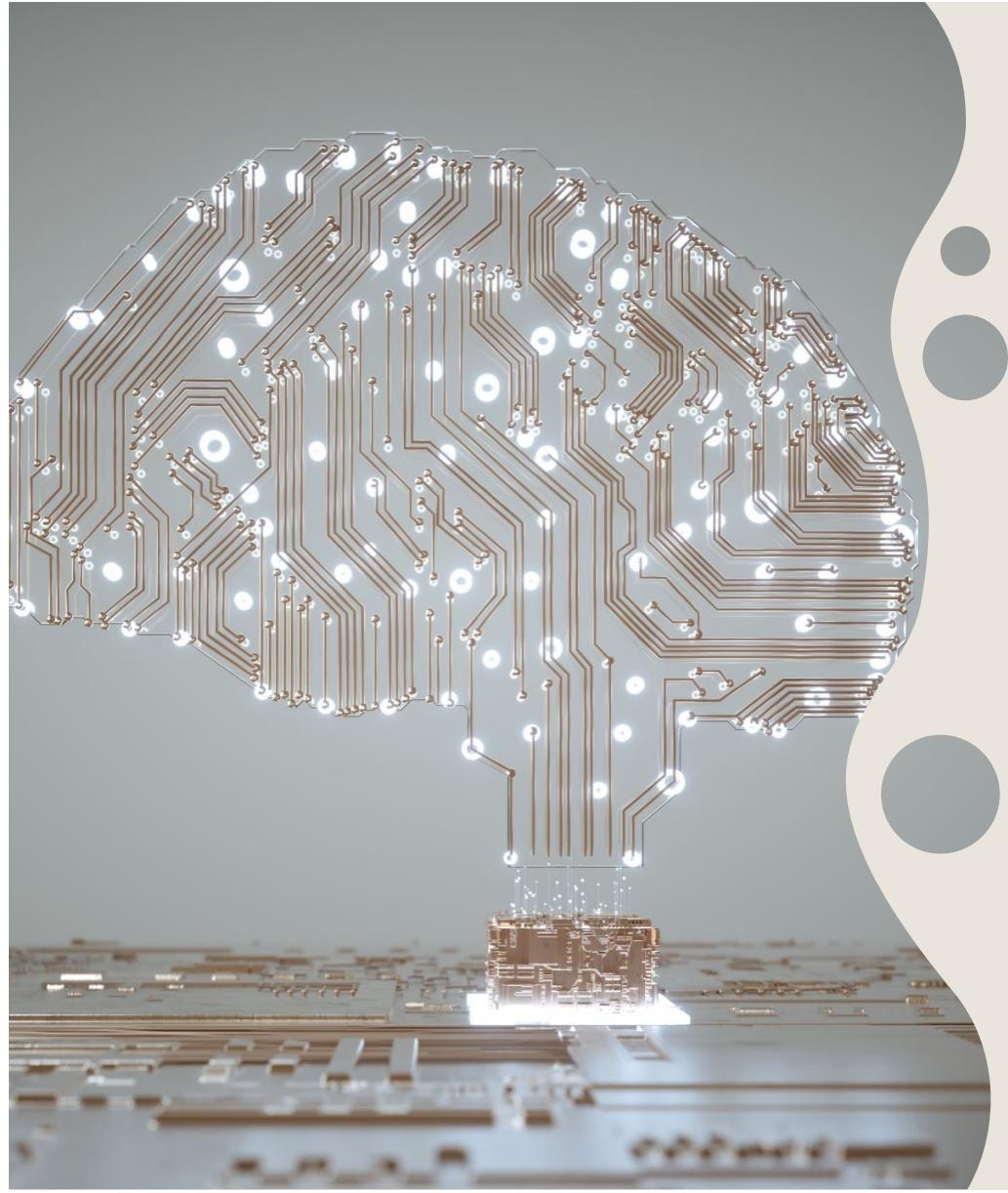


INTRODUCTION TO DATA SCIENCE IN SPACE PLASMAS



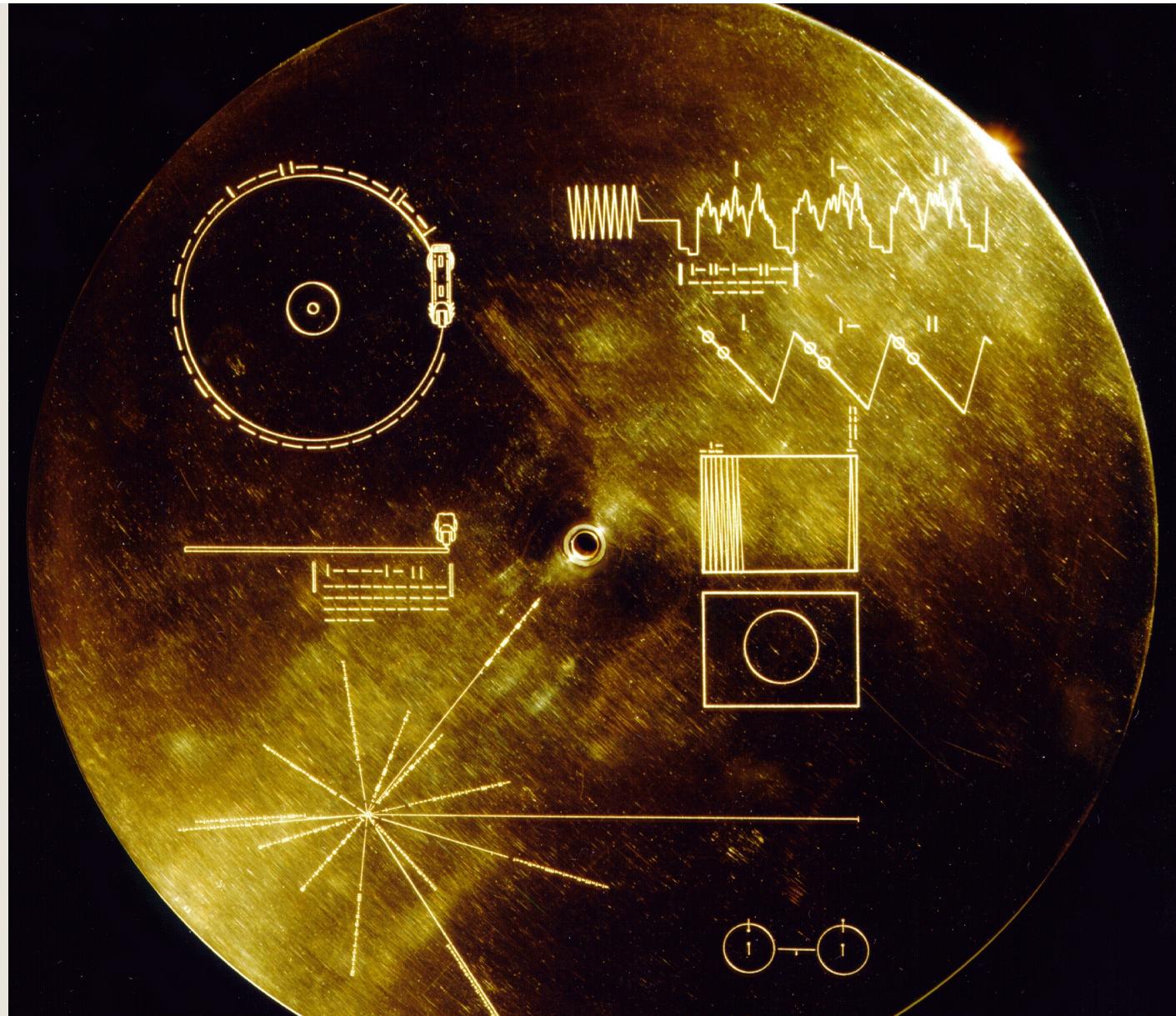


OUTLINE

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

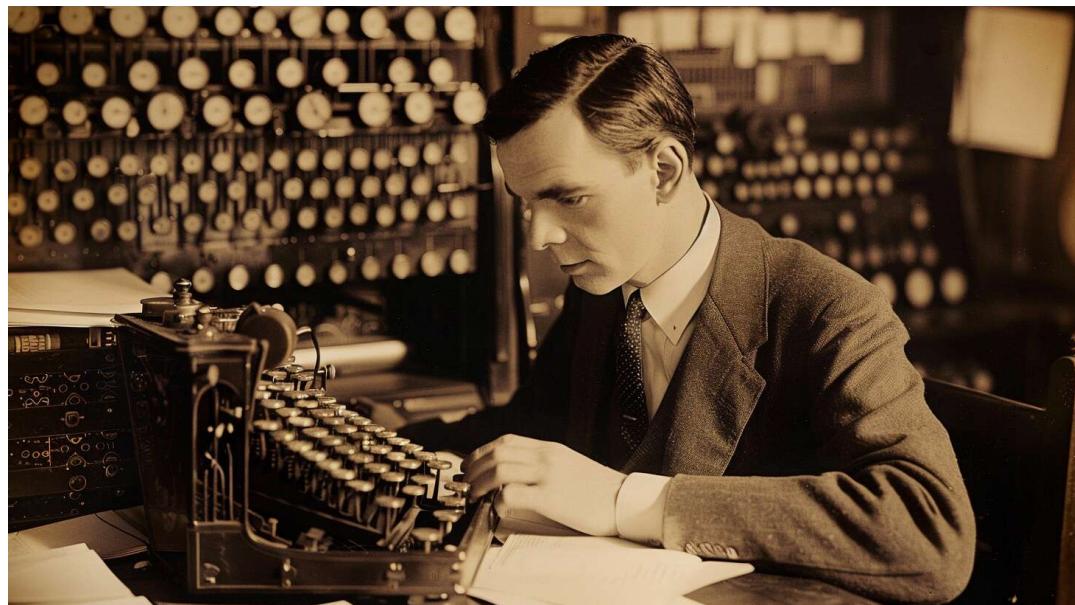
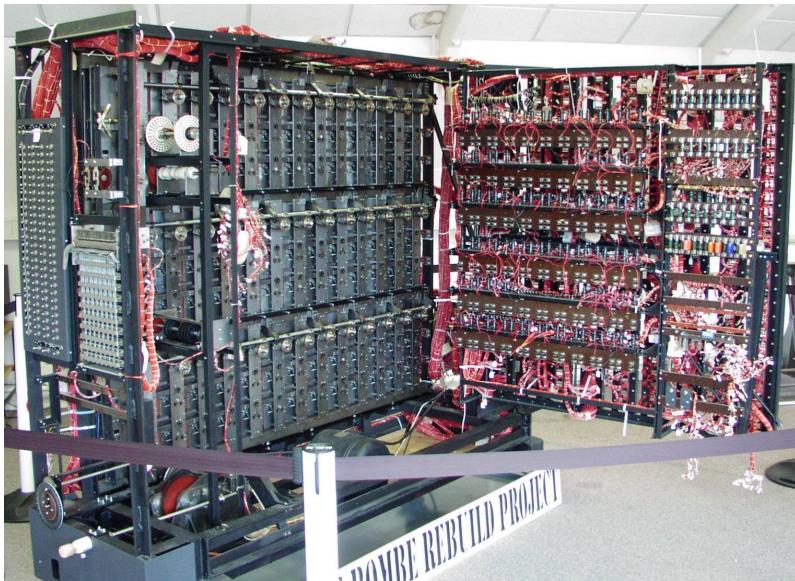


WHAT IS INFORMATION?



