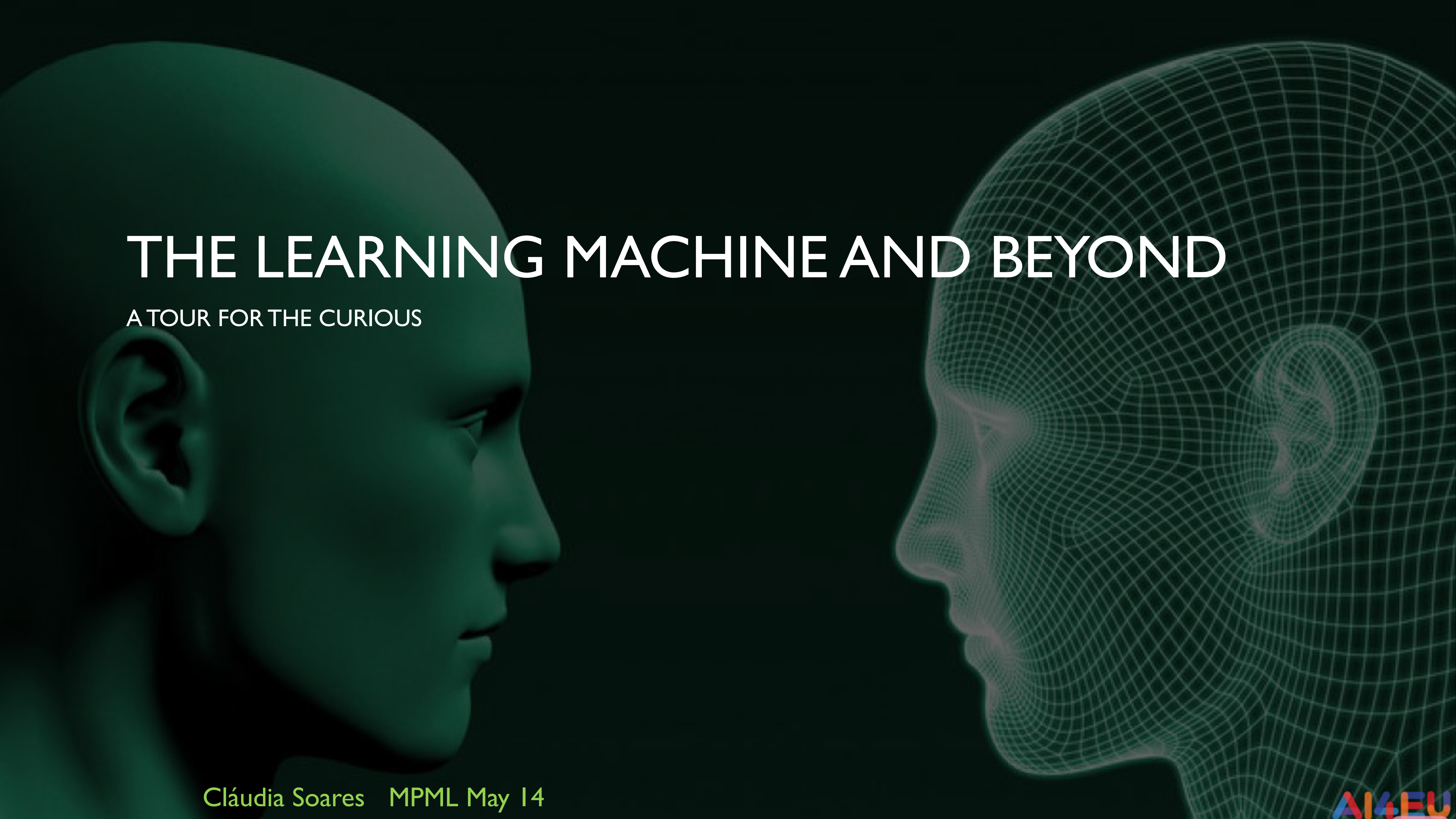
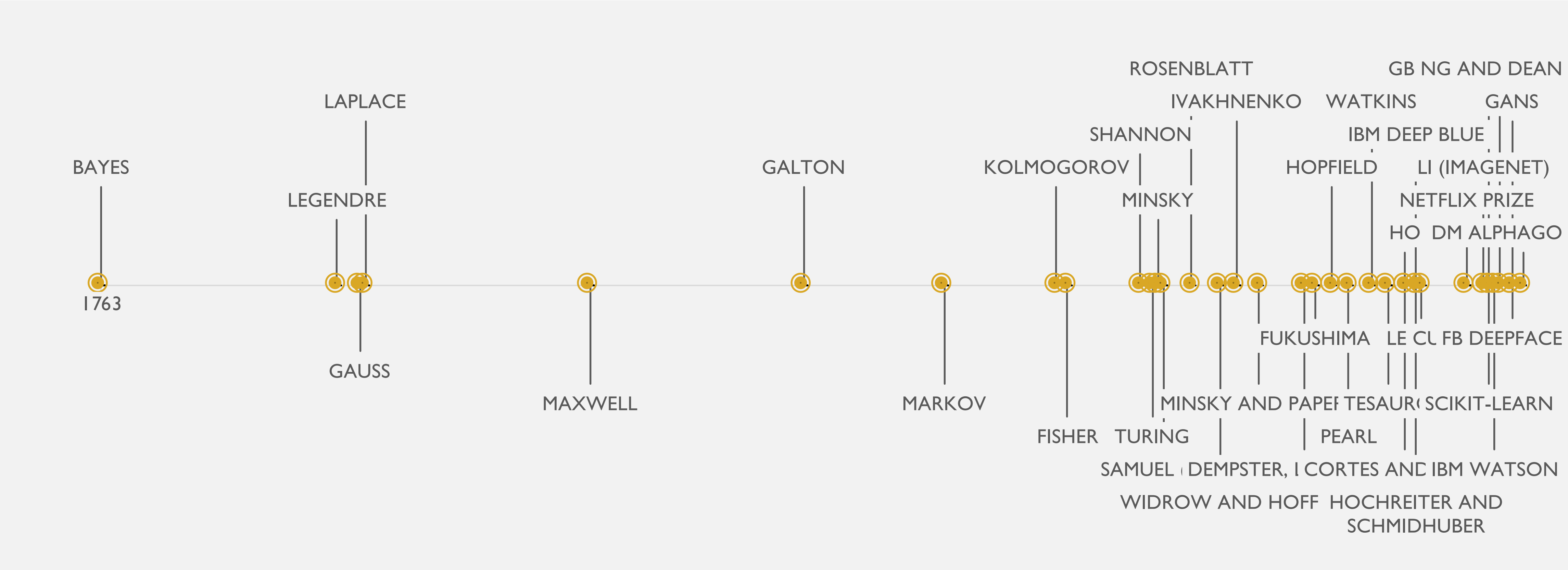


