Mondragon Unibertsitatea

Faculty of Engineering

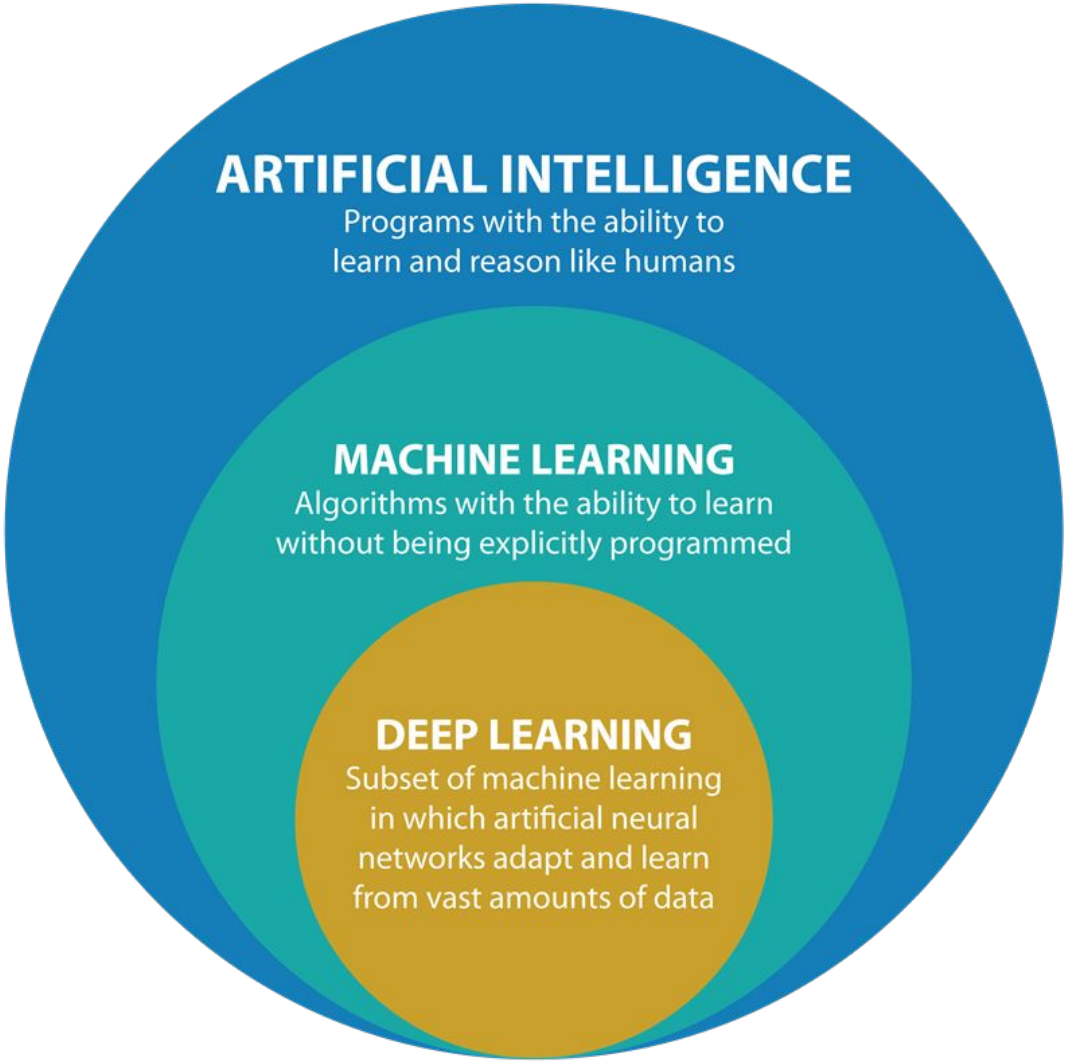
Advanced Machine Learning

Lecture 0: Introduction

Context

¿Machine Learning VS Deep Learning?

Context



Source:Argility.com

Machine Learning ≠ Deep Learning

Deep Learning = Machine Learning

Context

Mondragon Unibertsitatea
Faculty of Engineering

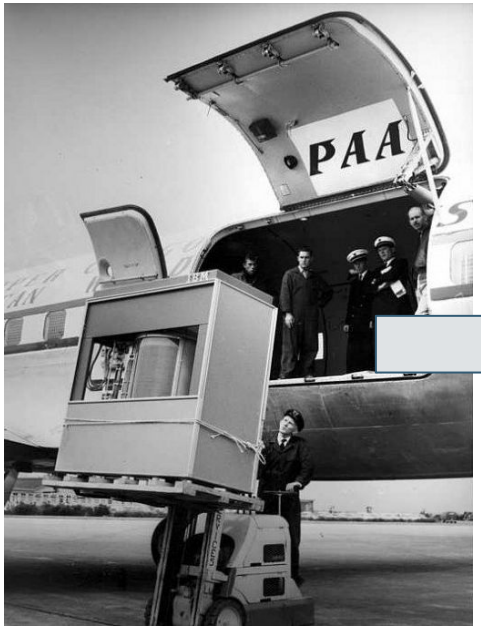
- Deep Learning ~ Neural Networks (multiLayer)
- Neural Networks
  - Phase 1: 1950 - 1960 (Hessian Learning)
  - Phase 2: 1975 - 1990 (introducción de la *“backpropagation”*)
  - Phase 3: 2005 - … (Deep Learning)
- Why the DeepLearning has not appeared before?
  - Lack of data (storing data was expensive)
  - Lack of Computation brute force

Context

Mondragon Unibertsitatea
Faculty of Engineering

- Lack of data:

1956 - 5 MB



X2

1980 - 10 MB

X51200

2020 - 512 GB

Pasa el ratón por encima de la imagen para ampliarla

SanDisk SDSQUAR-512G-GN6MA, Tarjeta de Memoria Ultra Android microSDXC UHS-I con Adaptador SD, Velocidad de Lectura hasta 100 MB/s, Clase 10, U1 y A1, 512 GB, Rojo/Gris

de SanDisk

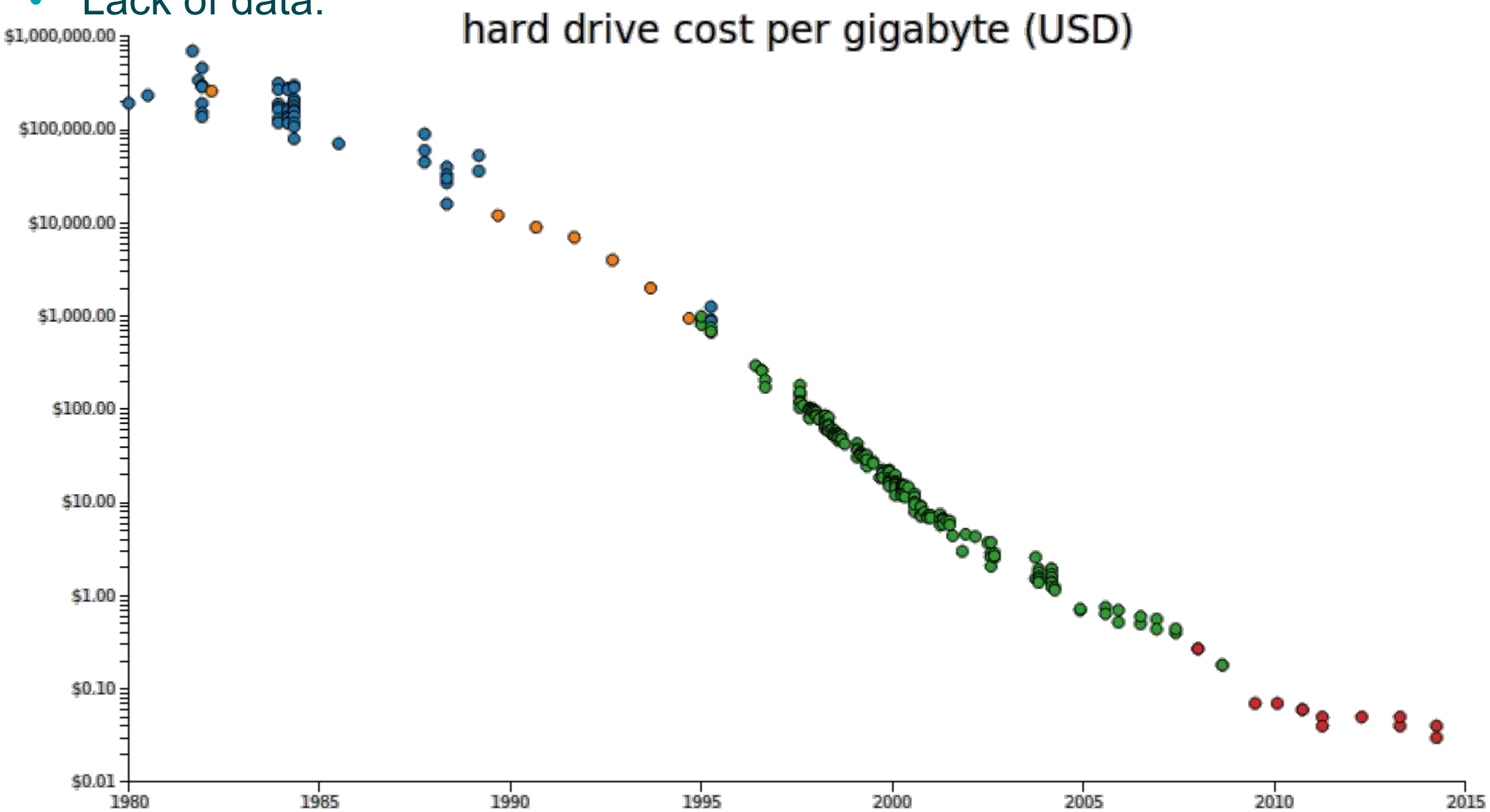
160.324 valoraciones | 701 preguntas respondidas