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

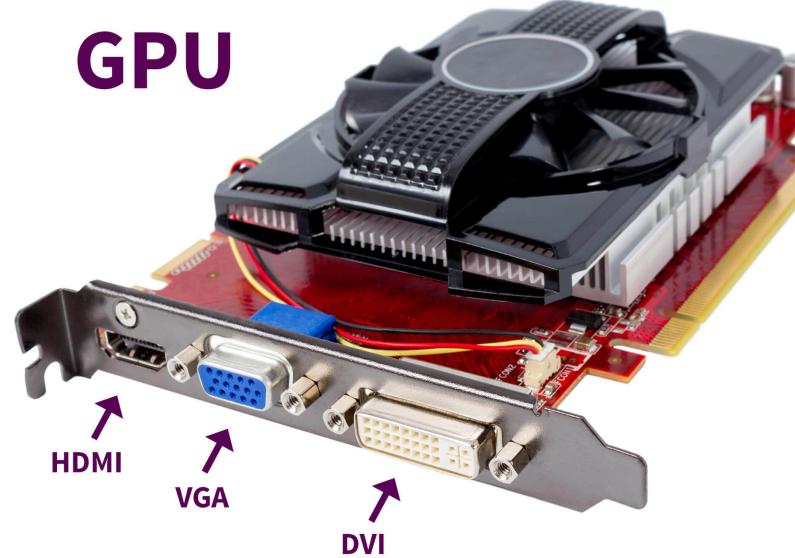
- Programming
- Human developed algorithms
- Challenge achieving human-level performance



EMERGENCE OF MACHINE LEARNING



Emergence of data sharing



© TechTerms.com

Availability of massively parallel hardware



ARTIFICIAL INTELLIGENCE

A program that can sense, reason, act, and adapt

MACHINE LEARNING

Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING

Subset of machine learning in which multilayered neural networks learn from vast amounts of data

MACHINE LEARNING VS ARTIFICIAL INTELLIGENCE



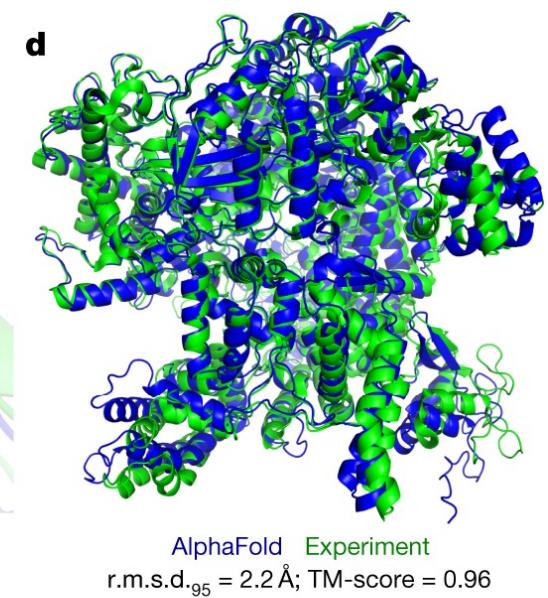
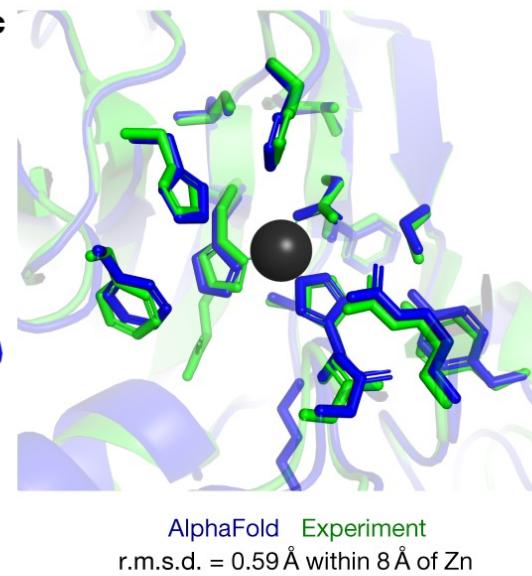
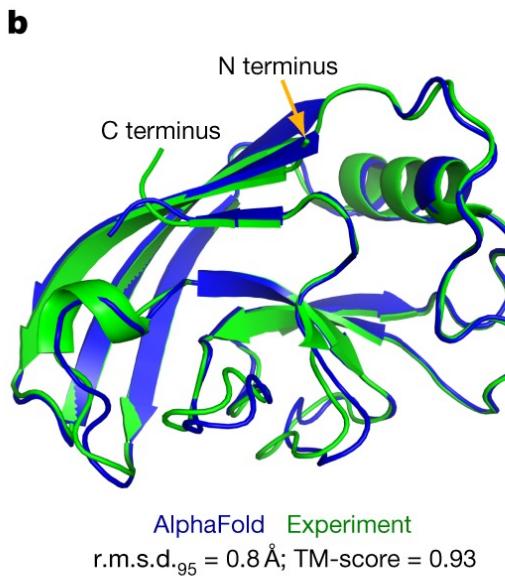
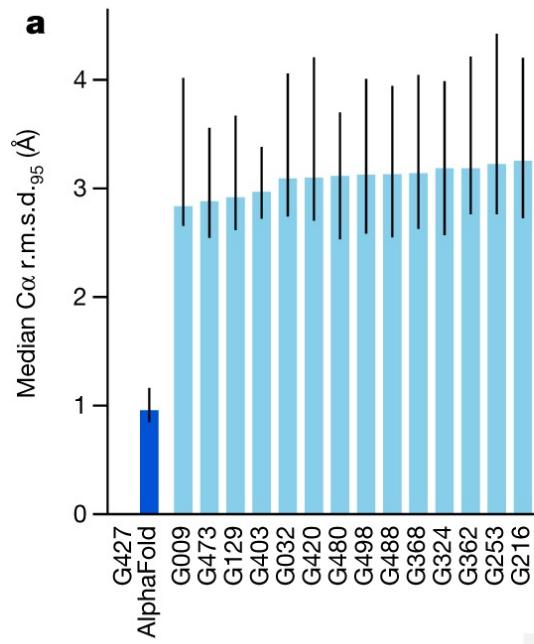
ALPHA GO: HUMAN CHAMPION BEATEN



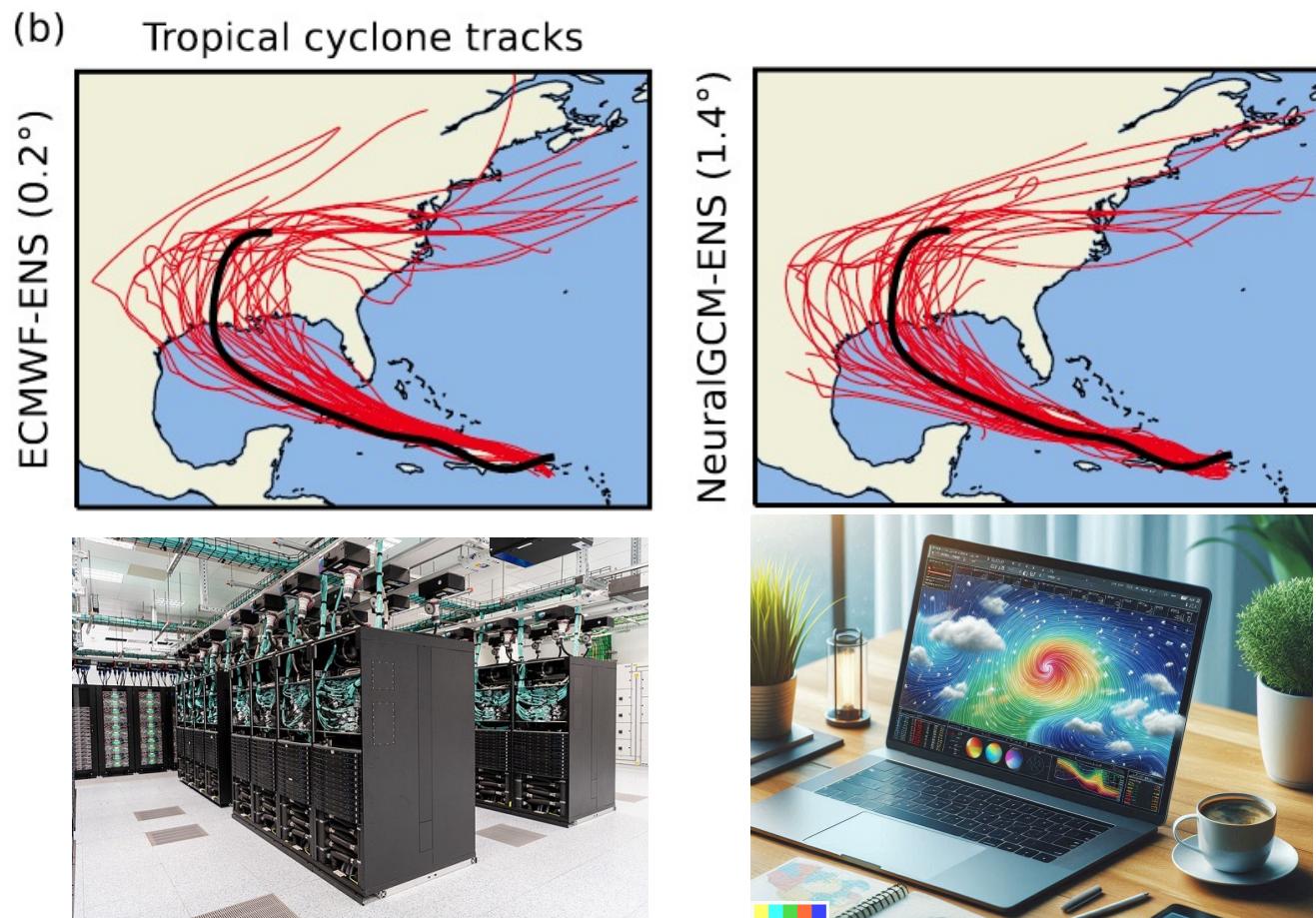
Lee Sedol facing AI opponent in Go



APPLICATION OF ML TO PROTEIN FOLDING



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

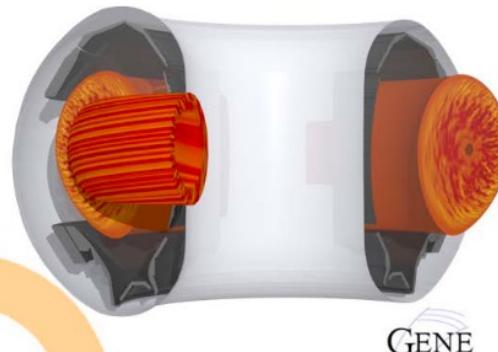
Neural general circulation models for weather and climate

[Dmitrii Kochkov](#) [Janni Yuval](#) [Ian Langmore](#), [Peter Norgaard](#), [Jamie Smith](#), [Griffin Mooers](#), [Milan](#)

1 second of plasma in...



