Machine Learning Intro
ML. 0 Organization

Datalab CSIC

Brief description

- Machine learning (or automated statistical learning) studies algos which improve in an automated manner through experience when performing a task.
- Such algos mainly build up statistical models based on data (and expert opinión), the training data, to make forecasts and support decisions.
- Numerous applications.
- The course introduces key ML models and techniques and showcase their implementation in relevant problems.
- Focus on R

Objectives

- Present main ML methods through three major modalities:
  supervised,
  non supervised,
  reinforcement.

- Face some complex ML projects.

- Illustrate some current developments in ML.

- Illustrate R implementations

Stats vs ML?

- ML, coated stats
- When you're fundraising, it's AI. When you're hiring, it's ML. When you're implementing, it's logistic regression
- Statistical learning
- ML, glorified stats
- Computer Age Statistical Inference

Stats vs ML
Jargon

Stats	ML
Estimation	Learning
Data	Instances
Regression	(continuous response) supervised learning
Classification	(discrete response) supervised learning
Response	Label
Regressor	Feature
….	…

Stats vs ML?
Some ‘differences’

Stats	ML
Emphasis on inference	Emphasis on prediction
Quantify uncertainty	Optimize
Understand underlying principles	Predict accurately
Maths	Computer Science
More human intervention	More machine intervention

ML and Stats

- J. Friedman. Data Mining and Statistics, What's the connection (1998)
- L. Breiman. Statistical Modeling,. The two cultures (2001)
- Cross Validated. What's the difference between data mining, statistics, machine learning and AI (2010)
- S.D. Sekar What's the difference between Artificial Intelligence, Machine Learning, Statistics and Big Data (2014)
- Cross Validated. What exactly is Big Data? (2015)
- David Donoho. 50 years of data science (2015)
- B. Efron, T. Hastie. Computer Age Statistical Inference (2016)
- *D. Dunson Statistics in the Big Data era: Failures of the Machine* (2019)
- D. Spiegelhalter The Art of Statistics (2020)
- M. Hernan, J. Hsu, B. Healy A second chance to get causal inference right: a classification of data science tasks (2019)

ML

A computer program learns from experience E with respect to task T and performance measure P,

If its behaviour with respect to T, measured according to P, improves with experience E

Representation-Evaluation-Optimization

Goodfellow et al DL book

ML: Tasks

- Classification, possibly with missing entries
- Regression
- Transcription
- Automatic translation
- Anomaly detection
- Synthesis and sampling
- Filtering
- Density estimation
- Visualization
- Compression
- ….

