





Gyires-Tóth Bálint

Deep Learning a gyakorlatban Python és LUA alapon
Bevezetés

SmartLab
Intelligent Interactions

http://smartlab.tmit.bme.hu

NVIDIA
GPU
EDUCATION
CENTER

Jogi nyilatkozat

Jelen előadás diái a „*Deep Learning a gyakorlatban Python és LUA alapon*” című tantárgyhoz készültek és letölthetők a http://smartlab.tmit.bme.hu honlapról.

A diák nem helyettesítik az előadáson való részvételt, csupán emlékeztetőül szolgálnak.

Az előadás diái a szerzői jog védelme alatt állnak. Az előadás diáinak vagy bármilyen részének újra felhasználása, terjesztése, megjelenítése csak a szerző írásbeli beleegyezése esetén megengedett. Ez alól kivétel, mely diákon külső forrás külön fel van tüntetve.

Bemutatkozás



Gyires-Tóth Bálint, Ph.D.

- Deep Learning, gépi tanulás
- Szekvenciális adatok, idősor modellezés

- BME TMIT
- NVidia Deep Learning Institute Certified Instructor & University Ambassador

Oktatók, gyakorlatvezetők, segédoktatók

Bálint



Tamás



Csaba



Róbert



Gergely



Mohammed



Sevinj



Hamdi



SmartLabs (TMIT)

Deep Learning oktatás és kutatás

- Oktatás:
  - Deep learning a gyakorlatban Python és LUA alapon
  - BSc, MSc, PhD
- Kutatás
  - Alap és alkalmazott kutatás
  - Hazai és nemzetközi projektek

http://smartlab.tmit.bme.hu

Laborvezető: Dr. Németh Géza

Tárgy célkitűzése

- Gépi tanulás új paradigmája. Óriási potenciál!
- Gyakorlati tudás vállalkozáshoz vagy munkavállaláshoz.
- A tárgy elvégzése 5-ös érdemjeggyel presztízs értékű! Vonzó a cégek számára!
- „Junior” deep learning mérnök:
  - 1 komplex feladat,
  - 20-80 tanítás.
- „Senior” deep learning szakember:
  - legalább 8-10 komplex feladat,
  - 2000-3000 tanítás.

Tárgyról infók

Előadás: kedd 12-14h online

Gyakorlat:

csütörtök 12-14h online

csütörtök 16-18h online

Konzultáció előzetes egyeztetés esetén online vagy gyakorlat alatt a szabad gyakorlatvezetőkkel.

http://smartlab.tmit.bme.hu/oktatas-deep-learning



Ajánlott tárgyak

- Neurális hálózatok (VIMIJV07)
  Dr. Pataki Béla, tavaszi félév (✝ Dr. Horváth Gábor)

- 'Big Data' elemzési eszközök nyílt forráskódú platformokon (VITMAV15)
  Prekopcsák Zoltán, Gáspár Csaba
- Média- és szövegbányászat (VITMM275),
  Dr. Szűcs Gábor, Dr. Tikk Domonkos
- Adatbányászati technikák (VISZM185),
  Dr. Katona Gyula, Dr. Pintér Márta Barbara, Dr. Sali Attila
- A lineáris algebra numerikus módszerei (VIMAD041)
  Dr. Pach Péter Pál, tavaszi félév
- Mátrixanalízis (doktori képzés, VIMAD569)
  Dr. Pach Péter Pál, őszi félév

Tárgy követelményei

- Előadáson, gyakorlaton való részvétel
- Gyakorlatok 70%-án való részvétel
- Kis házi feladatok (+1 jegy)
- Nagy házi feladat (aláírás, +1 jegy, megajánlott jegy)
- Kahoot! (+1 jegy)
- Vizsga: írásbeli + szóbeli



Kis házi feladat

- Nem kötelező, de erősen ajánlott
- Félév során várhatóan 5 db kis házi feladat.
- Eredmények: Github.com-ra Jupyter notebook formátumban.
- Az eredmények és az oda vezető út legyen egyértelmű!
- Határidő: 2 hét. Késve leadott feladatokat nem fogadunk el.
- Az öt házi feladatra kapható pontok 70%-nak elérése esetén: +1 jegy
- A kiértékelésig maradjon secret gist, utána lehet publikus.
- Beküldés: http://smartlab.tmit.bme.hu/oktatas-deep-learning

Nagy házi feladat I.

- Aláírás feltétele.
- Komplexebb házi feladat esetén lehetőség van 4-es vagy 5-ös megajánlott jegyre. Rosszabbra nem!

Benyújtandó anyagok:

1. Jól kommentezett forráskód Github-ra feltöltve. (Bármelyik deep learning keretrendszerben készülhet.)
2. Github-on readme.md a csapatról és a tanítás és tesztelés módjáról (a reprodukáláshoz mit futtassunk, milyen paraméterekkel, milyen szoftver környezetben).
3. Beszámoló dokumentum (min 4, max 8 oldal).
4. Vizualizáció Jupiter notebookban, ha indokolt.

Nagy házi feladat II.

- 3 fős csapatok, 2-2 fős „vetésforgó” az alábbi szerepkörökkel:
  1. Tudományterület feltérképezése, cikkek olvasása.
  2. Adatok beszerzése, tisztítása, előkészítése.
  3. Hálózat és tanítás.
  4. Hiperparaméter optimalizálás és tesztelés.
  5. Interfész csatolás, opcionális (pl. Android).

- Tehát a csapat minden tagja legalább 3 ponttal foglalkozik.
- Mindenkinek kötelező a (3)-as vagy (4)-es pontok valamelyikével foglalkoznia.

Nagy házi feladat III. - Mérföldkövek

KIZÁRÓLAG a mérföldkövek határidőre történő teljesítése esetén van lehetőség a megajánlott (vagy +1) jegyre! Nincs utólagos pótlásra lehetőség!

I. mérföldkő: 7. hét (október 25. 23:59)

Adatok beszerzése, adatfeltárás, vizualizáció (ha szükséges) és előkészítés tanításhoz.

Eredmény: tanító, validációs és teszt adatbázisok

II. mérföldkő: 11. hét (november 22. 23:59)

Alaptanítás, ahol a tanítást végző kód és a kiértékelés első változata már rendben lefut, de még nem kötelező, hogy használható pontosságot adjon.