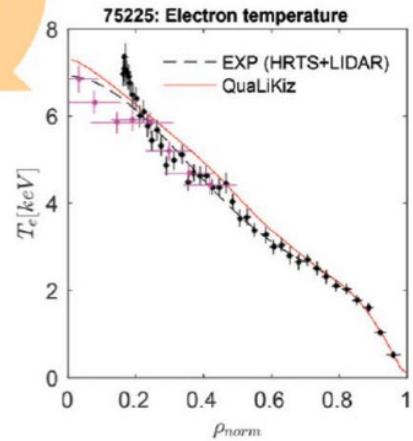
1 year on 1000 cores...



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



HOME

PROJECT VISION

WORKPACKAGES

IMPACT

PARTNERS

EV

AI ON-BOARD OF SPACE MISSIONS

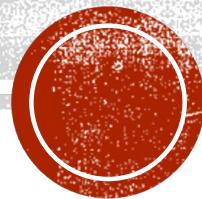
- ASAP – Automatics of SpAce exPloration
- ML algorithms for next generation space missions
- Automation of operations
- Predicting space weather



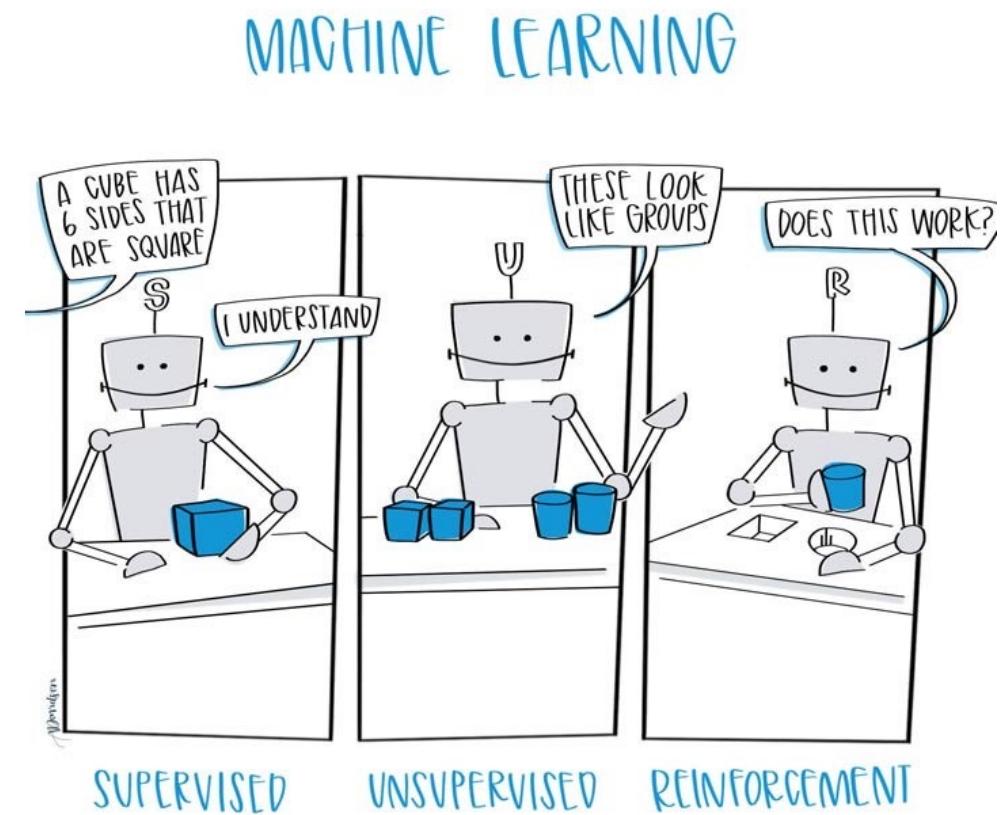
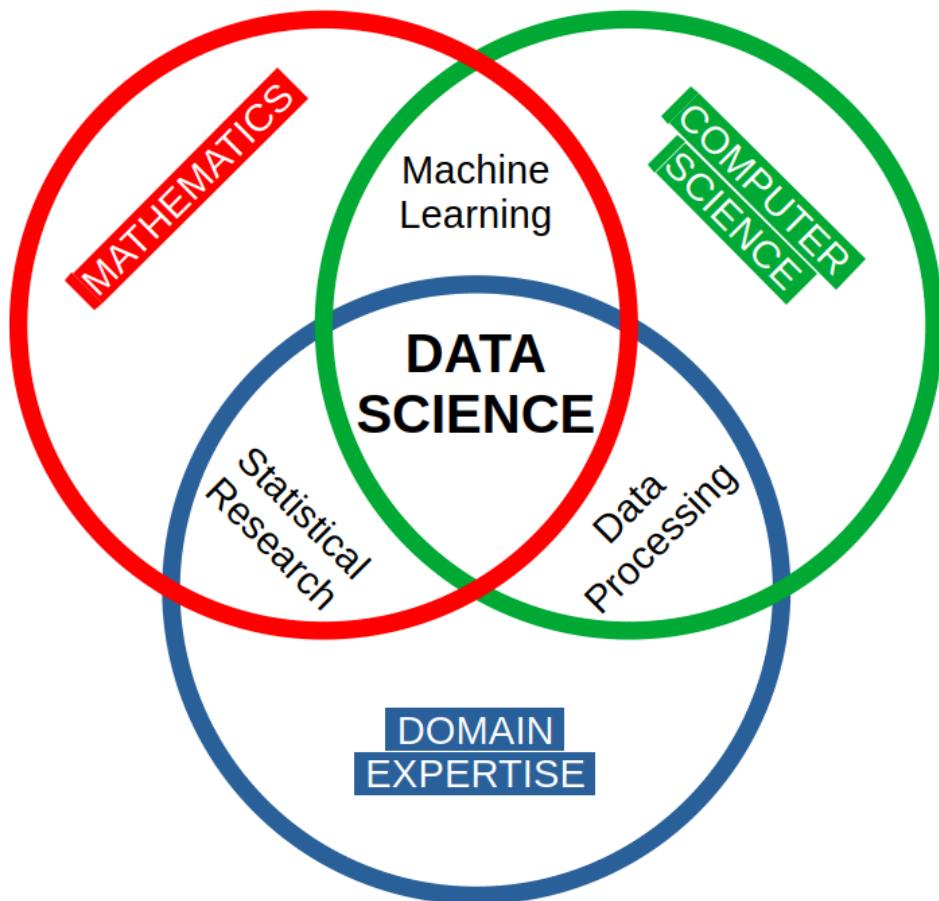


Input image

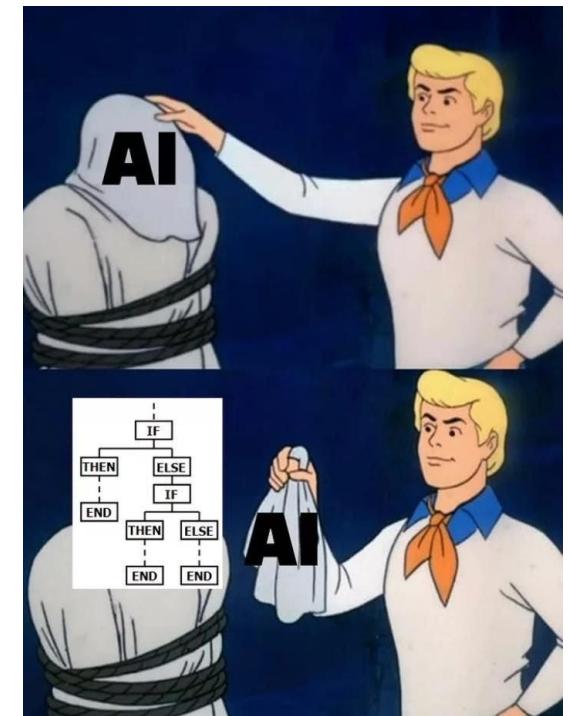
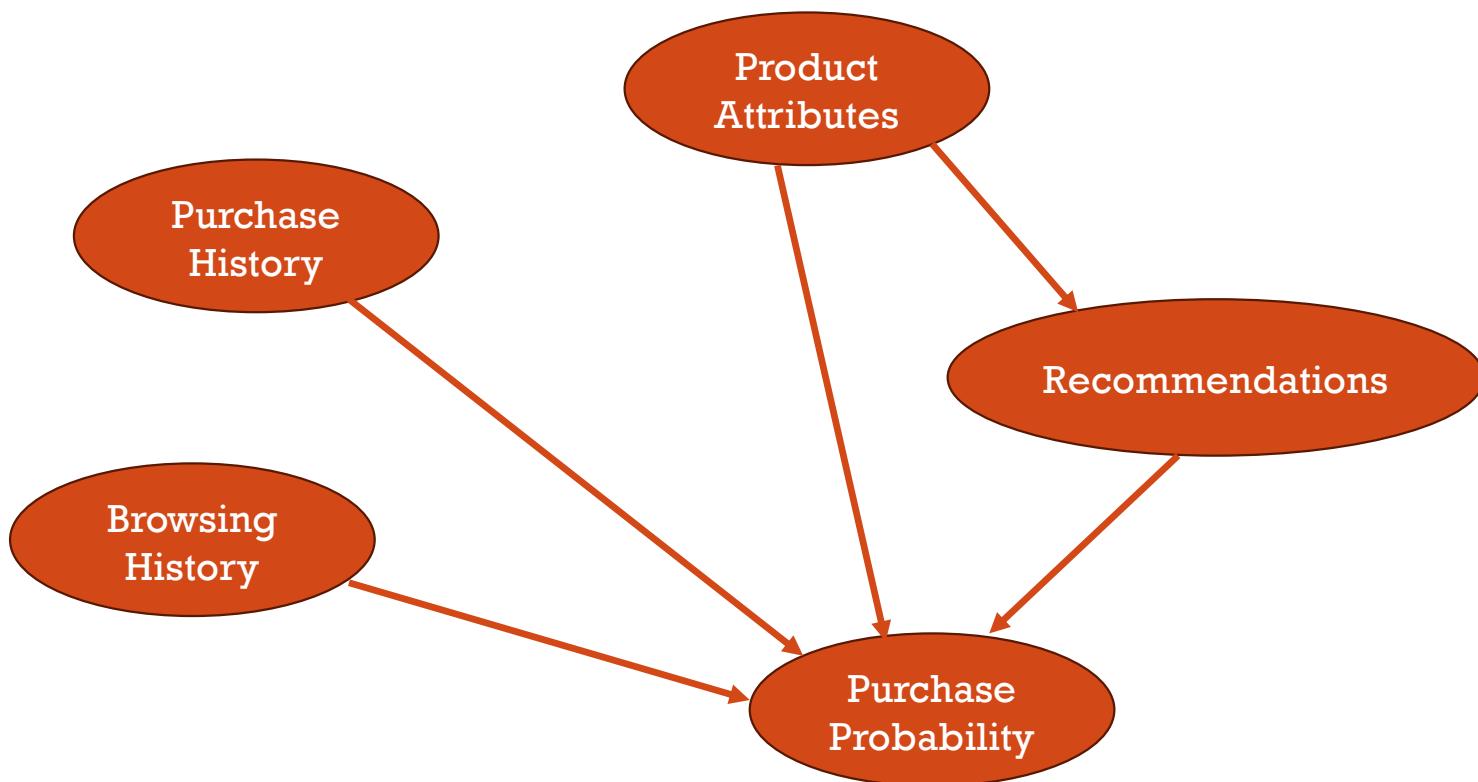
WHAT IS ARTIFICIAL INTELLIGENCE?