ML: applications

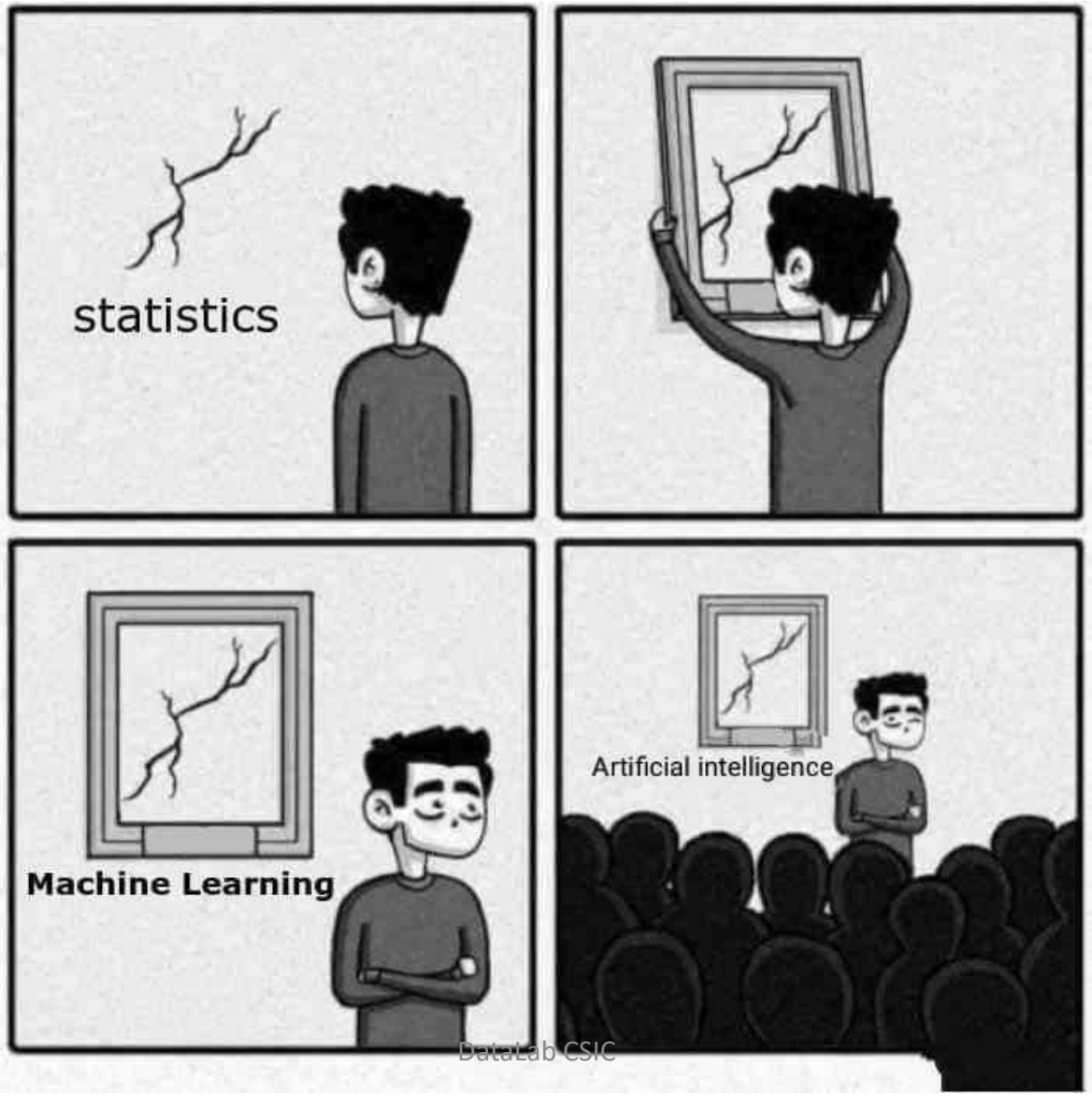
- Autonomous driving systems
- Cancer screening
- Natural language processing
- AlphaGo
- AlphaFold
- Image, video and conversation generation
- Recommender systems
- Content filters
- ….