Leadás: december 11. 23:59 (aláírásért és megajánlott jegyért is)

+prezentáció vizsga helyett/mellett: megajánlott jegyért

Nagy házi feladat IV. - beszámoló

- Magyar és angol nyelvű cím
- Magyar és angol nyelvű kivonat (abstract): 100-200 szó.
- Tartalom (angol vagy magyar):
  - Bevezető
  - Tématerület ismertetése, korábbi megoldások
  - Rendszerterv: milyen hálózatot használtok
  - Megvalósítás
    - Adatok beszerzése, előkészítése
    - Tanítás
    - Kiértékelés (tanítási, validációs és teszt hiba), hiperparaméter optimalizálás, stb.
    - Tesztelés
  - Jövőbeli tervek, összefoglalás
  - Hivatkozások
    - Legalább 10 db KORREKT irodalmi hivatkozás (nem webes!).
    - scholar.google.com, ArXiv.org
- Minimum 4, maximum 8 oldal.
- Formázási irányelvek:
  - NIPS formátum (LaTex): https://nips.cc/Conferences/2016/PaperInformation/StyleFiles
  - IEEE Transactions (Word, LaTex): https://www.ieee.org/publications_standards/publications/authors/author_templates.html
- PDF + forrás (Word vagy LaTex) feltöltése a GitHub-ra.

Nagy házi feladat V. - GPU

Ha van saját, akkor azon dolgozzatok.

Ha nincs:

- Google Colab kezdetben
- Ezt követően AWS és Google Cloud használata

Nagy házi feladat VI.

- A házi feladatok, forráskódja, dokumentáció és prezentáció publikus lesz és felkerül a tárgy honlapjára, illetve GitHub.com-ra GNU GPLv3 licensz alatt.
- A projectek és a vonatkozó linkek megjelenhetnek a tárgy honlapján.
- A kiemelkedő projektek esetén segítünk a témát továbbvinni TDK-ra, konferenciára/folyóiratba, startup-ba.

Nagy házi feladat témák

- Témaötletek: http://smartlab.tmit.bme.hu/oktatas-deep-learning
- Saját téma is lehet, sőt!
- Legkésőbb 4. hét végéig álljanak össze a csapatok és válasszatok témát. Mindekinek kötelező a 4. hét végéig regisztrálnia itt: http://smartlab.tmit.bme.hu/oktatas-deep-learning-nagy-hazi
- Bármelyik deep learning keretrendszerben készülhet.
- TDK, BSc & MSc diploma, publikációs és akár PhD lehetőség.
- És ez miért is jó?!

- Példa aláírásért: számfelismerő (kézi és gépi) készítése sárgacsekk automatikus digitalizálására.
- Példa megajánlott jegyért: automatikus árfolyammozgás becslés a részvényhez köthető hírek alapján.

Github.com

- **Kis házi feladat:** mindenkinek legyen egy saját „Secret Gist” -je a GitHub-on. Ha ez nem elég, GitHub repo. Kiértékelésig ne adjátok meg az URL-t senkinek!
- **Nagy házi feladat:** a csapatok készítenek saját GitHub repo-t, mely publikus GNU GPLv3 licensz alatt.
- Az eredeti copyright-al a BME rendelkezik!
- GNU GPLv3 licensz szöveg fel kell tüntetni a repo-ban: http://choosealicense.com/licenses/gpl-3.0/
- Link beküldése mindig a honlapon keresztül: http://smartlab.tmit.bme.hu/oktatas-deep-learning

Házi feladat nyelve

Python vagy LUA

+1 jegy és megajánlott jegy esetén ANGOL

- Beszámoló
- Kommentek
- README.MD

Nagy házi aláírásért: lehet magyar.

ACM SIGITE 2020

Interpretable Deep Learning for University Dropout Prediction

Máté Baranyi
Budapest University of Technology and Economics
baranyim@math.bme.hu

Marcell Nagy
Budapest University of Technology and Economics
marcessz@math.bme.hu

Roland Molontay
MTA-BME Stochastics Research Group
molontay@math.bme.hu

ABSTRACT

The early identification of college students at risk of dropout is of great interest and importance all over the world, since the early leaving of higher education is associated with considerable personal and social costs. In Hungary, especially in STEM undergraduate programs, the dropout rate is particularly high, much higher than the EU average. In this work, using advanced machine learning models such as deep neural networks and gradient boosted trees, we aim to predict the final academic performance of students at the Budapest University of Technology and Economics. The dropout prediction is based on the data that are available at the time of enrollment. In addition to the predictions, we also interpret our machine learning models with the help of state-of-the-art interpretable machine learning techniques such as permutation importance and SHAP values. The accuracy and AUC of the best-performing deep learning model are 72.4% and 0.771, respectively that slightly outperforms XGBoost, the cutting-edge benchmark model for tabular data.

CCS CONCEPTS

• **Applied computing** → **Education**; • **Information systems** → **Decision support systems**.

KEYWORDS

dropout prediction; higher education; neural networks; interpretable machine learning; explainable artificial intelligence

ACM Reference Format:
Máté Baranyi, Marcell Nagy, and Roland Molontay. 2020. Interpretable Deep Learning for University Dropout Prediction. In *The 21st Annual Conference on Information Technology Education (SIGITE '20), October 7–9, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3368308.3415382

1 INTRODUCTION

Student dropout and delayed completion are some of the most burning issues in higher education all over the world. Approximately every third student drops out from higher education worldwide [21, 31]. In Hungary, the graduation rate is particularly low, one of the lowest among OECD and EU countries [12, 38].

Understanding and identifying the factors affecting students' university performance, and detecting at-risk students are in the focus of educational research for decades [1, 5, 11, 23, 32]. Moreover, due to the rise of data-driven approaches and the vast amount of data stored in educational administrative systems, a tremendous number of predictive analytical educational research papers and artificial intelligence-based decision support systems have been published in the last few years, for systematic reviews, we refer to [10, 13, 19, 33].

The majority of the related works use traditional machine learning models (e.g. logistic regression, k-nearest neighbors, and decision tree-based ensemble models) for dropout prediction [6, 7, 9, 34]. Deep neural networks (DNN) are not frequently applied on tabular data, due to the fact that such datasets are usually not large and complex enough for these models to achieve high performance, moreover, DNNs are more computationally expensive and they require more fine-tuning and expertise, besides these models are less interpretable than the traditional ones. However, there is a growing number of deep learning models that are designed for tabular data (e.g. TabNet [4], RLNs [36], and NODE [29]), hence many researchers started to incorporate these models in their analyses. Also in educational data science some papers already successfully applied deep neural networks for student performance and dropout prediction. Xing and Du constructed a dropout prediction model for identifying at-risk students in MOOCs [39]. Kim *et al.* proposed a deep learning-based model, called GritNet, that handles student performance prediction as a sequential event prediction problem [16]. A probabilistic neural network was introduced by Mason *et al.* to predict engineering student retention [24]. Recently, Agrusti *et al.* trained a convolutional neural network to predict student dropout [2]. Various other network architectures (e.g. denoising autoencoders) have been used to predict student retention and identify at-risk students [3, 20, 28], nevertheless a common conclusion can be drawn: deep learning models frequently outperform traditional machine learning models for student dropout prediction.