Precio anterior: ~~118,82 €~~
Precio: 114,99 € **Envío estándar GRATIS** para clientes Prime Detalles
Ahorras: 3,83 € (3%)
Precio final del producto

Context

Mondragon Unibertsitatea
Faculty of Engineering

- Lack of data:



source: mkomo.com

Context

- Lack of data:

DNA Storage

STORAGE LIMITS

Estimates based on bacterial genetics suggest that digital DNA could one day rival or exceed today's storage technology.

	Hard disk	Flash memory	Bacterial DNA
Read–write speed (μs per bit)	~3,000–5,000	~100	<100
Data retention (years)	>10	>10	>100
Power usage (watts per gigabyte)	~0.04	~0.01–0.04	$<10^{-10}$
Data density (bits per cm^3)	$\sim10^{13}$	$\sim10^{16}$	$\sim10^{19}$

WEIGHT OF DNA NEEDED TO STORE WORLD'S DATA

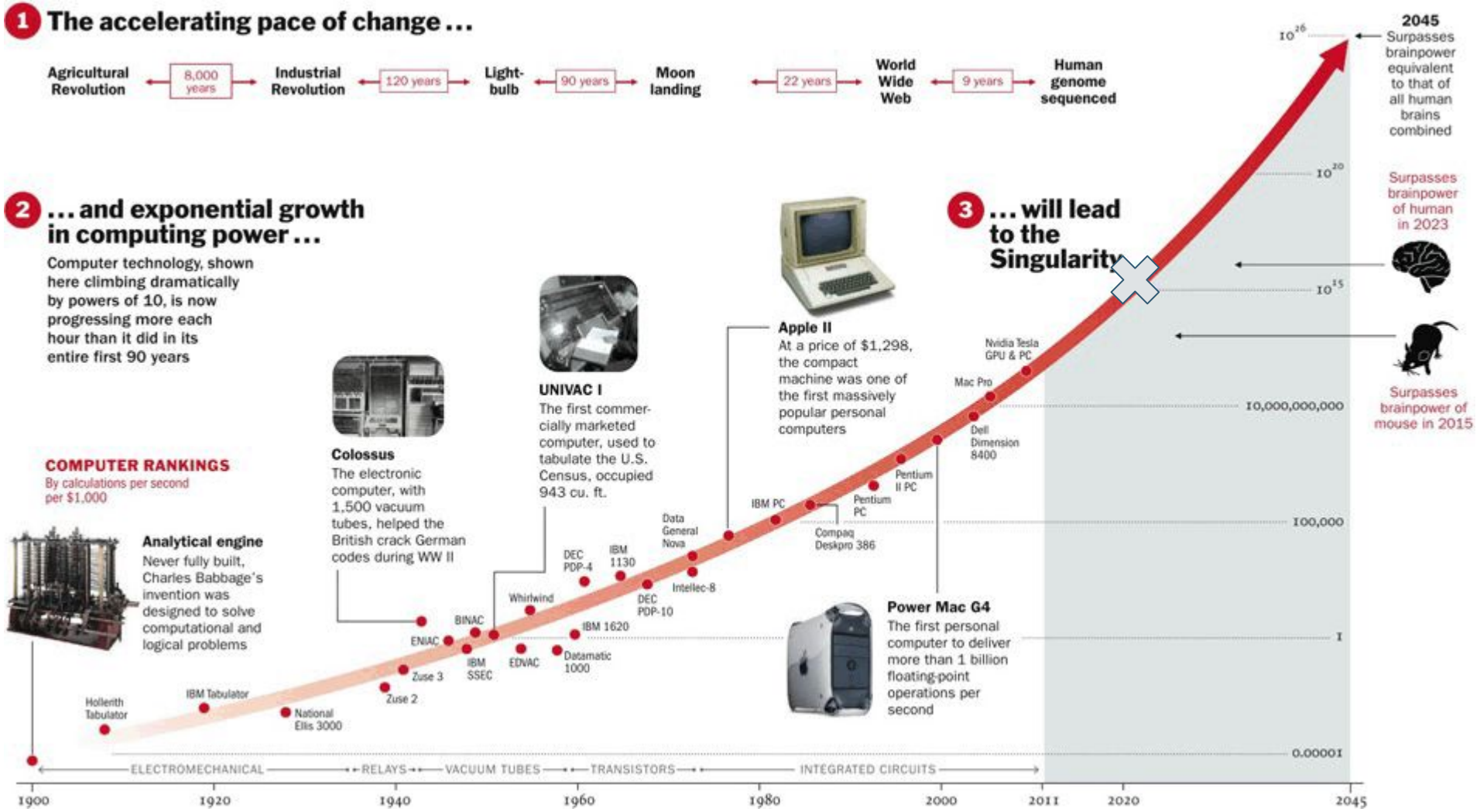
~1 kg

©nature

source: nature.com

Context

Mondragon Unibertsitatea
Faculty of Engineering



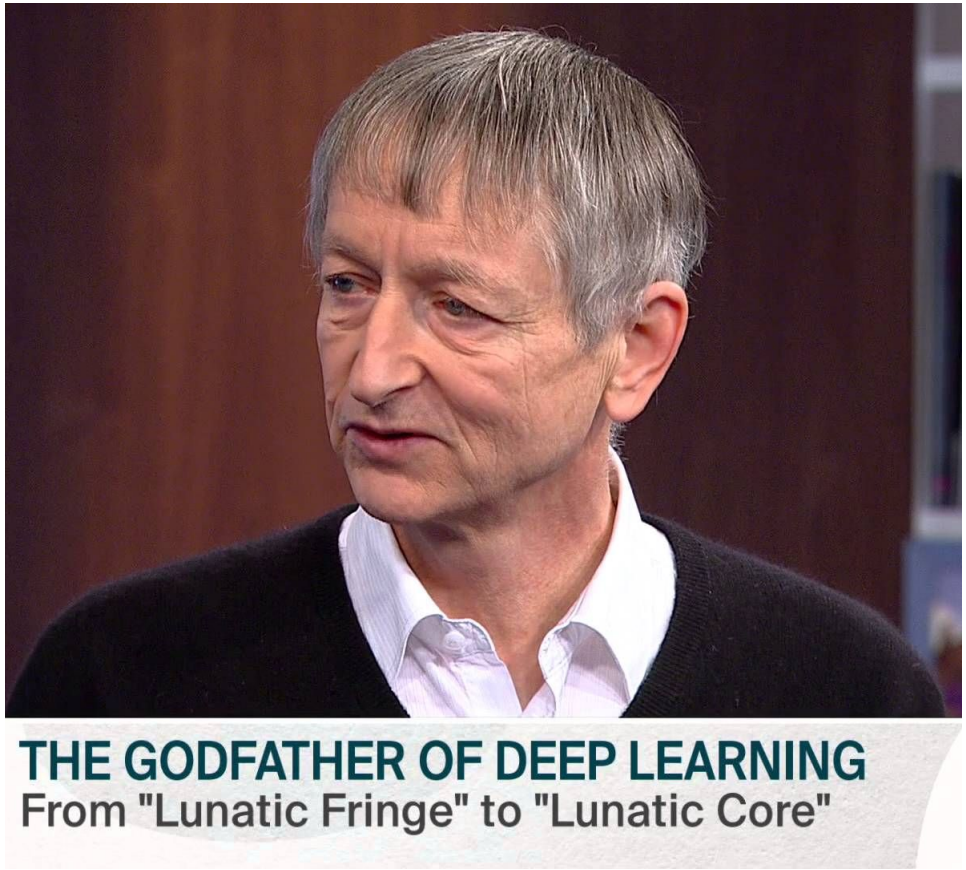
source: Time magazine

We have achieved a good state for Neural Networks and Deep Learning

Context

Mondragon Unibertsitatea
Faculty of Engineering

- The “Father” of Deep Learning: Geoffrey Hinton

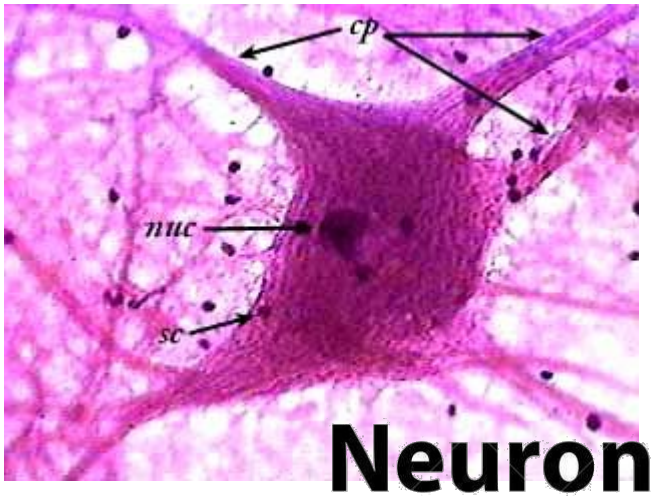


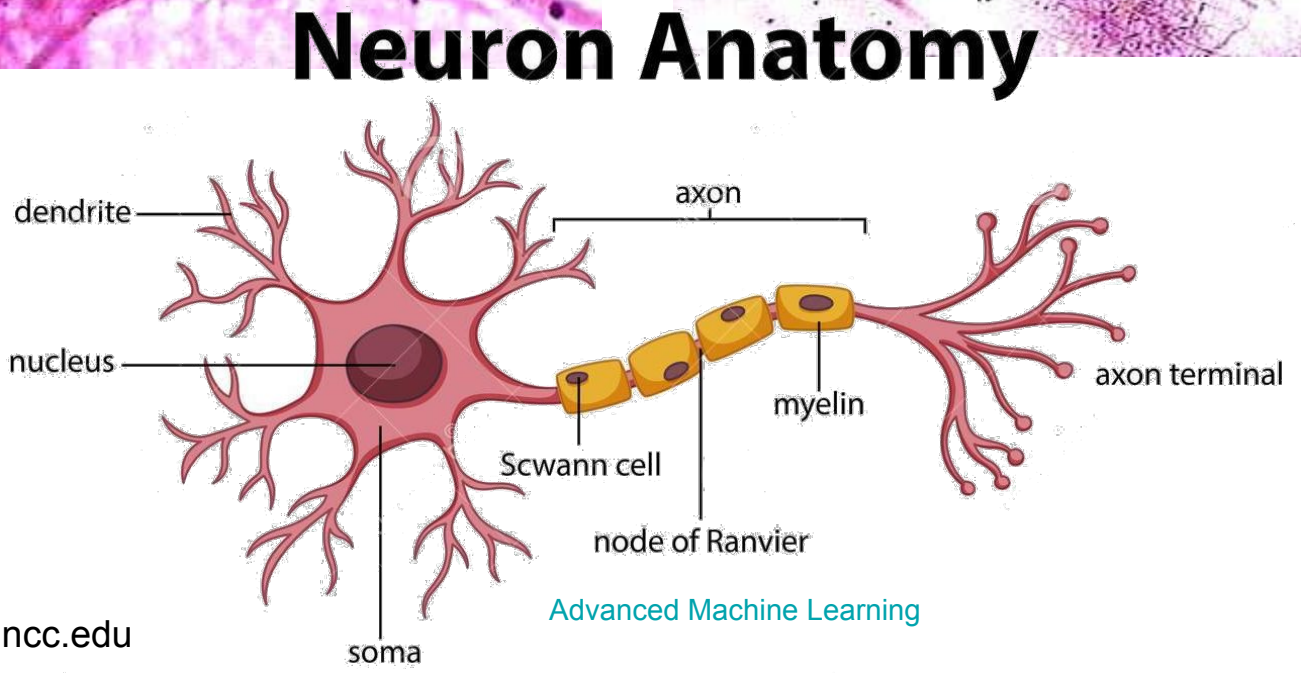
What is Deep Learning??

Context

Mondragon Unibertsitatea
Faculty of Engineering

- How does the human brain “work”?

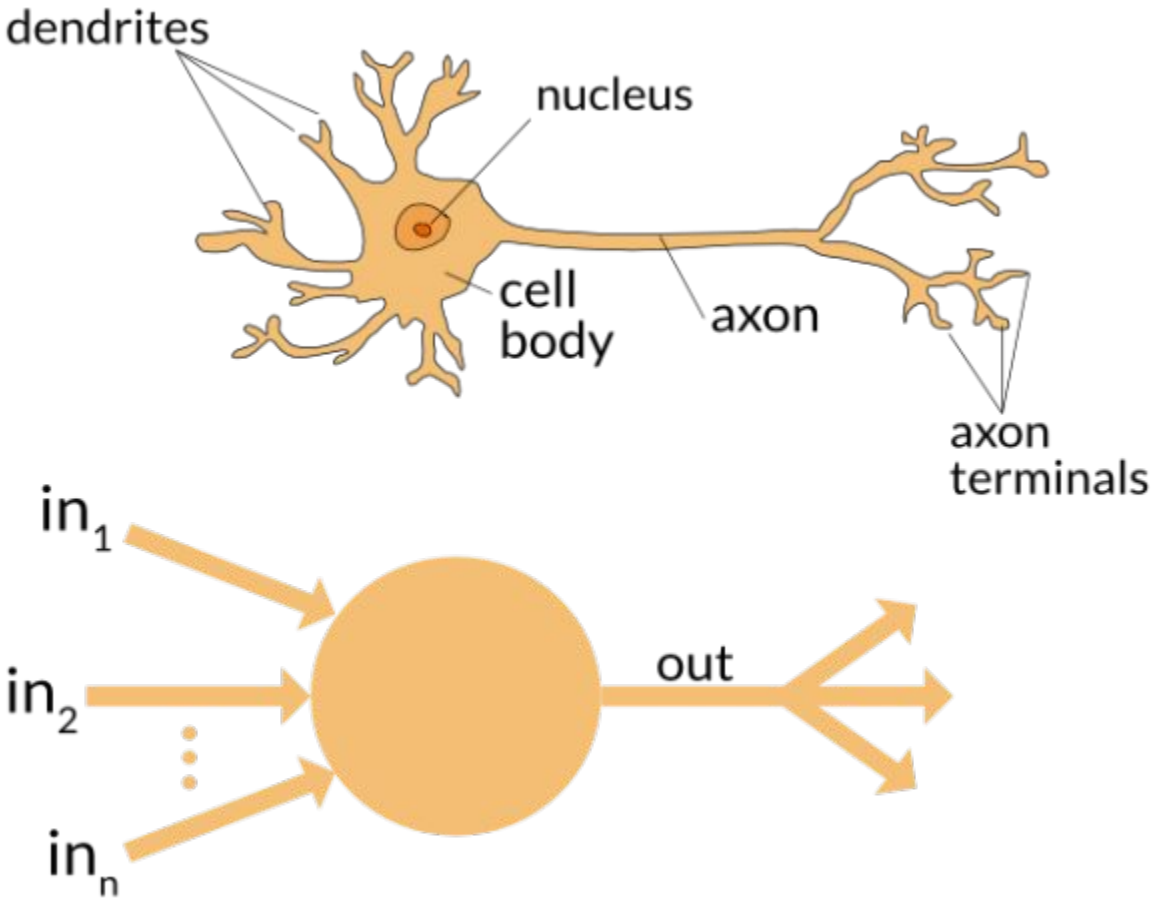




source: austincc.edu

Context

- Human Neuron VS ANN Neuron (or perceptron)