ML: Yes but…

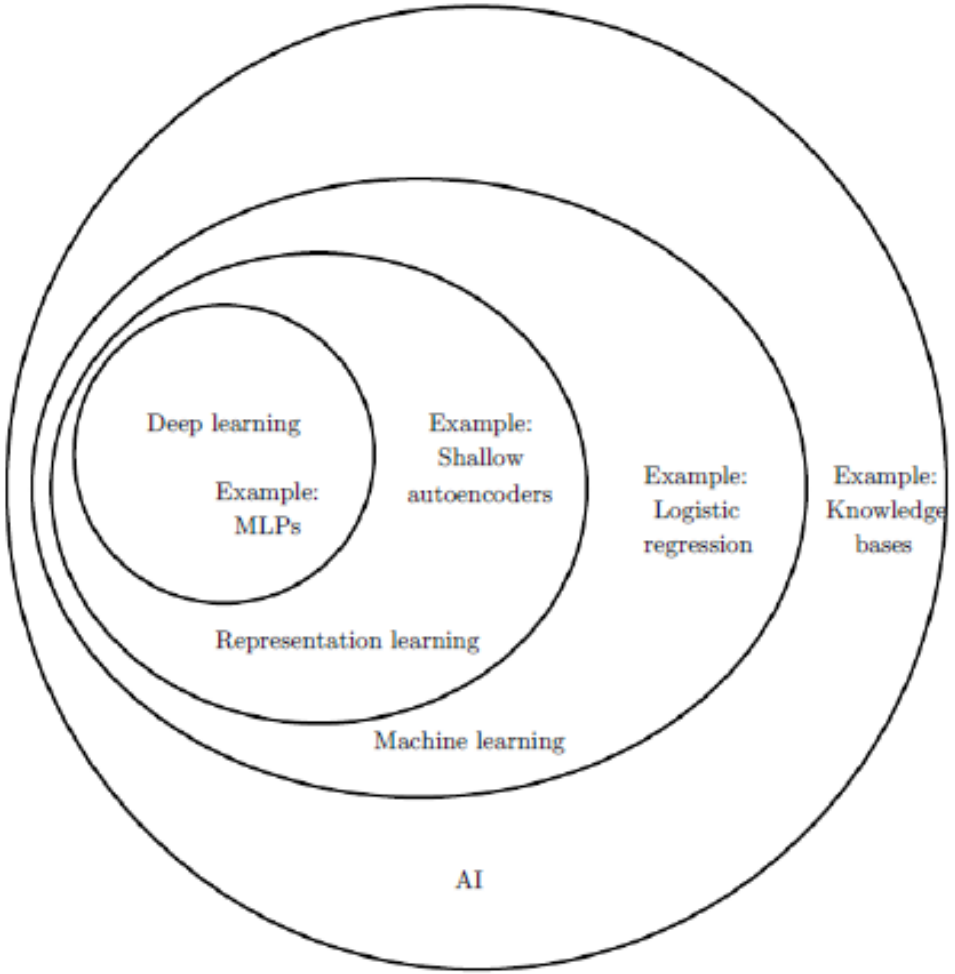
- Privacy
- Safety and security
- Equity
- Bias
- Explainability, Interpretability
- ….

GDPR
EU AI Act (Appendix 1)
UE vs China vs USA

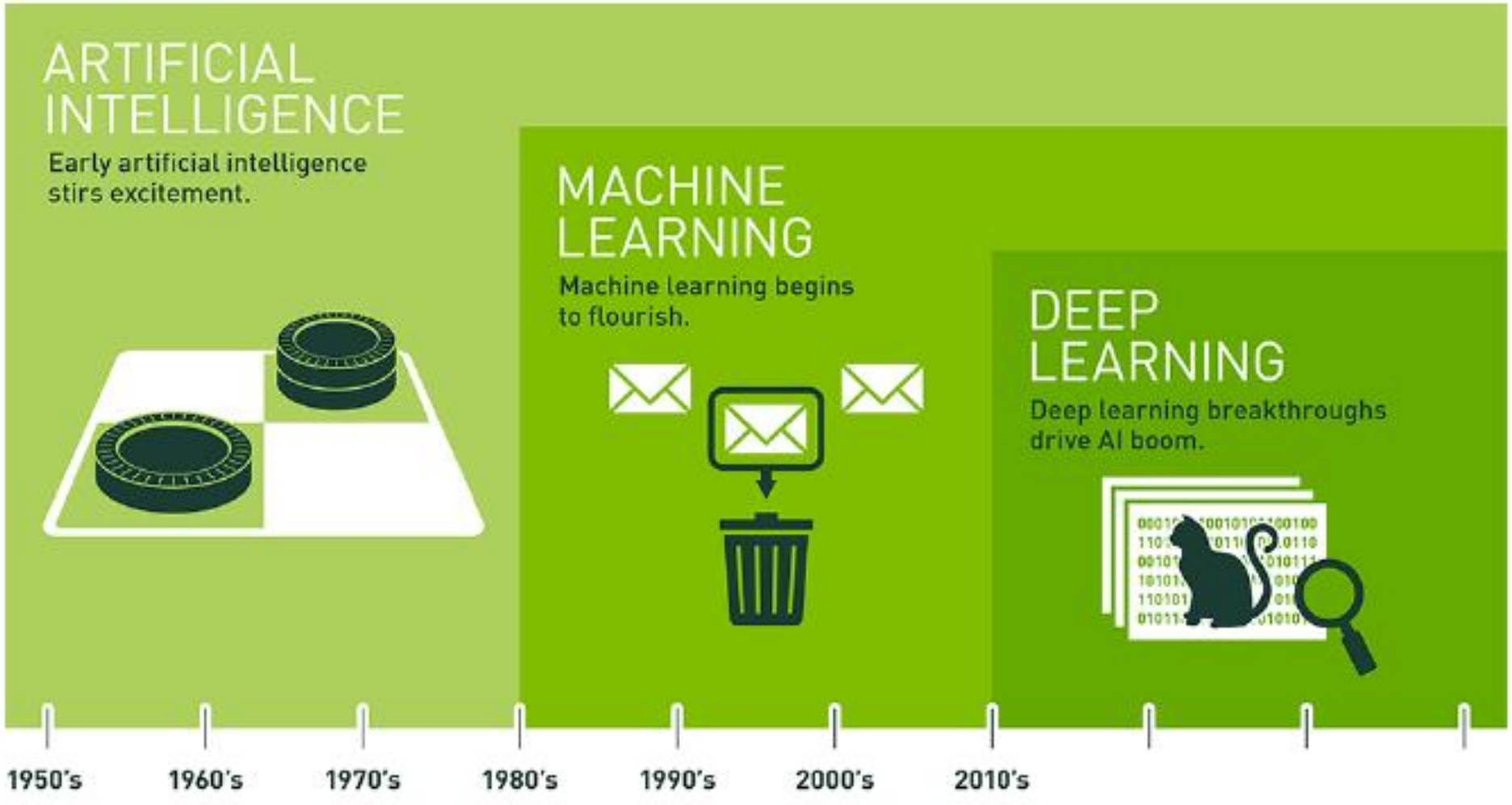
Stats vs ML vs AI?