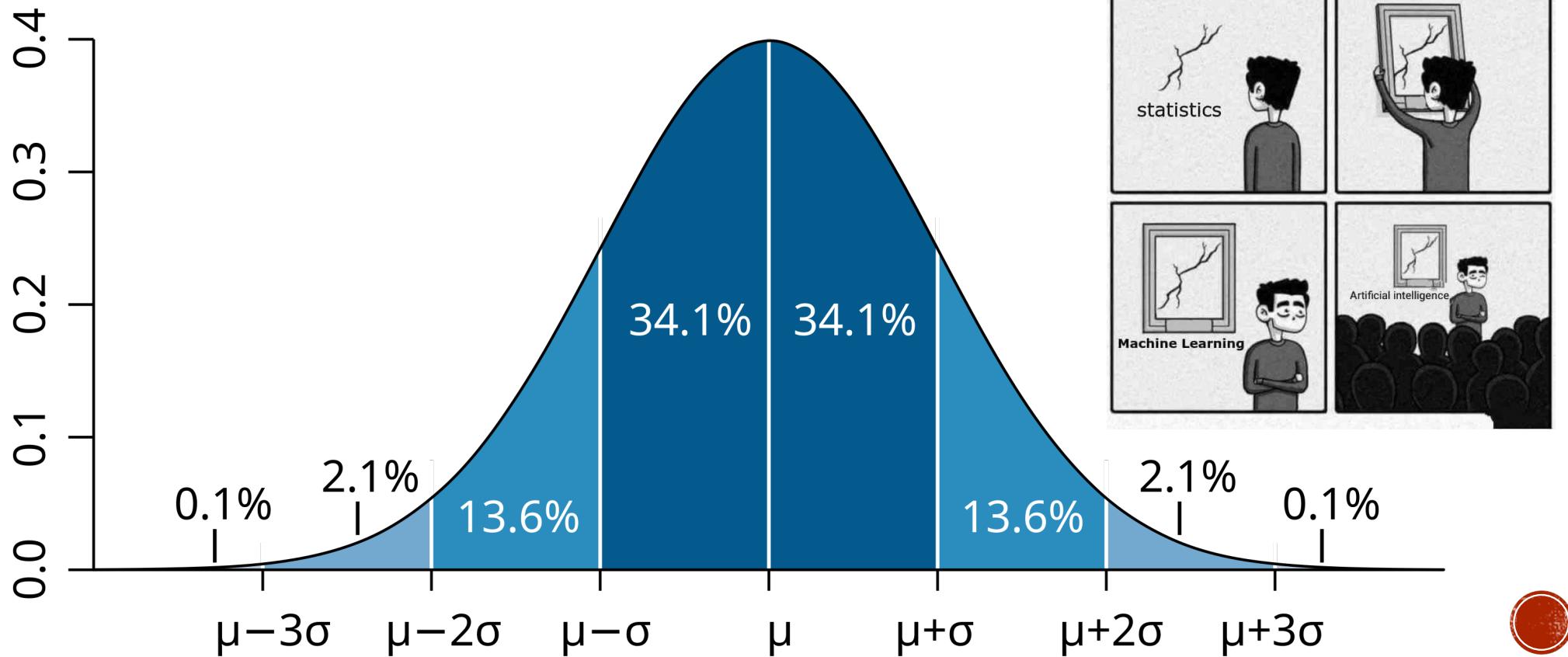
DIFFERENT TYPES OF MACHINE LEARNING



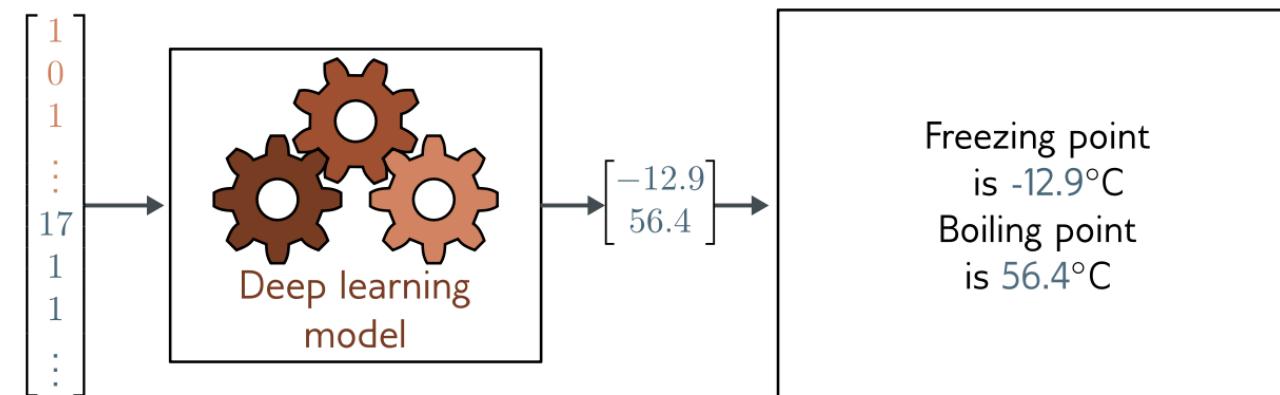
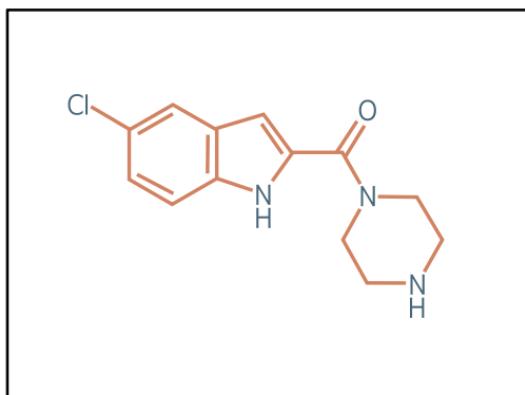
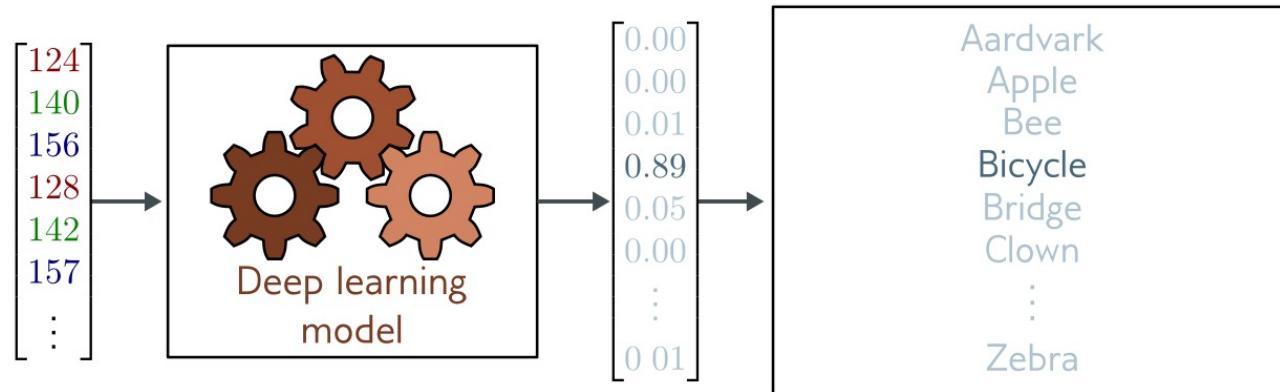
WHAT IS MACHINE LEARNING AFTER ALL?