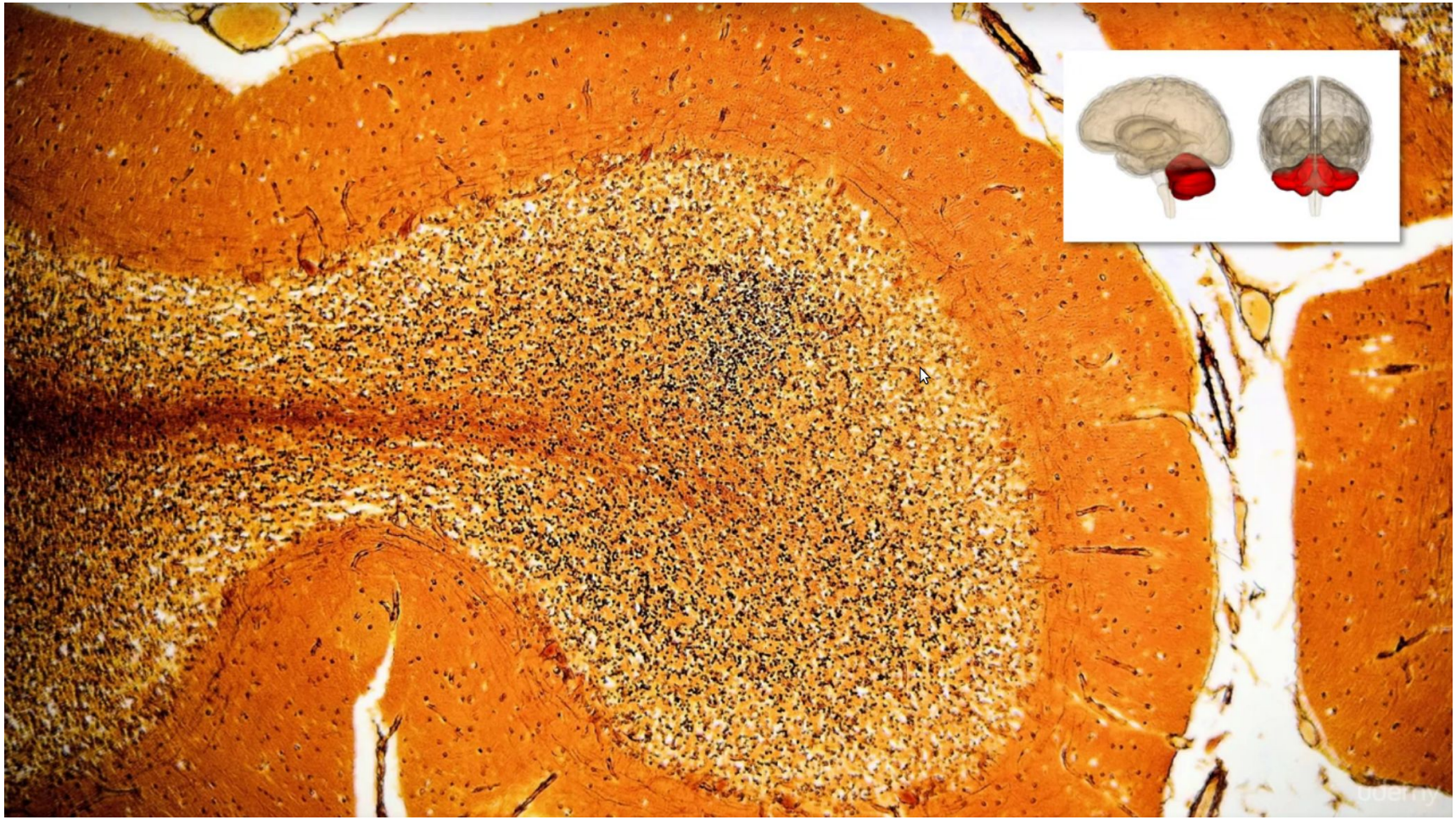


source: analyticsvidhya.com

Context

Mondragon Unibertsitatea
Faculty of Engineering

- neurons interconections

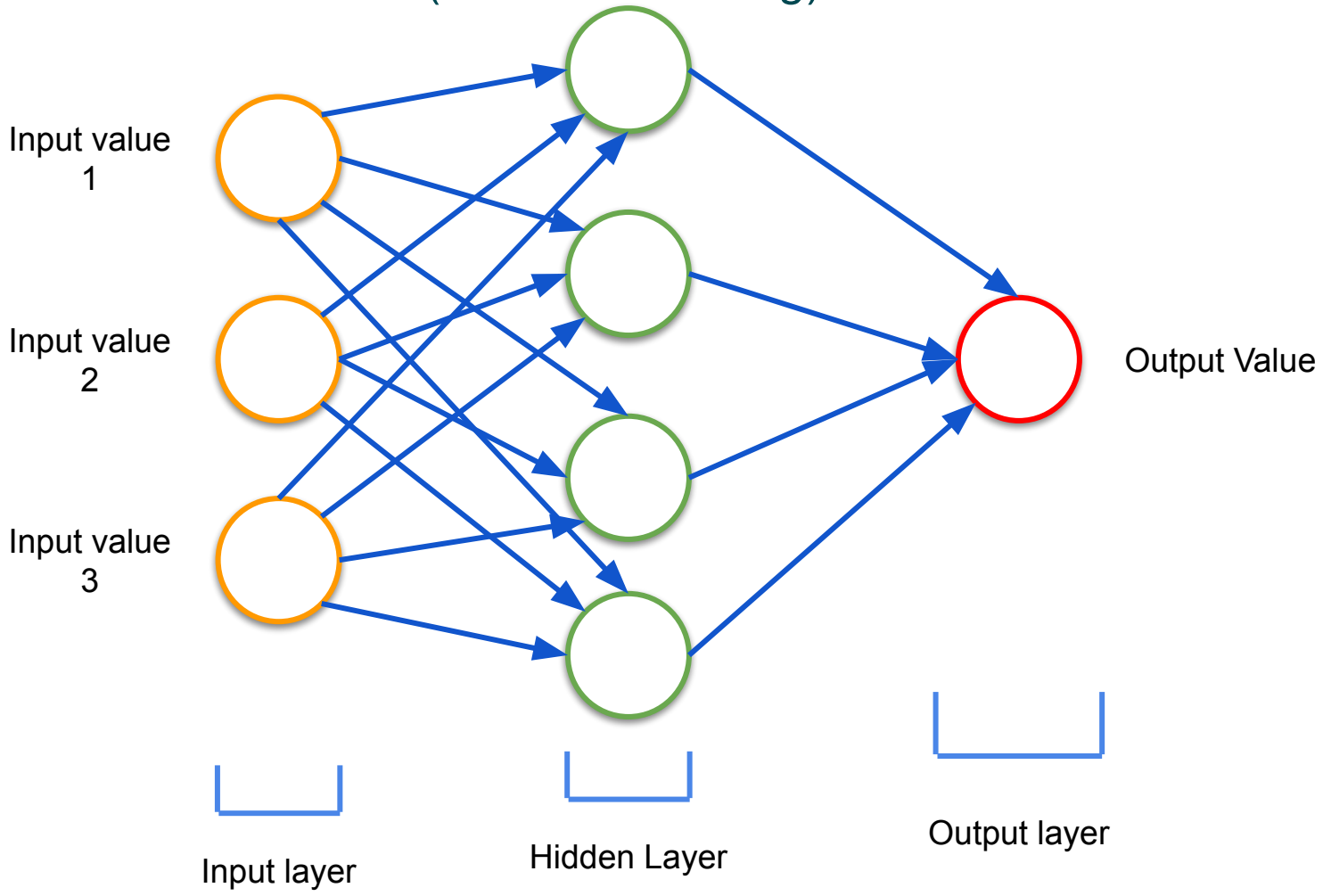


source: science photo library

Context

Mondragon Unibertsitatea
Faculty of Engineering

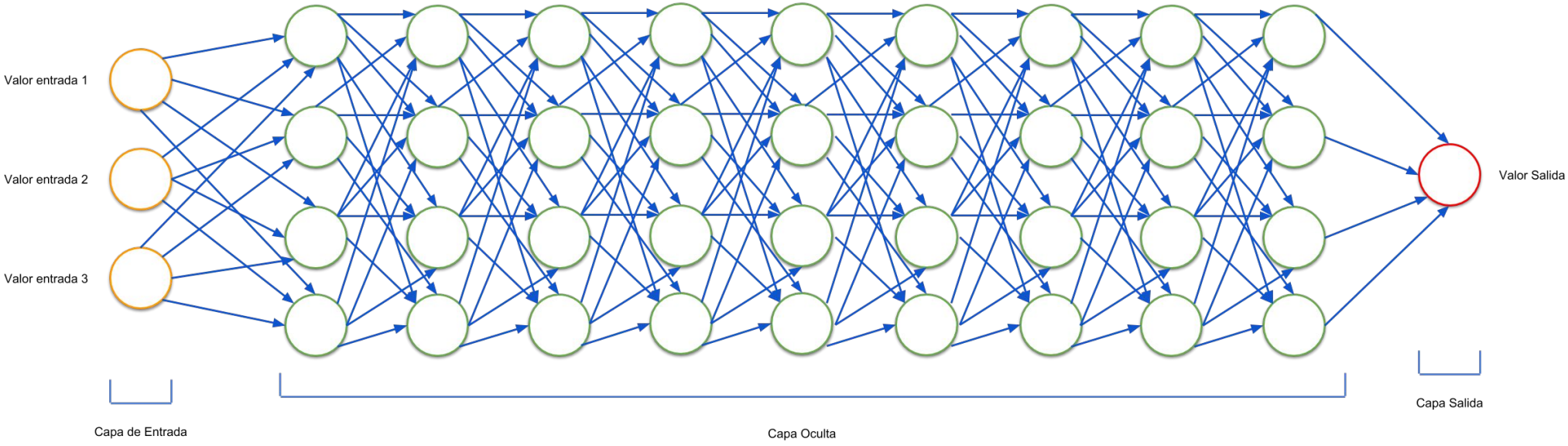
- Neurons interconnection (Shallow Learning)



Context

Mondragon Unibertsitatea
Faculty of Engineering

- Neuron Interconnection (Deep Learning)

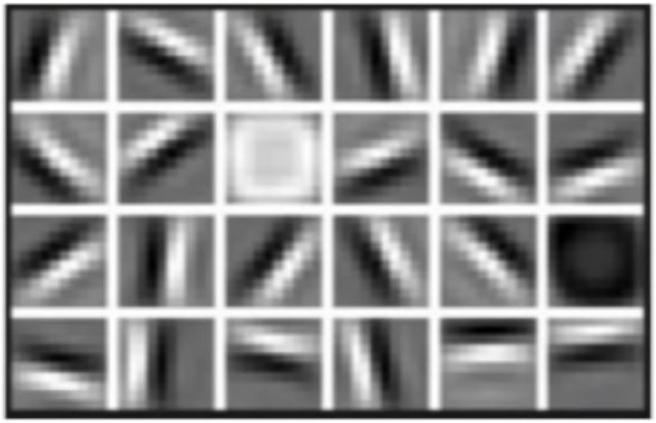


Context

Mondragon Unibertsitatea
Faculty of Engineering

- Why Deep Learning?
  - Traditional ML needs hand engineered features
    - Time consuming
    - not scalable
  - Can we learn the **underlying features** directly from data?

Low Level Features



Lines & Edges

Mid Level Features

Eyes & Nose & Ears

High Level Features

Facial Structure

Context

Mondragon Unibertsitatea
Faculty of Engineering

- Conclusions:
  NNs date back decades, so why the resurgence?

1952 Stochastic Gradient Descent

1958 Perceptron
- Learnable Weights

1986 Backpropagation
- Multi-Layer Perceptron

1995 Deep Convolutional NN
- Digit Recognition

1. Big Data

- Larger Datasets
- Easier Collection & Storage

IMAGENET

WIKIPEDIA
The Free Encyclopedia

2. Hardware

- Graphics Processing Units (GPUs)
- Massively Parallelizable

3. Software

- Improved Techniques
- New Models
- Toolboxes

TensorFlow

Frameworks

FrameWorks for DeepLearning

- Community / University
- Private company
- Facebook
- Microsoft
- Google