The present study joins the series of works aiming to build accurate dropout prediction models at the Budapest University of

Security

Egymásról való másolás gyanúja esetén az időben későbbi feltöltés tulajdonosa minden plusz lehetőségből ki lesz zárva (megajánlott jegy, +1 jegy) és kötelező számára a szóbeli vizsga.

Bizonyosság esetén -2 jegy.

Aki nem tudta felvenni vagy külsős

- A tárgy hivatalos hallgatóinak a hely biztosítás elsődleges.
- Házi feladatot javítani csak a tárgy hallgatóinak tudunk.
- Szünnapokat kövessétek: https://kth.bme.hu/
- Témában szívesen beszélgetünk.


KEEP CALM
MAKE MAGIC HAPPEN
SAY CHEESE

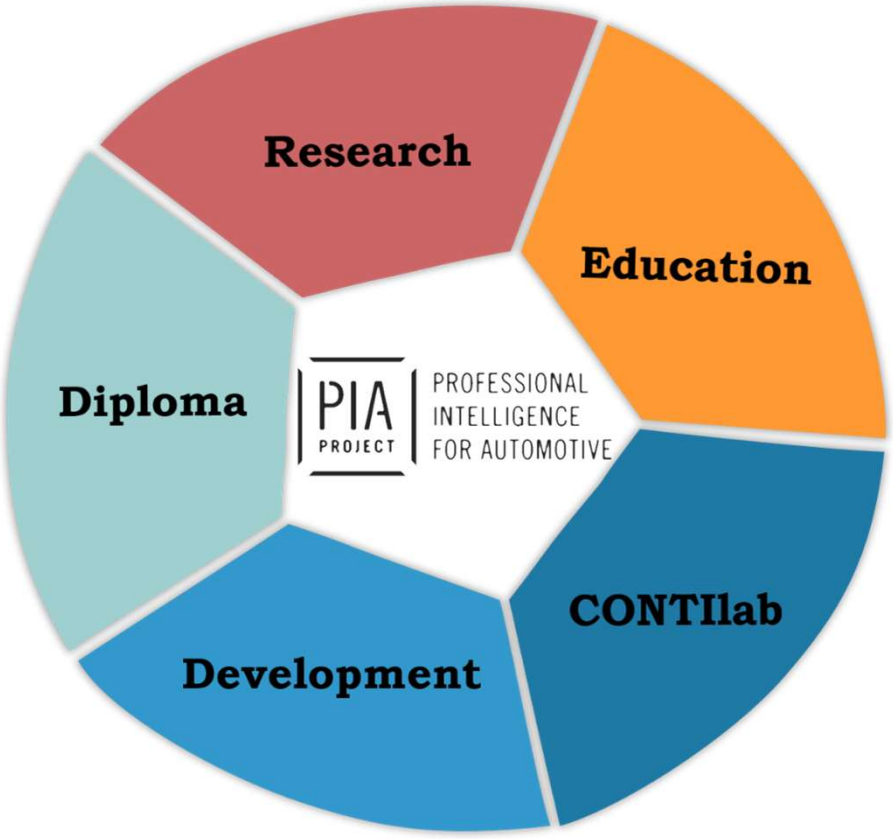
Kahoot!

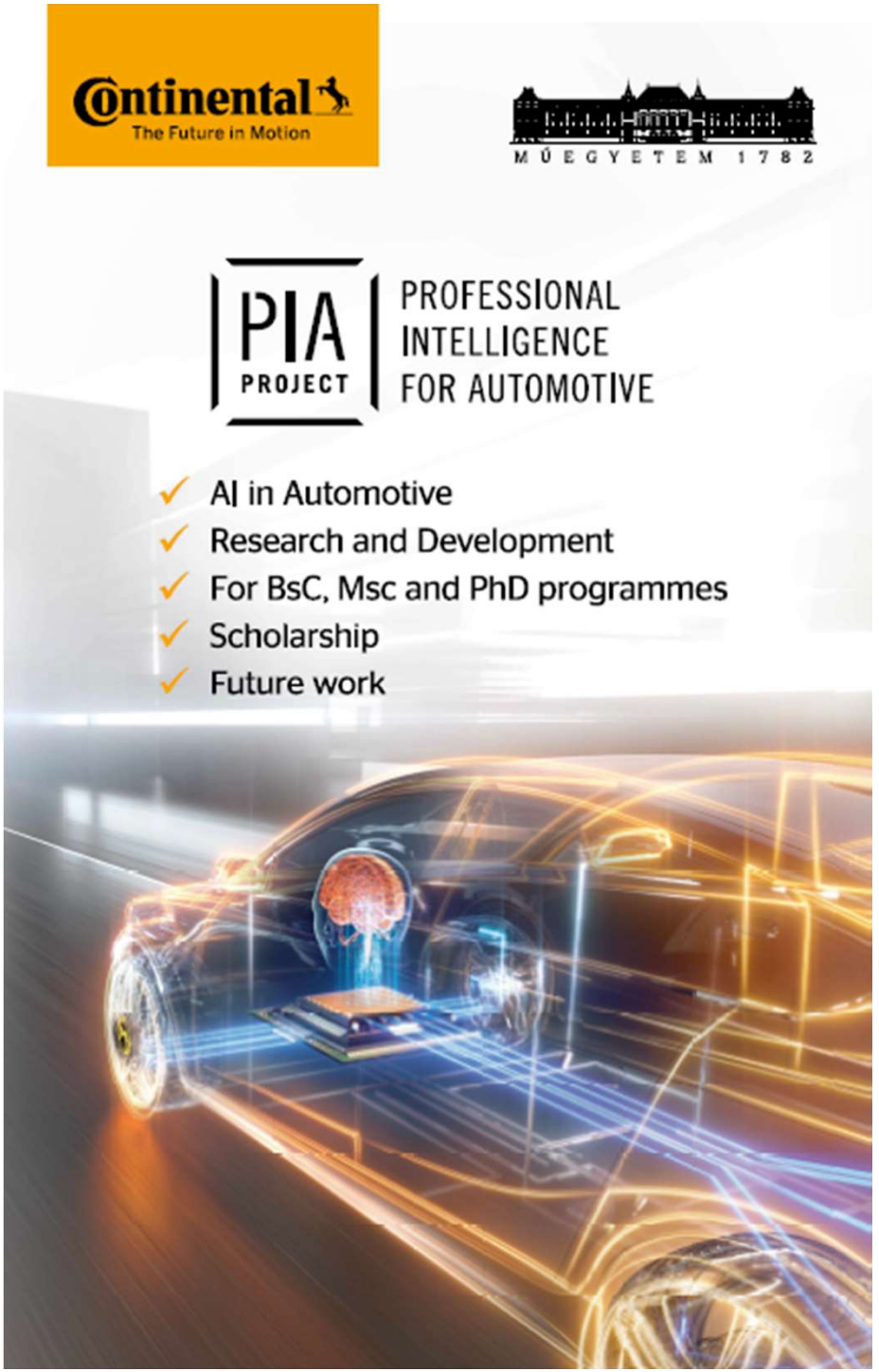
- Kahoot.it
- Mindig a NEPTUN kódodat add meg! (egyébként később nem tudjuk összerendelni Veled)
- TOP 5 (legtöbb jó válasz a félév végén): +1 jegy
- Minden előadáson (keddenként).
- Lehetséges témák:
  - előző előadás második fele,
  - a csütörtöki gyakorlat,
  - az aktuális előadás első fele.