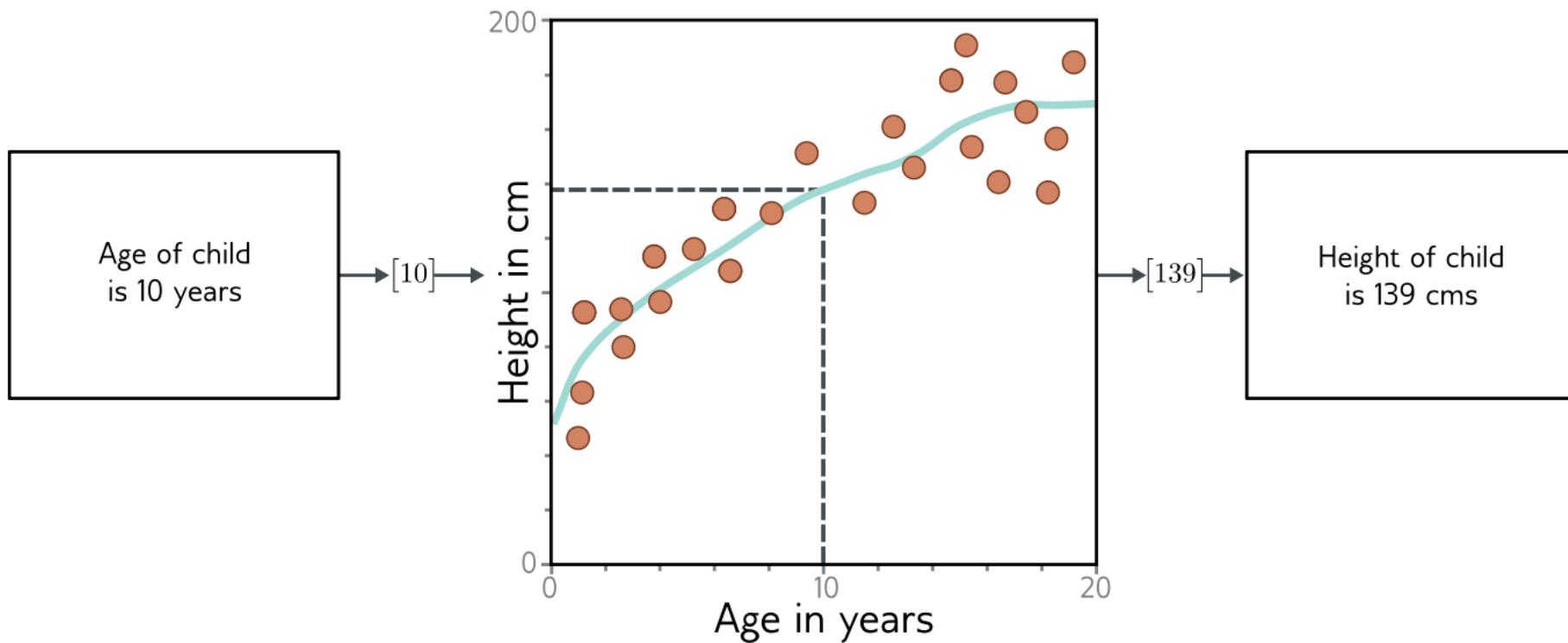
MACHINE LEARNING AND STATISTICS



SUPERVISED LEARNING

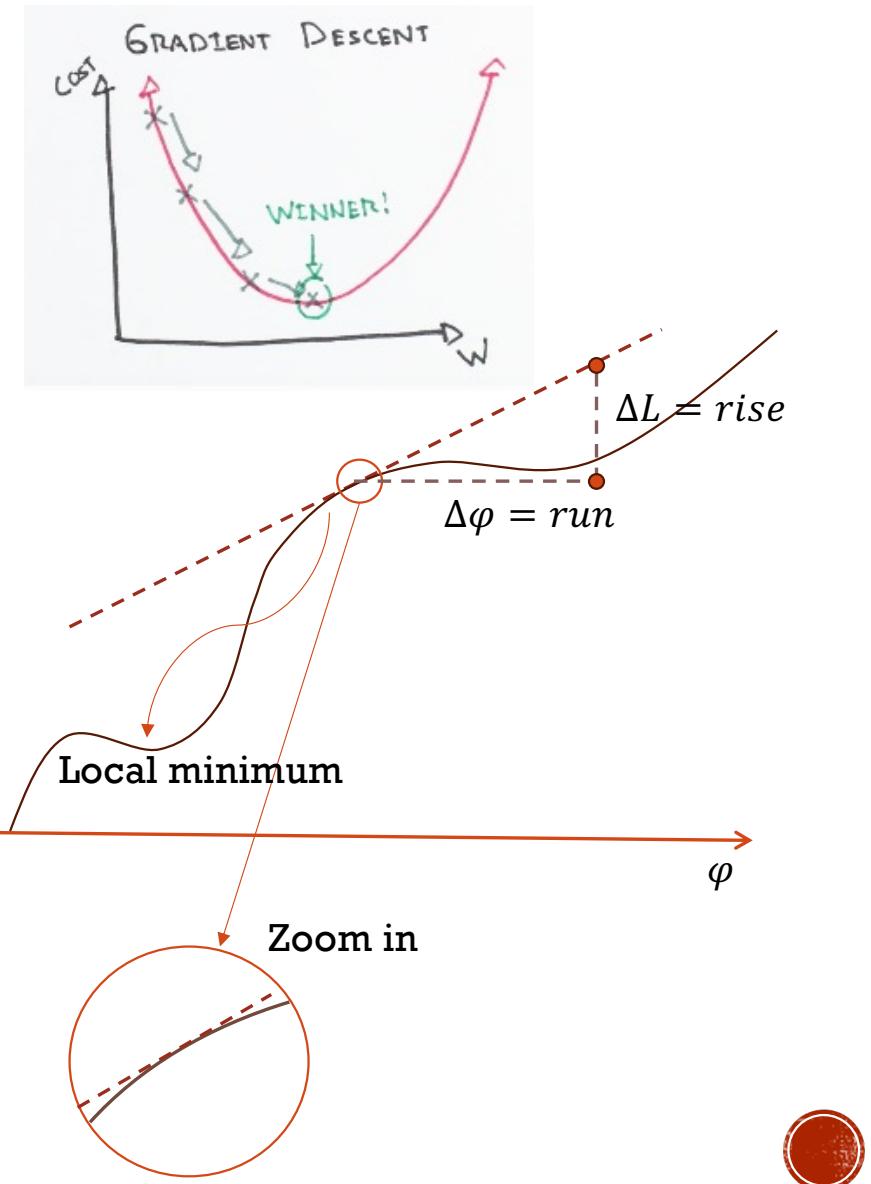


REGRESSION



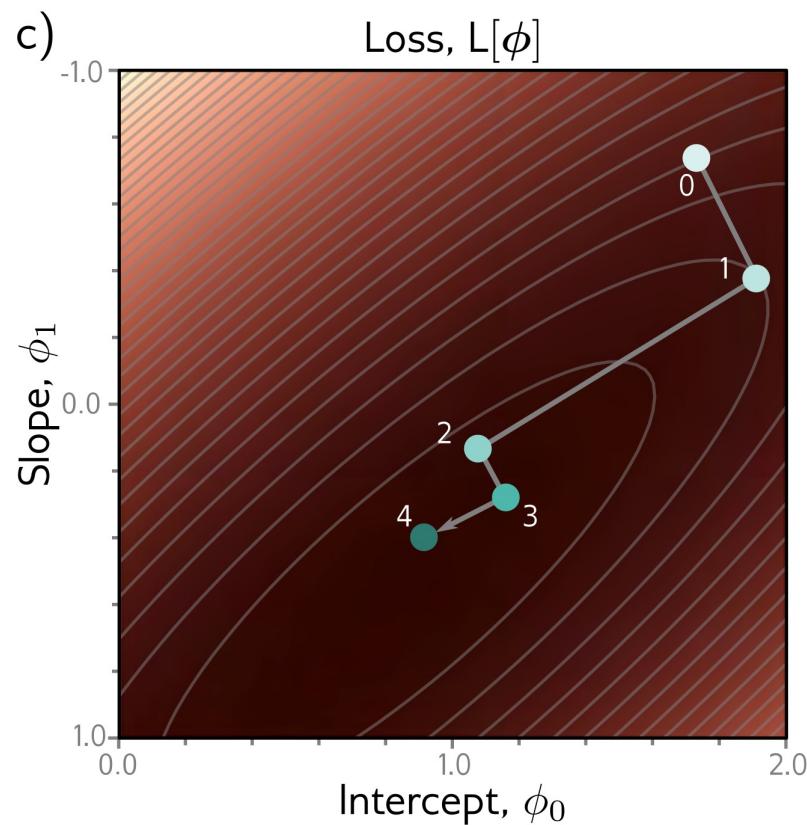
OPTIMIZATION PROBLEMS

- How do we find a minimum?
 - What is a derivative?
 - How do I choose optimal step?
-
- $\frac{\partial L}{\partial \phi} \rightarrow \frac{L(\phi + \Delta\phi) - L(\phi)}{\Delta\phi}$



GRADIENT DESCENT

$$\hat{\phi} = \operatorname{argmin}_{\phi} [L[\phi]]$$

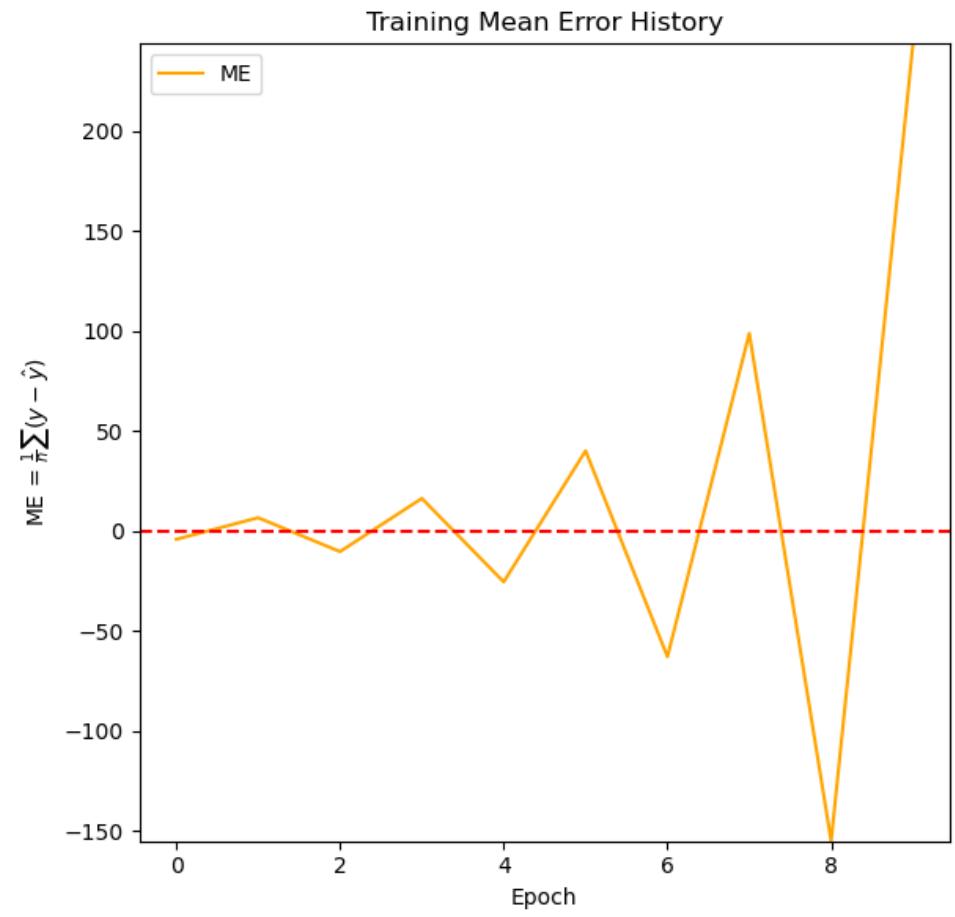
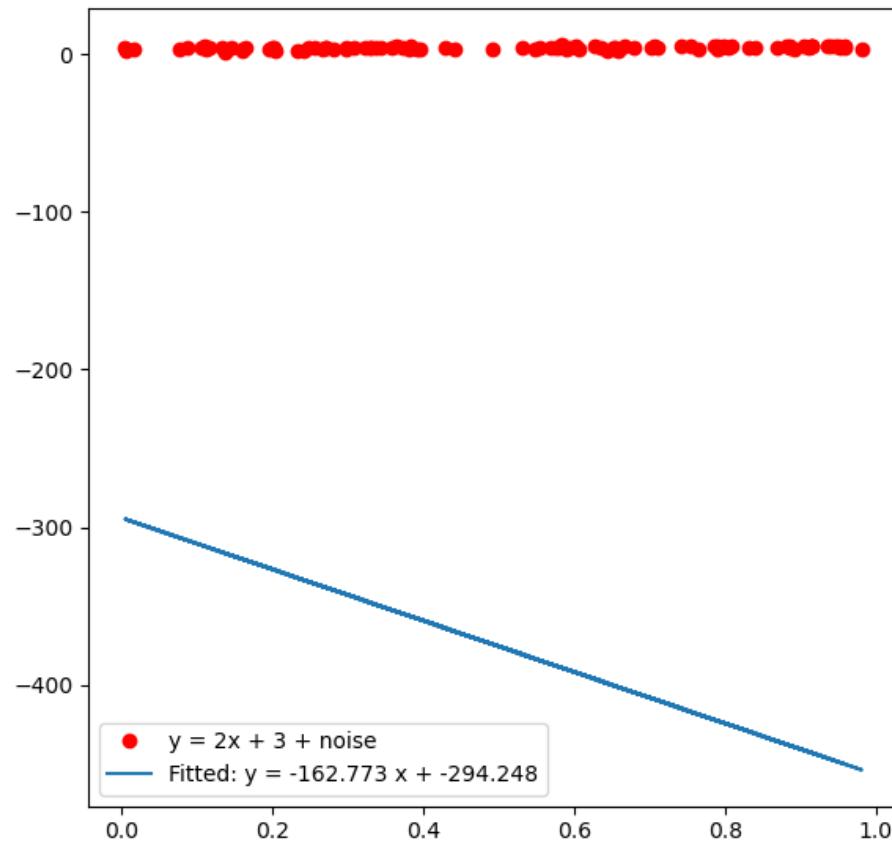


Step size (learning rate):