- Apache (AWS)

PYTHON

C / C++

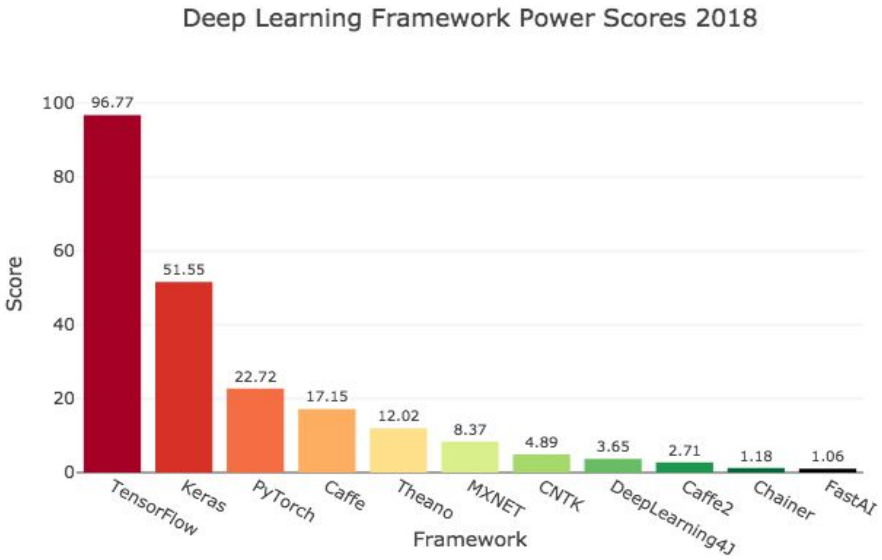
Java

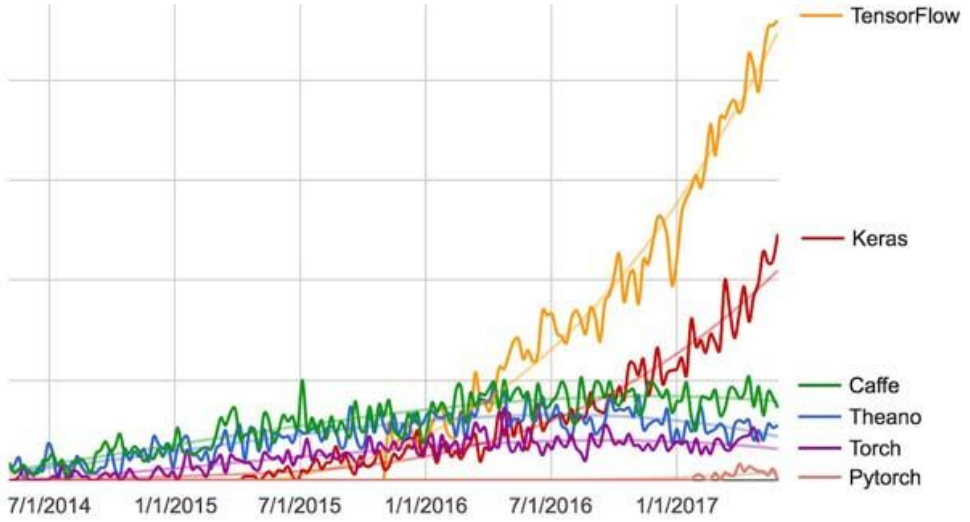
R

Matlab

FrameWorks for DeepLearning

Mondragon Unibertsitatea
Faculty of Engineering





* not percentages! weights that depend on:

- work offers
- google searches
- scientific papers
- Deep Learning books
- GitHub activity

Interest dates:

- Tensorflow - Nov. 2015
- Keras - Mar. 2015 (theano)
- PyTorch - Oct 2016
- Caffe - Abril 2017
- Theano - 2007

Source: towardsdatascience.com

FrameWorks for DeepLearning

theano

- First deep learning Framework (Open Source)
- Made by Montreal university
- python based
- discontinued since 2017

- Open source
- written in LUA with C, Lua, C++ interfaces
- flexible and easy model creation

- GOOGLE Open Source Framework
- One of the most used nowadays
- Interfaces for: python, C,Java,GO, R, Julia
- Complex but very flexible

FrameWorks for DeepLearning

Mondragon Unibertsitatea
Faculty of Engineering

Keras

- Interface for simplifying the DL model creation
- created by François Chollet (google engineer)
- based on python (interfaces: Python & R)
- Google has integrated Keras on tensorflow 2.0
- Can be used on top of Theano, Tensorflow o CNTK
- Simple but flexibility is lost as a tradeoff

Caffe2

- Caffe: Modularity and speed centered framework for CNN
- Written in C++ (interfaces: C++, Python & Matlab)
- If only CNN are going to be used, it is the best
- Caffe2: Facebook: Caffe + many pretrained models, in order to develop applications faster