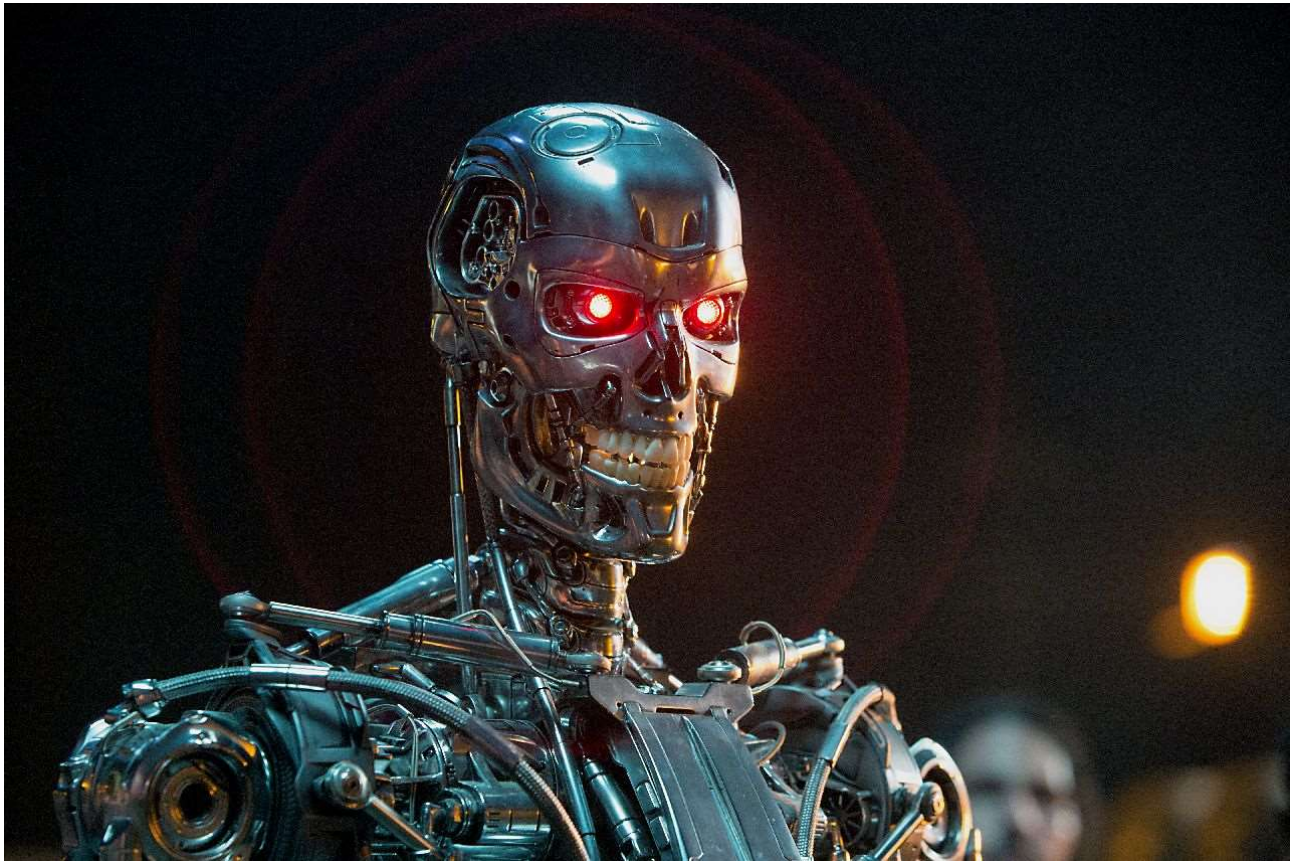
Névnek a NEPTUN kódodat add meg!

https://kahoot.it/


Research
Education
Diploma
PIA
PROJECT
PROFESSIONAL
INTELLIGENCE
FOR AUTOMOTIVE
CONTIlab
Development


Continental
The Future in Motion
MŰEGYETEM 1782
PIA
PROJECT
PROFESSIONAL
INTELLIGENCE
FOR AUTOMOTIVE
AI in Automotive
Research and Development
For BsC, Msc and PhD programmes
Scholarship
Future work

Facebook Shut Down AI That Developed Its Own Language: Is This What Experts Have Warned About?



A dead man was resurrected yesterday: is this the beginning of a zombie apocalipse?