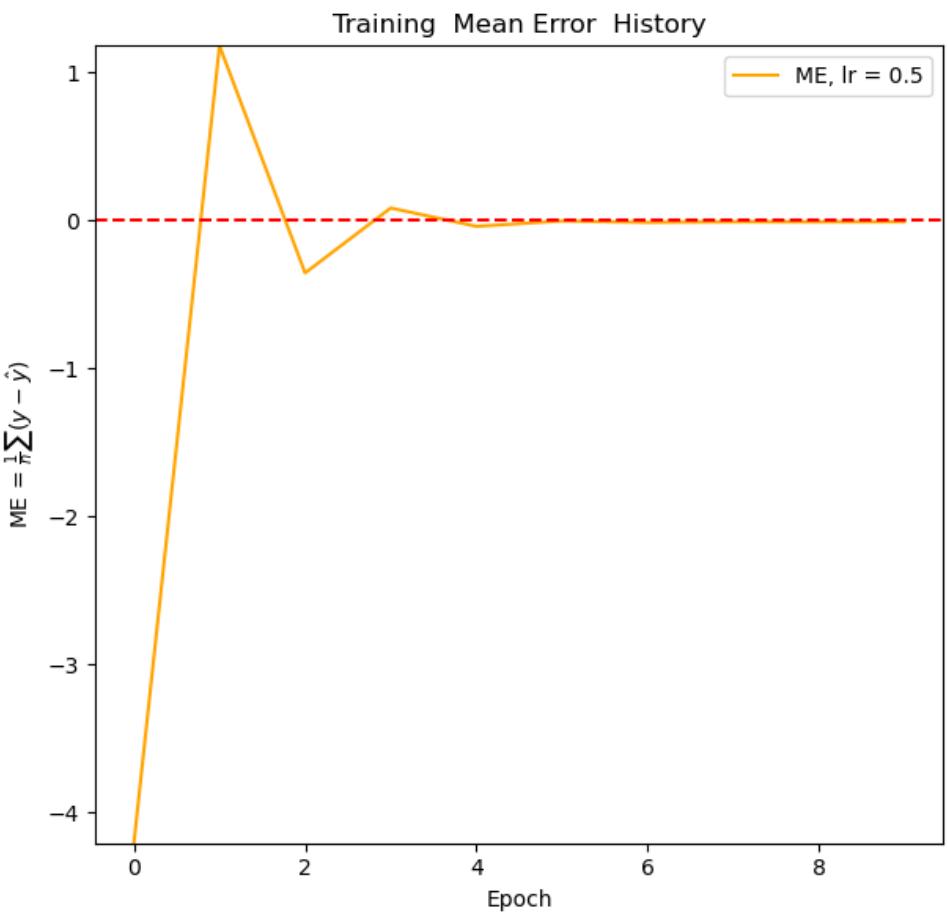
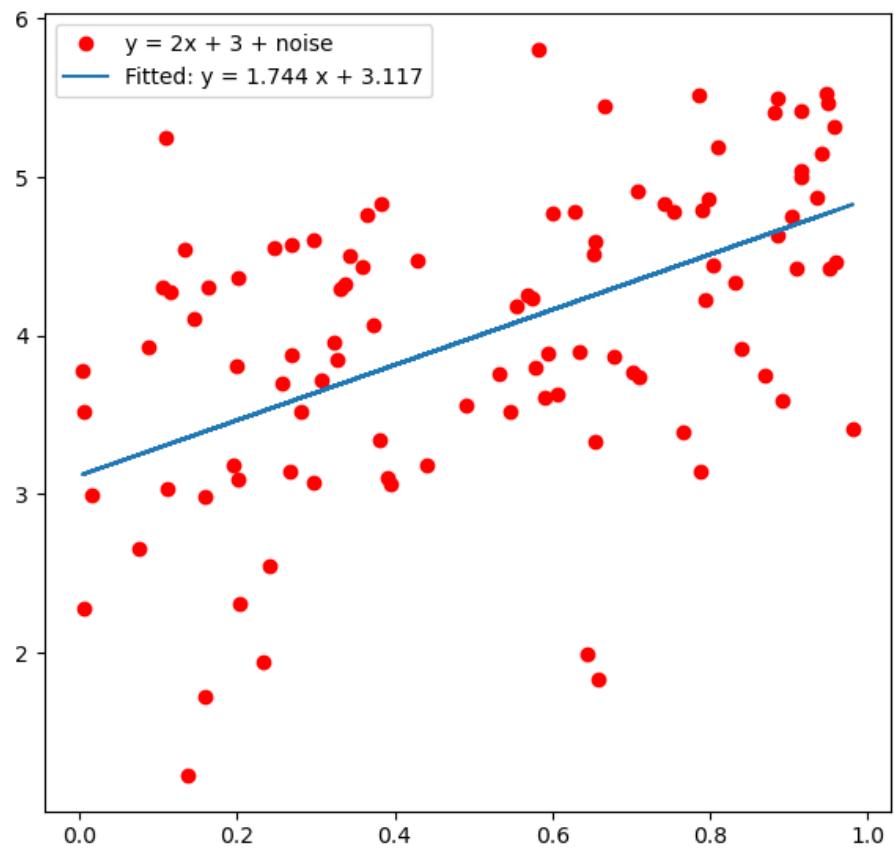
$$\phi \leftarrow \phi - \alpha \cdot \frac{\partial L}{\partial \phi}$$

$$\begin{aligned} L[\phi] &= \sum_{i=1}^I \ell_i &= \sum_{i=1}^I (f[x_i, \phi] - y_i)^2 \\ &= \sum_{i=1}^I (\phi_0 + \phi_1 x_i - y_i)^2 \end{aligned}$$

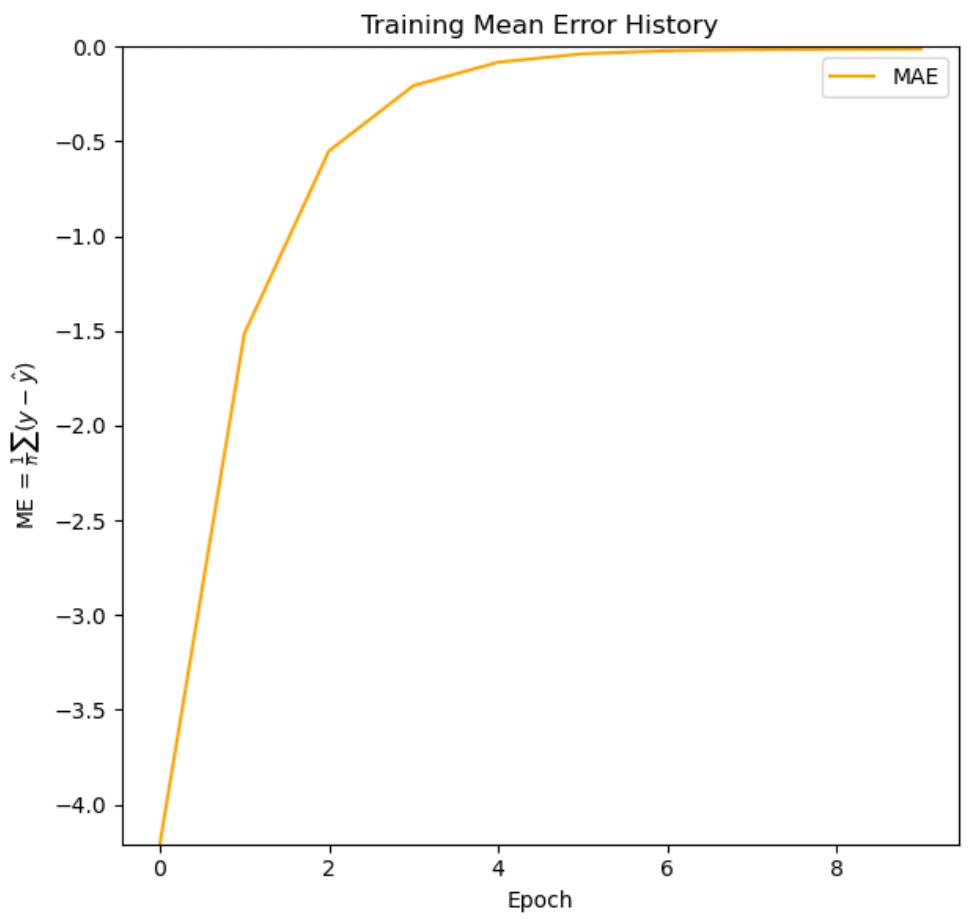
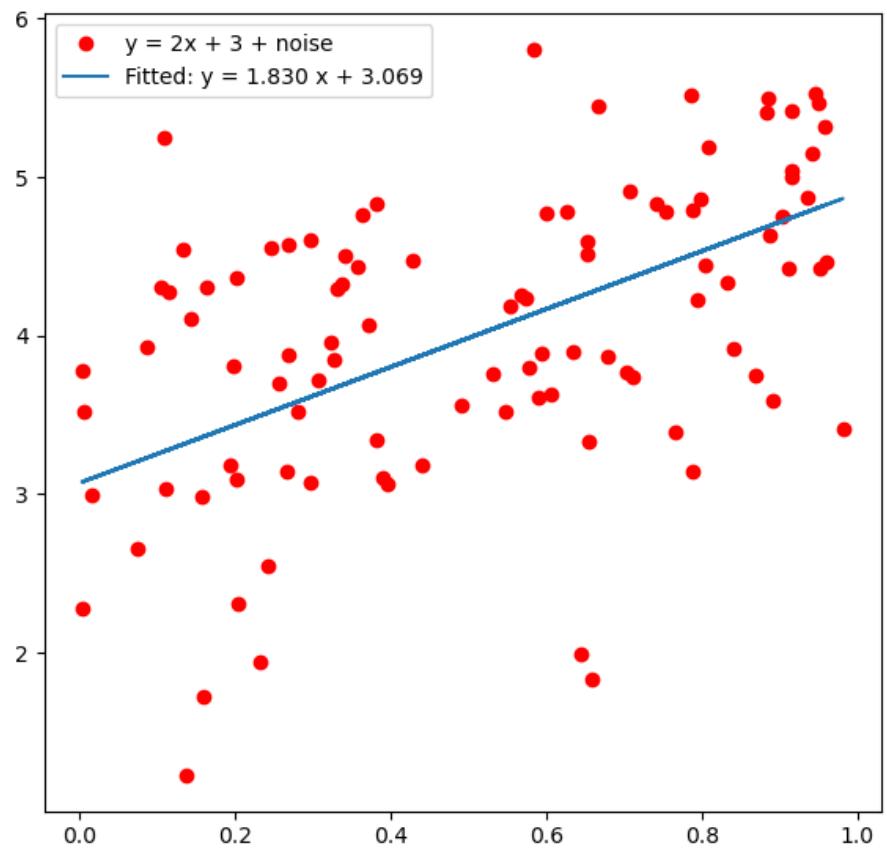
REGRESSION: $Y = 2X + 3 + \text{RANDOM}$



