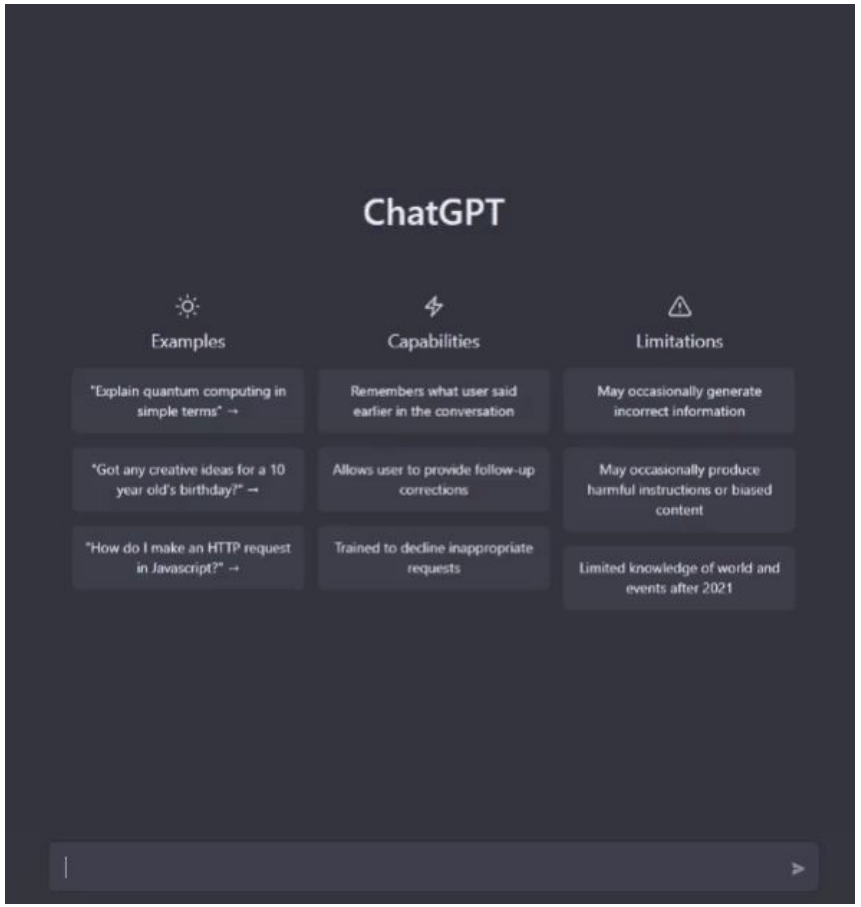
SPATIAL ECOLOGY

Machine Learning for Geodata

Antonio Fonseca

Machine Learning is everywhere

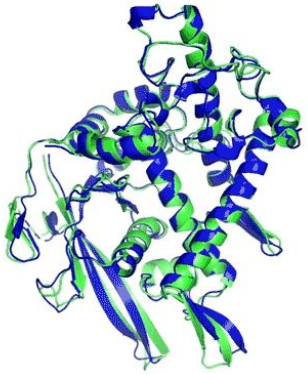
Text generation



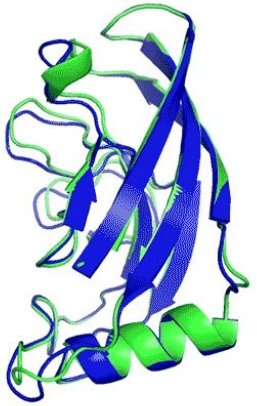
Video generation

How is ML currently being used for geocomp?