FrameWorks for DeepLearning

PYTORCH

- Torch has no python interface and Facebook has developed it
- Simple, fast and flexible
- along tensorflow+keras the most used framework

mxnet

- DL from Amazon
- Interfaces: JavaScript, R, Go, Python & C++
- Few developers, little documentation
- Gluon is the “Keras” of MXNET

FrameWorks for DeepLearning

Mondragon Unibertsitatea
Faculty of Engineering

DEEPLEARNING4J

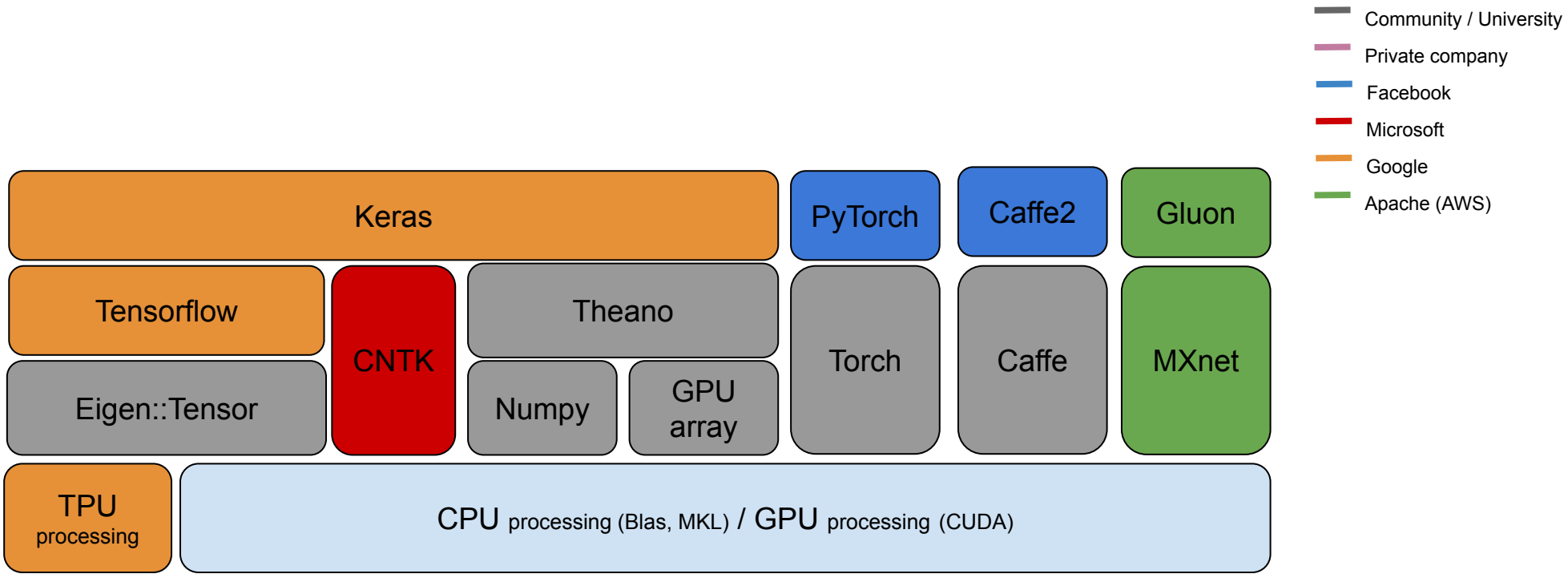
- open source and commercial
- can import almost every model created with the other frameworks
- best Java option
- can run on top of Spark

Microsoft CNTK

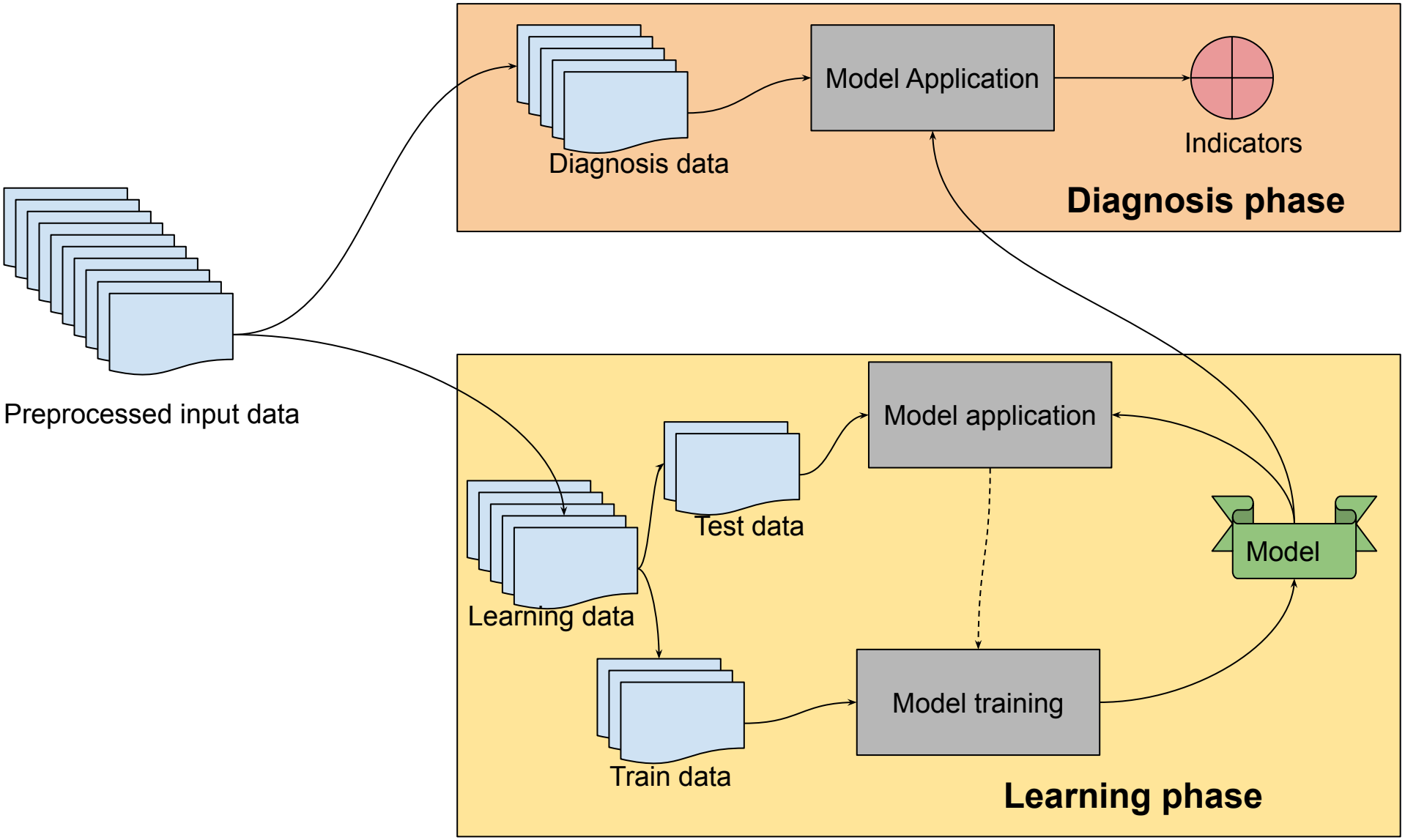
- DL from Microsoft (OpenSource y Comercial)
- written in C++
- Interfaces Python (keras), c++, cmd, .net
- very good performance under windows