THE LEARNING MACHINE AND BEYOND

A TOUR FOR THE CURIOUS

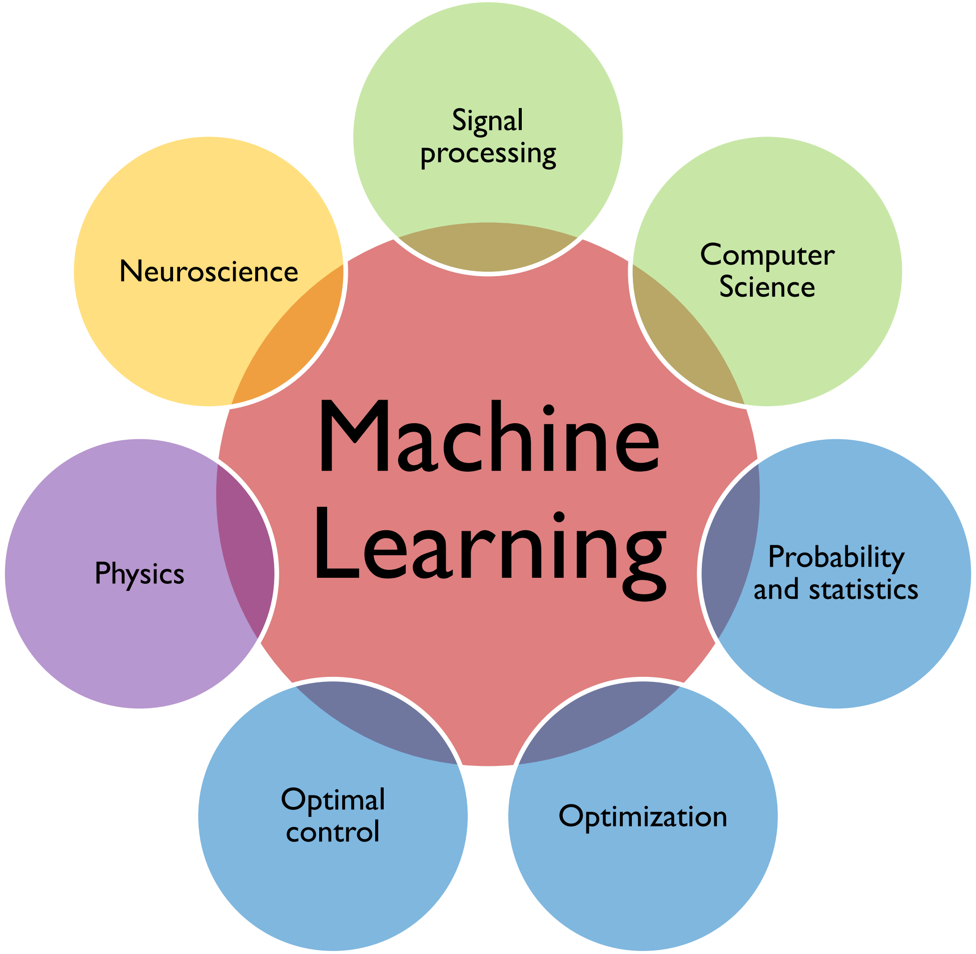
Cláudia Soares MPML May 14

CONTENTS

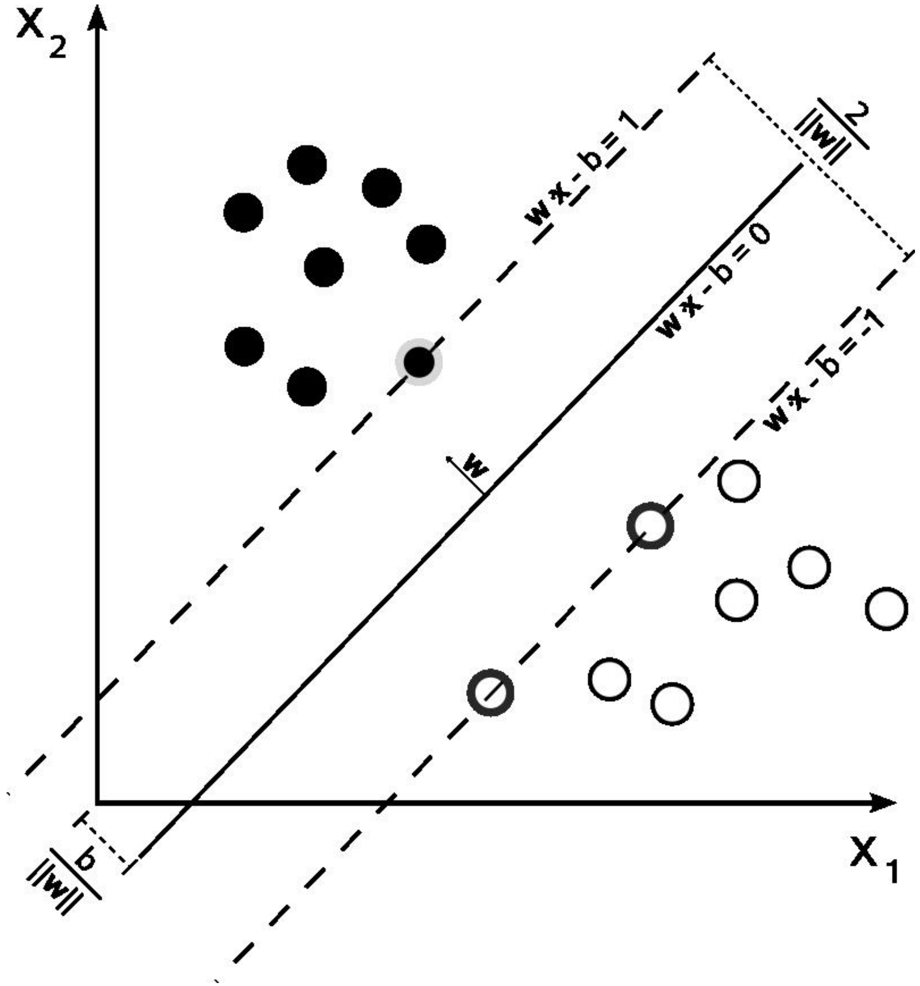


LEARNING MACHINES: A TIMELINE

ML AND (A FEW) OTHER FIELDS