Protein design and prediction



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



T1049 / 6y4f
93.3 GDT
(adhesin tip)

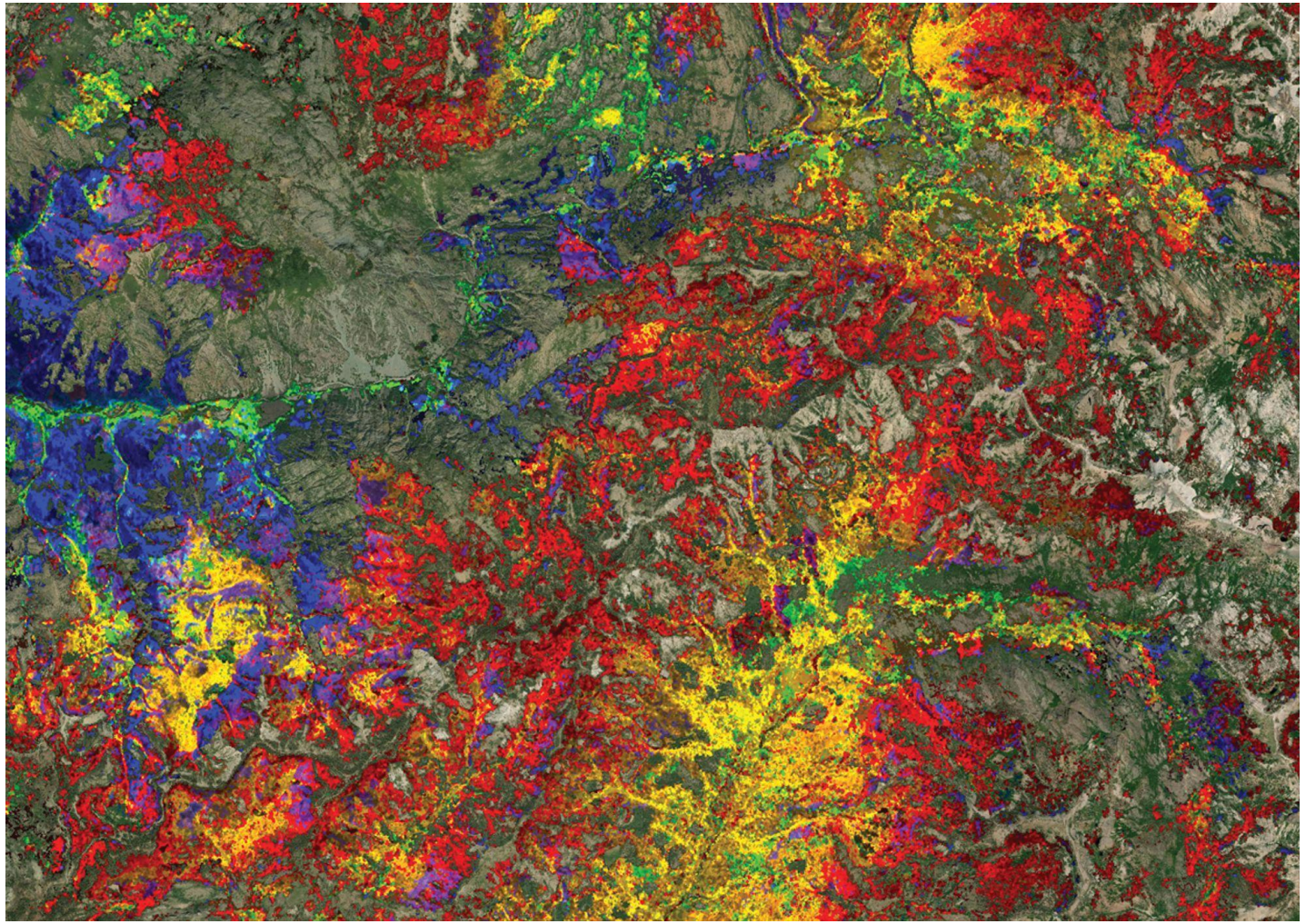
Machine Learning in the geo domain