Valóság…

—*Hey ****, can you add X to my to do list.*

—*Sure, Y added to your to do list.*

—*****, can you remove the last item from the to do list.*

—*Sorry I cannot do that.*

—*Sorry I don’t understand, please repeat slowly.*

—*…*

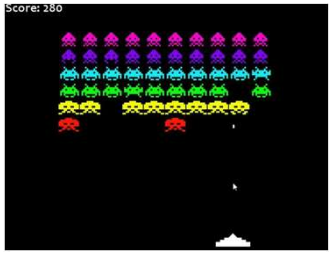
Valóság…

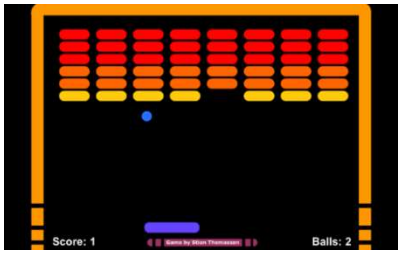


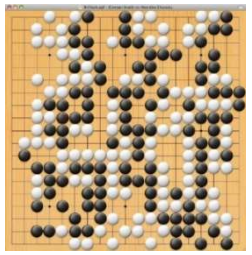
Deep Learning hírek a nagyvilágból

2014 –Deep Mind, ~500.000.000$ (https://deepmind.com)

- Space Invaders
- Breakout
- Go, 5-0 és 4-1







2015 december – OpenAI, ~1.000.000.000$ (https://openai.com)

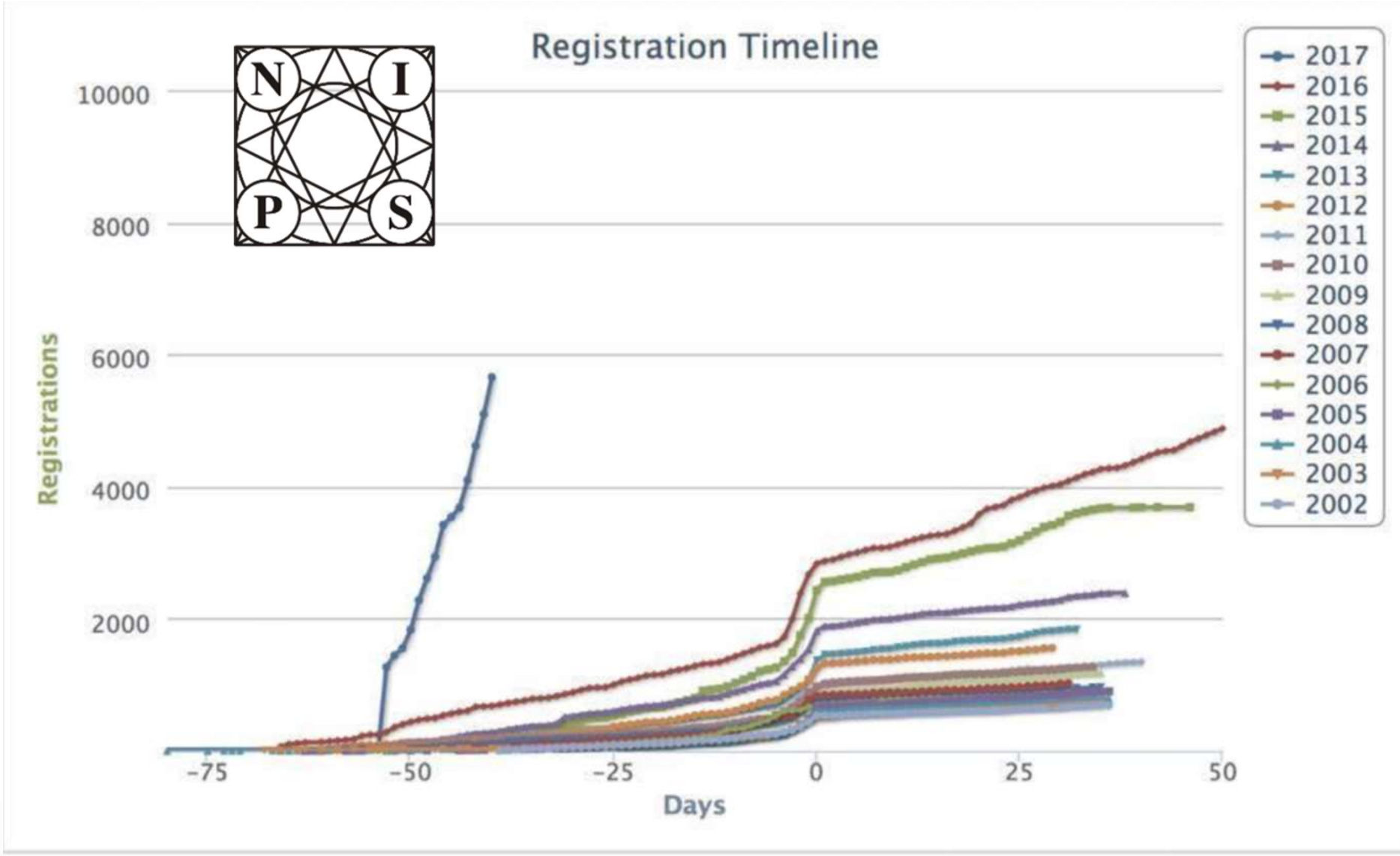
- Elon Musk (PayPal, SpaceX, Tesla Motors)
- Open source AI megoldások az emberiség szolgálatában
- Open AI five