REGRESSION: $Y = 2X + 3 + \text{RANDOM}$



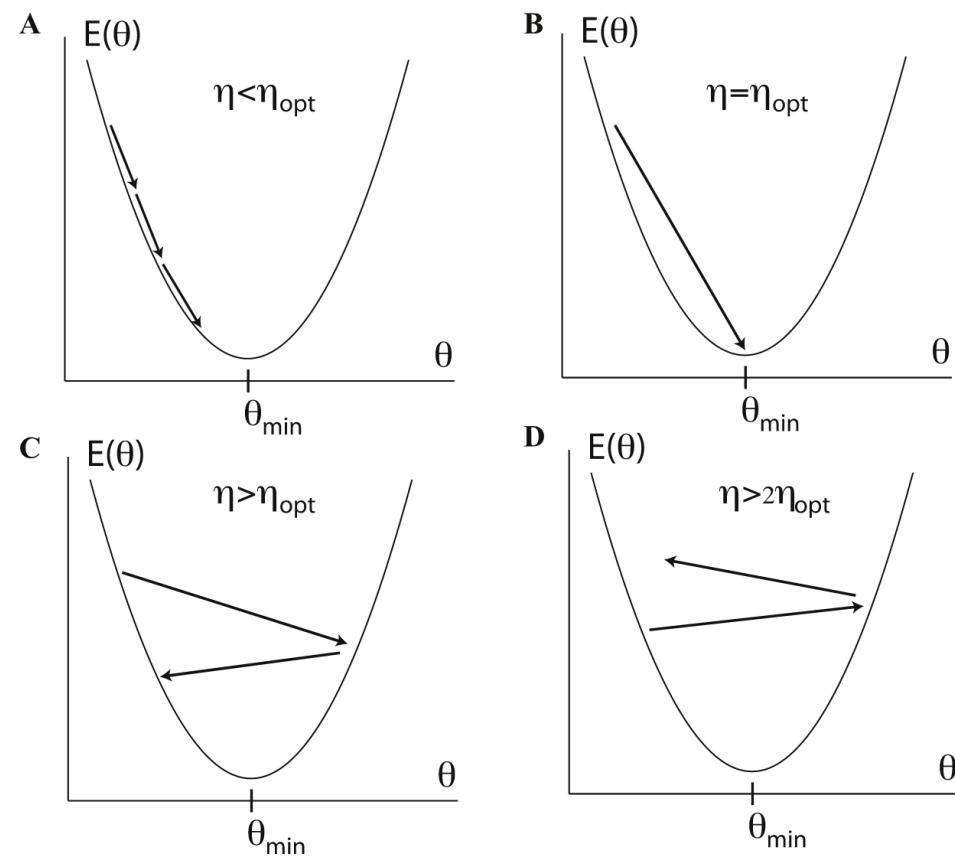
REGRESSION: $Y = 2X + 3 + \text{RANDOM}$



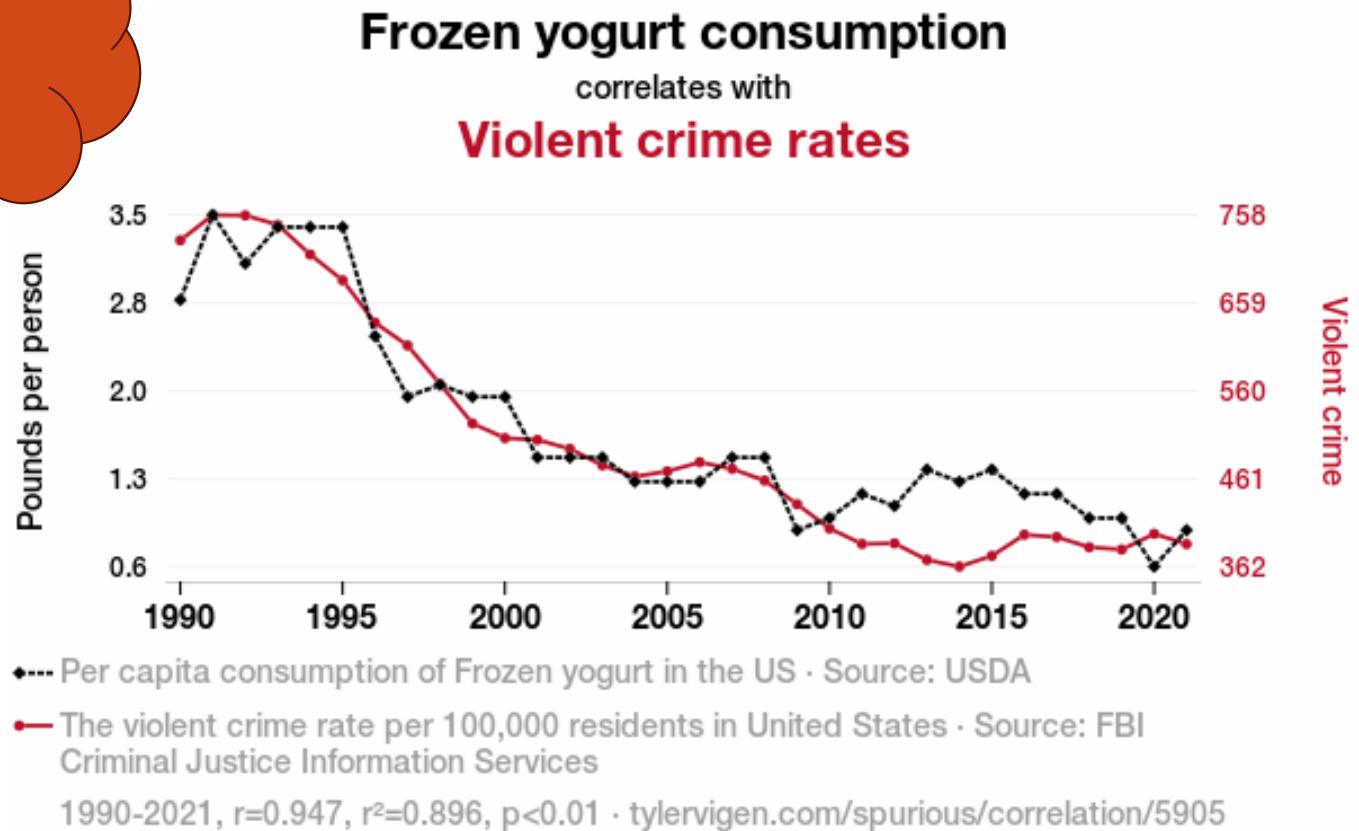
OPTIMIZATION PROBLEMS

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

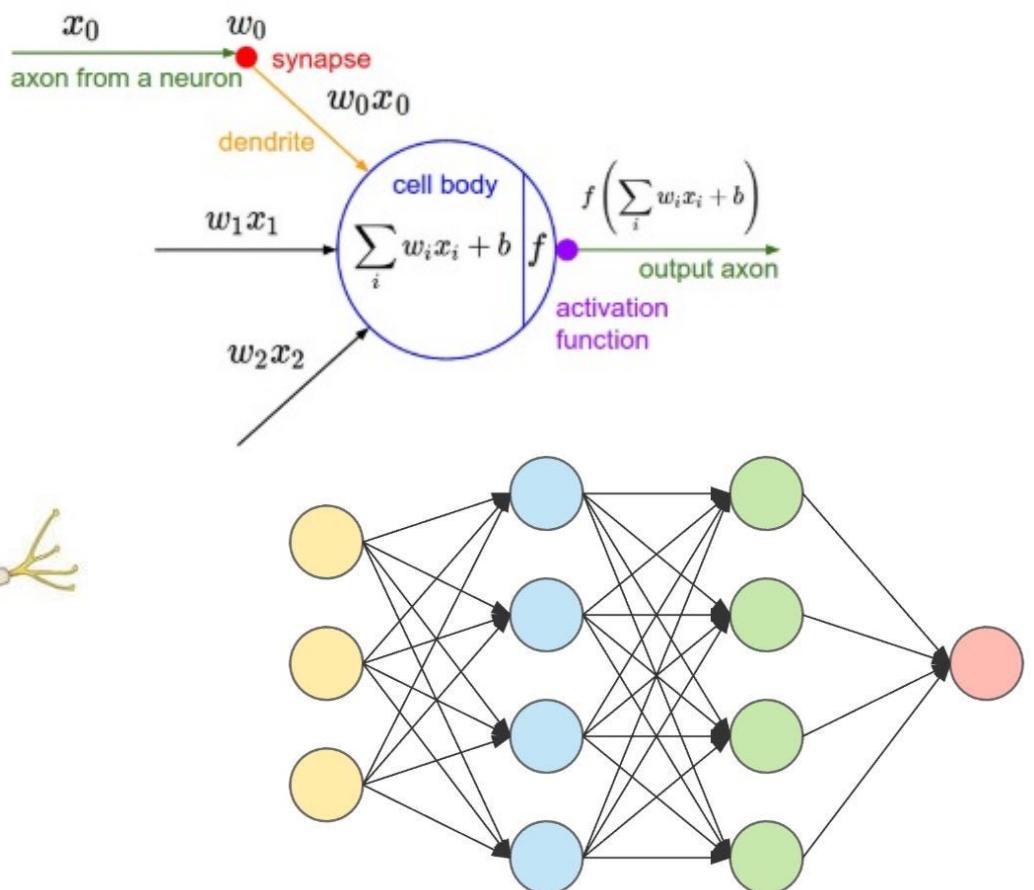
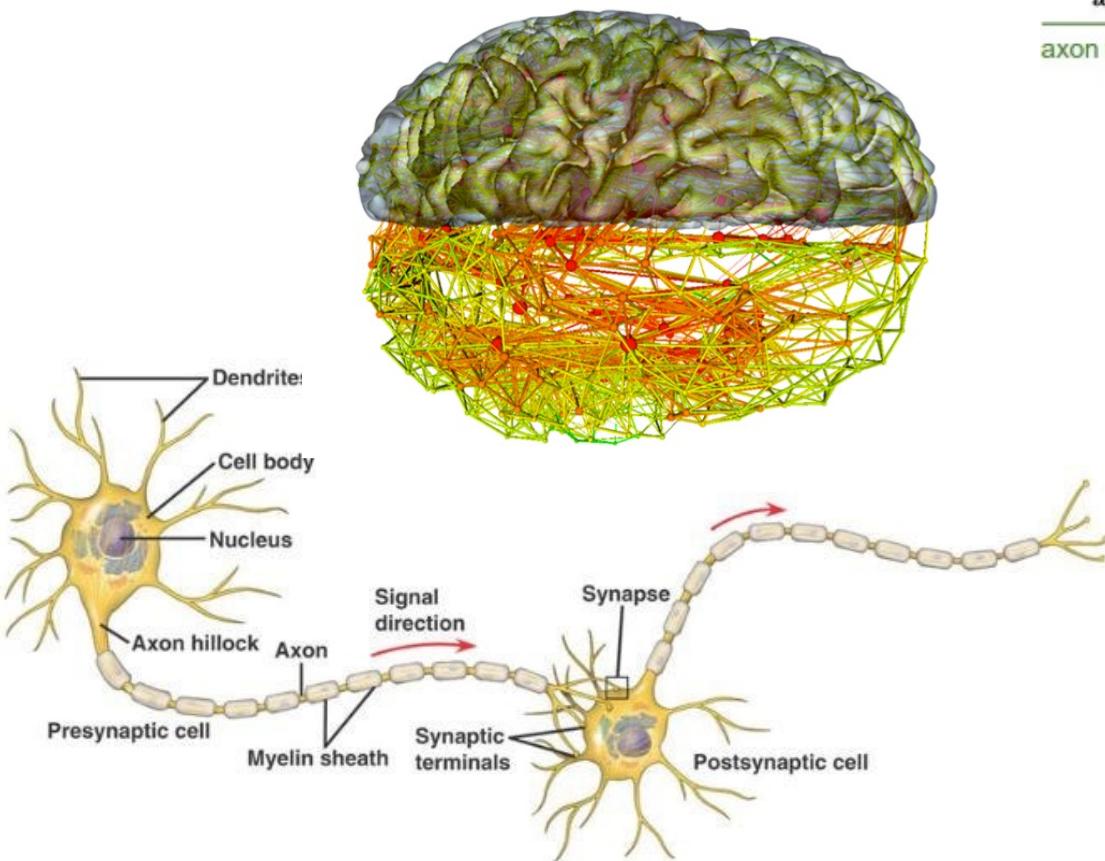
$$\blacksquare \frac{\partial L}{\partial \phi} \rightarrow \frac{L(\phi + \Delta\phi) - L(\phi)}{\Delta\phi}$$



FUNNY CORRELATIONS



REAL VS ARTIFICIAL NEURAL NETWORKS



HOW NEURAL NETWORKS WORK

Epoch 000,054 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: 50%
Noise: 0
Batch size: 10
REGENERATE

FEATURES Which properties do you want to feed in?
 X_1 
 X_2 
 X_1^2 
 X_2^2 
 $X_1 X_2$ 
 $\sin(X_1)$ 
 $\sin(X_2)$ 

+ - 1 HIDDEN LAYER
1 neuron
Click anywhere to edit.
Weight is 0.36.
This is the output from one neuron. Hover to see it larger.