Swimming pool detection

Mapping

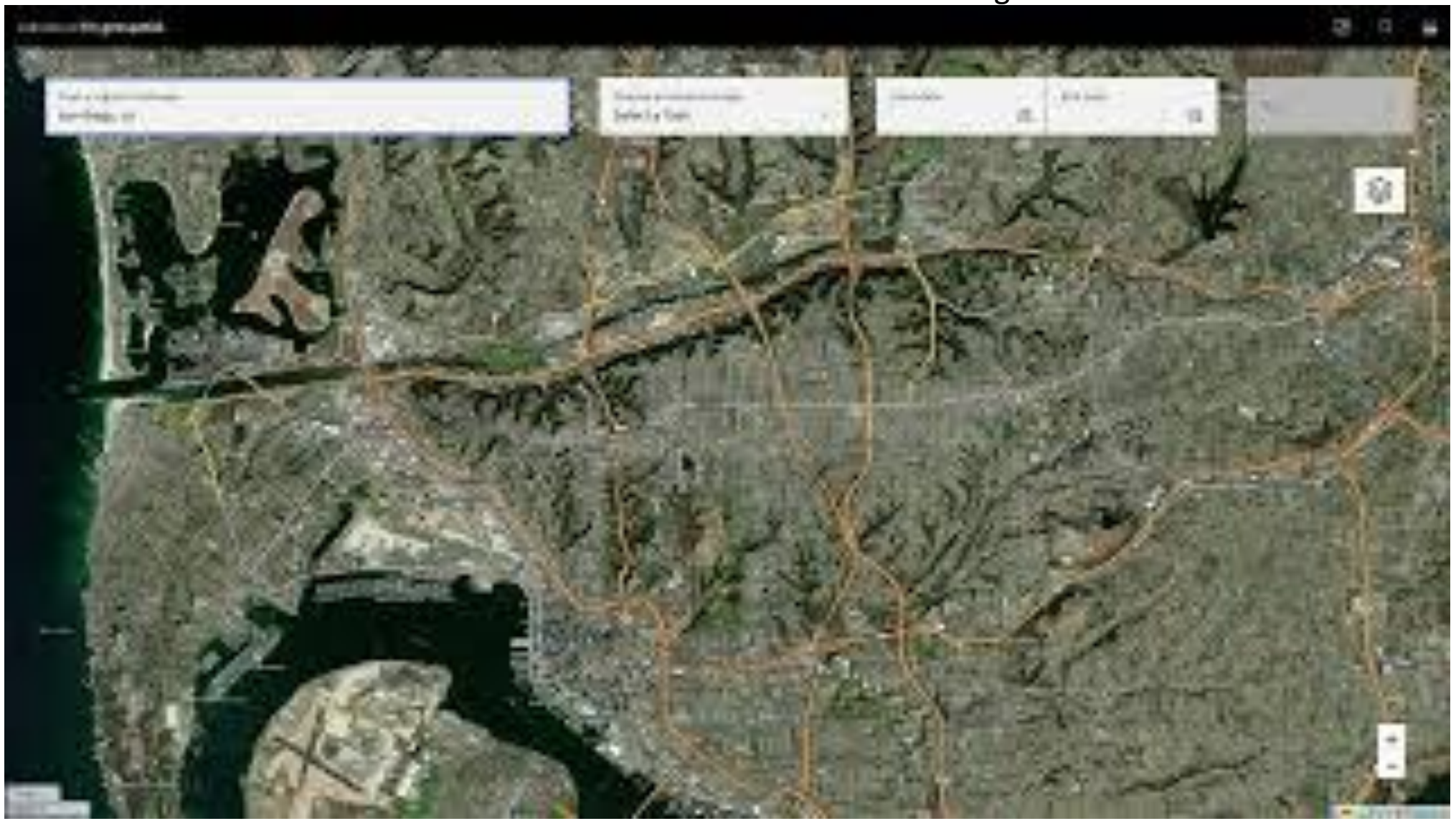
Machine Learning in the geo domain

Dominant tree species



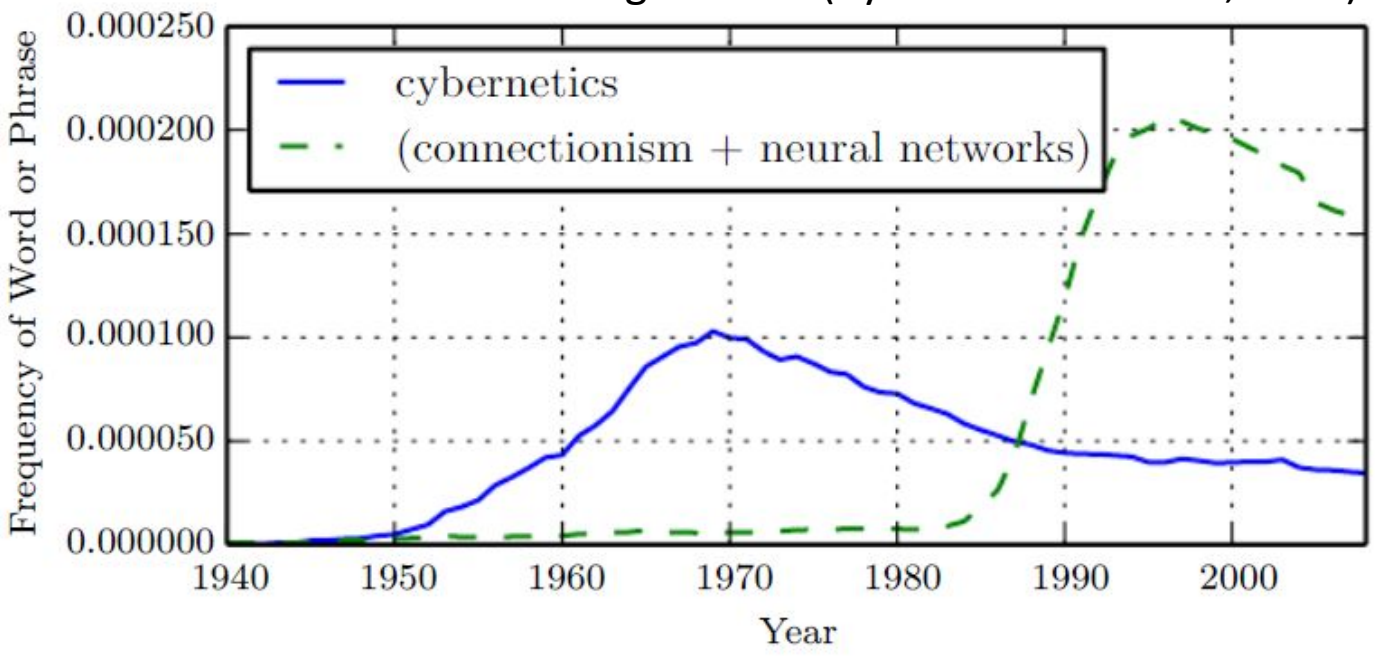