2016 augusztus– Nervana – Intel fúzió

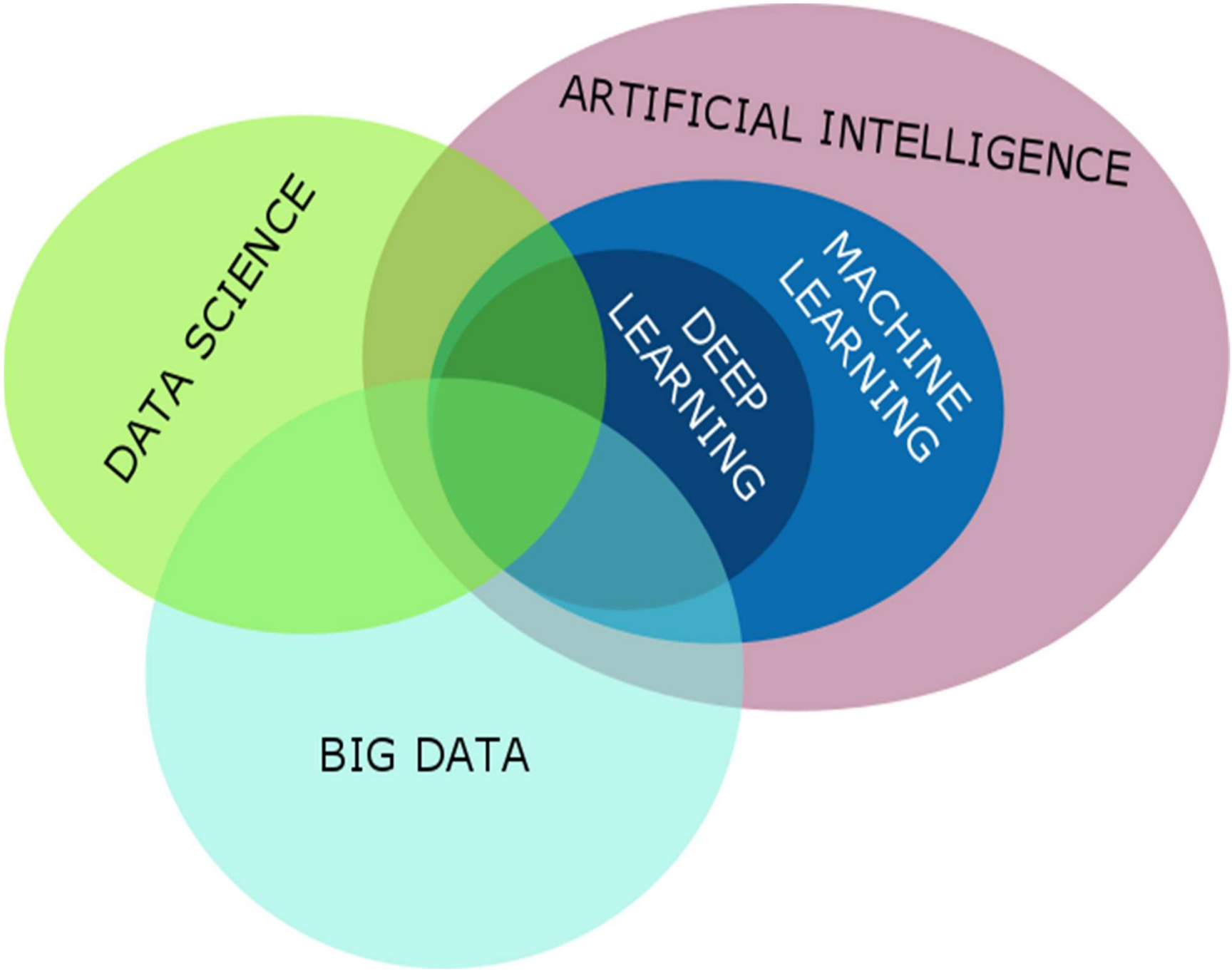
2017 január – Maluuba (Montreál, Kanada) – Microsoft fúzió

2018 március - Emmanuel Macron: AI kutatásra 1.850.000.000$

IBM Watson, Google Brain, Baidu, Facebook, Flickr, Twitter, DARPA, stb.


Registration Timeline
10000
8000
6000
4000
2000
Registrations
-75
-50
-25
0
25
50
Days
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002

AI az új szoftver


ARTIFICIAL INTELLIGENCE
DATA SCIENCE
MACHINE LEARNING
DEEP LEARNING
BIG DATA

Na jó, de mi is az a deep learning?

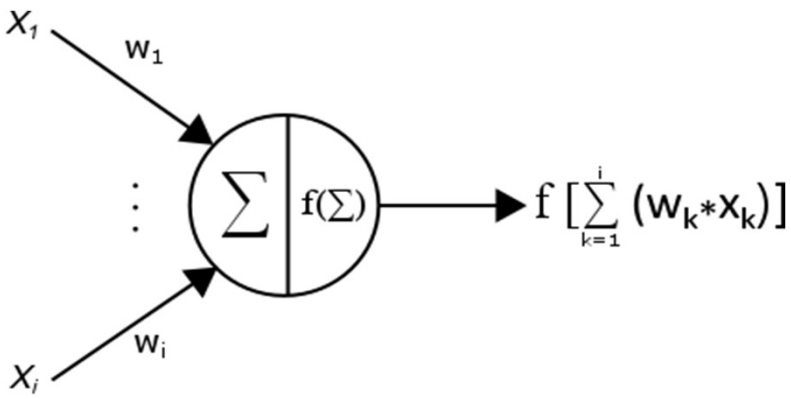
Gépi tanuló algoritmusok struktúrált összessége, ahol több rétegen keresztül próbáljuk az adatok különböző szintű absztrakcióit kinyerni és modellezni.

Gyakorlatban a deep learning kifejezést elsősorban a mély neuronhálókkal kapcsolatban használják.

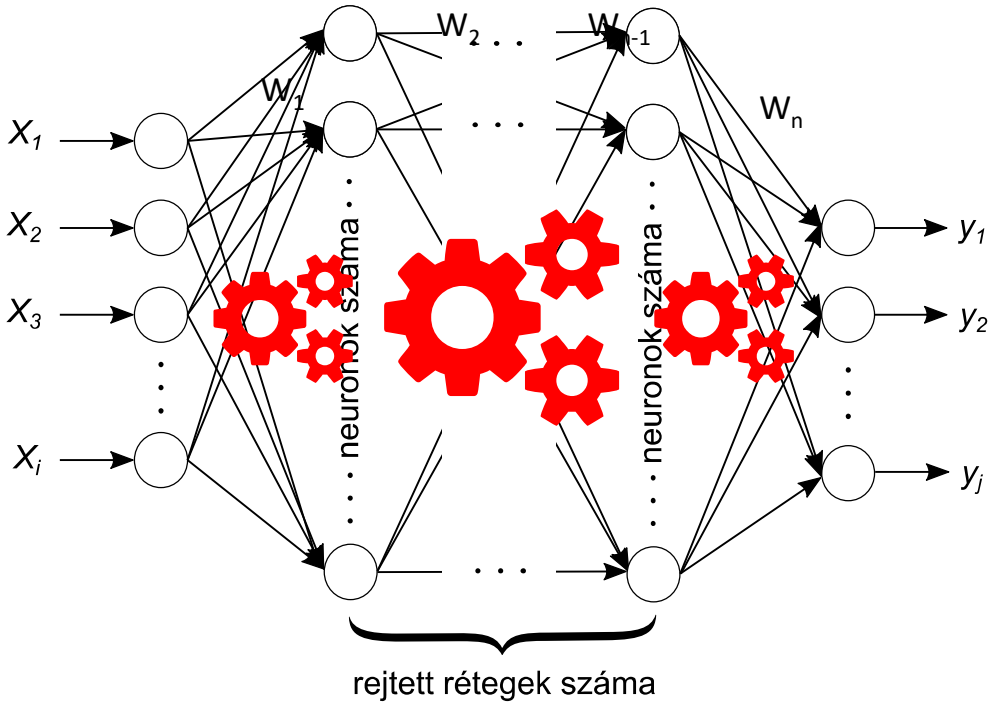
Korábbiaknál jobb modellek és nagyobb pontosság.