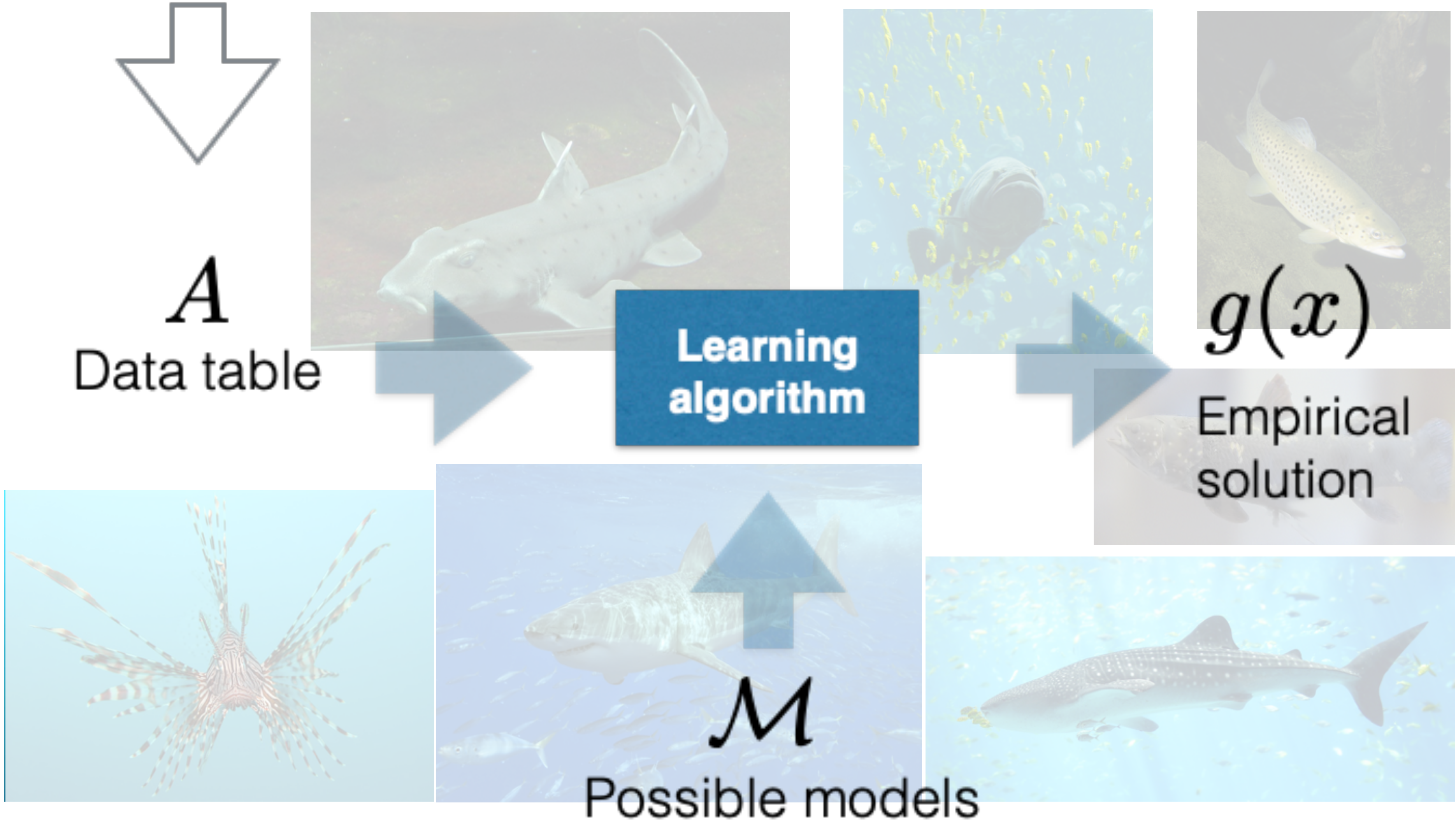


INCIPIT: LEARNING FROM DATA



WE ALL LEARN FROM DATA

$$f(x) =? \, f(x) =?$$

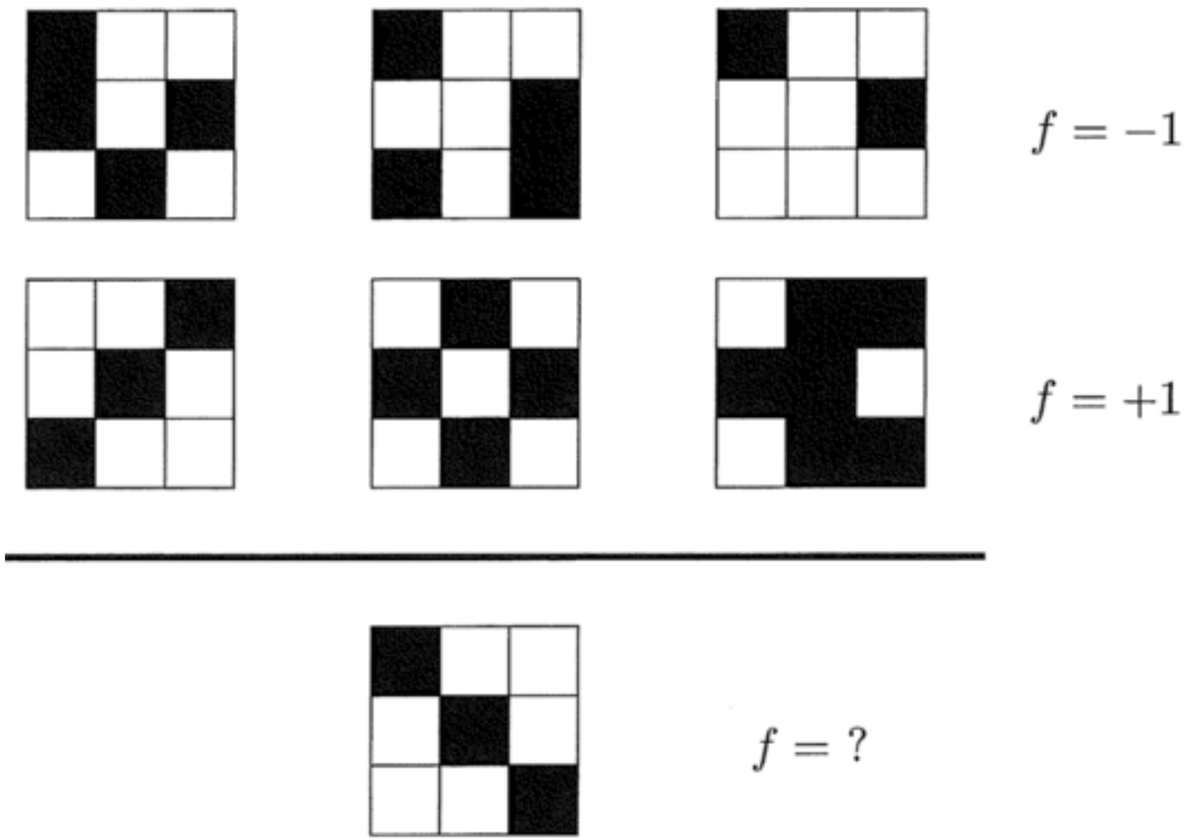


A
Data table
Learning algorithm
$g(x)$
Empirical solution
$\mathcal{M}$
Possible models

IS LEARNING FEASIBLE?