FrameWorks for DeepLearning

Mondragon Unibertsitatea
Faculty of Engineering



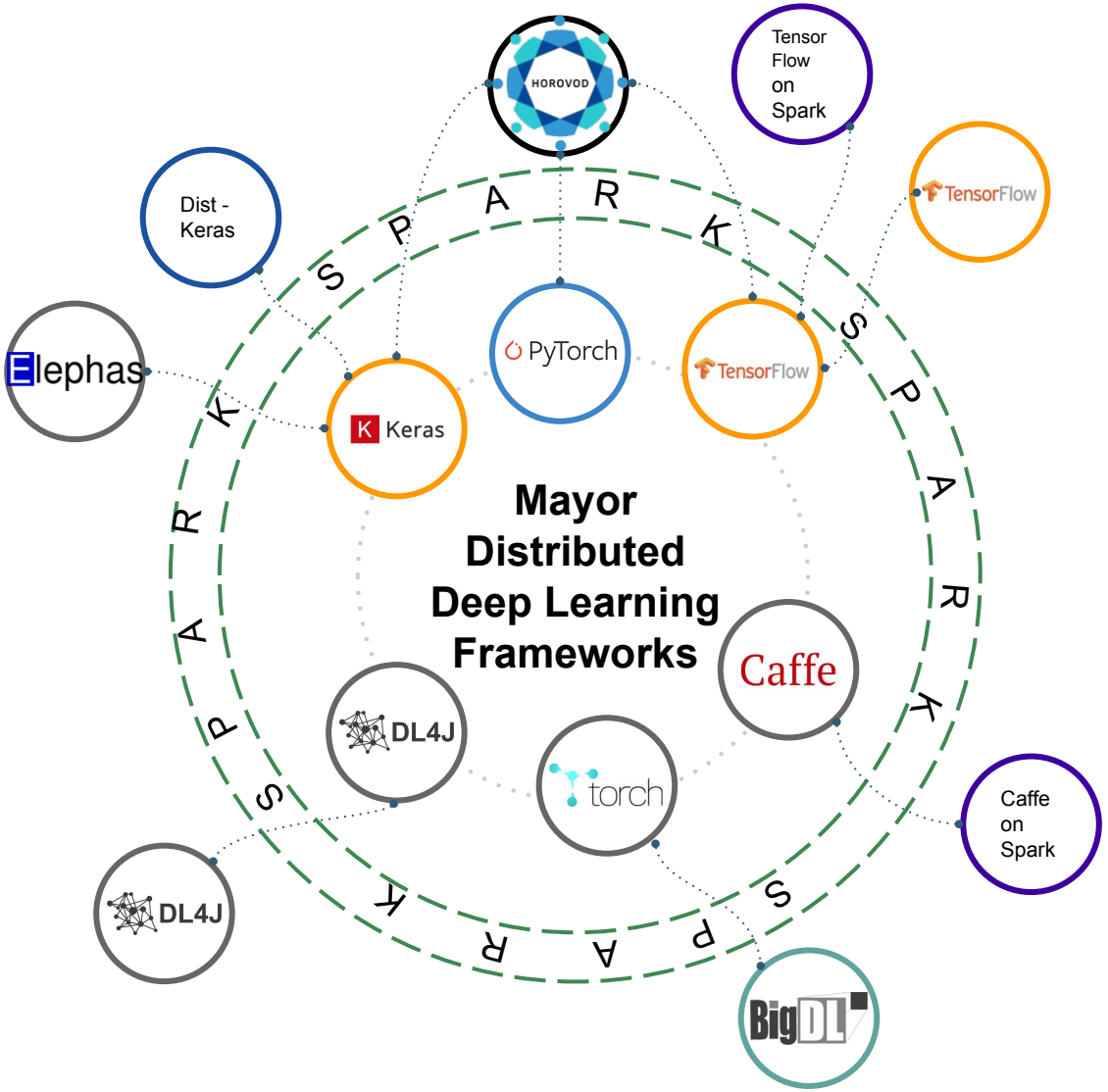
ML modeling


Diagnosis data
Model Application
Indicators
Diagnosis phase
Preprocessed input data
Model application
Test data
Model
Learning data
Model training
Train data
Learning phase

Distributed DL FrameWorks

- Almost every framework until now, is a single computer
- But new frameworks are appearing that bassically run on top of spak

DL FW with distributed learning



- Community / University
- Google
- Facebook
- Uber
- Yahoo
- CERN
- INTEL

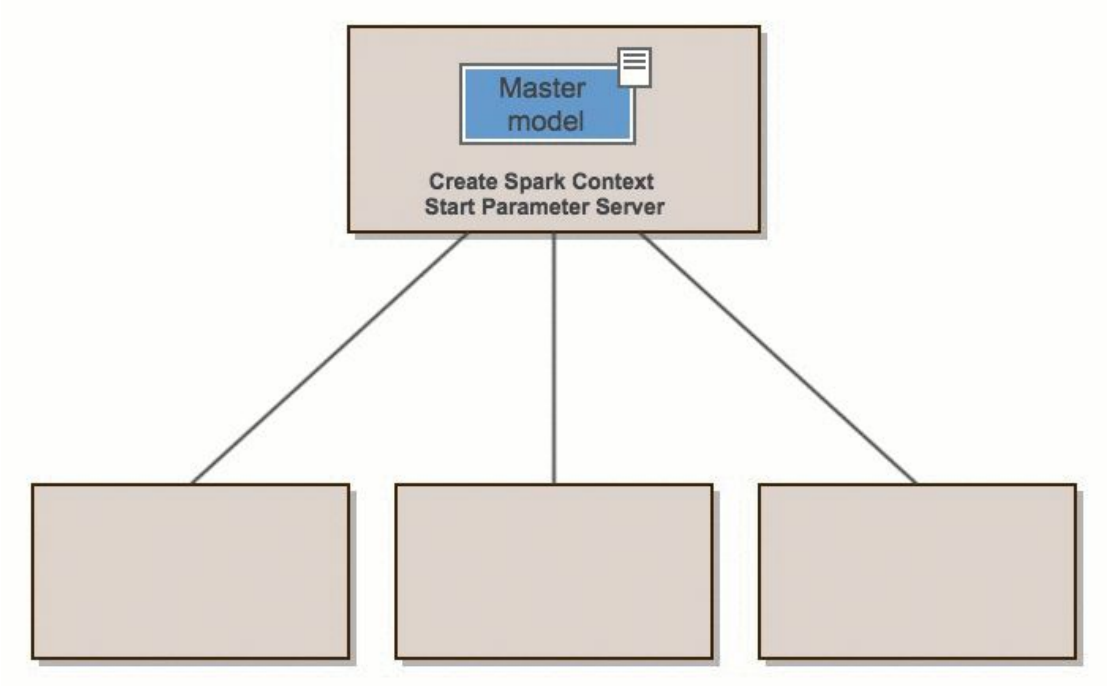
NOTE: All this is NEW!

- Horovod (Jun 2018)
- TensorFlowonSpark (Dic 2017)
- CaffeOnSpark (No hay release oficial)
- Dist-Keras (Oct 2016)
- BigDL (Apr 2017)
- Elephas (Jun 2016)

Distributed DL Frameworks

Mondragon Unibertsitatea
Faculty of Engineering

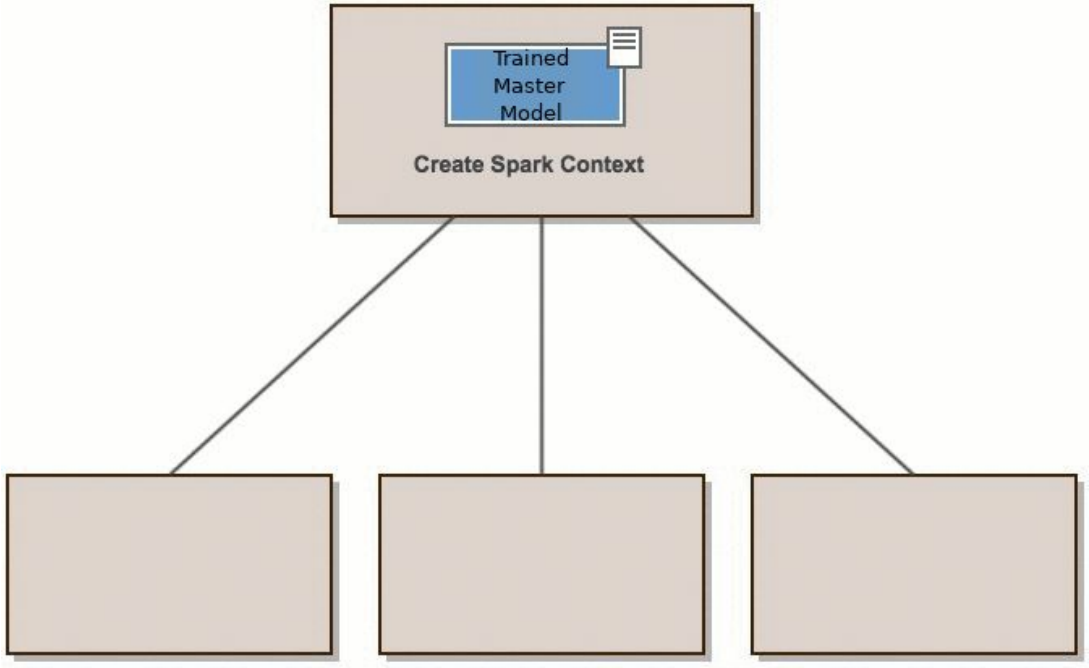
- Main Idea (**Learning phase**):
  a. pass subset of data to nodes (with hadoop, you skip this part)
  b. replicate master model in workers
  c. pass initial parameters
  d. Each node responds with upgraded parametros (per Epoch)
  e. Master model renews the parameters depending the response of the node