Új algoritmusok + nagymennyiségű adat + GPU

Előrecsatolt mély neuronháló



f(∑): nemlináris függvény
pl. f(∑)=max(0, ∑)



backpropagation

TANULÁS: súlyok „hangolása”

Hardware

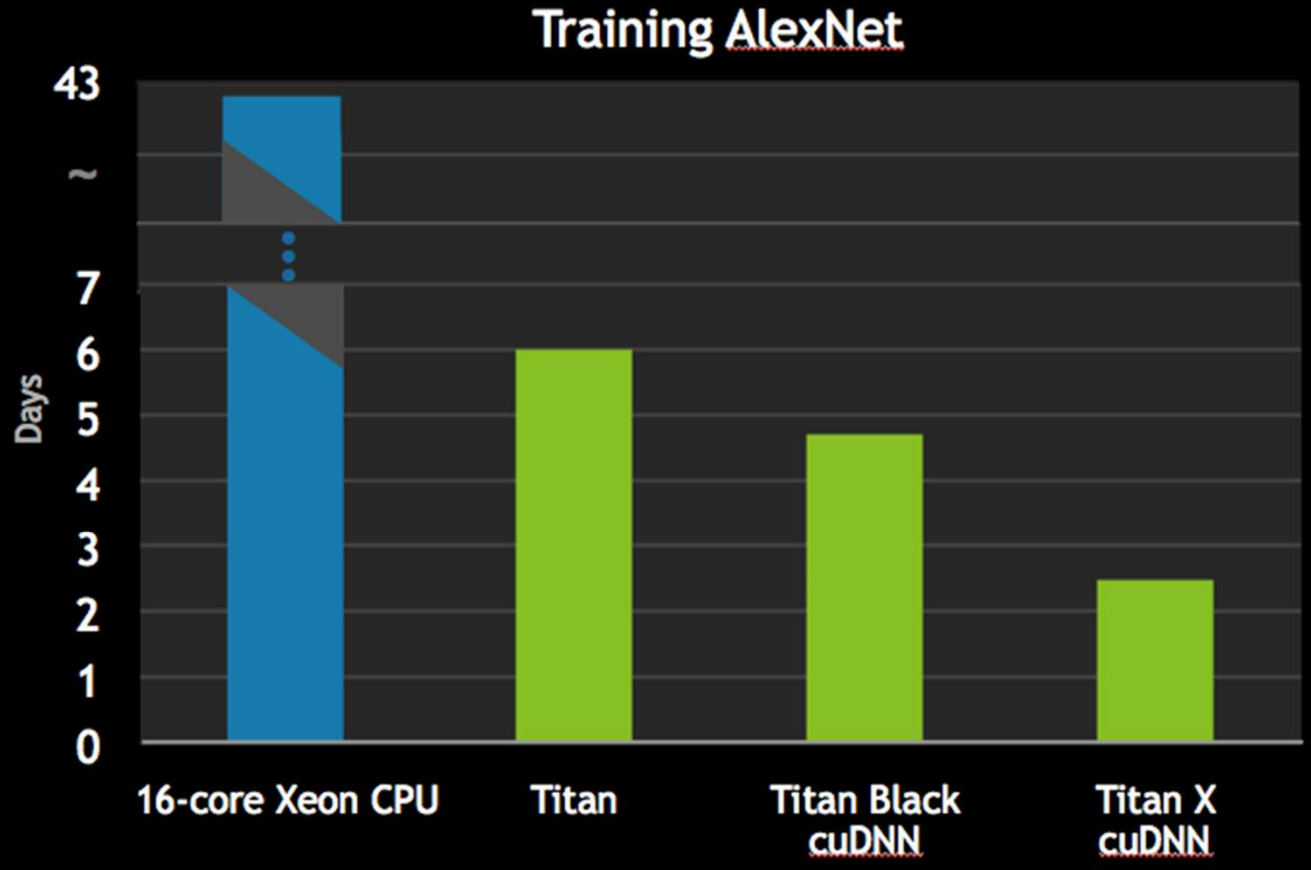
1-8 GPU workstation (min. 64 GB RAM, 2-4 TB HDD)

NVIDIA TITAN X, 1080Ti, TITAN Xp, Titan V, 2080Ti, RTX

- Titan V: 12 teraFLOP (FP32)
- Tensor Core: 110 Tflops (FP16, deep learning)
- Max. 12 GB GPU RAM

Cloud Services

TITAN X FOR DEEP LEARNING



http://blogs.nvidia.com/blog/2015/03/17/digits-devbox/



2020 szeptember:
RTX 3080 ($699, 10 GB RAM)
RTX 3090 (1499$, 24 GB RAM)

Hardware II.