- How can a **limited data set** inform us about an **entire unknown function**?
- We know all the values of the unknown function on the training examples.
- Does this mean we know f?
- Does the data set tell us anything about the function outside of the data set?

$f(x)$ Is completely unknown outside of the dataset



OUTSIDE THE TRAINING SET

Learn $f(x)$

Predict its values outside of the dataset

Will our prediction be **incorrect**?

Possible!

Will our prediction be **likely correct**?

With some probability, yes!

KINDS OF LEARNING

SUPERVISED LEARNING

$$A = \begin{bmatrix} x_{11} & \cdots & x_{1d-1} & y_1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & \cdots & x_{Nd-1} & y_N \end{bmatrix}$$

Choose a feature that you want to predict

You will be learning

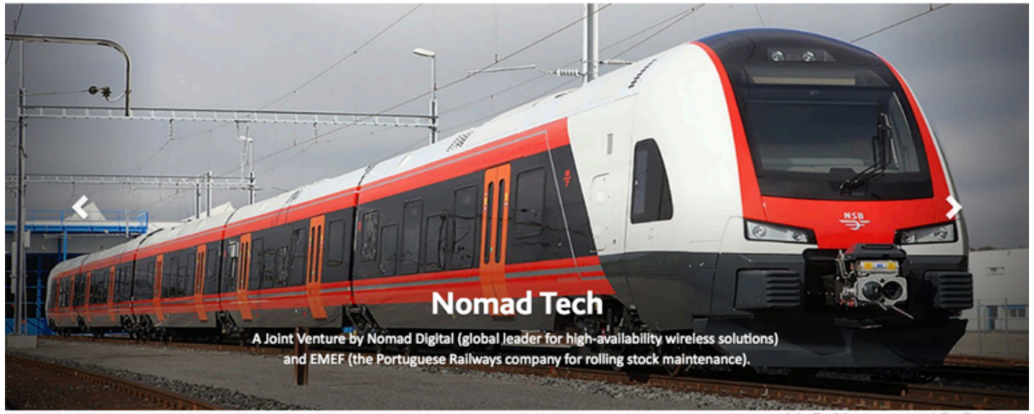
$$f(x_i) \approx y_i$$

SUPERVISED LEARNING TASKS

- Classification (discrete, not ordered outcome)
- Regression (real, ordered outcome)
- Preference learning (discrete, ordered outcome)

CLASSIFICATION

Faillure prediction of suburban trains



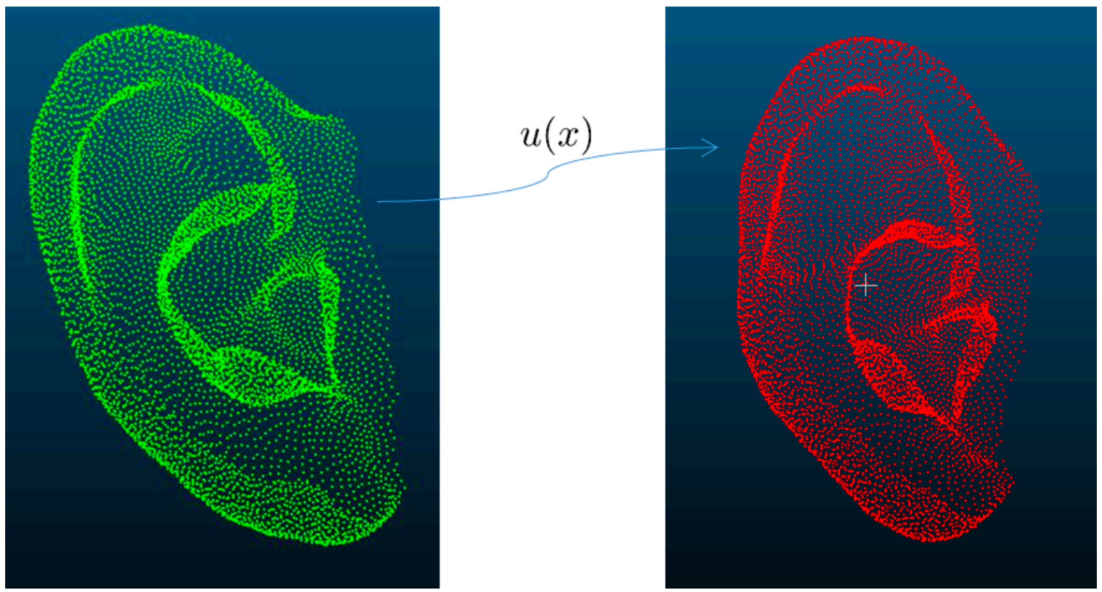
with R. Ventura and R. Borges

- ~1000 variables every 10mn
- Real data issues
  - Class imbalance
  - Noise
  - Missing data
- High-dimensional large data streams

REGRESSION

- Problem: face prosthetics of ears and noses, such that the created face part fits well with the face and respects border conditions
- Parametric Gaussian Process regression

