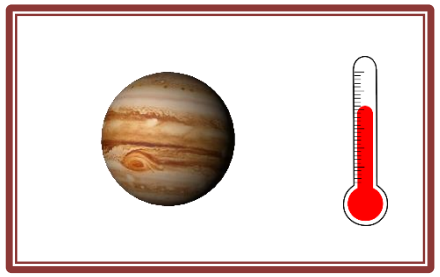
Backup diák

Scientific understanding and neural networks

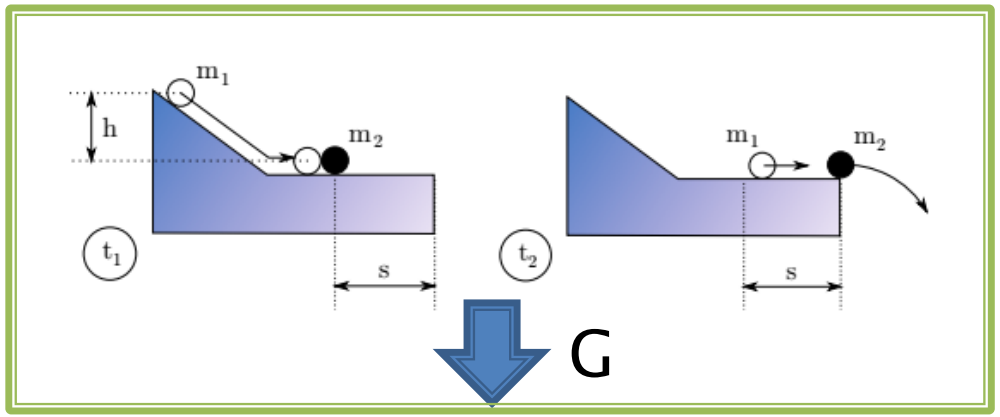
Understanding natural Phenomena

- **Key Concepts in natural sciences**
  - Experiment, Phenomena
  - Environment, measurement
  - Isolation of measurement
  - Neglecting certain aspects of the measurements (idealization)
- **These can be more precisely understood in this framework**
  - X: all possible configurations of the world
    - There are infinitely many facts and possible measurements
  - Subset C: singles out the phenomena we are interested in (experimental setup)
    - Defined by relevant coordinates over C
    - Neglecting irrelevant coordinates – idealizations, isolation, defining the measurement

Irrelevant



Relevant



Understanding natural Phenomena

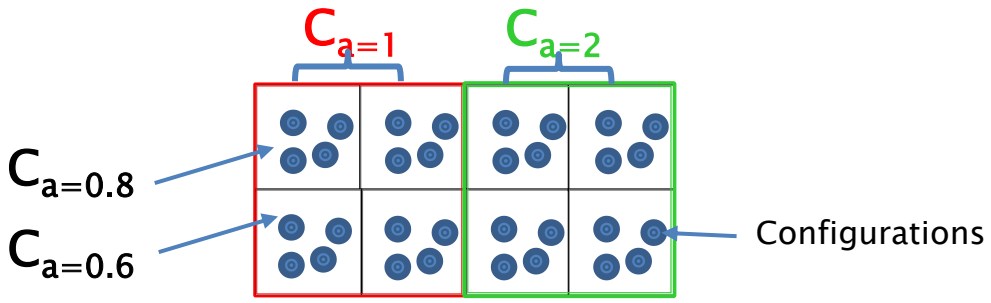
- **More precisely:**
  - Phenomena is defined by the **relevant (deterministic) coordinates** over C
    - They define the experimental setup (the slope, balls parameters etc..)
  - **Irrelevant coordinates**: Do not influence the experiment (temperature, pressure etc..)
  - $C = \cup_a C_a$ where C_a -s collect all the states of the world where the experiment results in measurement with a value a.
  - **Partially relevant coordinates**: These are the most important ones, they decide that a given C_a corresponds to which parameter value, the fix the connection between measurement and the state of the world.

- Scientific method:
  - A lot of experiments of the same kind to generate C
  - Search for **relevant coordinates**: any measurement can be expressed as their function
  - **Non-independent** relevant coordinates: their connections are physical laws
  - Typically scientific theories have a **small number of relevant parameters**

Connection to renormalization

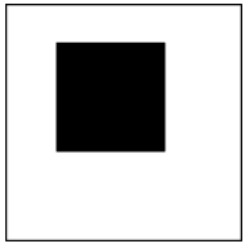
- **Measurements have finite precision:**
  - At a given precision we can differentiate a given number of different outcomes
  - C_a -s correspond to **interval of measurement results** (not a discrete value).
  - C_a collects configurations of the system which produce **„similar" measurements**
  - Changing the resolution changes which configurations are in one interval (C_a) and how many intervals (C_a -s) we have.
- **Renormalization view**:
  - Precision is taken into account by a **scale: k**
  - We **coarse-grain** the configurations at a given scale: We average configurations which are „similar": differences are smaller than the scale
  - This results in scale dependent C_a (k)-s which form series as the scale changes
  - At a given scale we have different relevant and irrelevant coordinates as C_a -s change.