Saját összeállítású Titan Xp workstation

DGX Station (480 Tflop), $69.000

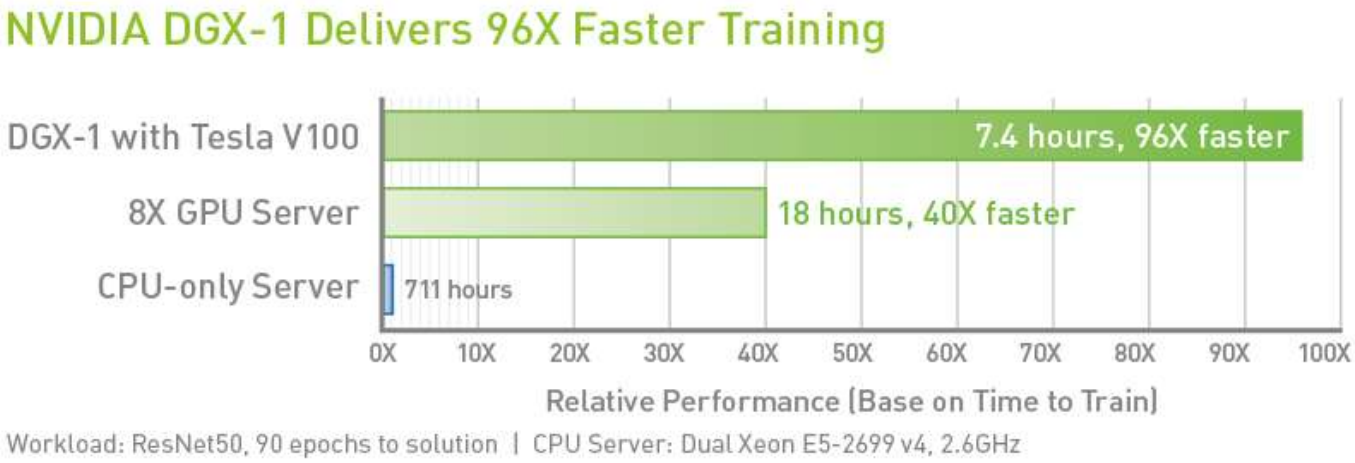
DGX-1 V100 (960 Tflop), $149.000

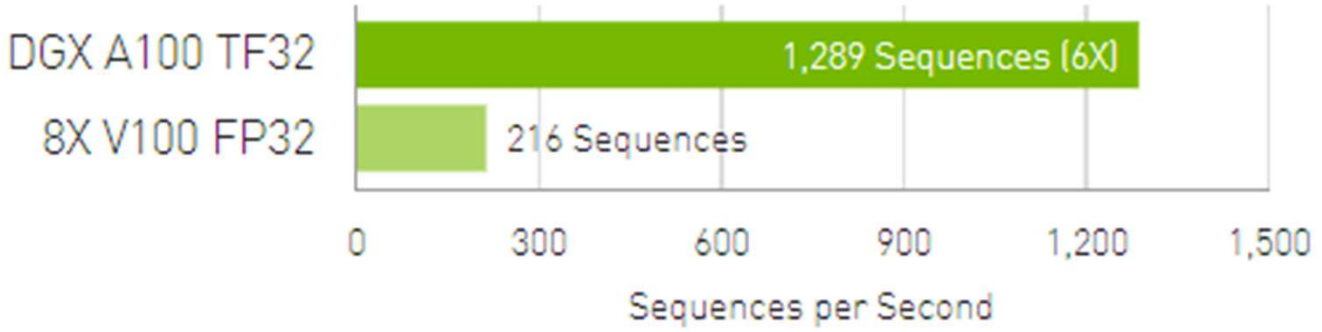
DGX-2 V100 (2 Pflops), $399.000

DGX A100 (5 Pflops DL), $199.000



DGX-1 teljesítmény





BERT Pre-Training Throughput using PyTorch including (2/3)Phase 1 and (1/3)Phase 2. Phase 1 Seq Len = 128, Phase 2 Seq Len = 512. V100: DGX-1 with 8X V100 using FP32 precision. DGX A100: DGX A100 with 8X A100 using TF32 precision.


DGX SATURNV
NVIDIA.


summit
IBM

Szoftver

Python alapú fejlesztés

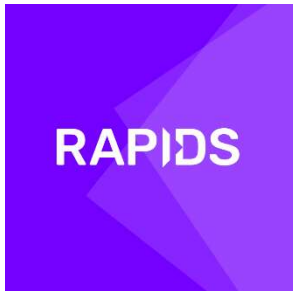






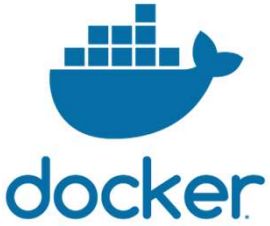
- TensorFlow (Google)
- Keras (Google)
- PyTorch (Facebook AI Research, Uber „Pyro” probabilistic programming)
- Microsoft CNTK, Chainer, Theano, Matlab DL Toolbox, etc.
- Caffe, Kaldi

- Rapids.ai



Szoftver architektúra - Skálázhatóság

Docker: Container alapú fejlesztés az egyszerűbb fejlesztés, tesztelés, éles futtatás és skálázhatóság céljából.



Kubernetes: a container alapú rendszerek automatizált menedzsmentje.



Apache Spark 2.3 + Kubernetes 1.7+: az elosztott adatfeldolgozás és modellezés menedzsmentje. Az MLLib könyvtárban van deep learning modul.



Szoftverkörnyezet telepítés

GPU:

NVidia videókártya driver

CUDA, cuDNN

```
pip install tensorflow-gpu
pip install keras
```

CPU:

```
pip install tensorflow
pip install keras
```

Kis házi feladat

1. Linux erősen ajánlott!
2. Ha van GPU-d, akkor telepítsd hozzá a szükséges drivereket, továbbá a CUDA, cuBLAS, cuDNN könyvtárakat.
3. Telepítsd a szükséges programokat a Keras+TensorFlow vagy PyTorch/Torch7 keretrendszerekhez.
4. Telepítsd a Jupyter notebook-ot Python/LUA nyelvekhez.
5. Python/LUA nyelven olvass be öt darab tetszőleges 128x128 pixeles színes képet, jelenítsd meg és írd ki az RGB csatornáit tartalmazó tömbök (mátrixok, tensorok) tartalmát.
6. Csatornánként számold ki az egyes pixelek átlagát és szórását az öt képre, majd alakítsd át ezeket 0 várható értékű, 1 szórású adathalmazzá. Ezt követően ellenőrizd a kapott adathalmaz várható értékét és szórását.
7. Python vagy LUA nyelven olvass be egy tetszőleges hangfájlt és jelenítsd meg a spektrogramját.
8. Python/LUA nyelven tölts le a https://en.wikipedia.org/wiki/Deep_learning oldal szöveges tartalmát, jelenítsd meg a szöveges tartalmat, továbbá hisztogramon jelenítsd meg a tartalomban a betűk előfordulásának gyakoriságát.

Határidő: 2020. szeptember 29. 23:59

Kapcsolódó linkek

- Tárgy honlapja (legfontosabb infók)
  http://smartlab.tmit.bme.hu/oktatas-deep-learning

- Tárgy Twitter csatorna (hírek)
  https://twitter.com/VITMAV45
  (saját: https://twitter.com/bapalto)

- Microsoft Teams csoport
- ~~Tárgy Google Groups (vita, kérdések)~~
  ~~https://groups.google.com/forum/#!forum/vitmav45~~

- Tárgy Google Form (anonim javaslatok, kritika)
  http://goo.gl/forms/sGHU5mVeQdnDN8gM2

Deep learning nálunk

- Önálló laboratórium
- Szakdolgozat
- Diploma
- PhD
- Cikkek
- Versenyek

Bővebb infó: email-ben.

Köszönöm a figyelmet!

toth.b@tmit.bme.hu



twitter.com/bapalto