Γ_R Reference shape Γ_T Target shape

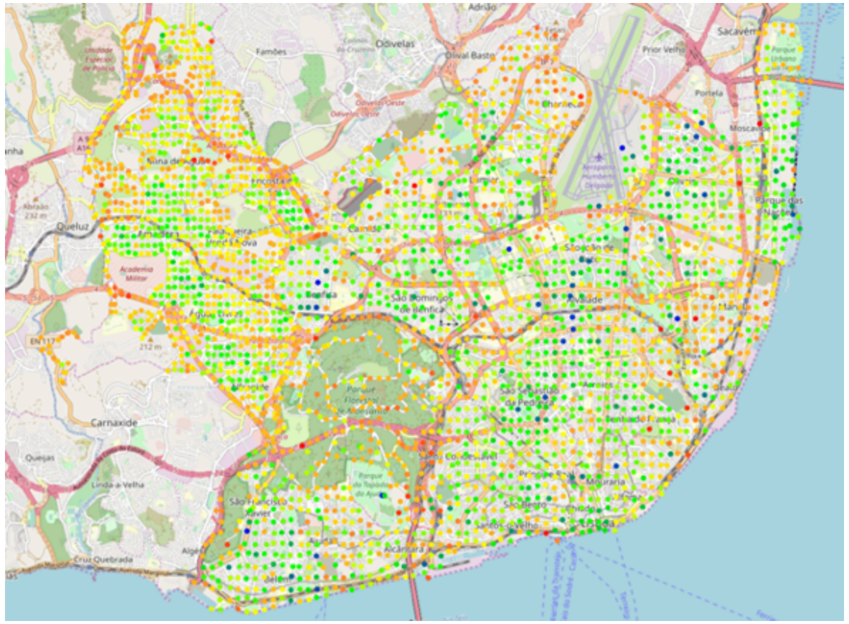


With F. Valdeira, A. Micheletti, R. Ferreira

BIGMATH

µROBOPTICS® TECHNICAL CONSULTING AND RESEARCH

PREFERENCE LEARNING: PERCEPTUAL SCALE FROM PAIRWISE JUDGEMENTS



With M. Marques and G. Costa

UNSUPERVISED LEARNING

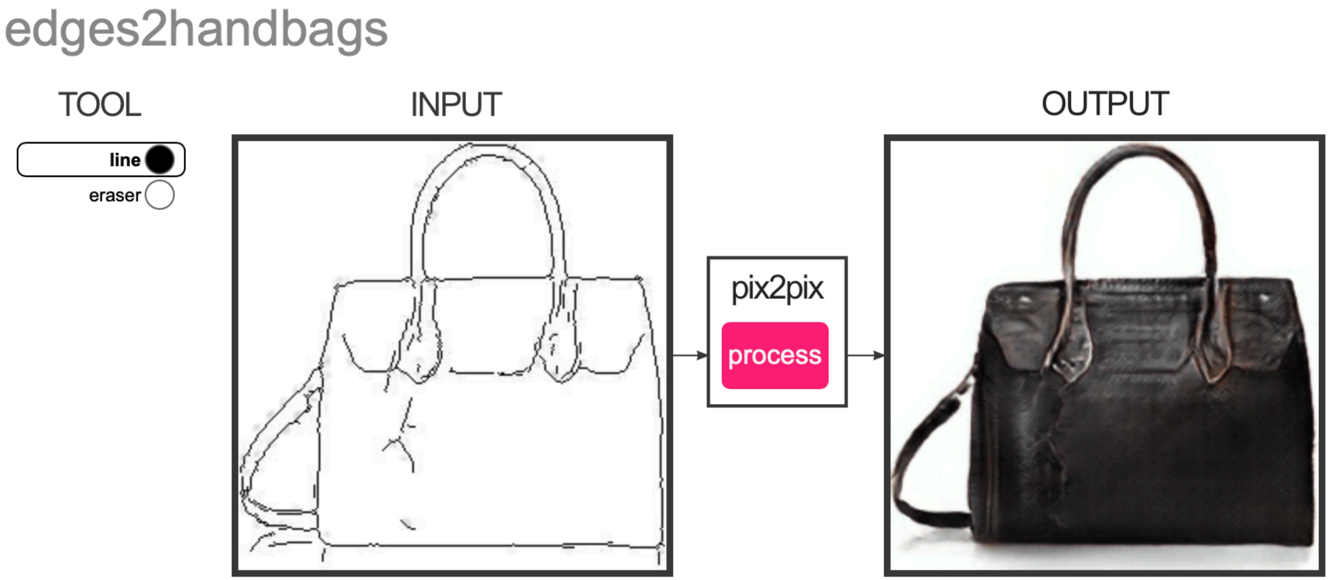
- Find **patterns** and **structure** in data
- Which features better represent the data? (Dimensionality reduction)
- Is data aggregated in subgroups? (Clustering)

UNSUPERVISED LEARNING TASKS

- Clustering (Classification)
- Dimensionality reduction (Regression)
- Generative models

GENERATIVE ADVERSARIAL NETWORKS (GAN)

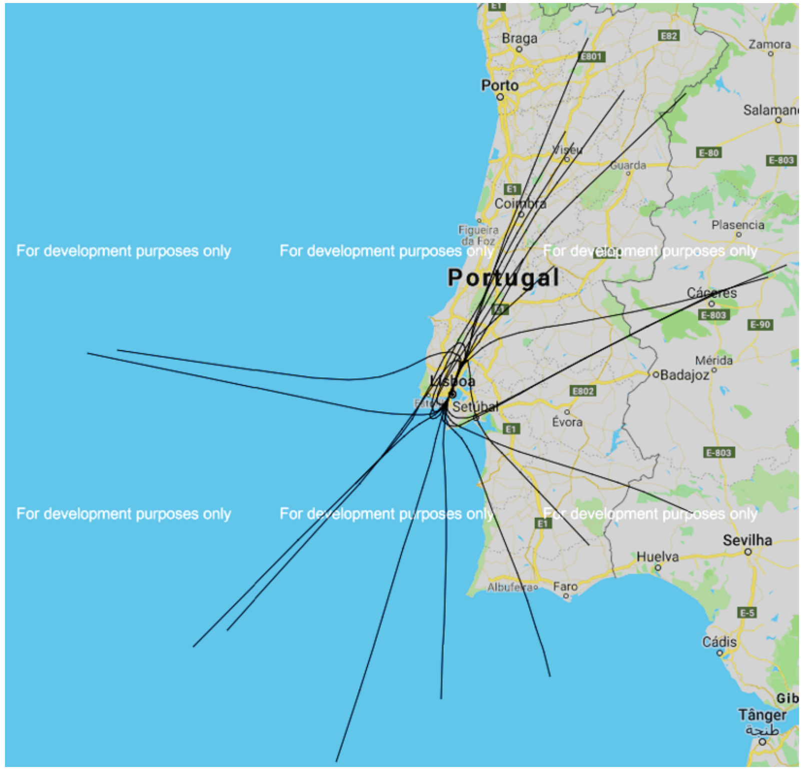
- Example: pix2pix Trained on a database of ~137k handbag pictures collected from Amazon and automatically generated edges from those pictures
- GANs are a game with 2 DNN as players. Nash equilibrium is a local minimizer of the generator and discriminator losses.



Isola et al, 2016

GENERATIVE MODEL FOR CLUSTERING AND DENSITY ESTIMATION

Landing time inference for the Lisbon airport

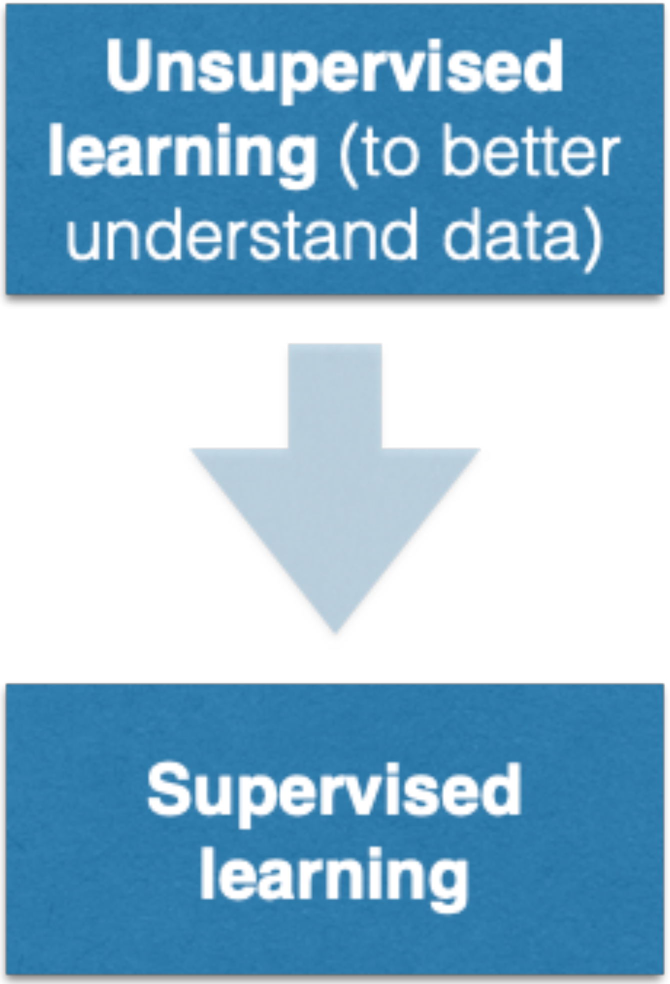


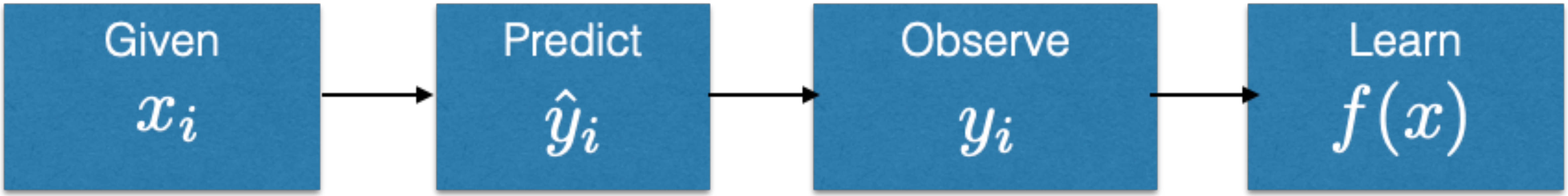
With R. Ventura and A. Fallah

TAP AIR PORTUGAL

- Fitted a GMM to aircraft trajectory data
- Discovered number of clusters
- Took advantage of the density model to compute the posterior given a small entry point trajectory piece

SUPERVISED AND UNSUPERVISED LEARNING





ONLINE LEARNING

- **Offline:** Learning and prediction are separated in time. Model is frozen after learning.
- **Online:** Algorithm sees the dataset one example at a time

ONLINE LEARNING USING TIME-VARIANT LOSSES (ONGOING!)

Goal: learn the model of the face as the animation is unrolling

Optimization of a nonconvex loss using online proximal gradient descent

With F. Caldas, D. Jakovetic (UNS), Z. Desnica (3L)

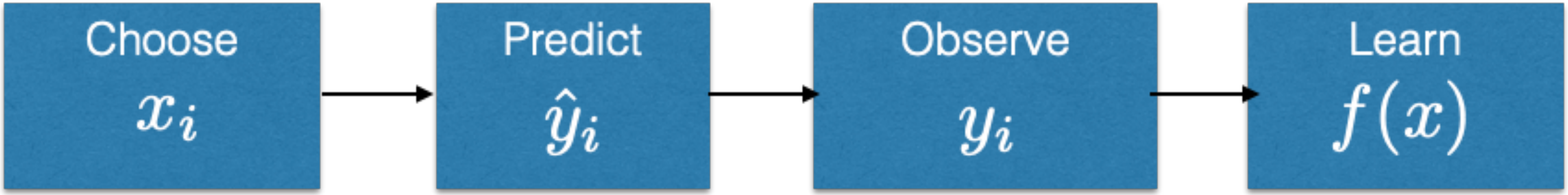
EPIC GAMES

3Lateral

BIGMATH

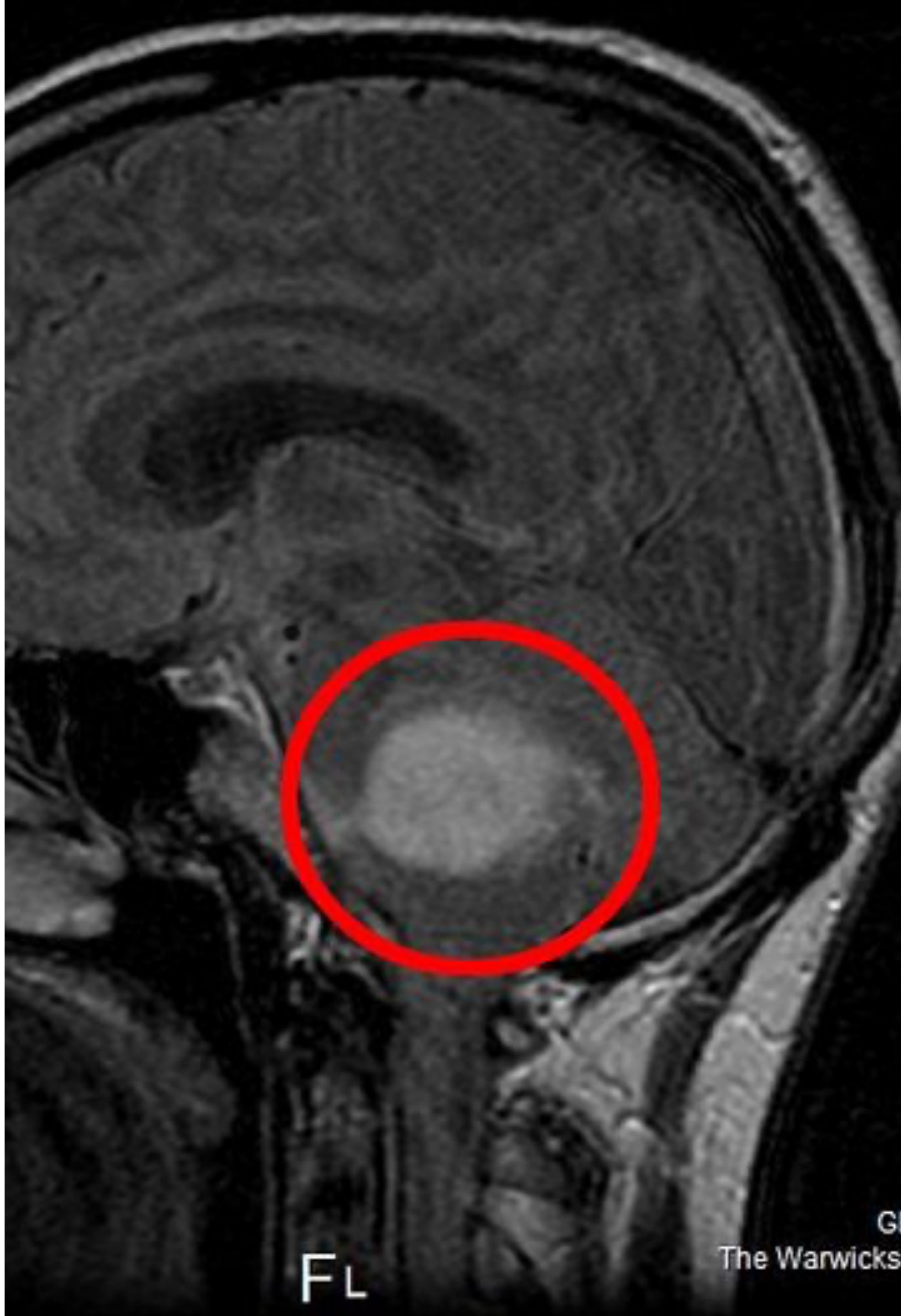
ACTIVE LEARNING

- We **choose** the input point to **maximize** its information value

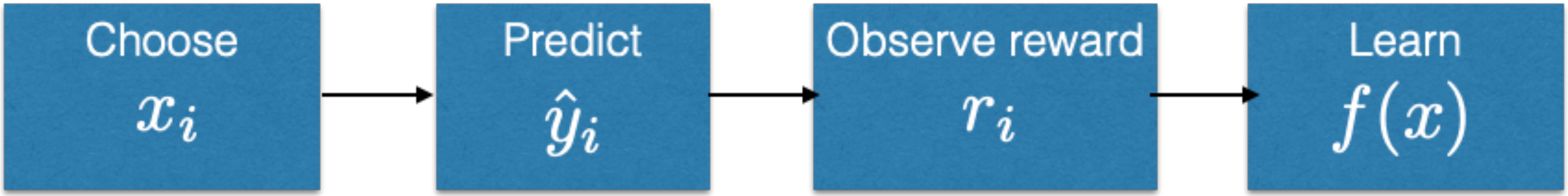


CHOOSING DATA LABELING STRATEGIES

- Large datasets for supervised learning are costly to label.
- Can machines learn with fewer labeled training instances if they are allowed to ask questions?
  - YES!



Konyushkova (2017)



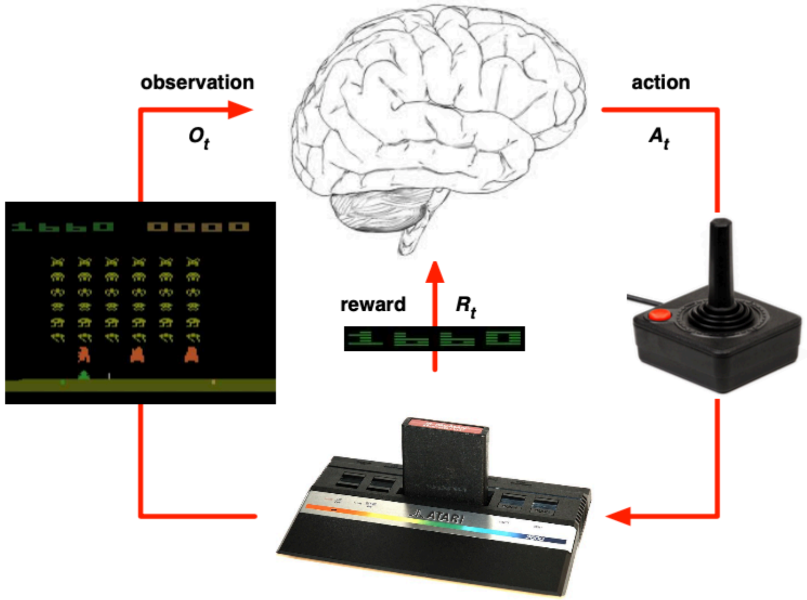
REINFORCEMENT LEARNING

- There is no supervisor, only a reward signal
- Feedback can be delayed, not instantaneous
- Time really matters (sequential, non i.i.d data)
- Algorithm's choices affect the subsequent data it receives

PLAYING ATARI GAMES

USE DNN AS A FUNCTION APPROXIMATOR FOR THE ACTION VALUE FUNCTION (Q LEARNING): THE DQN ARCHITECTURE

SUPERHUMAN PERFORMANCE ON MOST GAMES



- Rules of the game are unknown
- Learn directly from interactive game-play
- Pick actions on joystick, see pixels and scores

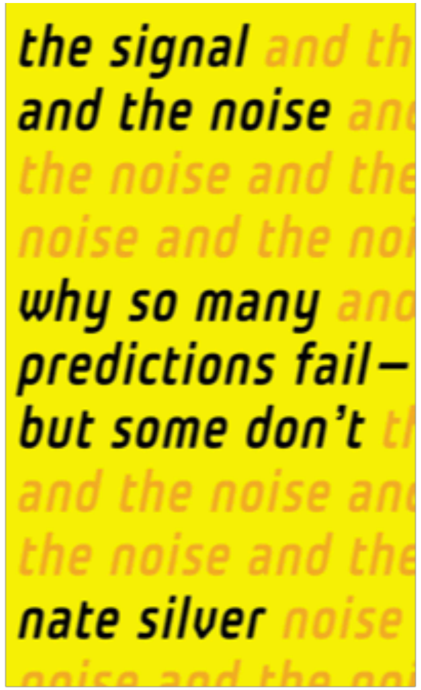
David Silver, Deepmind

ALGORITHMS FOR MODEL TRAINING AND INFERENCE

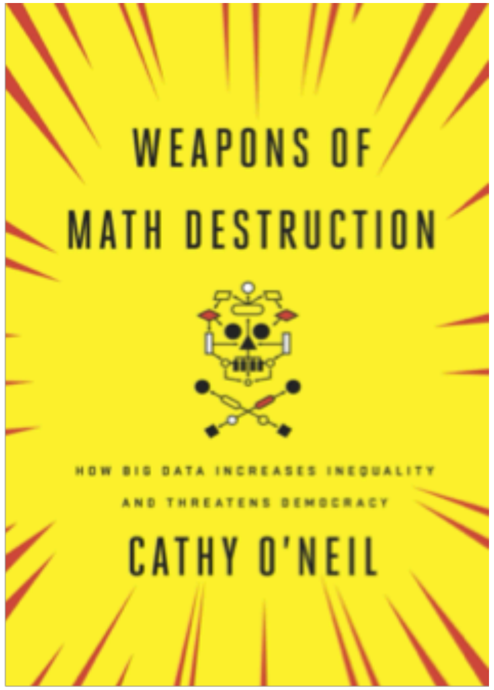
- Variational inference
- Stochastic gradient descent family
- Transfer learning of pre-trained model
- Federated learning / distributed learning

ETHICS

- Data Science is ML at the frontline in dealing with real data
- New ethical challenges
- Need of effective generalization theory

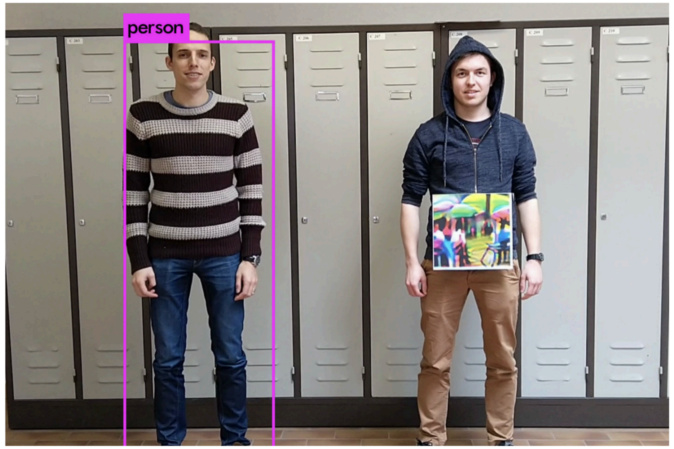


2012



2016

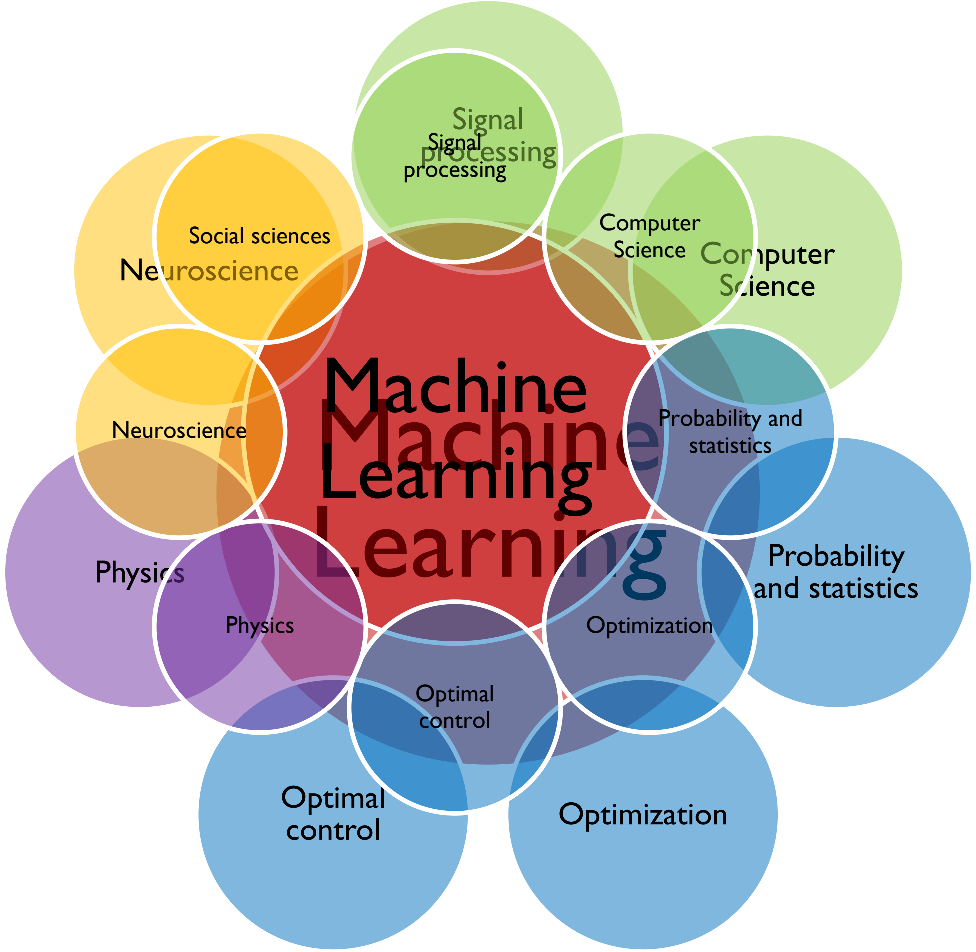
LIMITATIONS: ADVERSARIAL EXAMPLES



- Adversarial examples on deep Q Networks for Atari games: 1 px leads to 50% of human performance
- Pixel attacks for image classification on ImageNet
- Physical patch attack to object recognizer
- Catastrophic forgetting

IS LEARNING FEASIBLE (IN THE REAL WORLD)?

- Robust Machine Learning by helping algorithms to train in a more generalizable way
- Devising models that are both accurate and explainable
- Ethical learning machines



MPML SEMINAR SERIES GOALS: REJOINING FORCES

- Rejoin forces from the fields of Mathematics, Physics, Machine Learning to address some of these difficult problems in ML
- Rejoin those forces to look at strong theory-backed results in Mathematics and Physics to successfully predict and interact with the real world

RECENT EXAMPLES OF SYNERGIES

- Theodorou et al., (2010) Learning Policy Improvements with Path Integrals
- Carlsson et al., (2018) Topological Approaches to Deep Learning
- Hofer, Christoph, et al. (2017) Deep learning with topological signatures.
- Chmiela, Stefan, et al. (2017) Machine learning of accurate energy-conserving molecular force fields.
- …