ML vs AI?


Deep learning
Example:
MLPs
Example:
Shallow
autoencoders
Representation learning
Machine learning
Example:
Logistic
regression
Example:
Knowledge
bases
AI

An evolution….



Contents

1. Introduction and basic concepts. 2. Regression algos.
3. Classification algos. 4. Trees and forests.
5. SVMs. 6. PGMs.
7. Nnets. Deep Nnets.
8. Non supervised learning. 9. Reinforcement learning. 10 Ethics and Security

Mainly based on MLE (+regulariser). Sometimes based on Bayesian methods

Follow up course on Bayesian data science

R, keras

Slides in english+dictionary. Images from bibliography

Basic bib

- Efron, Hastie (2017) Computer Age Statistical Inference. Cambridge UP
- Goodfellow, Bengio, Courville (2017) Deep Learning, MIT Press.
- Hastie, Tibshirani, Friedman (2009) Elements of Statistical Learning. Springer
- James, Hastie, Witten, Tibshirani (2013) An introduction to Statistical Learning. Springer.
- Sutton, Barto (2018) Reinforcement learning: An introduction, MIT Press.
- French, Ríos Insua (2000) Statistical Decision Theory, Wiley

Complem bib

- Barber (2020) Bayesian reasoning and machine learning.
- Bishop (2006) Pattern Recognition and Machine Learning. Springer
- Some papers to be mentioned

Bib ML and R

- Chollet (2018) Deep Learning with R, Manning.
- Ciaburro (2020) Reinforcement Learning with R, Packt
- And many of the earlier books!!!

Meeting

- F 10:15-11:45
- Quest+Concept+Lab+(Case study)+(Problem proposal)
- Email for doubts. introml@icmat.es
- Course list. mlcsic@googlegroups.com
- Course zoom. https://us06web.zoom.us/j/81640739811?pwd=RFNCQmRWNFMvZ3YzSWttUFlNMVppQT09
- Slides and R stuff, datasets, etc. https://datalab-icmat.github.io/courses_stats.html