Distributed DL Frameworks

Mondragon Unibertsitatea
Faculty of Engineering

- Main Idea (**Diagnosis**):
  a. pass subset of data to each node
  b. Replicate **Trained Model** in workers
  c. every node evaluates the data



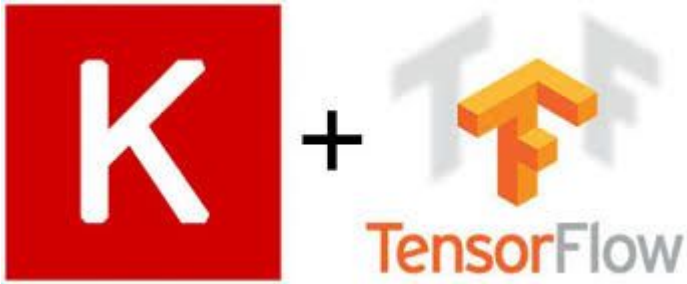
Distributed DL Frameworks

- The use of these, depends on the existing amount of data
- The learning phase gets complicated with Distributed DL
- The evaluation phase is easier to deploy

Theory in this Lecture:

Mondragon Unibertsitatea
Faculty of Engineering

0. Context and Frameworks
1. Deep Learning basics
   1.1. Artificial Neural Networks
   1.2. ML basics for NN (RECAP)
2. Supervised learning with DL:
   2.1. Recurrent Neural Networks
   2.2. Convolutional Neural Networks
3. Unsupervised / Self-Supervised learning with DL:
   3.1. Self Organizing Maps
   3.2. Generative Modelling
4. Conclusions



Evaluation of course

Mondragon Unibertsitatea
Faculty of Engineering

IKASTE-EMAITZAK

RA141 **Datu aurreratuen analisi eta aurreprozesamendu kontzeptuak identifikatzen ditu**

FORMAZIO-AKTIBITATEAK	IO	IG	OG
Banaka zein taldean egindako POPBL/proiektuei lotutako memoriak, txostenak, ikusentzunezko materiala, etab., garatzea, idaztea eta aurkeztea	3 h.	8 h.	11 h.
Banakako lana eta ikasketa, probak eta azterketak eta/edo kontrol-puntuak	1 h.	2 h.	3 h.
Ikasgelan aurkeztea klase parte-hartzaileetan ikasgaiekin loturiko kontzeptu eta prozedurak	4 h.		4 h.
Ariketa, problema nahiz praktikak egin eta ebaztea bakarka eta taldean	5 h.	7 h.	12 h.

EBALUAZIO-SISTEMAK	P	ERREKUPERAKETA-MEKANISMOAK
Froga idatziak, kodifikazio/programaziokoak eta ahozko indibidualak ikasgaiari buruzko konpetentzia teknikoak ebaluatzeko	%100	Banakako proba idatziak eta ahozkoak gaiari buruzko gaitasun teknikoak ebaluatzeko

IO - Irakastorduak: 13 h.
IG - Irak. gabekoak: 17 h.
OG - Orduak guztira: 30 h.

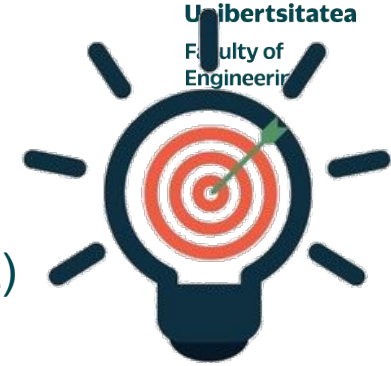
RA142 **Bere kabuz zein taldean, soluzioak proposatu eta garatzen ditu, zeintzuen oinarrian datuen analisia dagoen, eta, betiere, ikaskuntza automatiko aurreratuaren kontzeptuak erabiliz**

FORMAZIO-AKTIBITATEAK	IO	IG	OG
Banaka zein taldean egindako POPBL/proiektuei lotutako memoriak, txostenak, ikusentzunezko materiala, etab., garatzea, idaztea eta aurkeztea	5 h.	11 h.	16 h.
Banakako lana eta ikasketa, probak eta azterketak eta/edo kontrol-puntuak	1 h.	4 h.	5 h.
Ikasgelan aurkeztea klase parte-hartzaileetan ikasgaiekin loturiko kontzeptu eta prozedurak	9 h.		9 h.
Ariketa, problema nahiz praktikak egin eta ebaztea bakarka eta taldean	4 h.	11 h.	15 h.

EBALUAZIO-SISTEMAK	P	ERREKUPERAKETA-MEKANISMOAK
Txostenak ariketak egiteari, kasuen azterketari, ordenagailuko praktikei, simulazio praktikei, eta laborategiko	%40	Banakako proba idatziak eta ahozkoak gaiari buruzko gaitasun teknikoak ebaluatzeko
Gaitasun teknikoa, PBL/proiektuan inplikatzea, egindako lana, lortutako emaitzak, entregatutako dokumentazioa, aurkezpena eta defentsa teknikoa	%40	
Froga idatziak, kodifikazio/programaziokoak eta ahozko indibidualak ikasgaiari buruzko konpetentzia teknikoak ebaluatzeko	%20	

PBL Hours -> 8 + 19 = 27h
Exam Hours -> 2 + 6 = 8h
Theory Hours -> 13h
Project Hours -> 9 + 18= 26h

Installing environment for this course



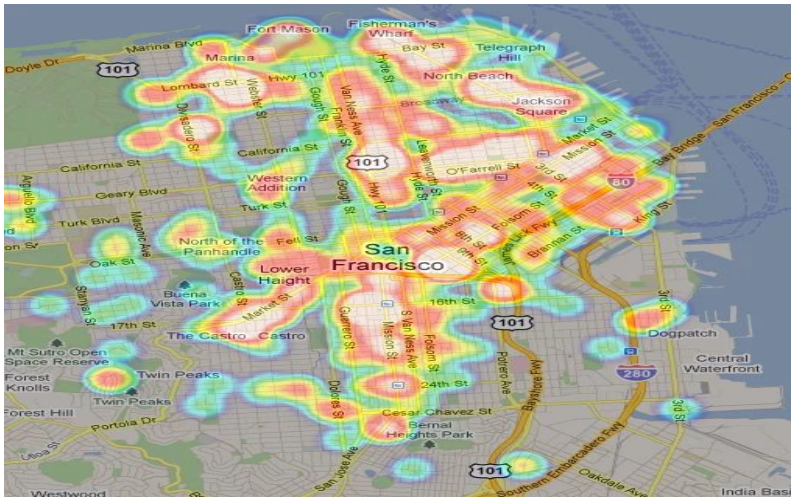
- Basically, tensorflow 2.x already packs Keras. (CPU)
  - create python environment with python 3.7 (3.8 not supported yet)
  - install either using conda or pip tensorflow
- If your system has an NVIDIA GPU:
  - Install Nvidia Drivers
  - Install Nvidia CUDA
  - create python environment with python 3.7
  - CONDA install tensorflow-GPU (automatically installs cudNN)
- OR simply….
  - Use google COLAB
    - Colab gives you free GPU and TPU access

Colaboratory (also known as *Colab*) is a free Jupyter notebook environment that runs in the cloud and stores its notebooks on Google Drive.

Deep Learning in Real Life

Mondragon Unibertsitatea
Faculty of Engineering

- Google maps Prediction system:
  - How long will it take to arrive in bilbao, on april 20th at 3:00pm?
  - How crowded is Eroski today at 15:00?
  - How is traffic right now?
- Emails:
  - Smart Spam filtering
  - Smart Email Categorization
- Grading and Assessment
  - Plagiarism checkers
- Banking/Fintech
  - Fraud Prevention
  - Credit Decisions
- Social Networking
  - Face identification
  - Nudes blocking
- Shopping
  - Recommender systems
- Voice Related
  - Text to speech
  - translators
  - Personal assistants
- Games
  - AI playing Agents alpha zero
- Autonomous vehicle
- Misc:
  - Deep Fakes
- VEEERY LONG ETC.



Mondragon Unibertsitatea

Faculty of Engineering

Eskerrik asko

Muchas gracias

Thank you

Ekhi Zugasti
ezugasti@mondragon.edu

Loramendi, 4. Apartado 23
20500 Arrasate – Mondragon
T. 943 71 21 85
info@mondragon.edu