Machine Learning in the geo domain

Classification of satellite and remote sensing data



Evolution of ANNs

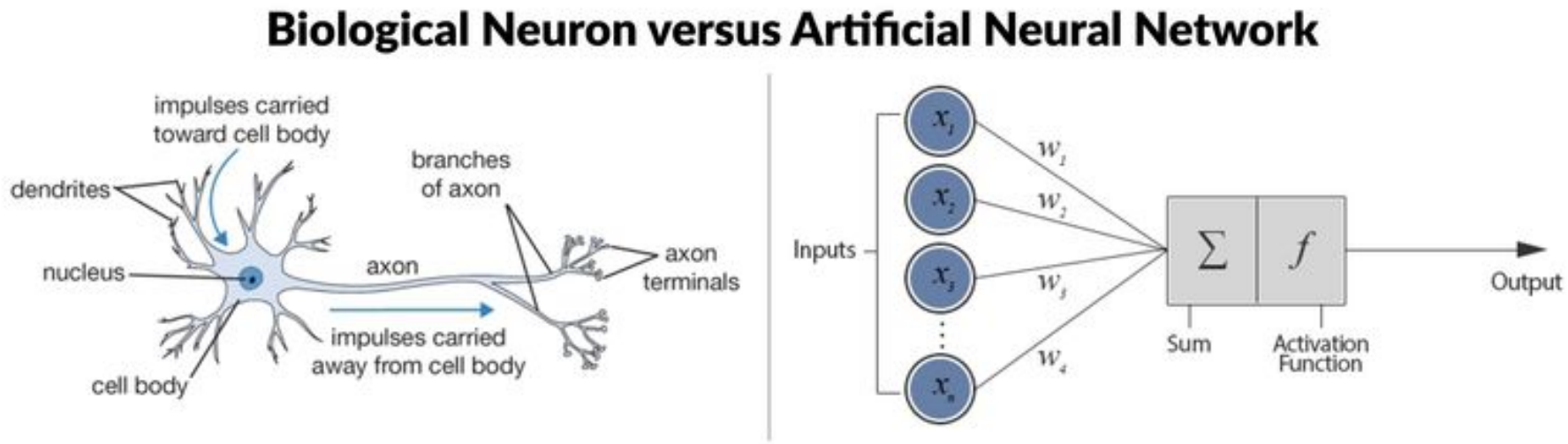
Google Books (by Goodfellow et. al, 2016)



First wave — Second wave — Third wave →

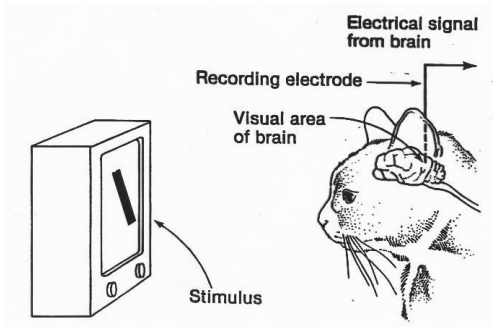
1) Biological Learning Theory (1943, 1949)
2) Perceptron (1958)
3) Backpropagation (1986)
4) Deep Learning (2006, 2007)

ANNs architecture

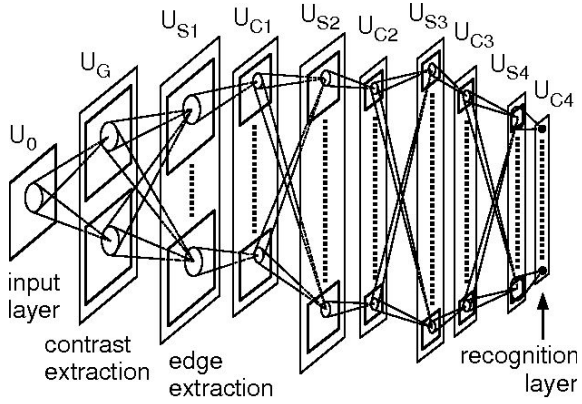


Brain "inspired" model

- Not enough info about brain processing…
- But we know the basics:



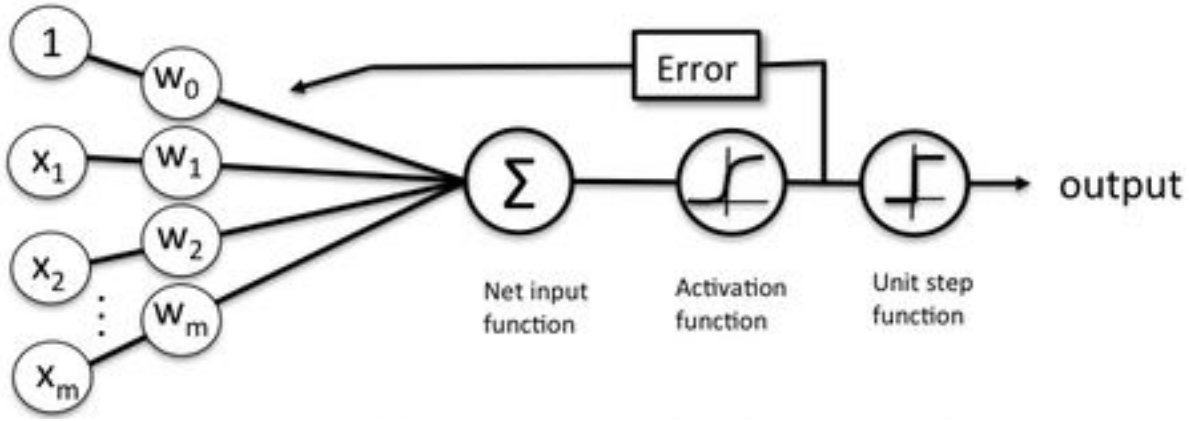
Hubel and Wiesel, 1959-1968



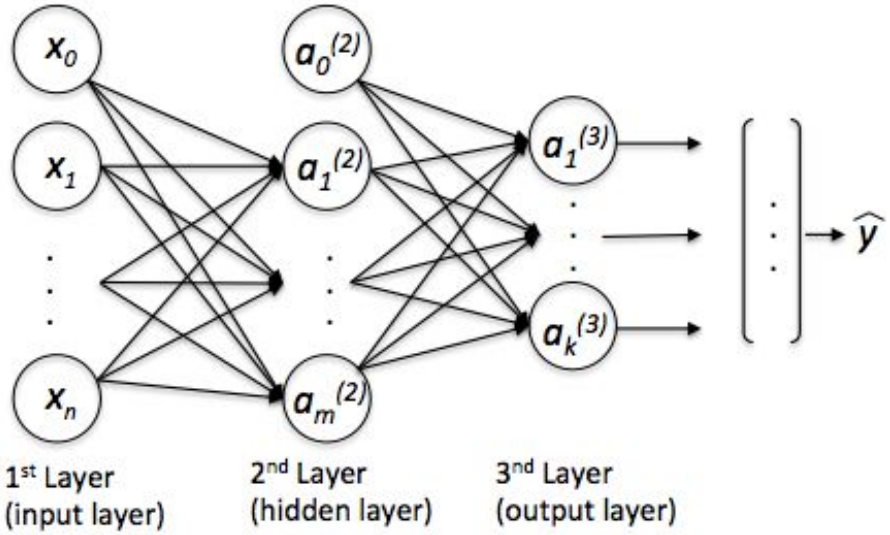
Fukushima, 1980

From logistic regression to Deep Nets

Perceptron

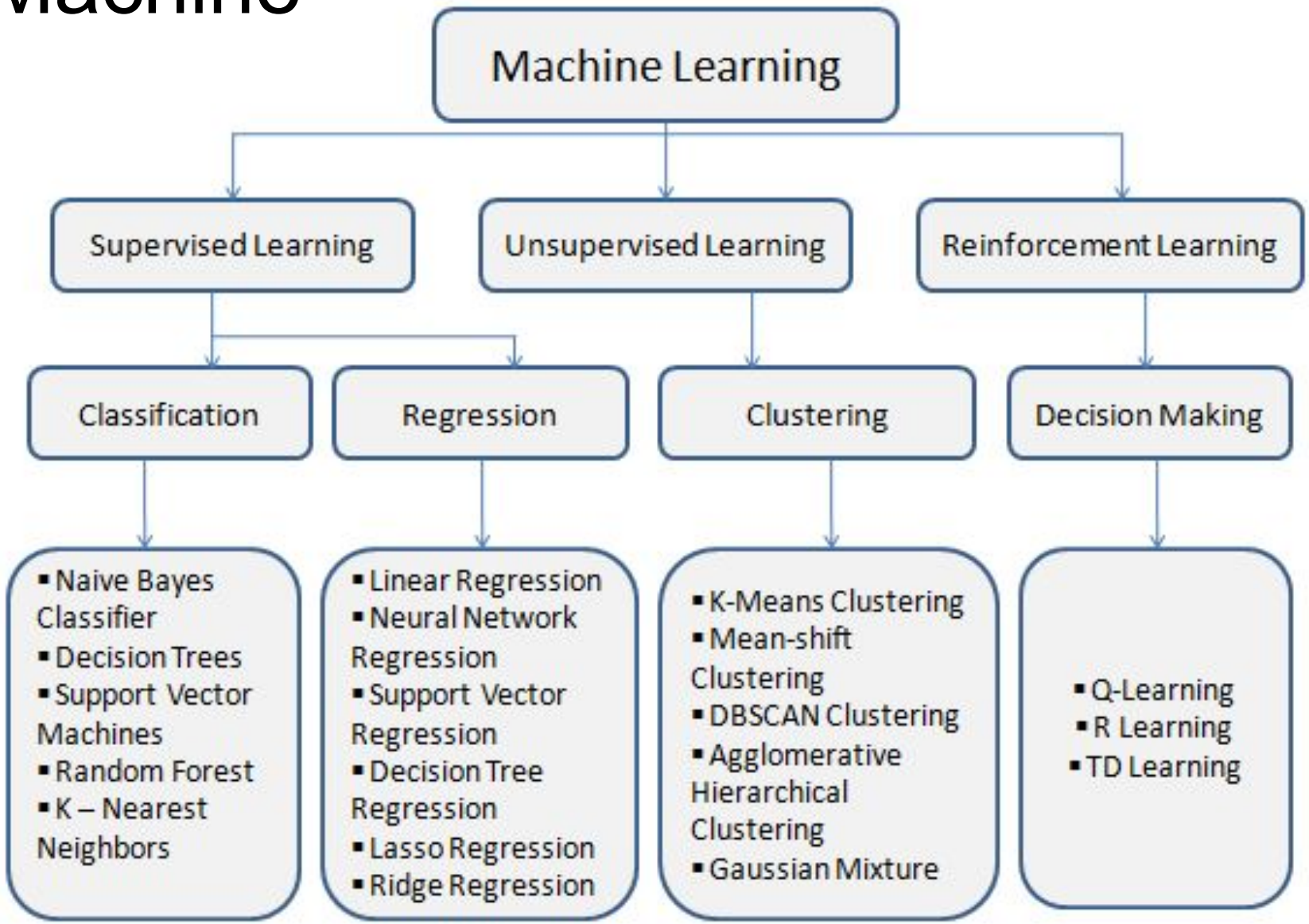


Multi-Layer Perceptron



How do we train Deep Nets?

Types of Machine Learning



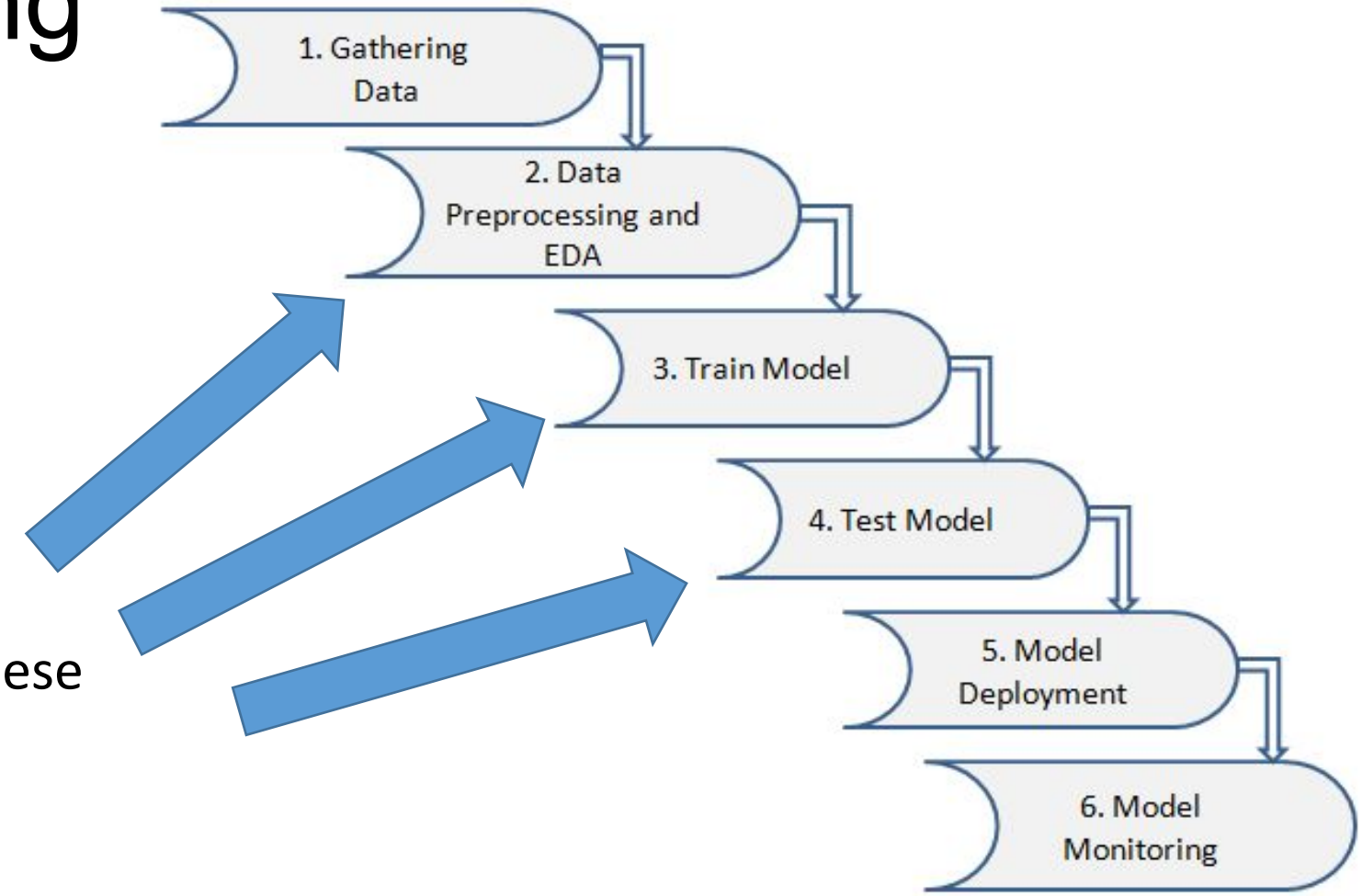
Putting these frameworks in perspective

- **"Pure" Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Unsupervised/Predictive Learning (cake)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- **(Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)**

Original LeCun cake analogy slide presented at NIPS 2016

Machine Learning Life cycle



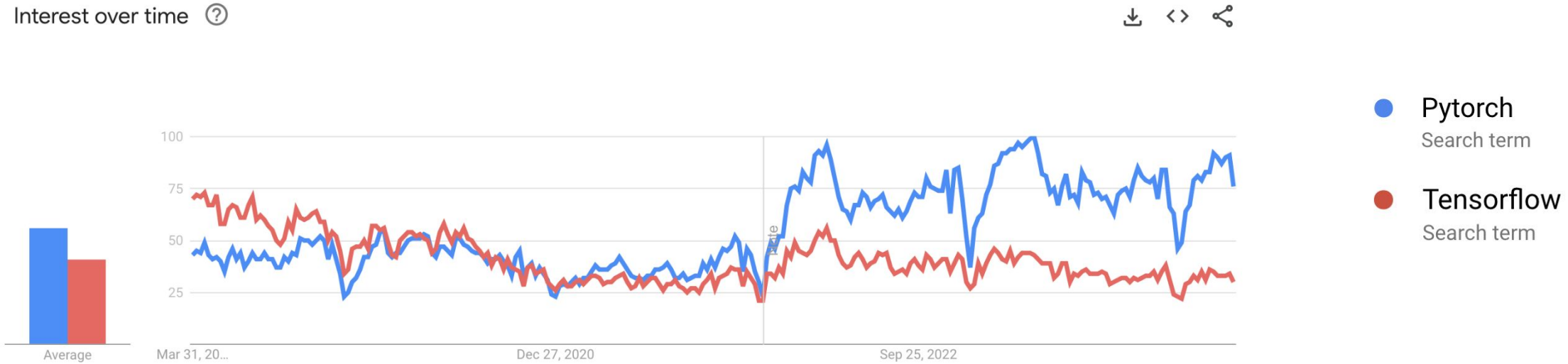
We will focus on these

Hands on!

Keras :
- Rapid prototyping
- Small dataset
- Multiple back-end support

TensorFlow :
- Large dataset
- High Performance
- Functionality

PyTorch :
- Flexibility
- Short training duration
- Debugging capabilities



Expected background

How much CS do I have to know?

- Must be comfortable programming in Python
- Algorithms knowledge helps but is not essential
- Good to know basics of machine learning (SVM, PCA, regression, etc)

How much math do I have to know?

- Linear algebra would help (eg: matrix multiplication, eigenvectors, eigenvalues)
- Calculus (eg: integral, derivative, ordinary differential equations)

Not sure if you attend the requirements?

- Talk to the instructor about it and I will provide extra reading material

Thank you!

How to reach me:

- Slack or whatever we decide to use
- Email: antonio.fonseca@yale.edu
- Twitter: @aho_Fonseca
- Website: https://ahof1704.github.io/