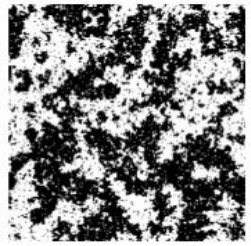
Configurations in one box produce the same measurement value (a) at a given precision



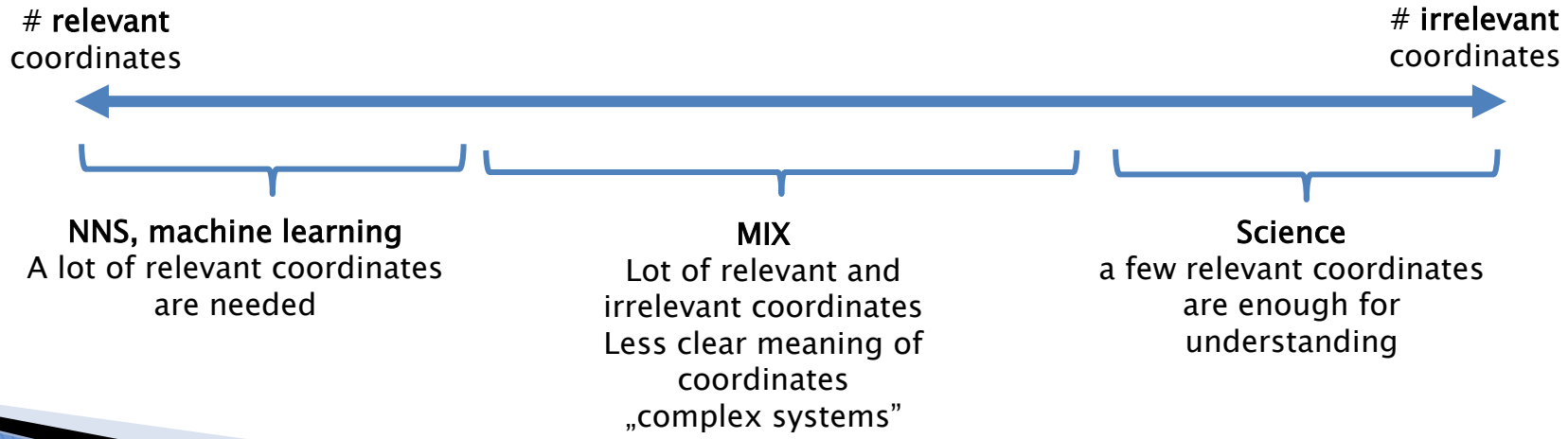
Scientific understanding vs neural networks

- **Image classification as a typical task:**
  - Take all **256x256** pixel binary images
  - A general picture has 65536 bits of information (**irrelevant bits**)
  - Single out a subset of pictures which depict a **black square**
    - **3 irrelevant coordinates**: square coordinates and size 3x8 = 24 bits
    - **All other 65512 bits are relevant!**





Scientific problems has only few relevant parameters!



Example: 3-bit image

- X = {000, 001, 010, 011, 100, 101, 110, 111} $|X| = 8$
  - Coordinatization: B = {1,0}, N = 3
- C = {001, 010, 100, 111} $|C| = 4$
- Probability distribution of the first coordinate over C:
  - $p_1^C(\xi_1 = 1) = \frac{2}{4} = \frac{1}{2} = p_1^C(\xi_1 = 0)$ (This is true for all other coordinates too i = 2,3)
- Expectation values
  - $\mathbb{E}(\xi_i \mid C) = \frac{1}{2}$
  - $\mathbb{E}(\xi_1\ \xi_2 \mid C) = \frac{1}{4} = \mathbb{E}(\xi_1 \mid C)\ \mathbb{E}(\xi_2 \mid C)$
  - $\mathbb{E}(\xi_1\ \xi_2\ \xi_3 \mid C) = \frac{1}{4} \neq \mathbb{E}(\xi_1 \mid C)\ \mathbb{E}(\xi_2 \mid C)\ )\ \mathbb{E}(\xi_3 \mid C)$ NOT independent!

0 0 0 1

C = {001, 010, 100, 111} FACTORIZES

- It is easy to deal with cases when coordinates factorize
  - If coordinates factorize over a set, we can select them randomly and generate an element from the set
  - If they are not independent we this is not true, like above: