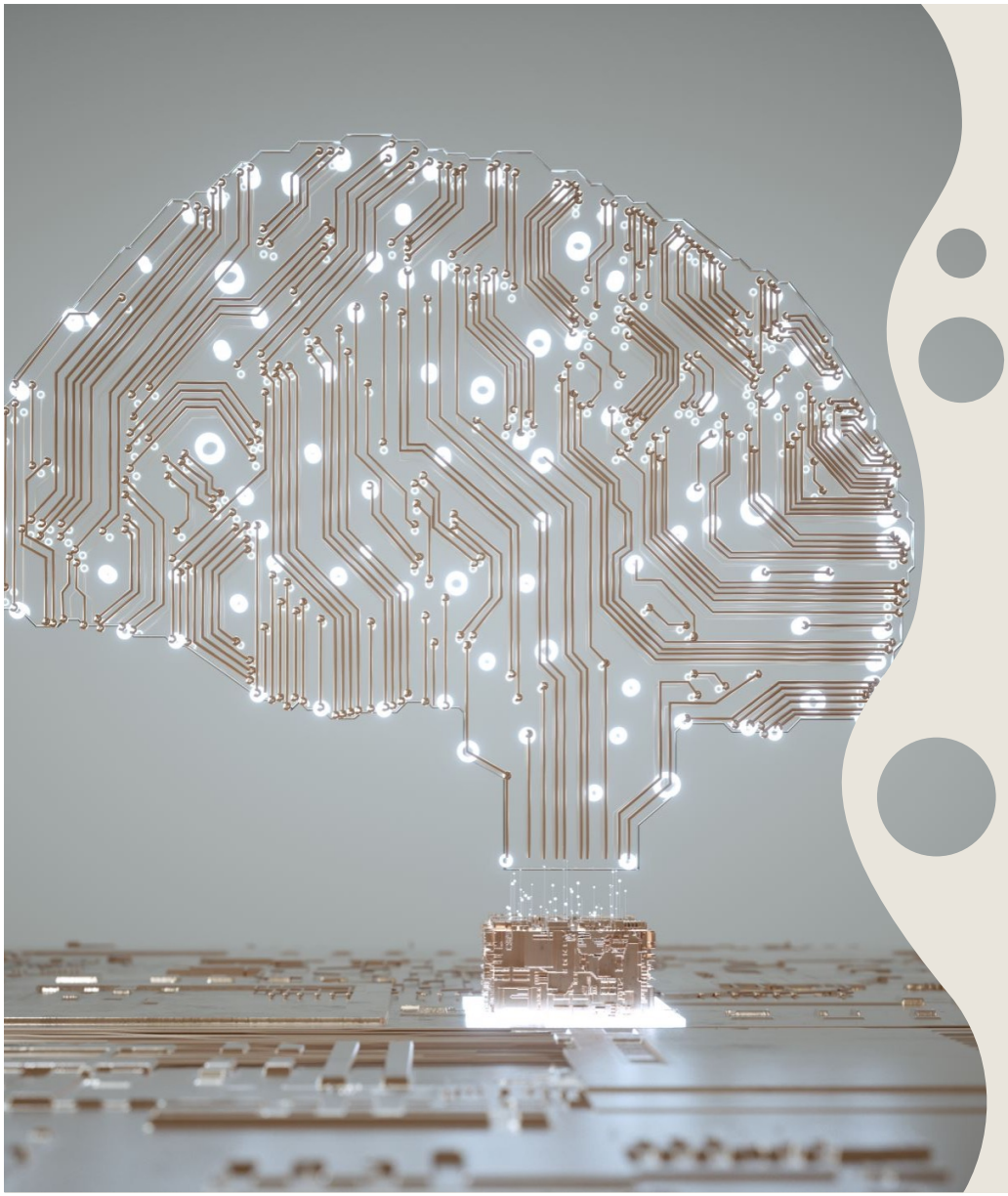
INTRODUCTION TO DATA SCIENCE IN SPACE PLASMAS

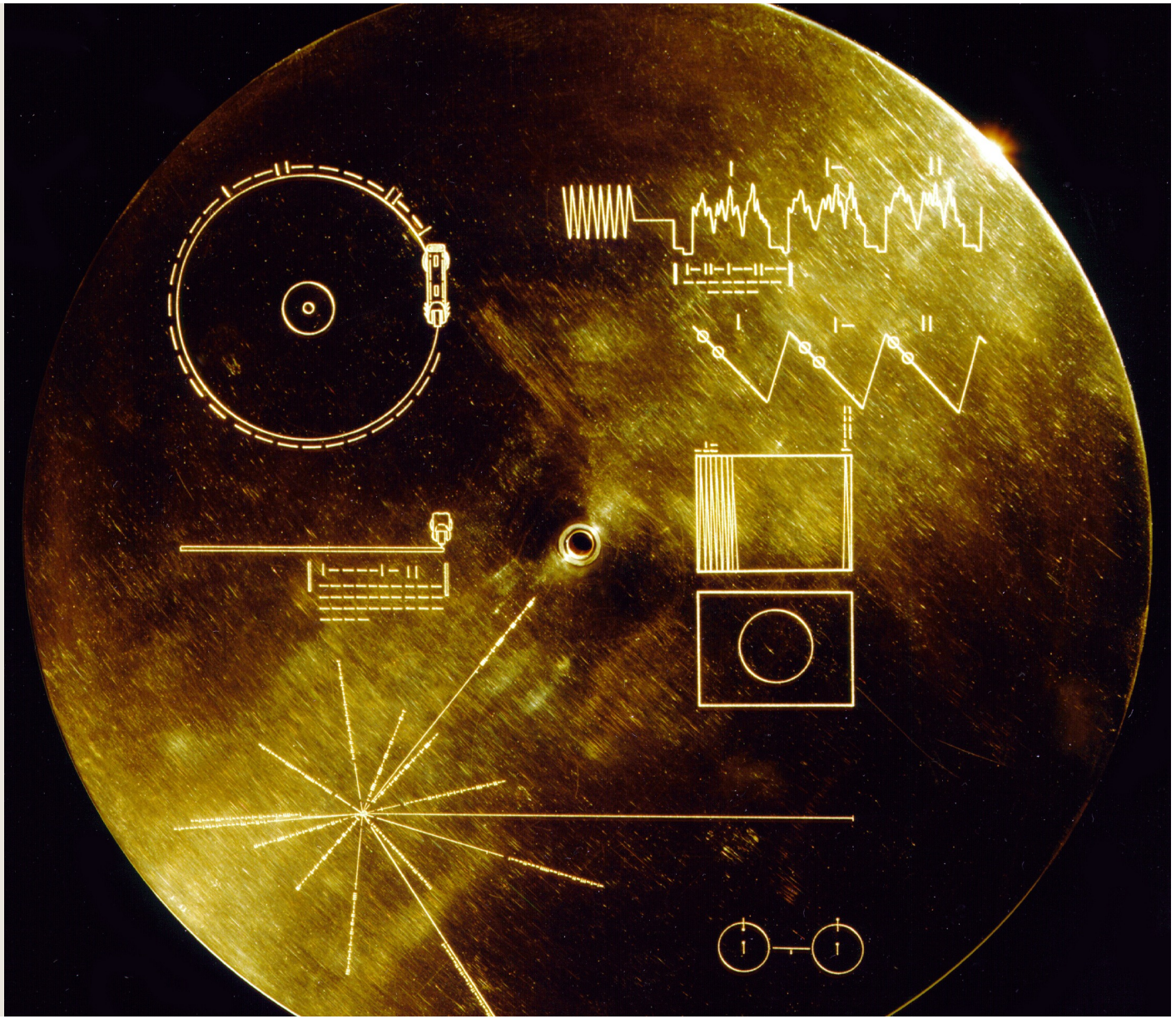
George Miloshevich and Francesco Carella



OUTLINE

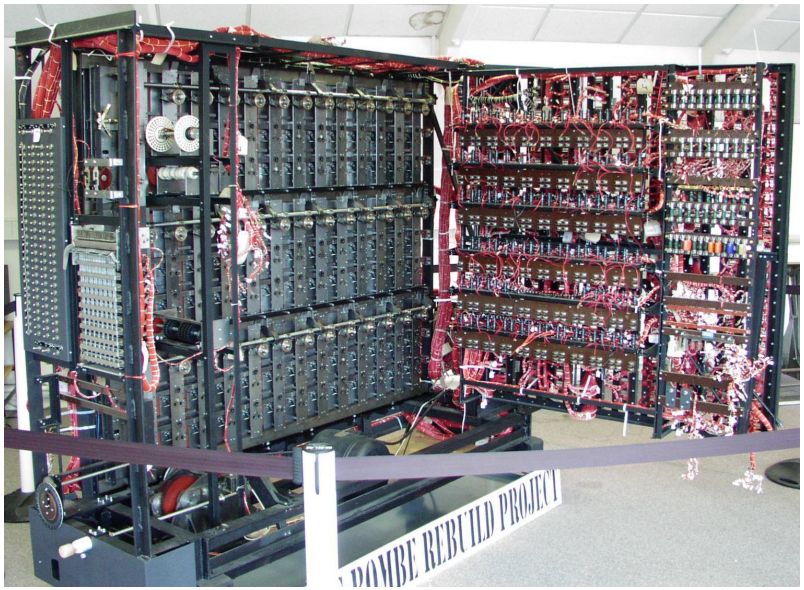
1. What is AI (Artificial Intelligence)?
2. History of data science
3. What is AI capable of?
4. How can we apply AI to space plasmas?
5. Examples of usage of AI in space
6. Data science hackathon

WHAT IS INFORMATION?



WHAT IS INFORMATION?

- Wikipedia: **Information** is an abstract concept that refers to something which has the power to inform.
- ChatGPT 4o: **Information** is data that has been processed or organized to make it meaningful and useful**.**

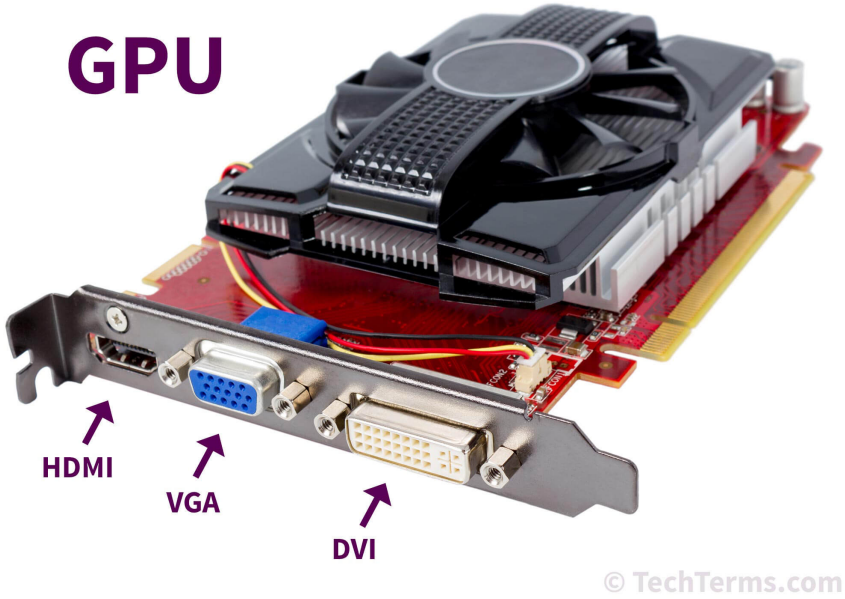


TRADITIONAL COMPUTING

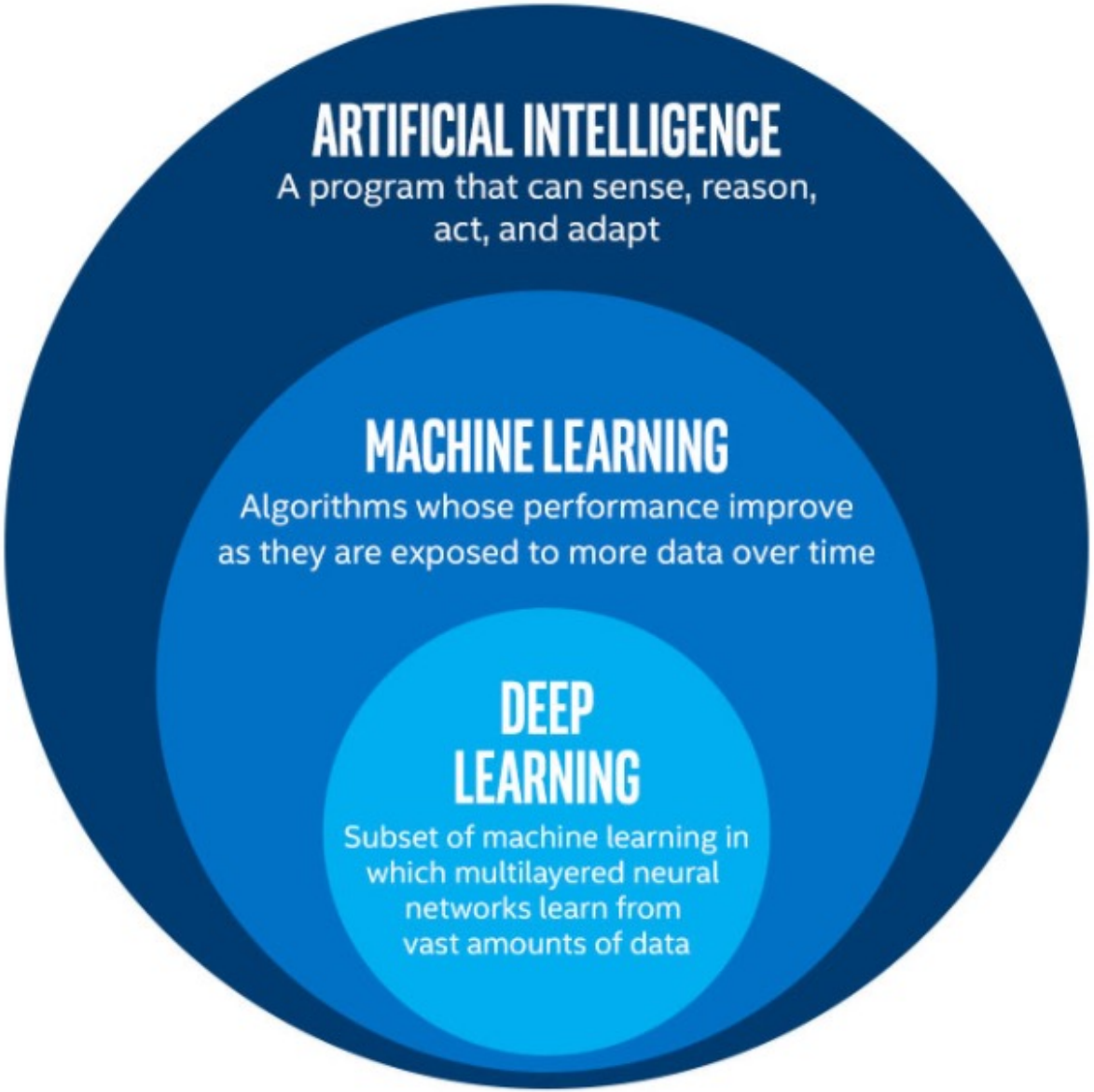
- Programming
- Human developed algorithms
- Challenge achieving human-level performance

EMERGENCE OF MACHINE LEARNING

Emergence of data sharing



Availability of massively parallel hardware



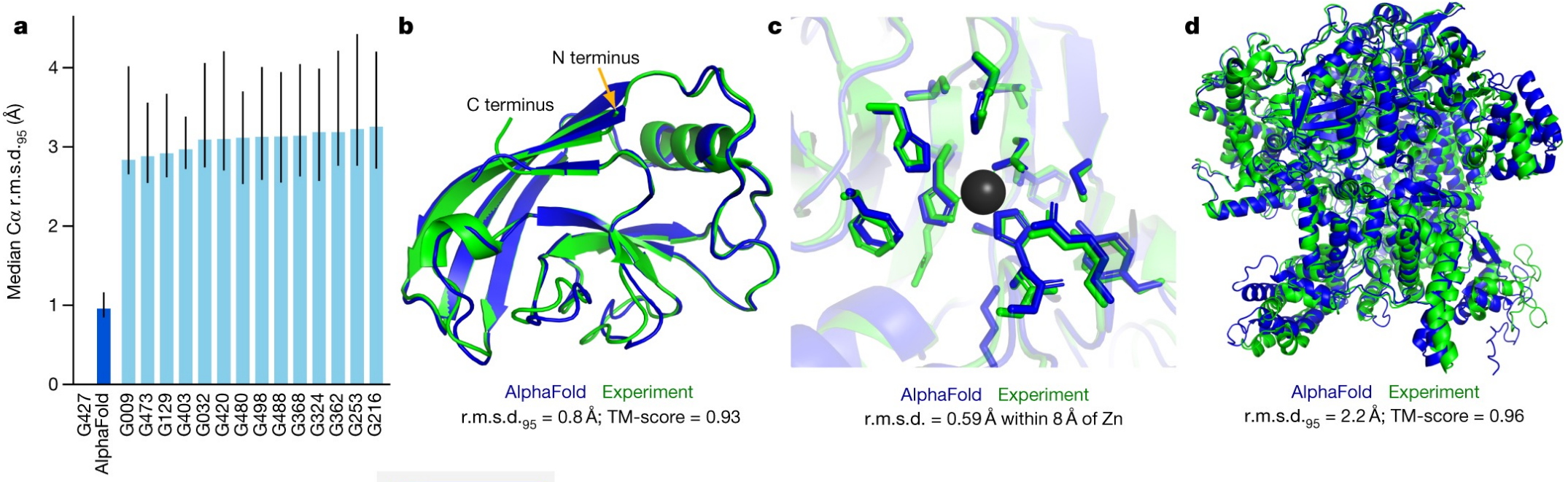
MACHINE LEARNING VS ARTIFICIAL INTELLIGENCE

ALPHA GO: HUMAN CHAMPION BEATEN

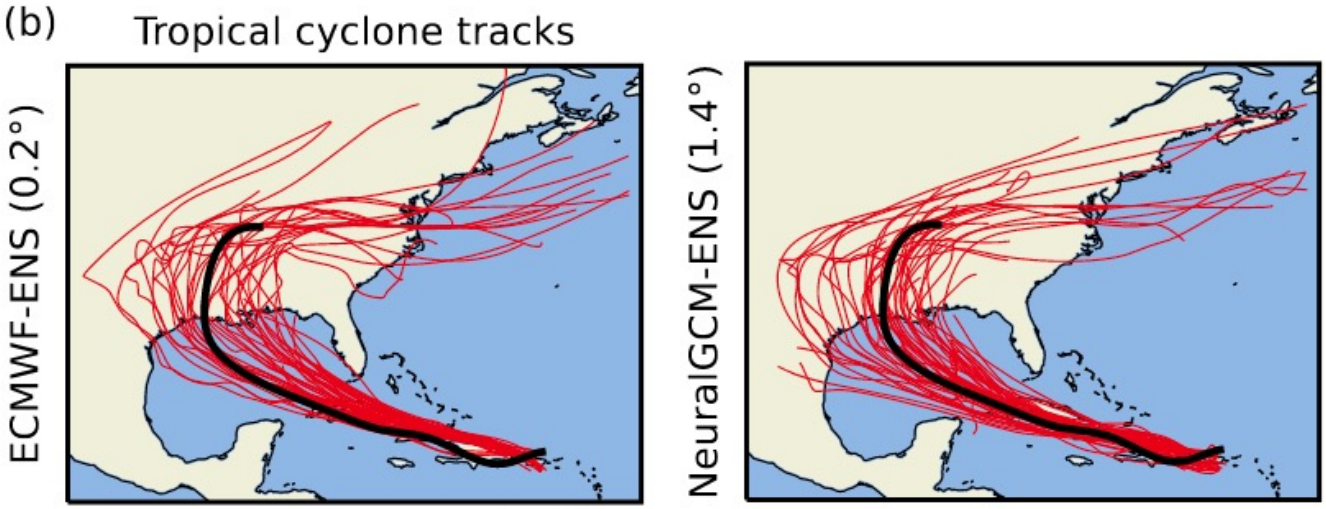
Lee Sedol facing AI opponent in Go

APPLICATION OF ML TO PROTEIN FOLDING

Amino Acid sequence → Google DeepMind → 3D Structure



USING AI TO PREDICT TROPICAL CYCLONES



nature

Explore content · About the journal · Publish with us

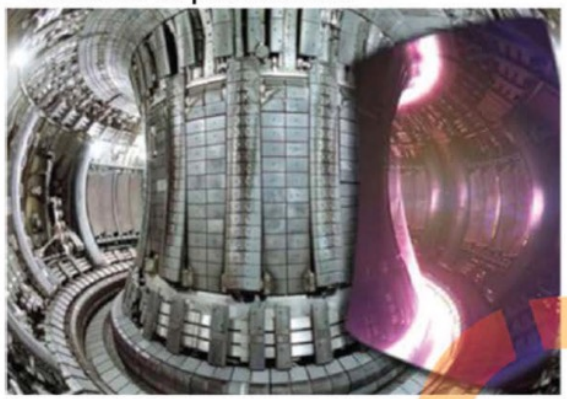
nature › articles › article

Article | Open access | Published: 22 July 2024

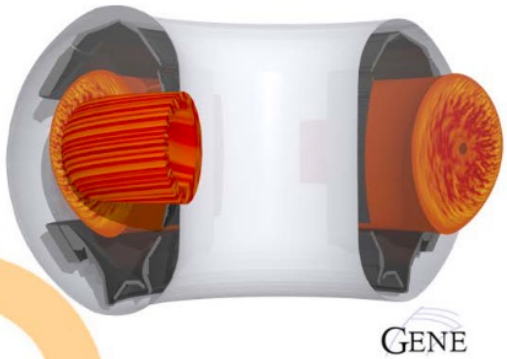
Neural general circulation models for weather and climate

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan

1 second of plasma in...

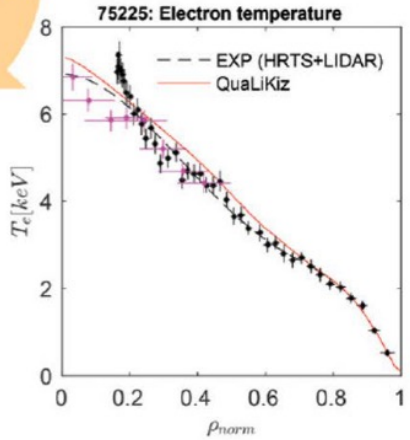


1 year on 1000 cores...



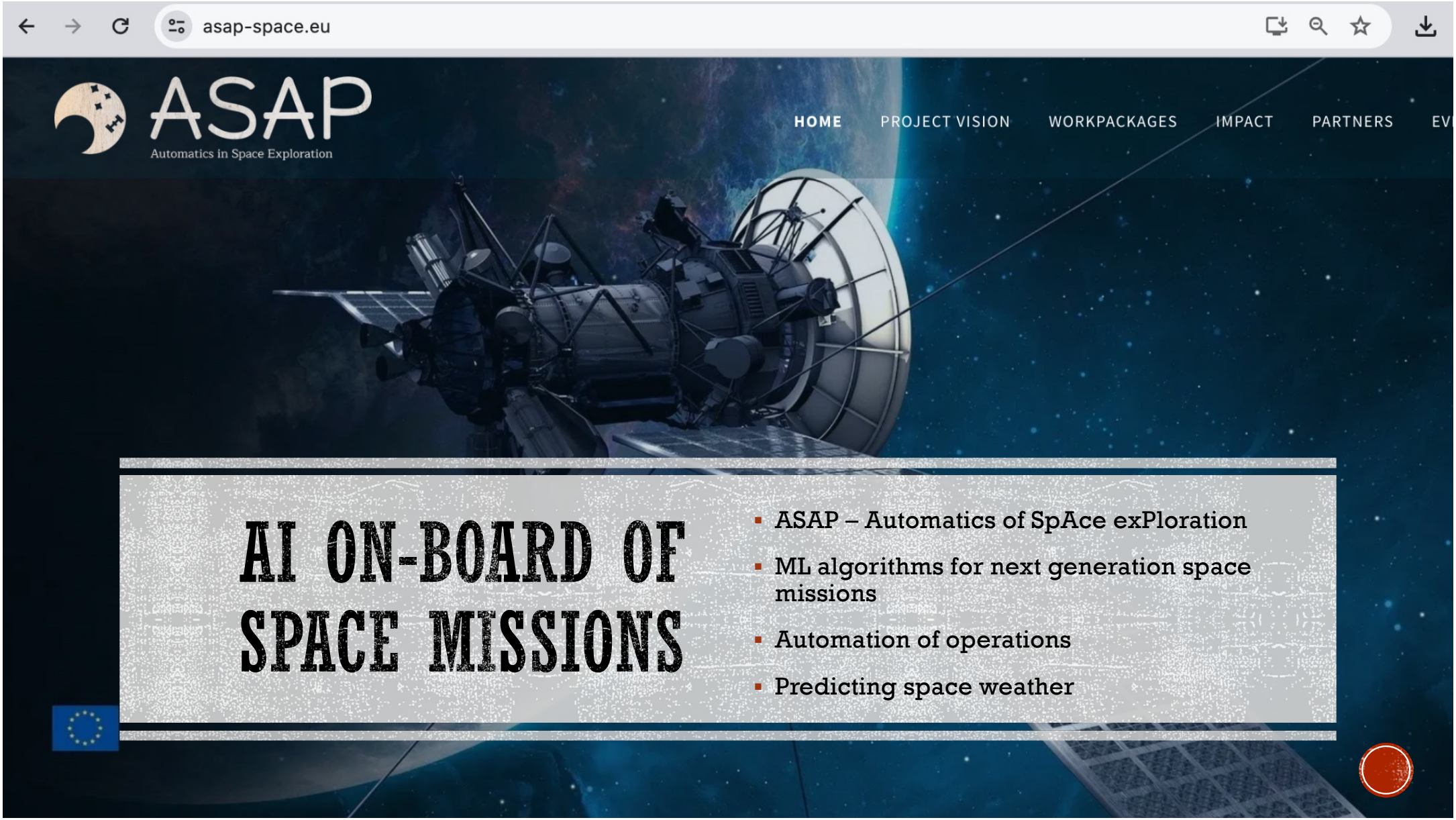
1 second on 1 core...

1 day on 24 cores...



ACCELERATE PREDICTIVE MODEL CONTROL

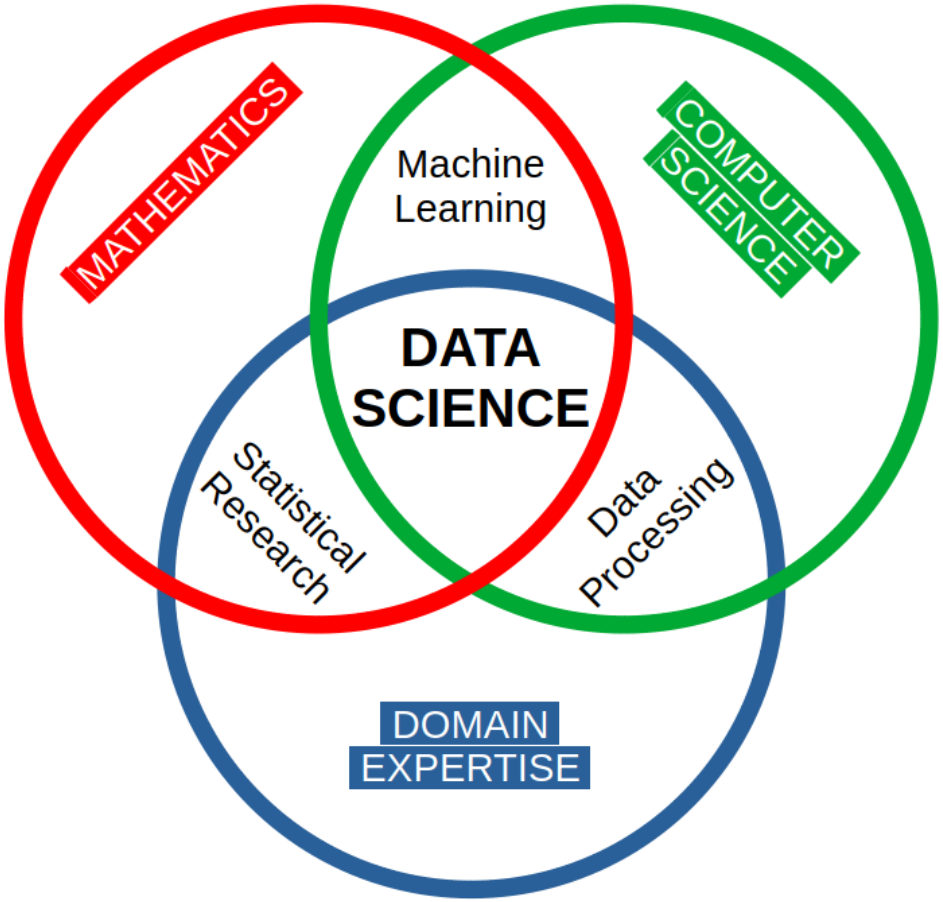
- direct and faster interpretation of complex experimental data
- Extract relevant information from measurements faster
- Automatic control

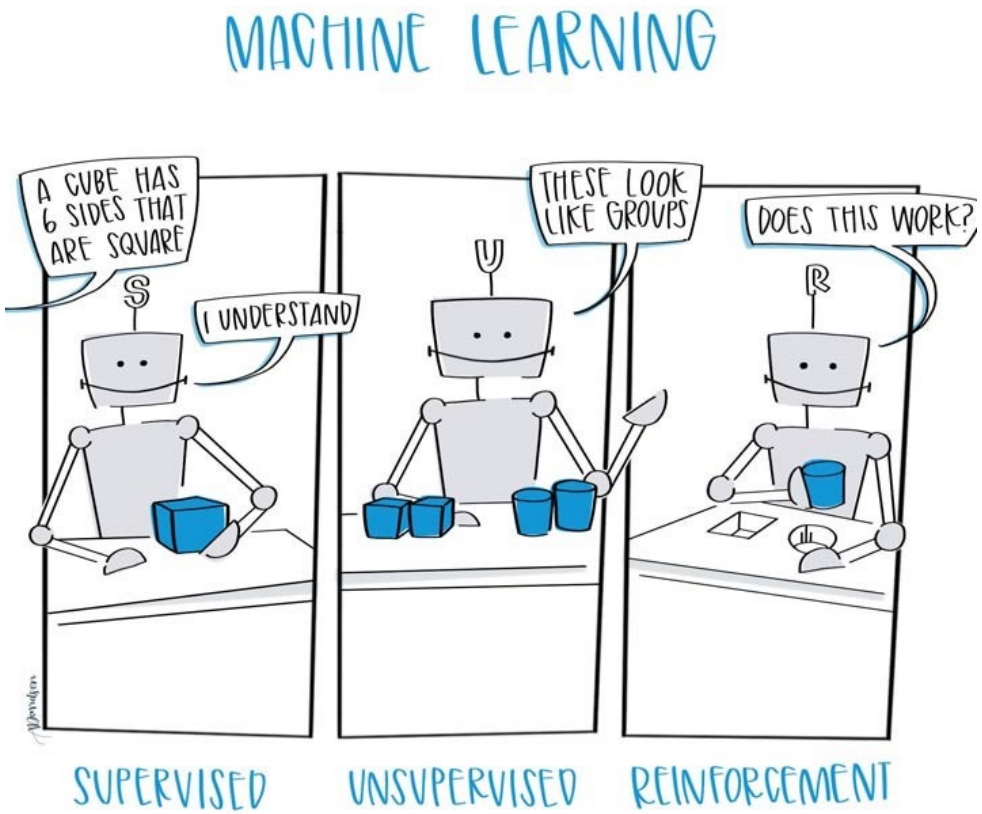

asap-space.eu
ASAP
Automatics in Space Exploration
HOME
PROJECT VISION
WORKPACKAGES
IMPACT
PARTNERS
AI ON-BOARD OF SPACE MISSIONS
ASAP – Automatics of SpAce exPloration
ML algorithms for next generation space missions
Automation of operations
Predicting space weather

Input image

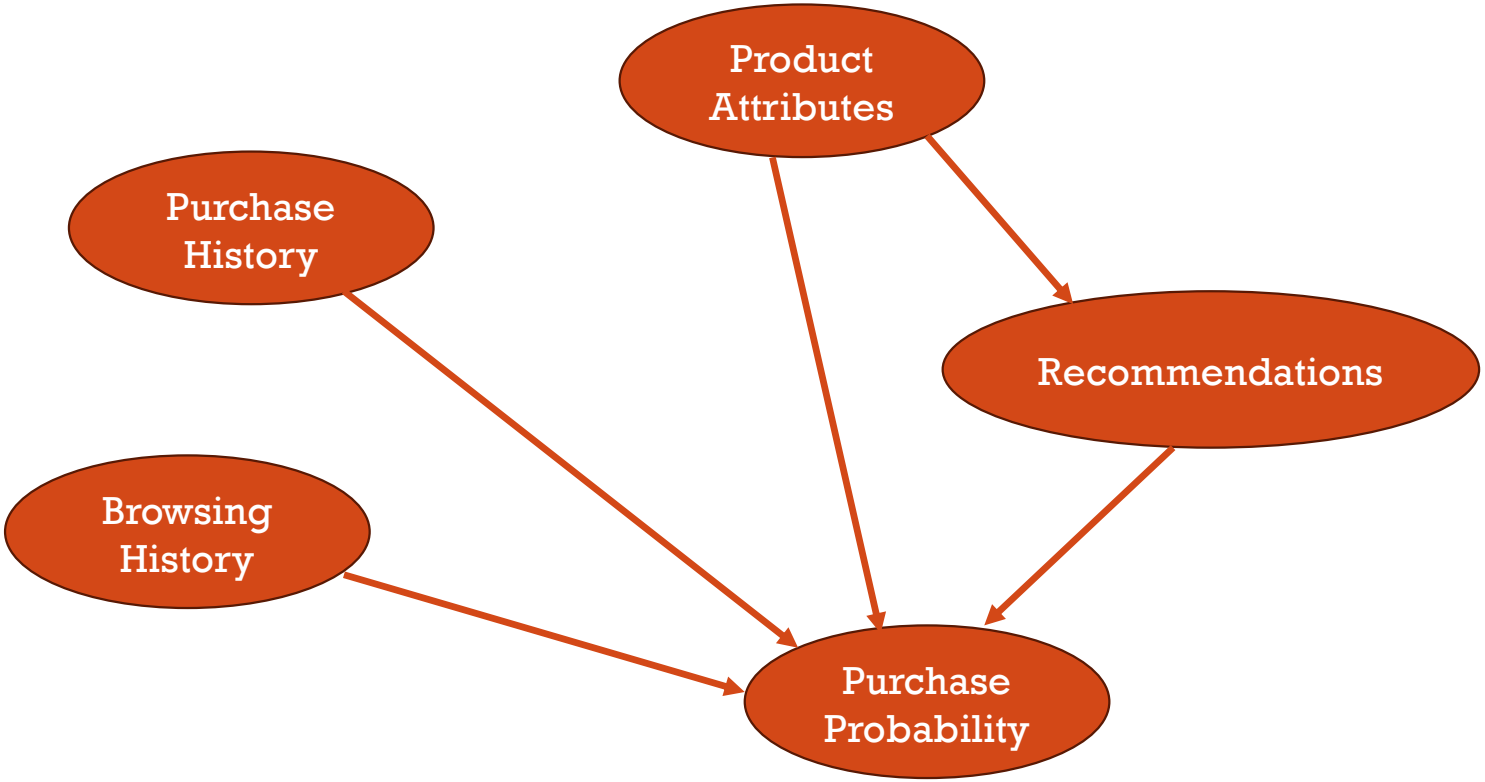
WHAT IS ARTIFICIAL INTELLIGENCE?

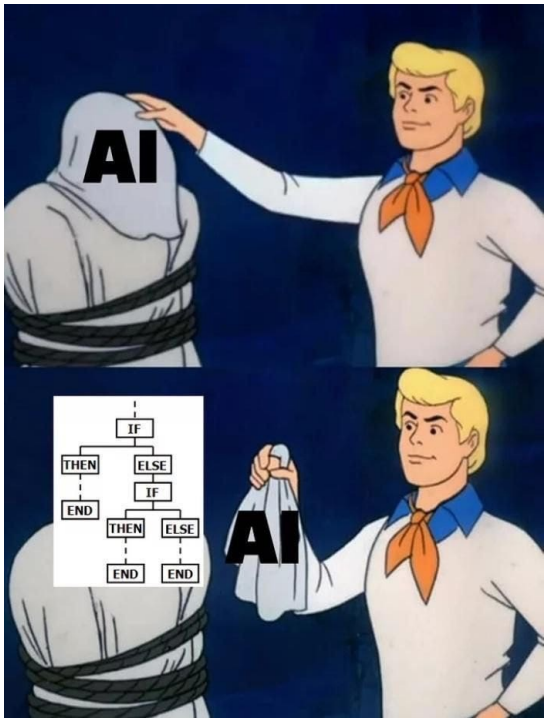
DIFFERENT TYPES OF MACHINE LEARNING



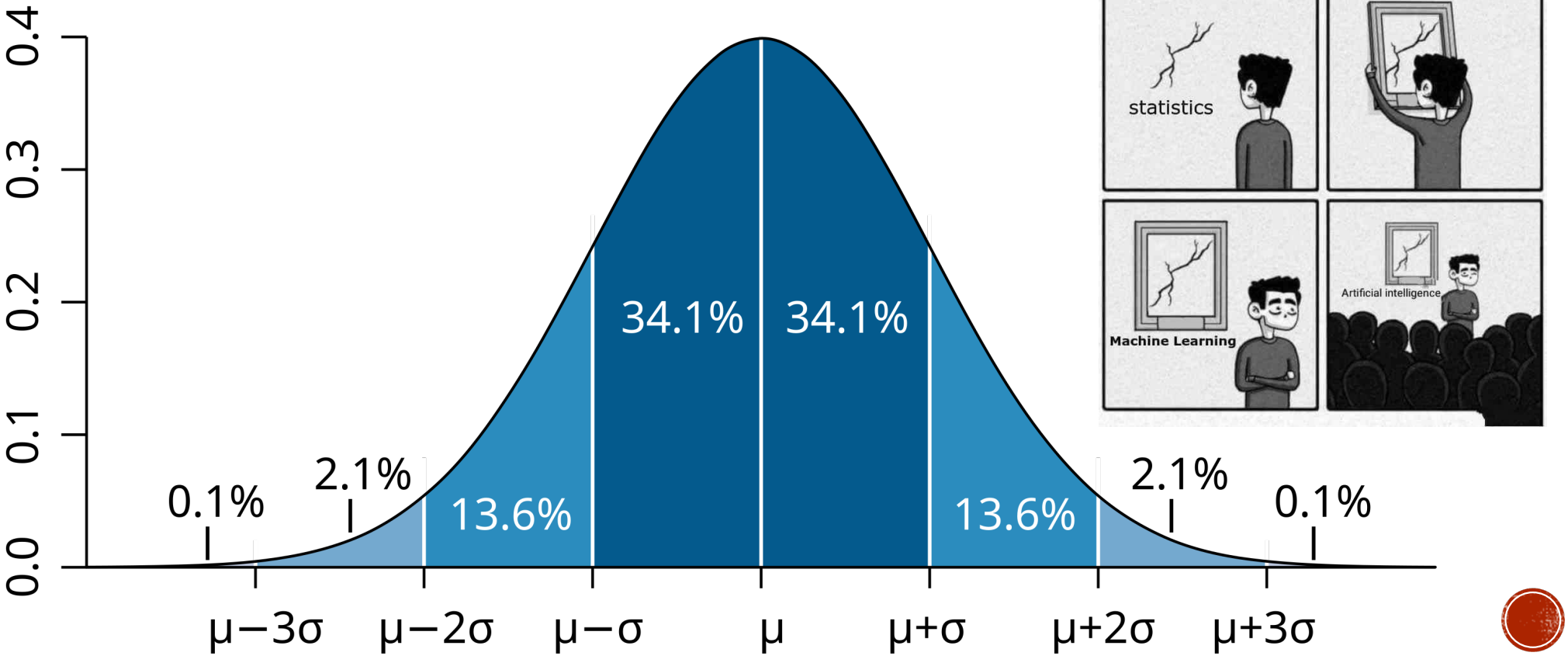


WHAT IS MACHINE LEARNING AFTER ALL?

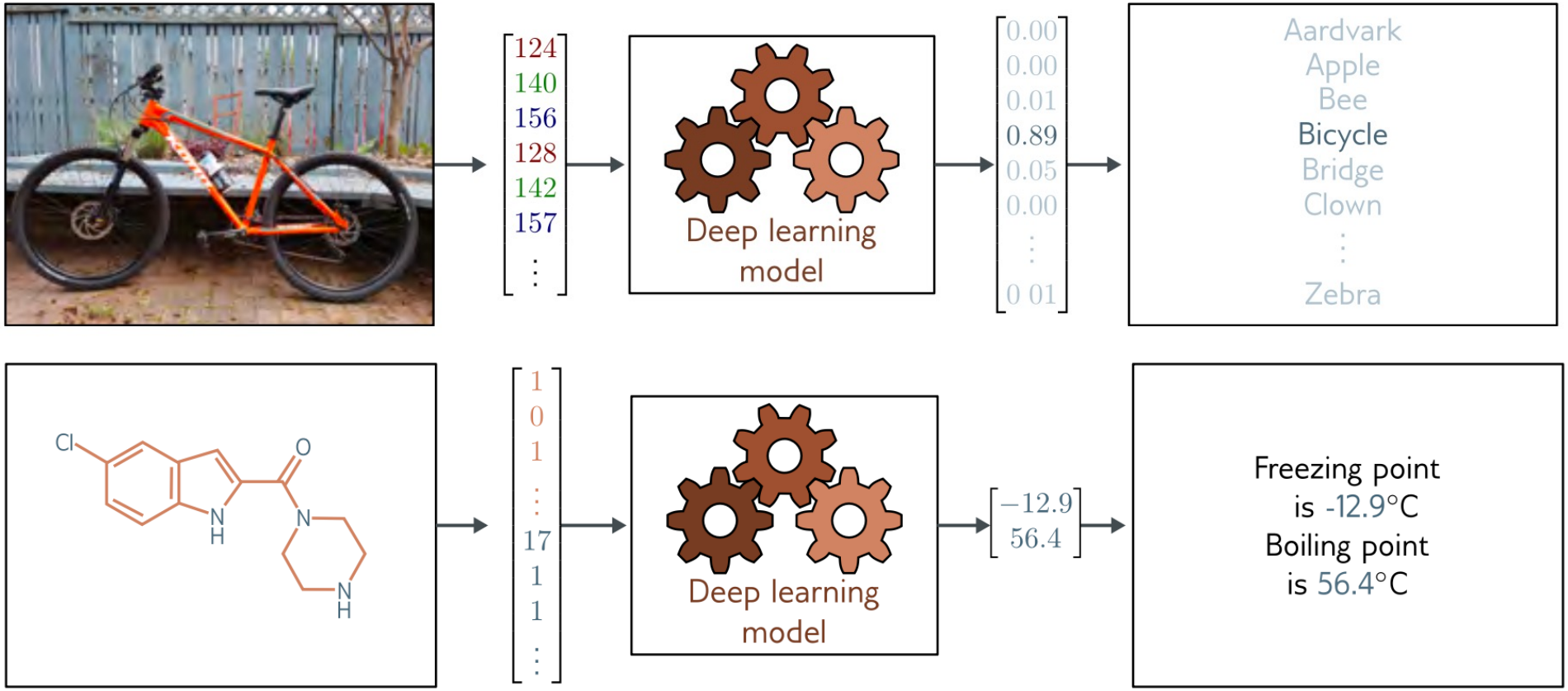




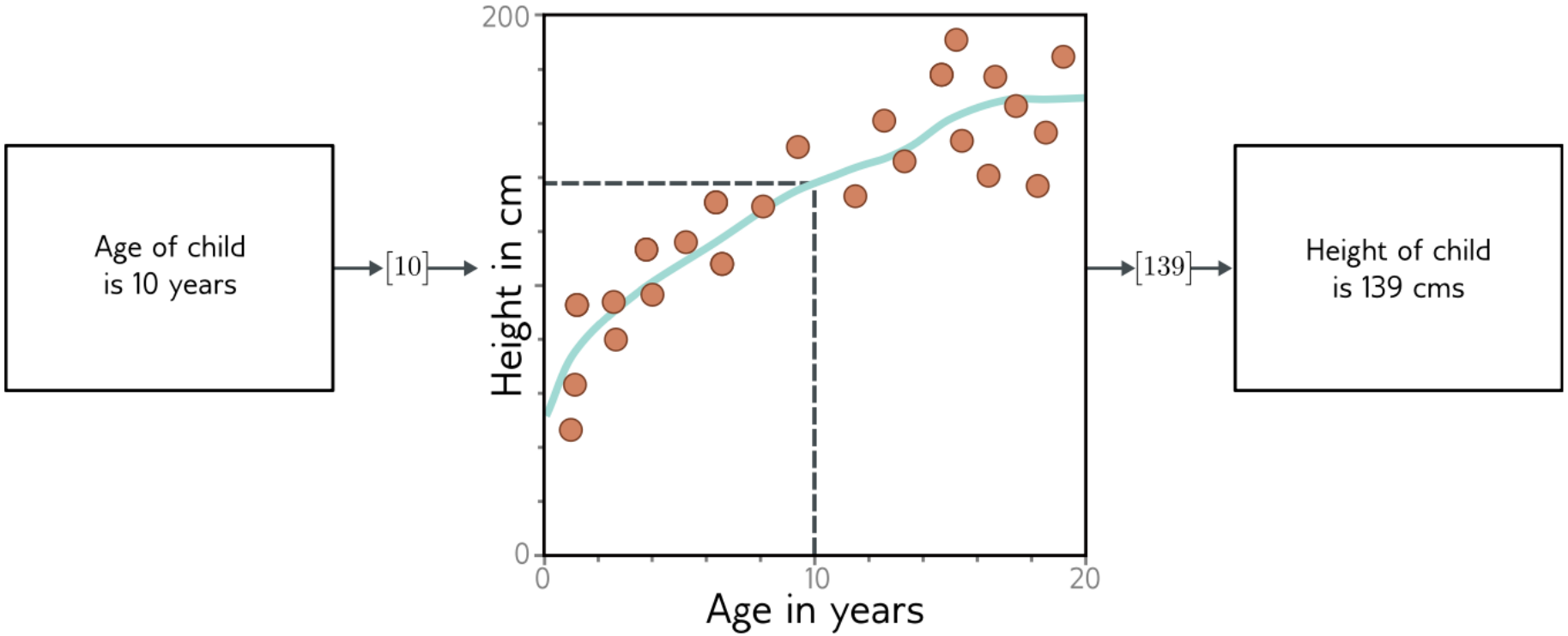
MACHINE LEARNING AND STATISTICS



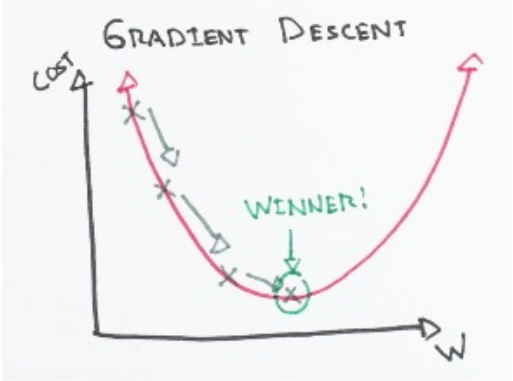
SUPERVISED LEARNING



REGRESSION

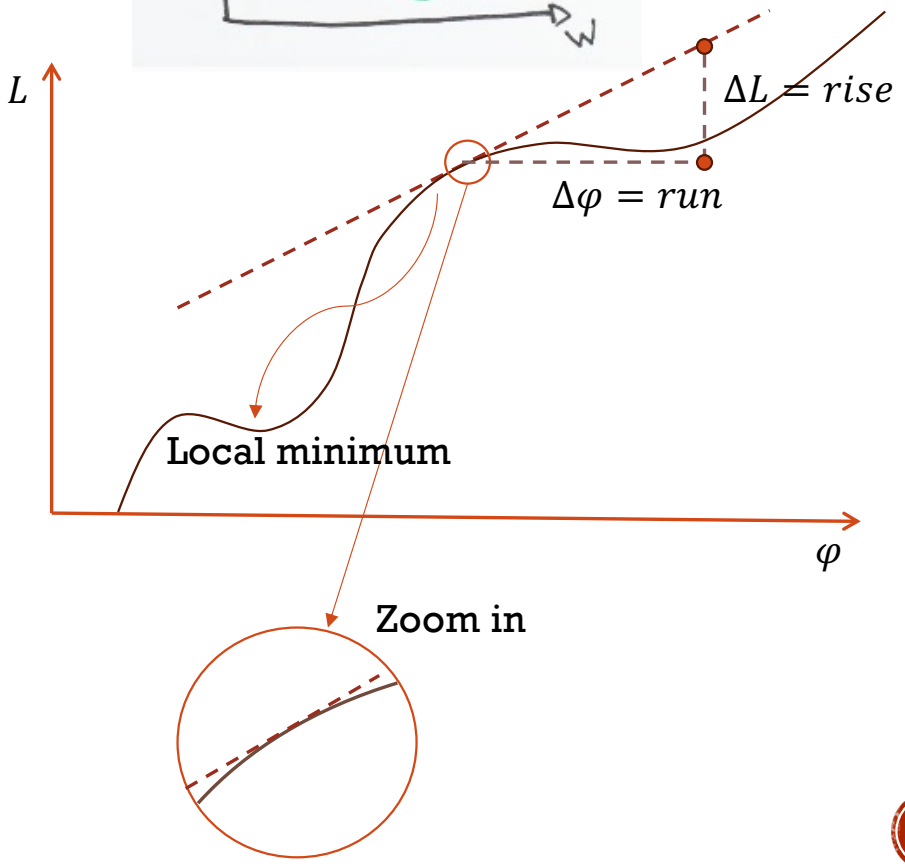

Age of child
is 10 years
[10]
200
0
Height in cm
0
10
20
Age in years
[139]
Height of child
is 139 cms

OPTIMIZATION PROBLEMS



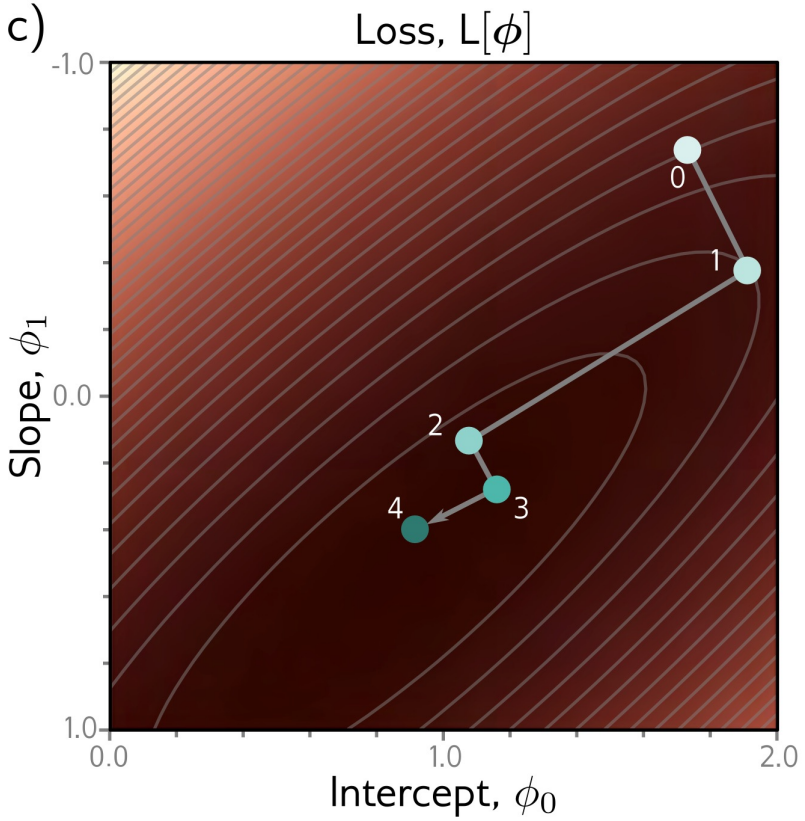
- How do we find a minimum?
- What is a derivative?
- How do I choose optimal step?

$$\frac{\partial L}{\partial \phi} \longrightarrow \frac{L(\phi+\Delta\phi)-L(\phi)}{\Delta\phi}$$



GRADIENT DESCENT

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\Big[\mathrm{L}[\phi]\Big]$$



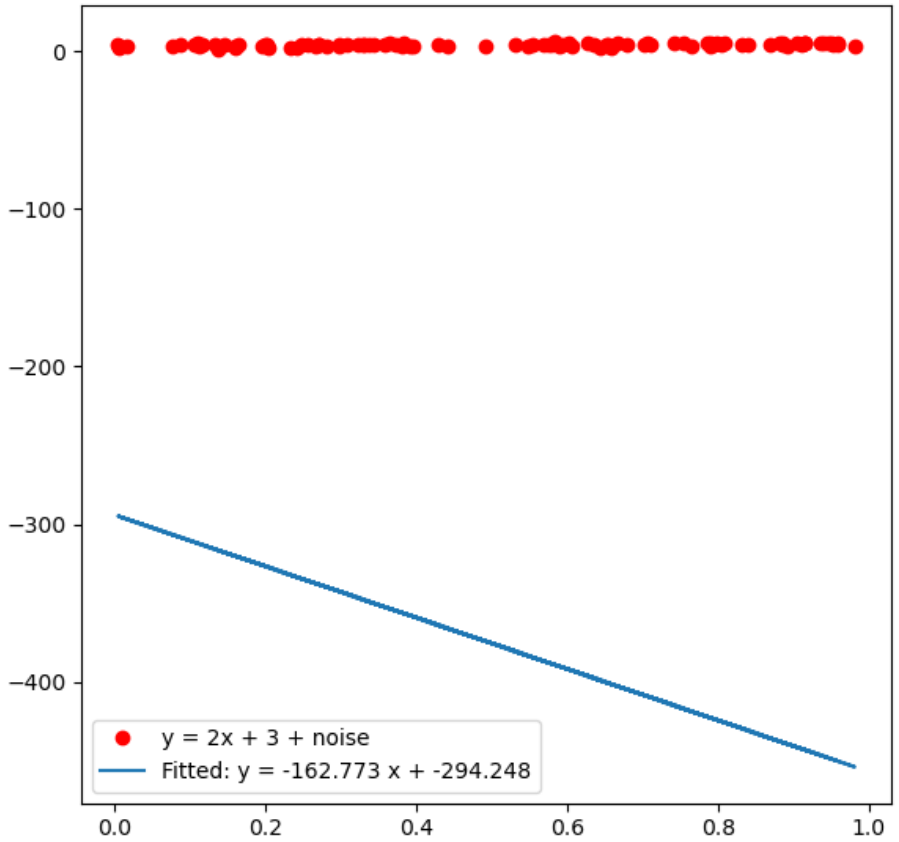
$$\phi \longleftarrow \phi - \alpha \cdot \frac{\partial L}{\partial \phi}$$

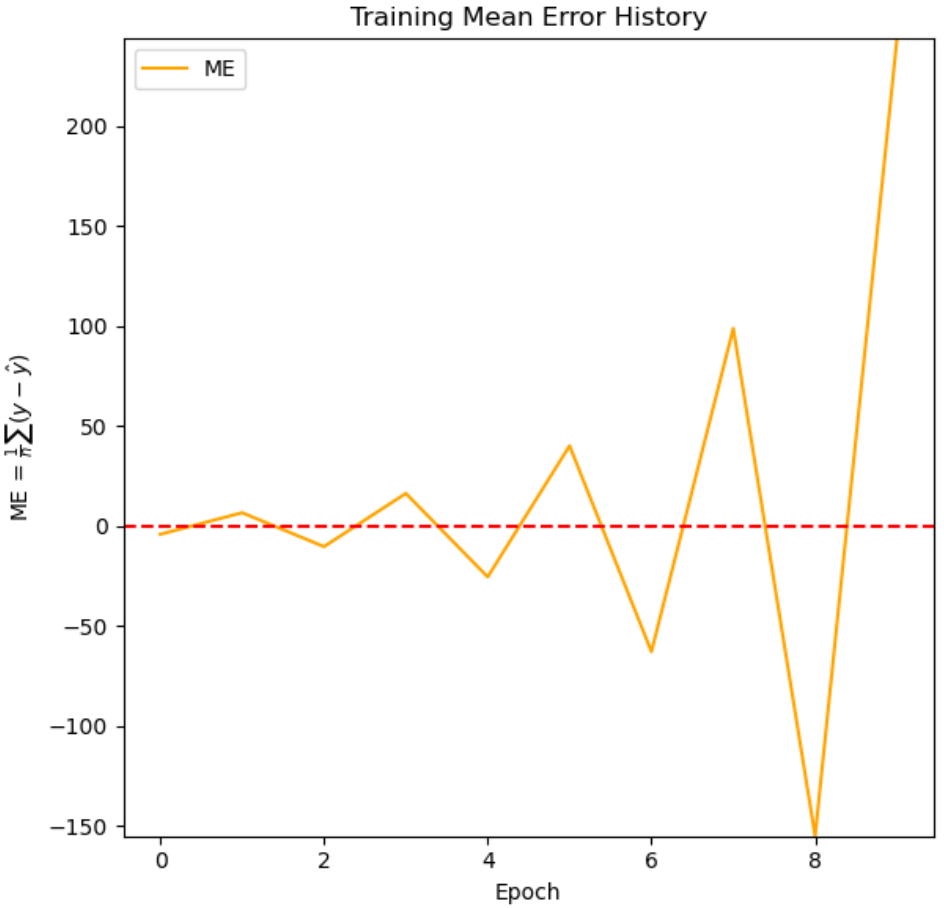
Step size (learning rate):

$$\mathrm{L}[\phi] = \sum_{i=1}^{I} \ell_i = \sum_{i=1}^{I} \left(\mathrm{f}[x_i, \phi] - y_i\right)^2$$

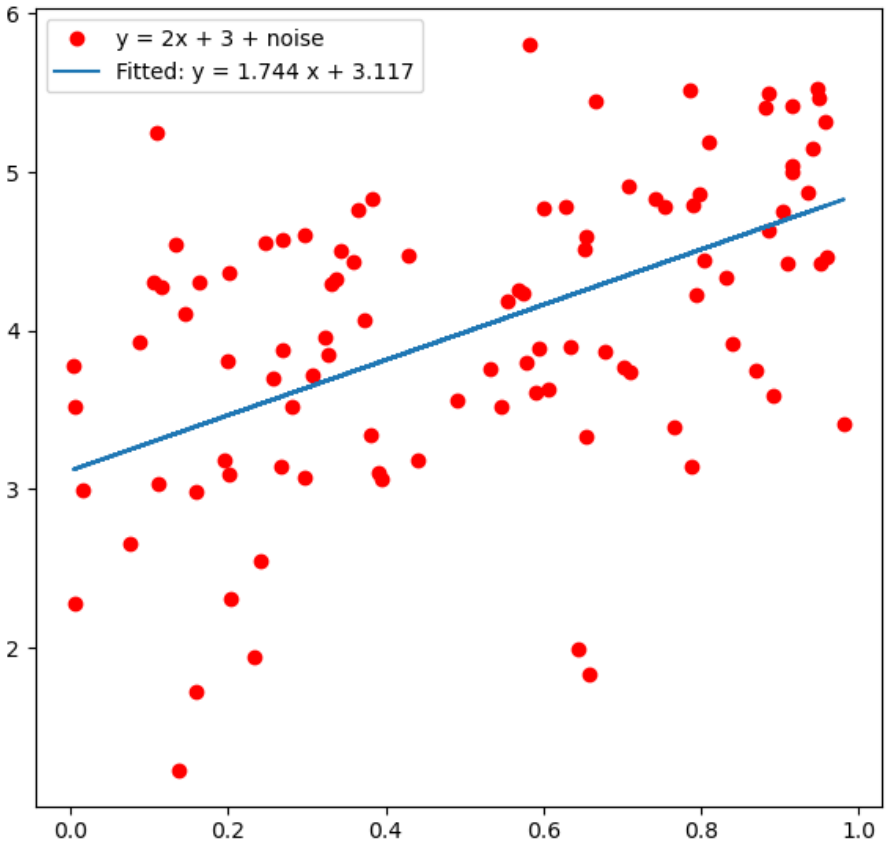
$$= \sum_{i=1}^{I} \left(\phi_0 + \phi_1 x_i - y_i\right)^2$$

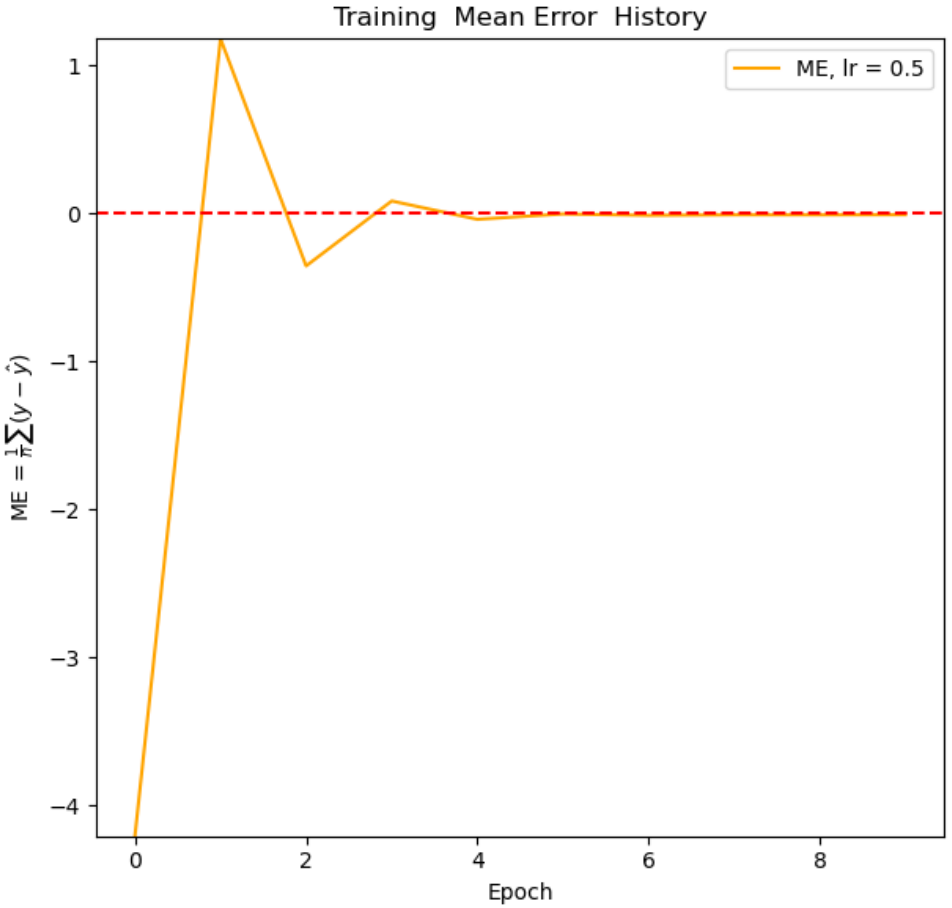
REGRESSION: Y = 2 X + 3 + RANDOM



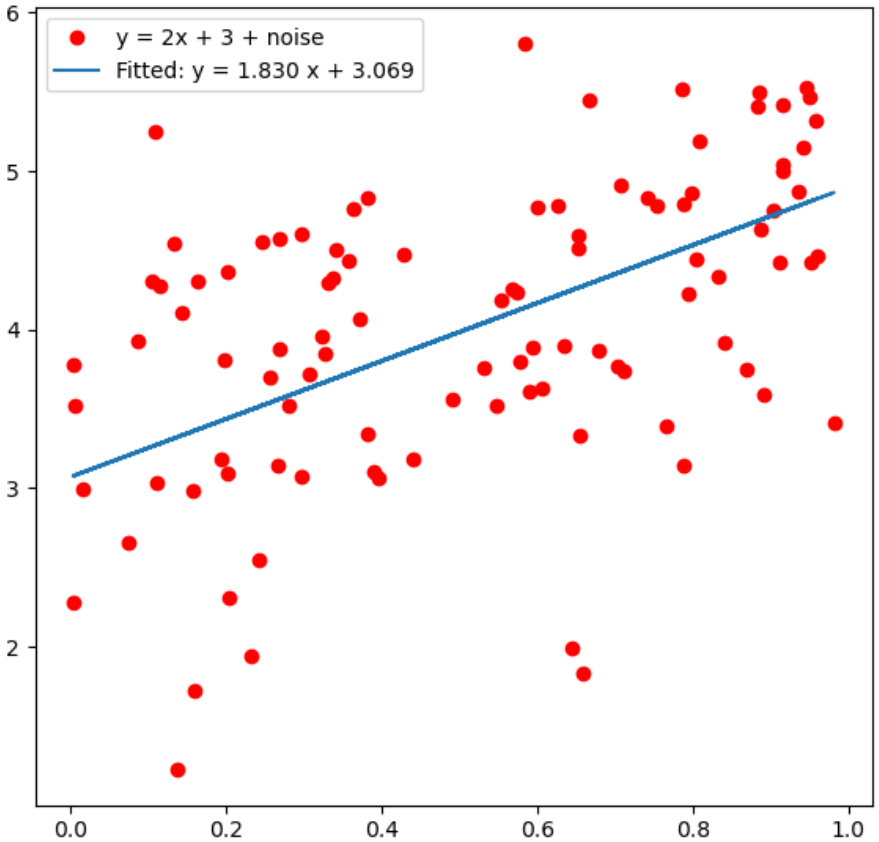


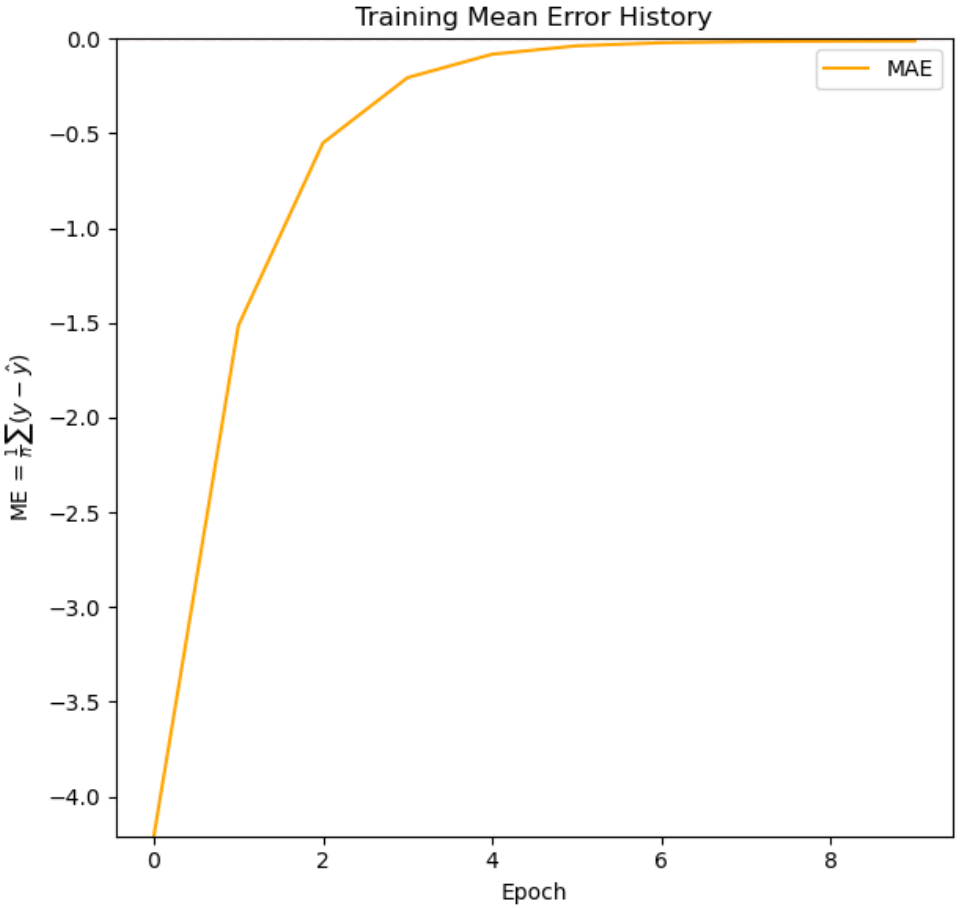
REGRESSION: Y = 2 X + 3 + RANDOM





REGRESSION: Y = 2 X + 3 + RANDOM

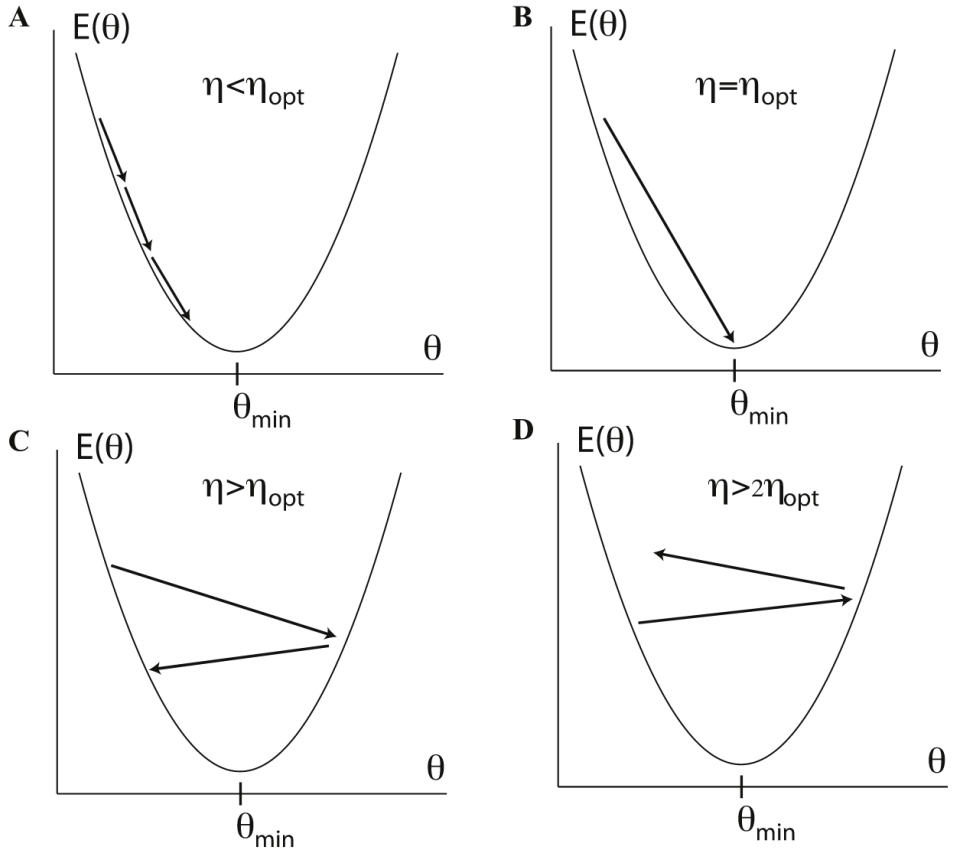




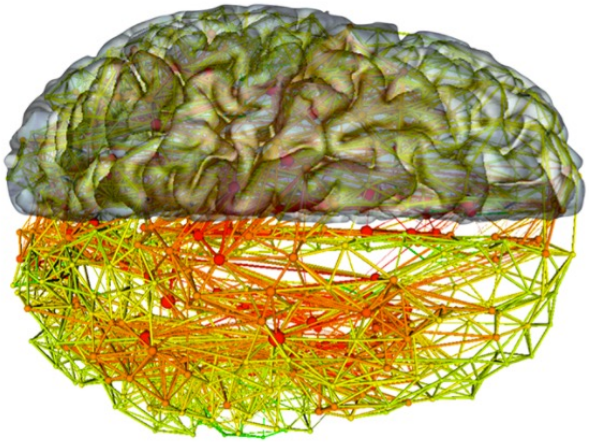
OPTIMIZATION PROBLEMS

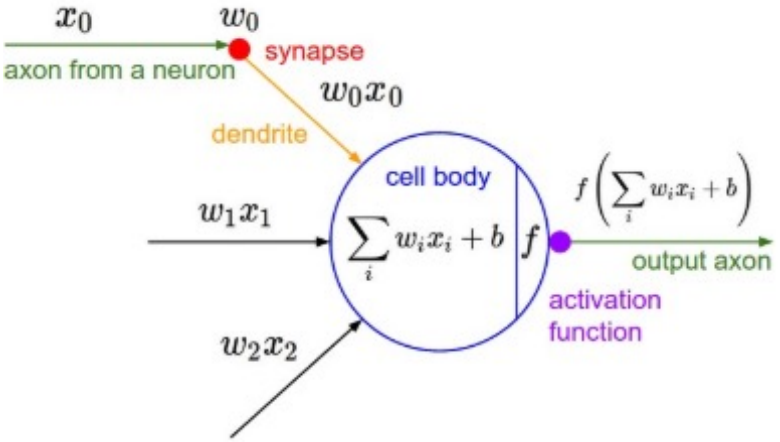
- How do we find a minimum?
- How do I choose optimal step?

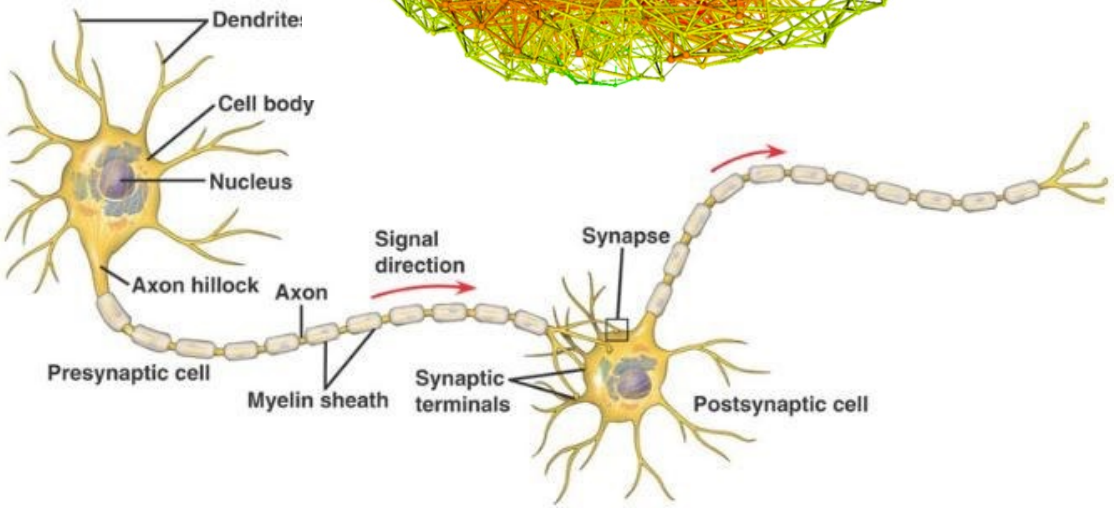
$$\frac{\partial L}{\partial \phi} \longrightarrow \frac{L(\phi+\Delta\phi)-L(\phi)}{\Delta\phi}$$

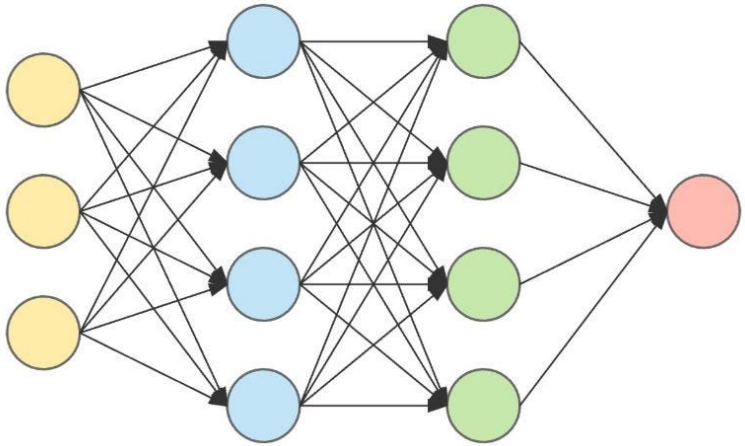


REAL VS ARTIFICIAL NEURAL NETWORKS

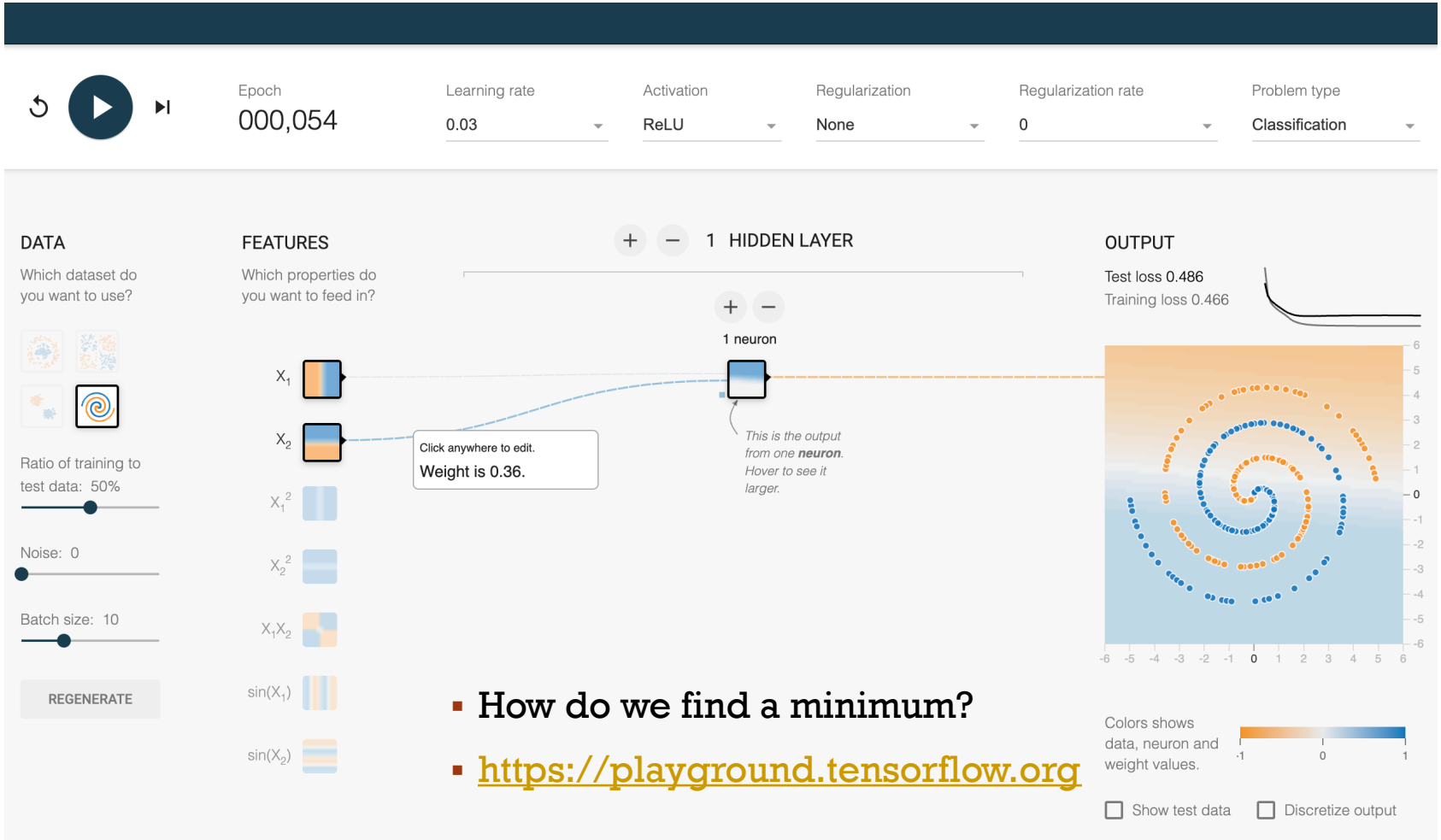




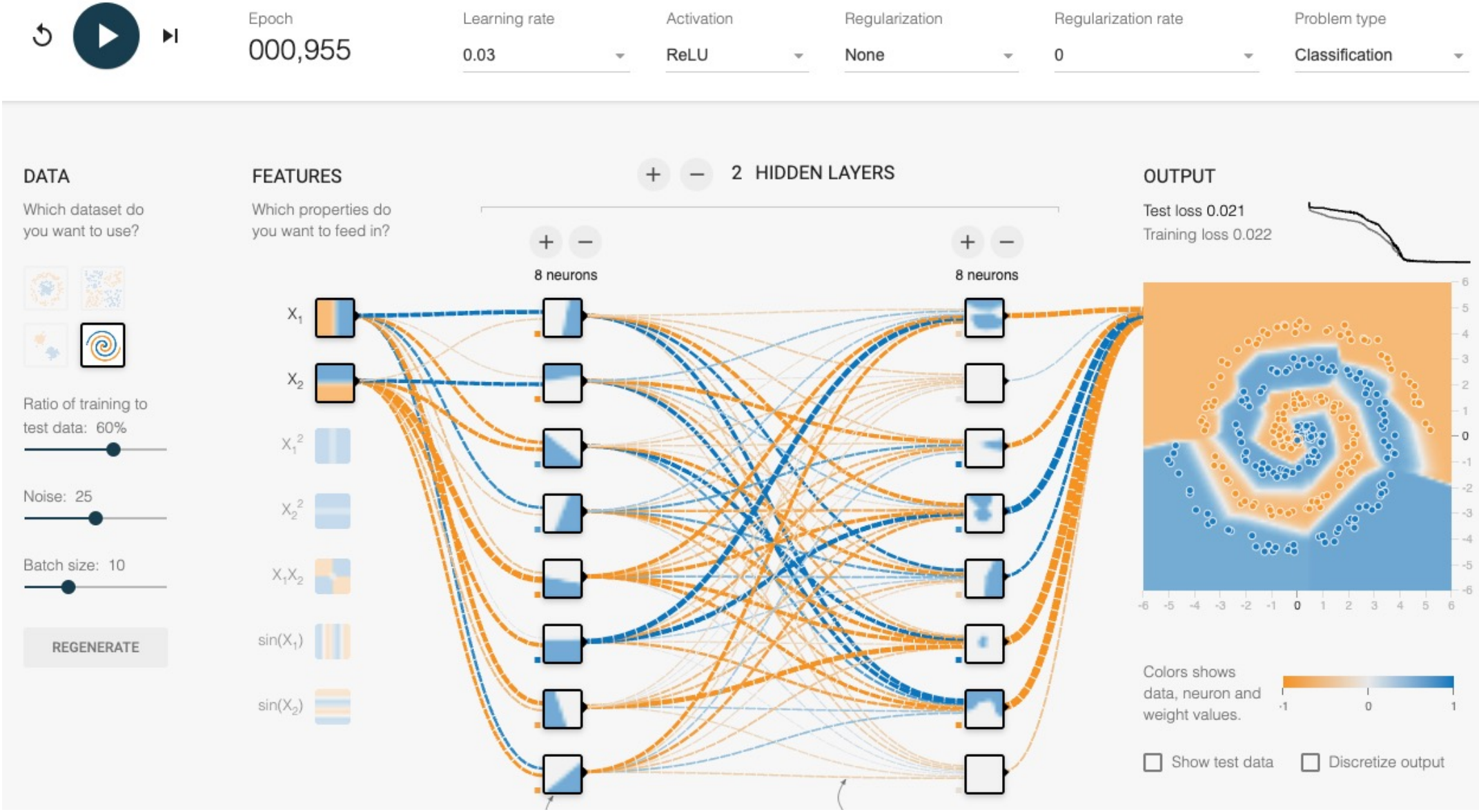




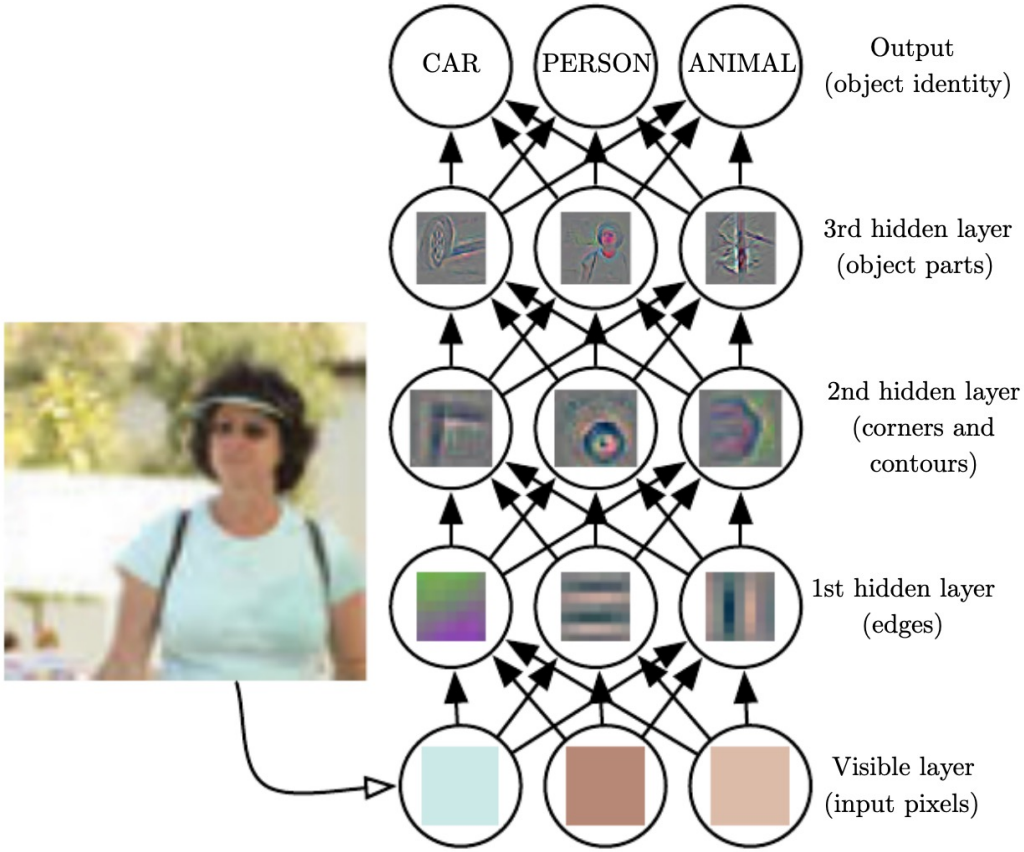
HOW NEURAL NETWORKS WORK

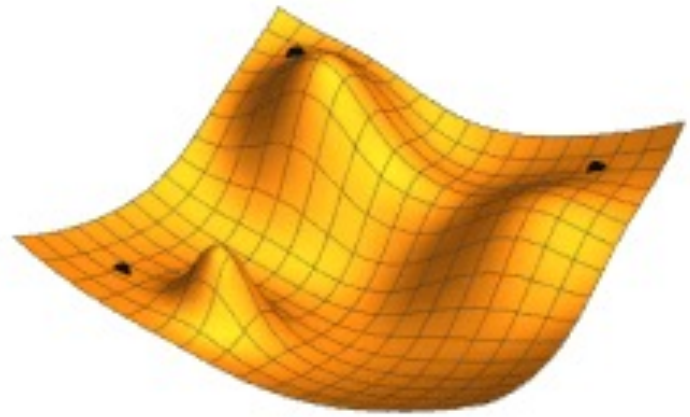


- How do we find a minimum?
- https://playground.tensorflow.org

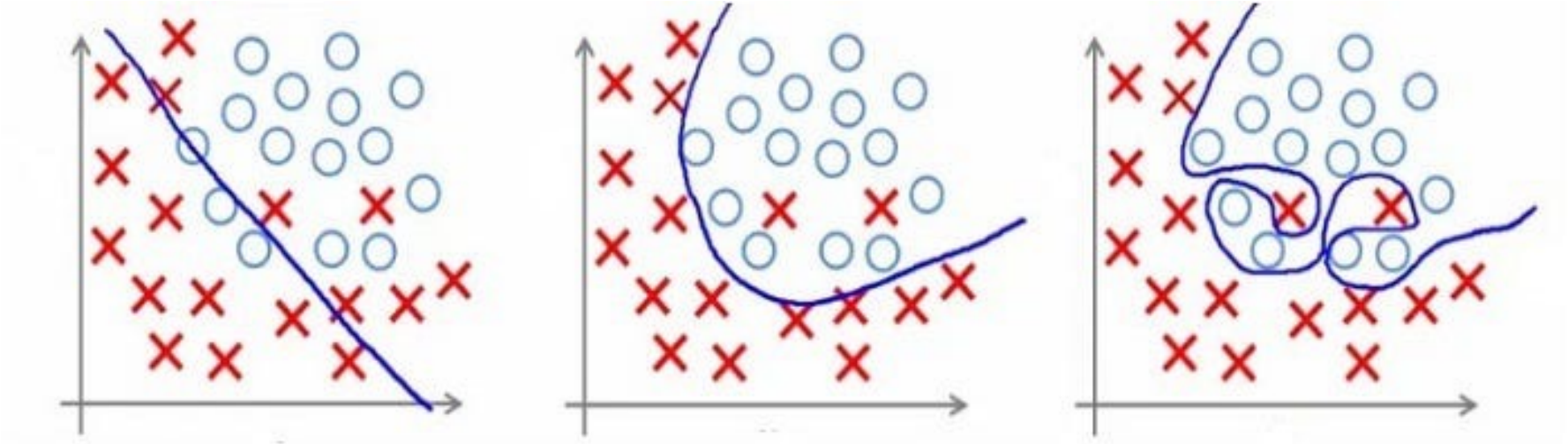

Epoch
000,955
Learning rate
0.03
Activation
ReLU
Regularization
None
Regularization rate
0
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 60%
Noise: 25
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
2 HIDDEN LAYERS
8 neurons
8 neurons
OUTPUT
Test loss 0.021
Training loss 0.022
Colors shows data, neuron and weight values.
Show test data
Discretize output

WHY DO ARTIFICIAL NEURAL NETS WORK?





OVERFITTING



Under-fitting

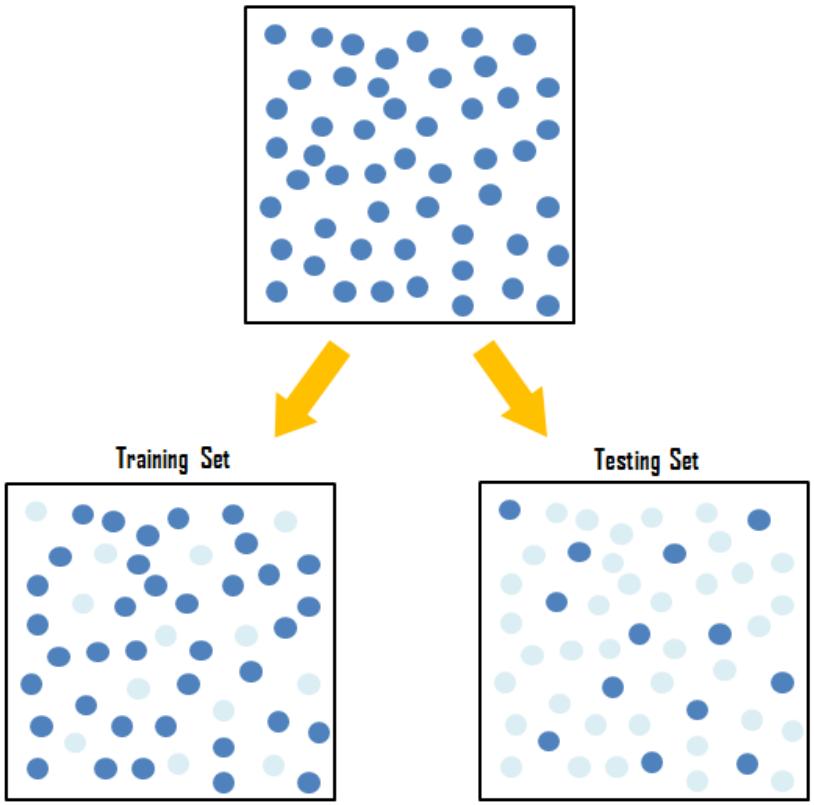
(too simple to explain the variance)

Appropriate-fitting

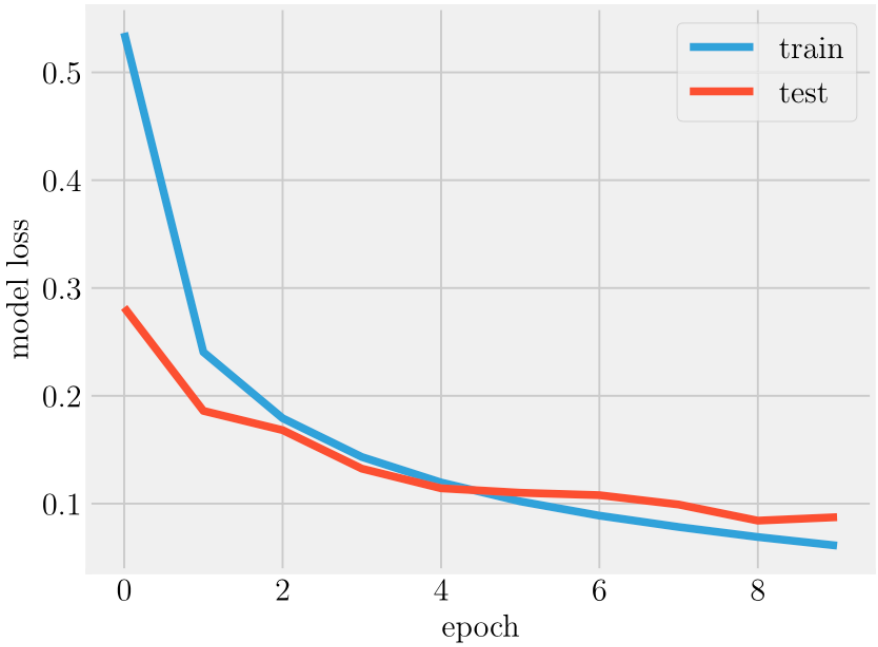
Over-fitting

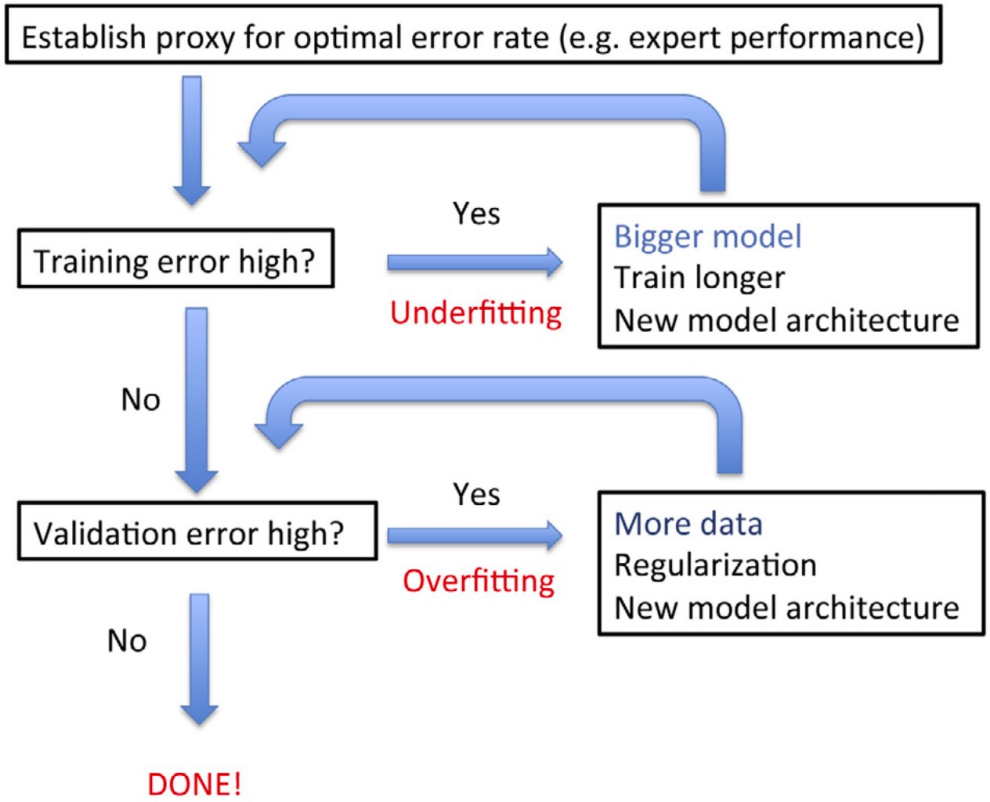
(forcefitting -- too good to be true)

TRAINING, VALIDATION, TESTING

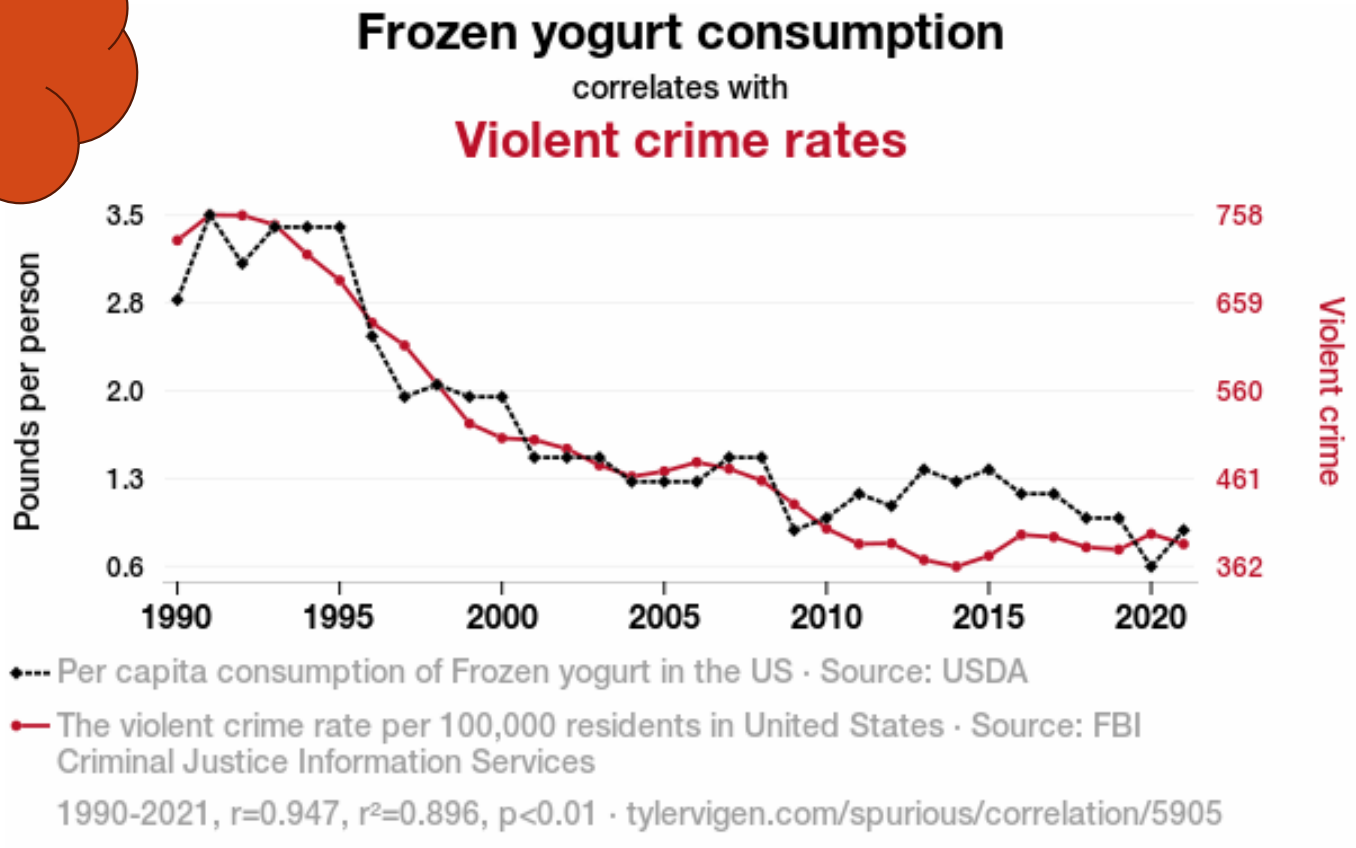


TRAINING: WHEN TO STOP?

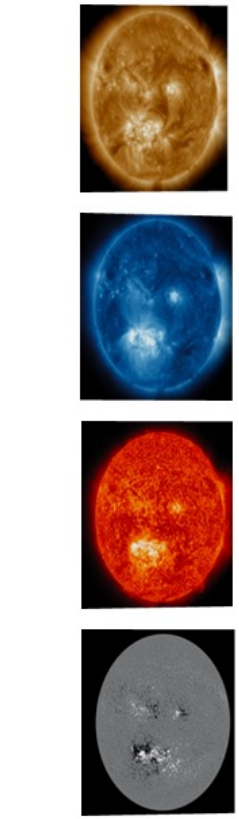




FUNNY CORRELATIONS



FORECASTING FLARES

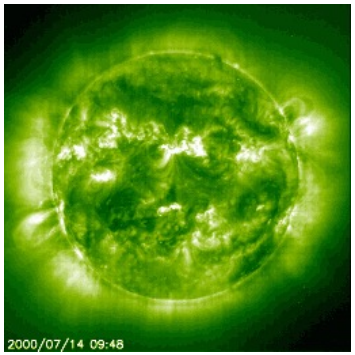


Inputs: 193Å, 304Å,171Å, HMI

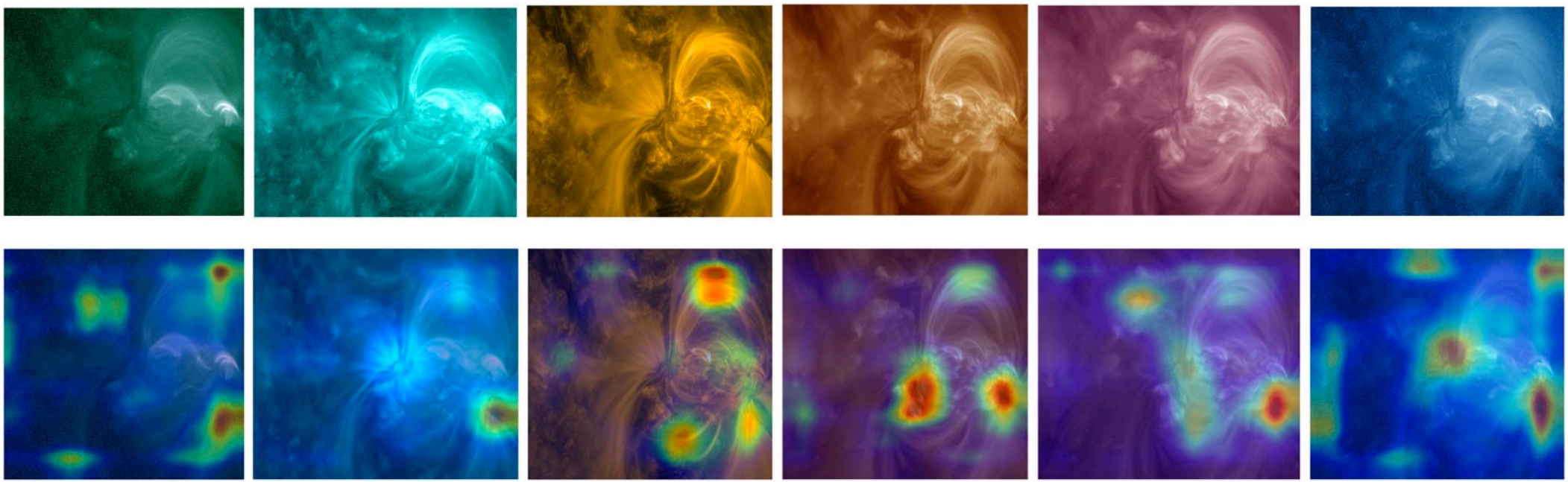
CNNs : Convolutions and poolings

Late feature map concatenation

Forecast result



FLARES: GRADIENTS TO CREATE HEATMAPS



SOMETHING IS WRONG WITH THE PANDA

$+ .007 \times$

$=$

$\boldsymbol{x}$

$y =$“panda”
w/ 57.7%
confidence

$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

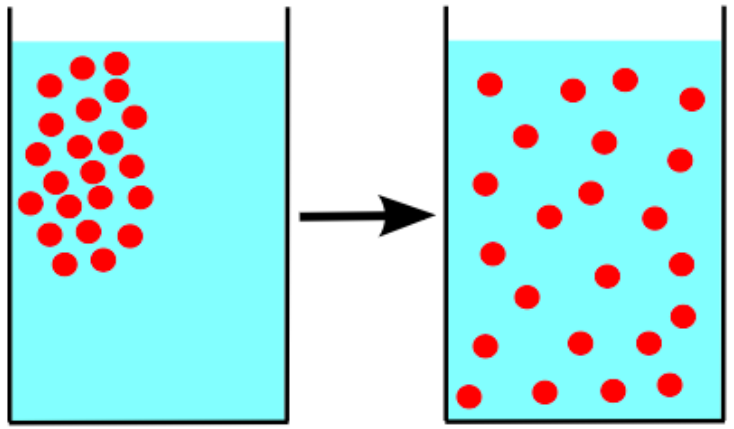
“nematode”
w/ 8.2%
confidence

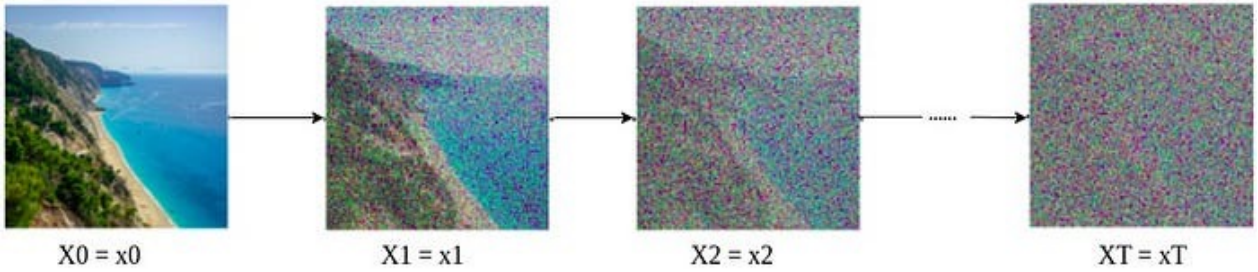
$\boldsymbol{x} +$
$\epsilon \, \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

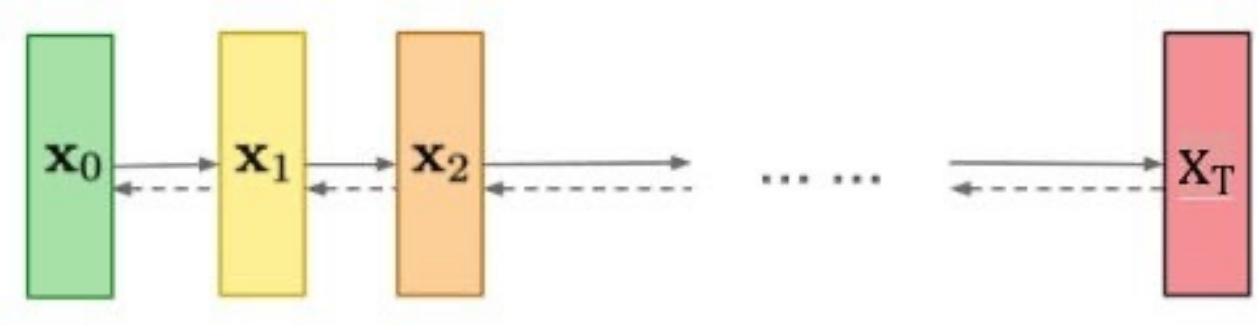
“gibbon”
w/ 99.3 %
confidence

GENERATIVE AI

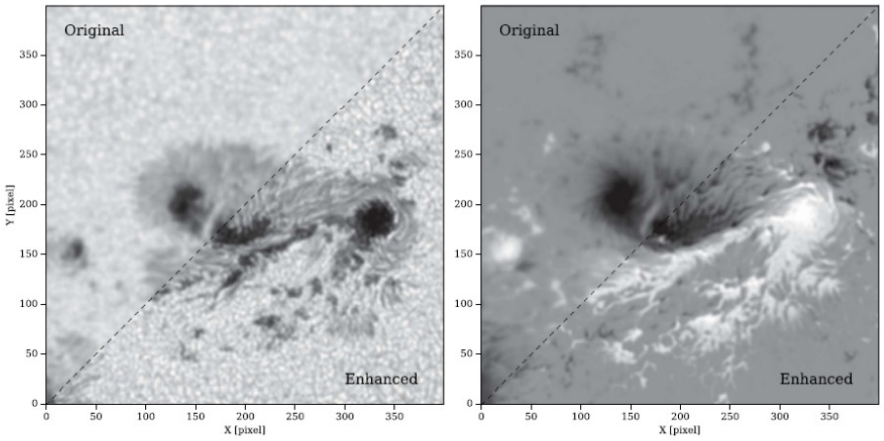
HOW DALL-E WORKS?

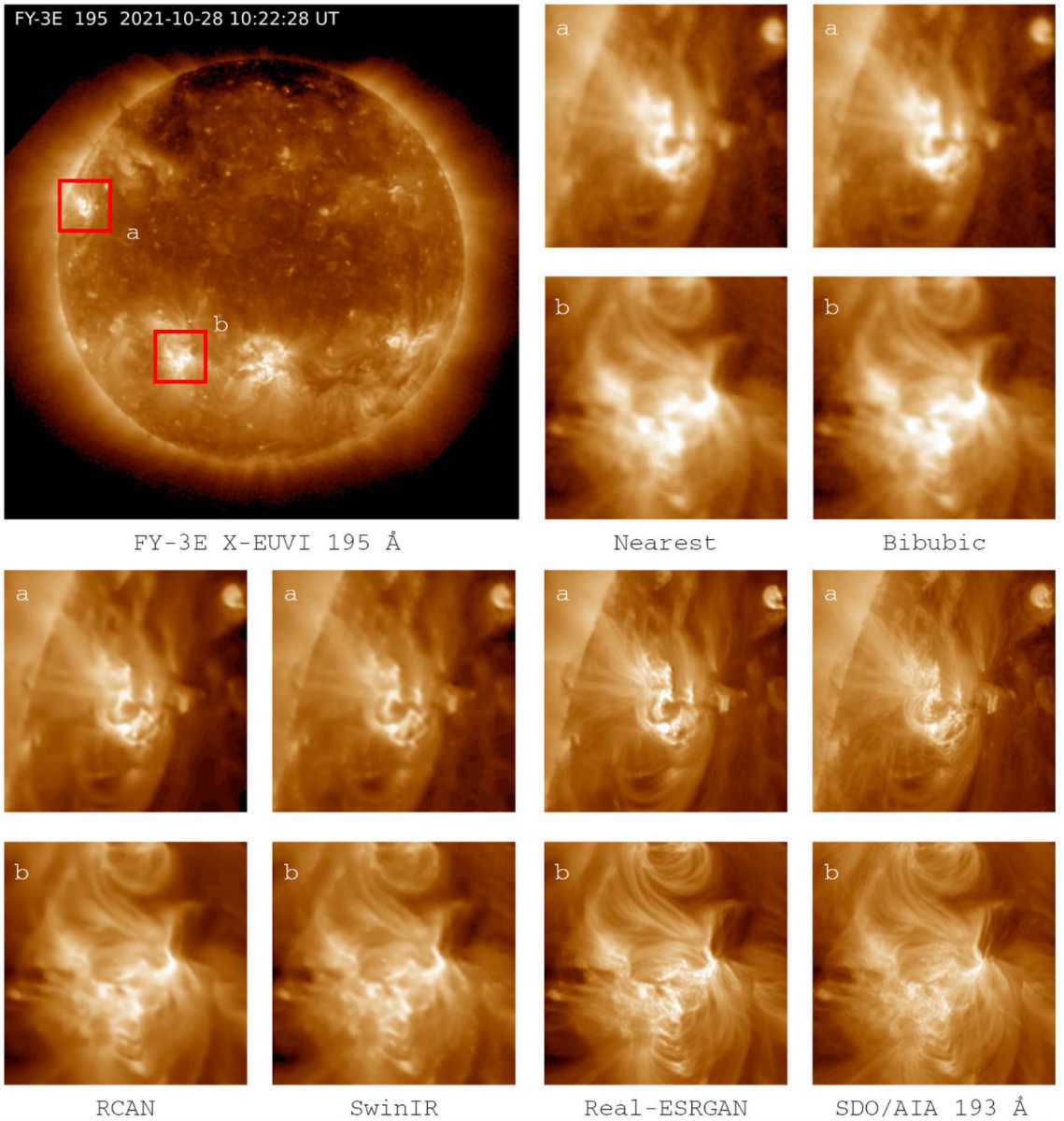






ENHANCING SOLAR IMAGES



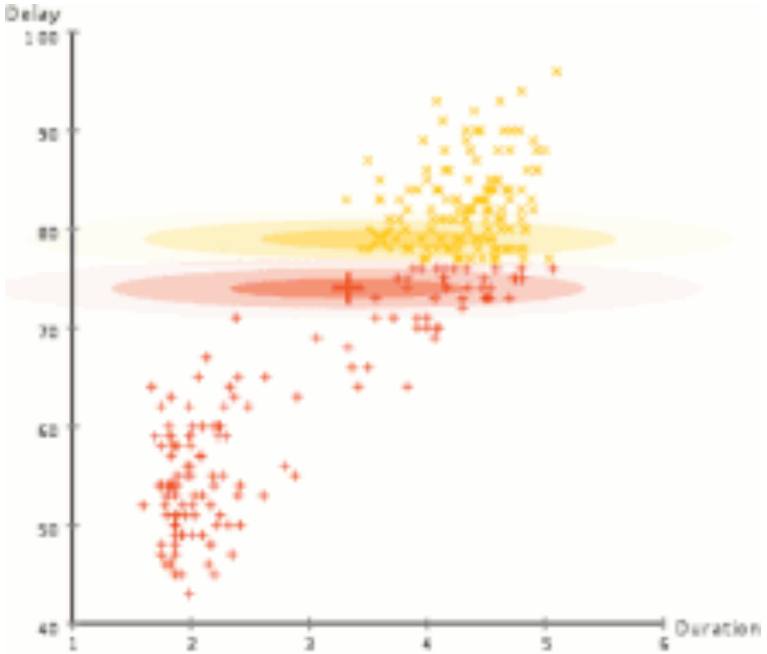


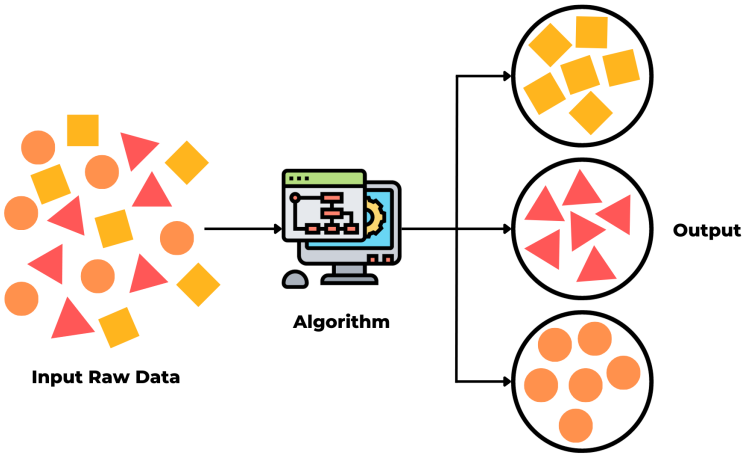
Prototypical Inversion Prompt in Midjourney Version 5.1

"An astronaut riding a horse"

"A horse riding an astronaut"

UNSUPERVISED LEARNING: CLUSTERING





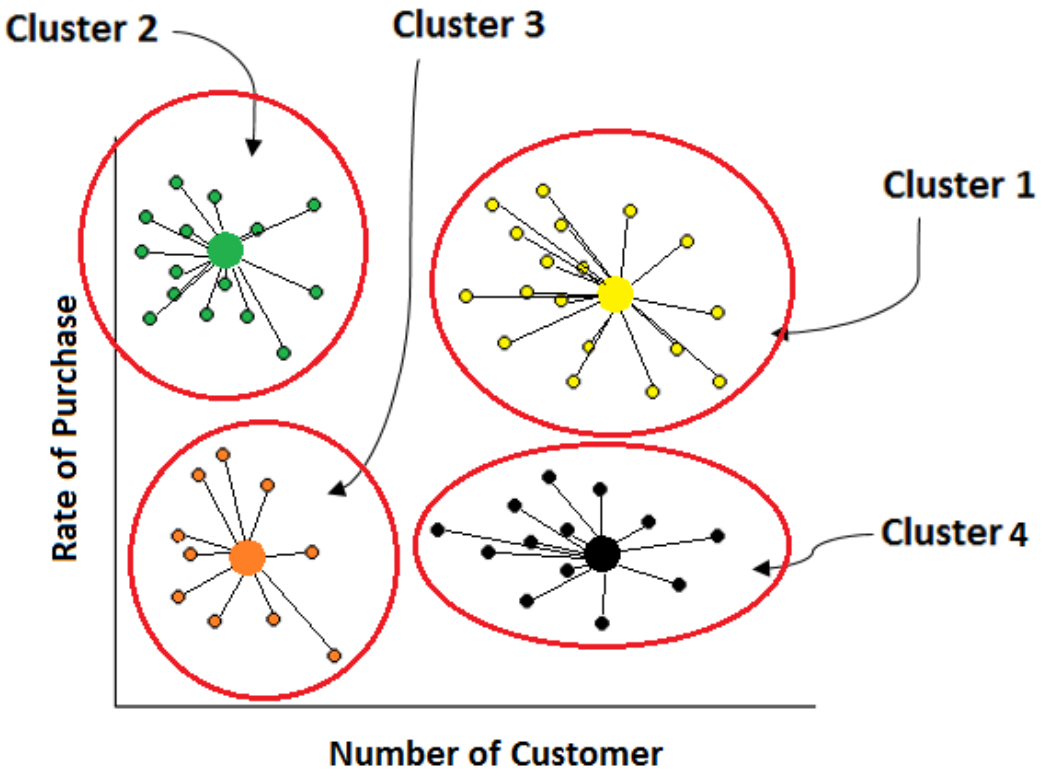
How clustering works

K-MEANS ALGORITHM

observations

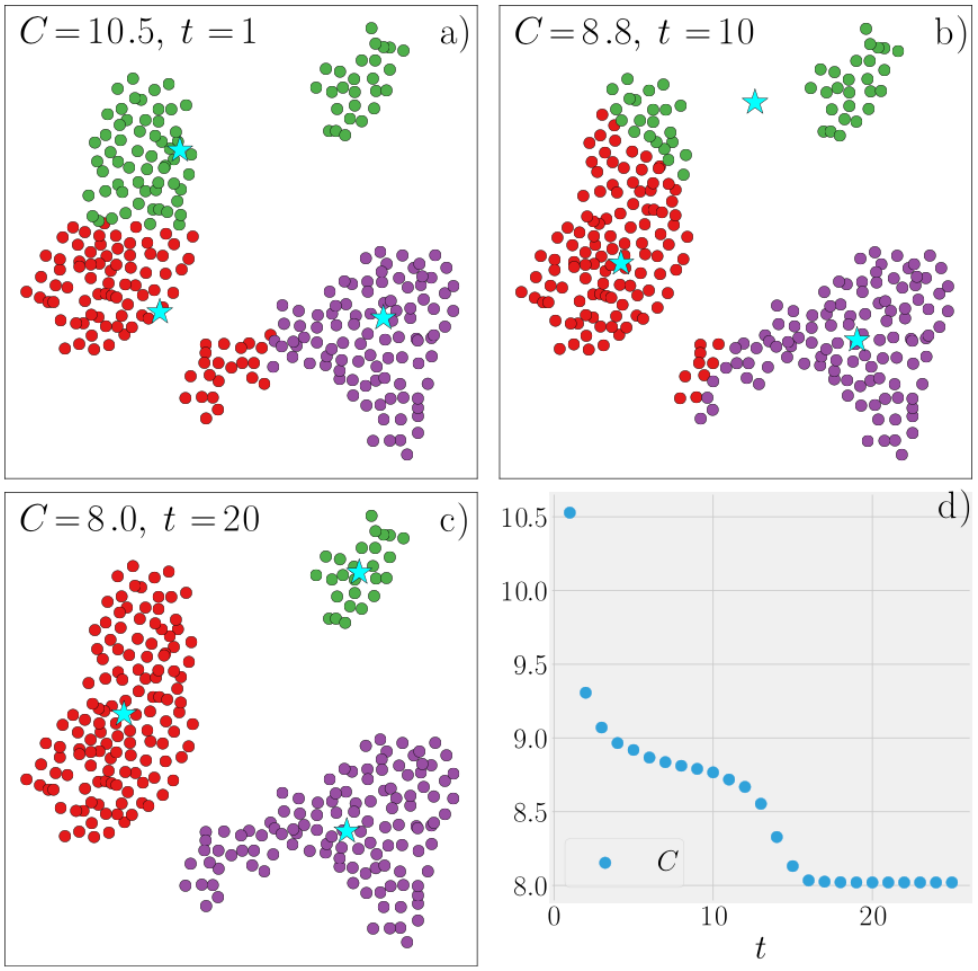
Cluster means: $\mu_k = \frac{1}{N_k}\sum_n^N r_{nk}\, x_n$

$$r_{nk} = \begin{cases} 1, & if\ k = \arg min_{k'}(x_n - \mu_{k'})^2 \\ 0, & otherwise \end{cases}$$



K-MEANS ALGORITHM

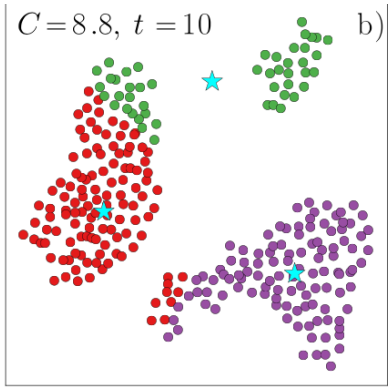
$$r_{nk} = \begin{cases} 1, & if\ k = \arg min_{k'}(x_n - \mu_{k'})^2 \\ 0, & otherwise \end{cases}$$

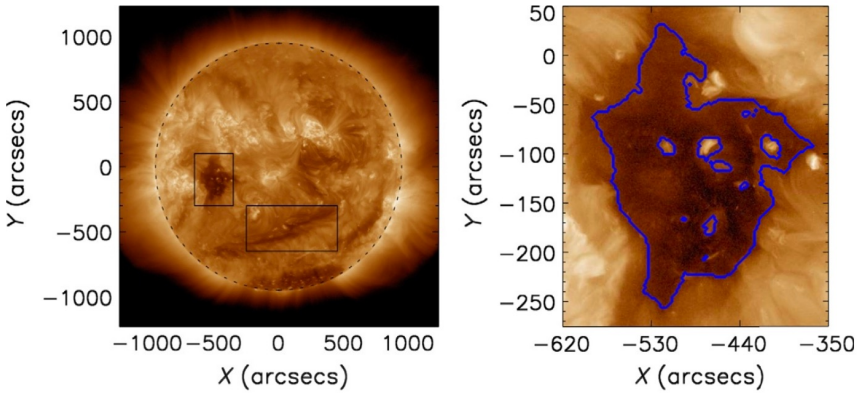


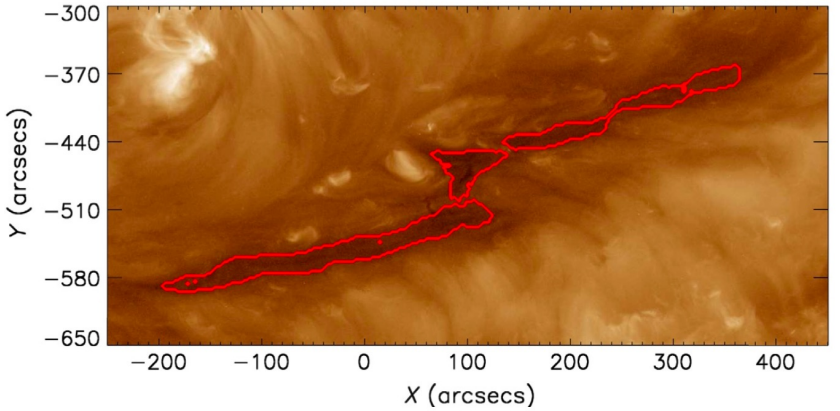
IDENTIFYING CORONAL HOLES

- Use k-means
- Minimize objective function:

$$C = \sum_{k=1}^{K} \sum_{n=1}^{N} r_{nk}(x_n - \mu_k)^2$$







PRACTICAL EXERCISE:

- Use k-means algorithm to identify active regions and/or other structures in solar images
  - https://github.com/f-carella/SoS-ML4-Sun