OUTPUT Test loss 0.486
Training loss 0.466

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Colors shows data, neuron and weight values. 
 Show test data Discretize output

- 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?

- X_1
- X_2
- X_1^2
- X_2^2
- $X_1 X_2$
- $\sin(X_1)$
- $\sin(X_2)$

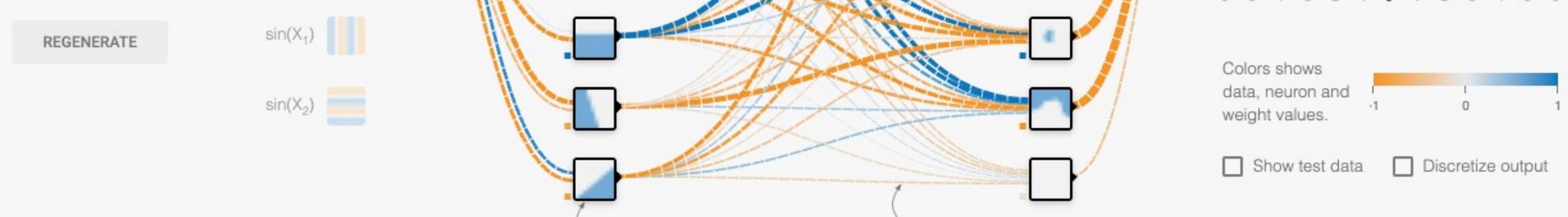
+ - 2 HIDDEN LAYERS

+ -

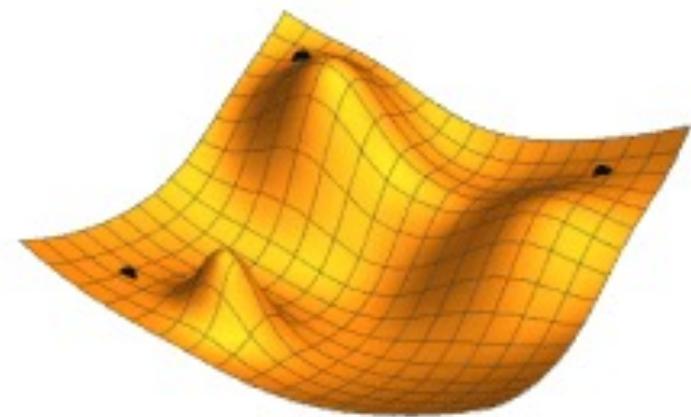
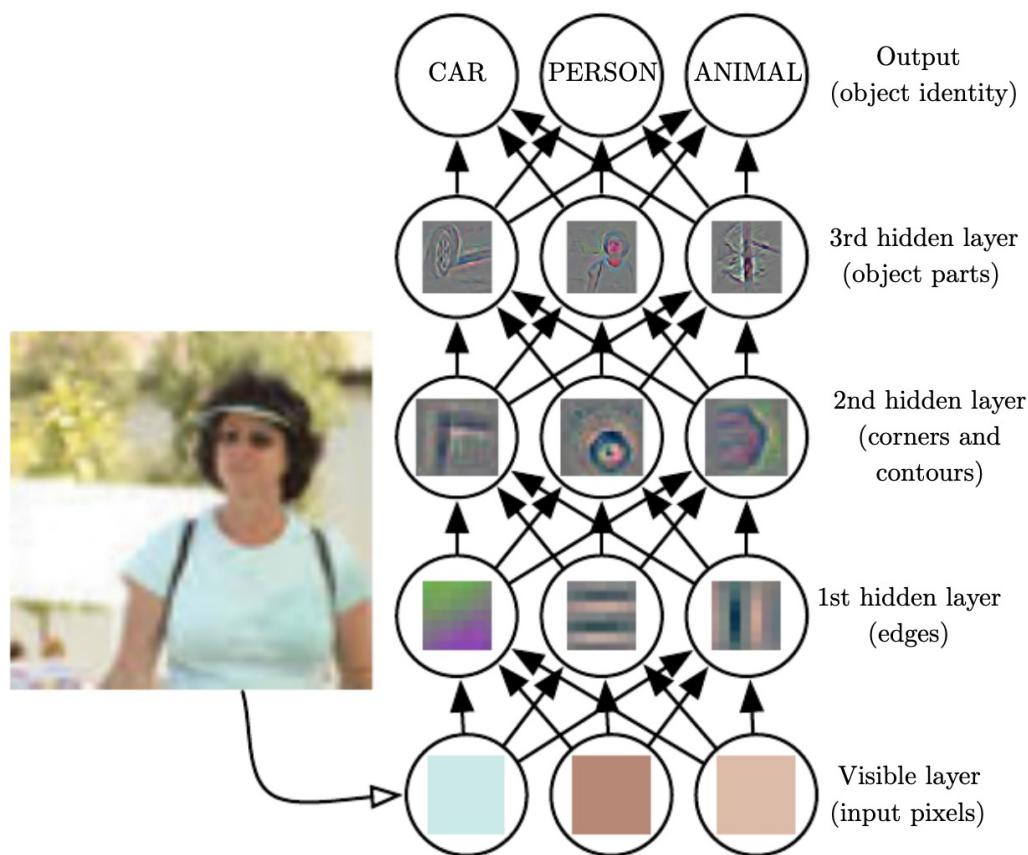
8 neurons

+ -

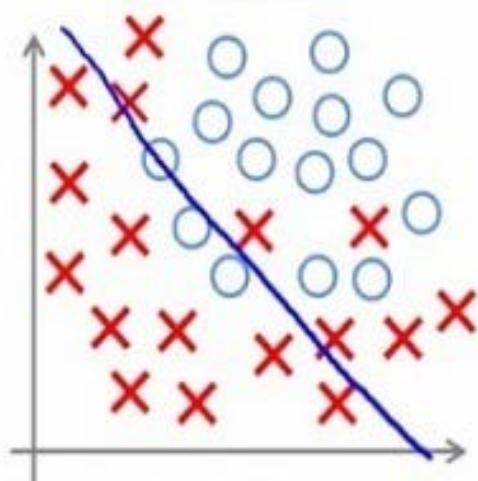
8 neurons



WHY DO ARTIFICIAL NEURAL NETS WORK?

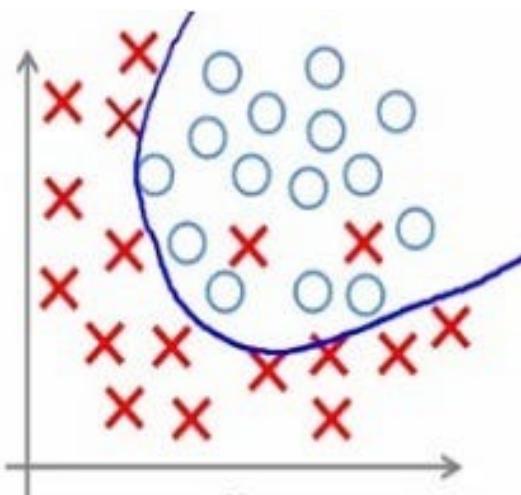


OVERFITTING

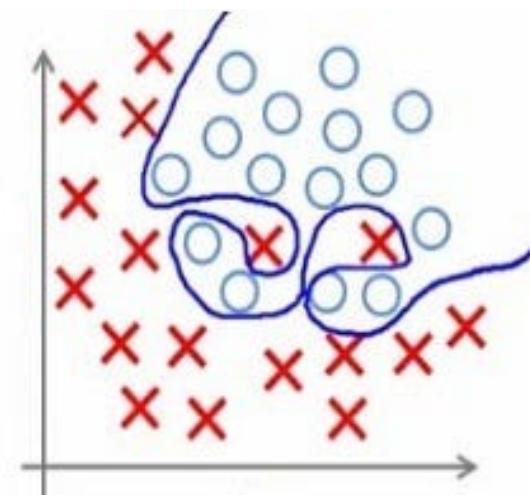


Under-fitting

(too simple to
explain the
variance)



Appropriate-fitting

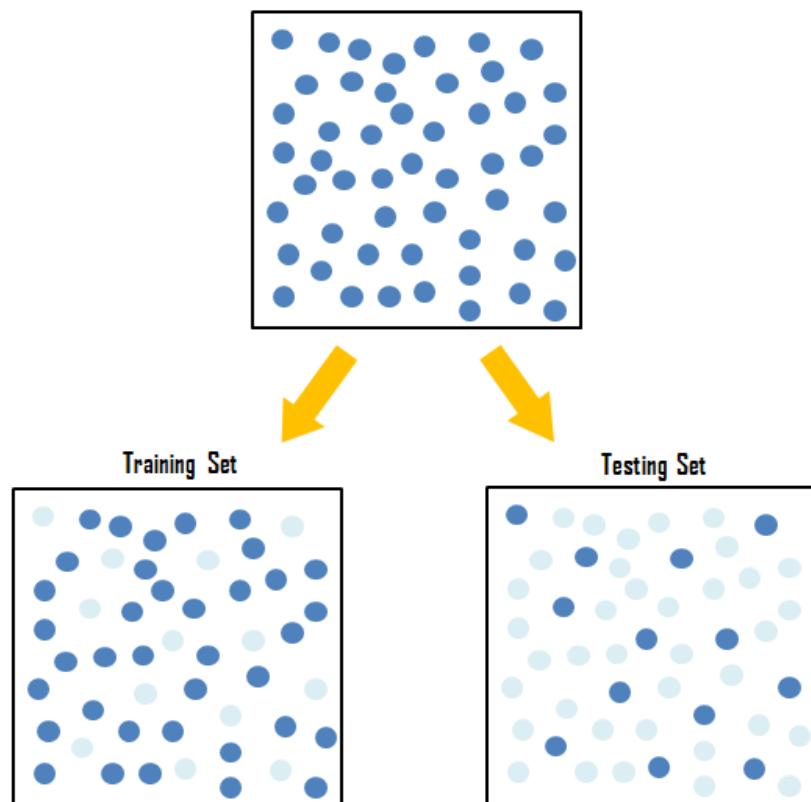


Over-fitting

(forcefitting – too
good to be true)



TRAINING, VALIDATION, TESTING



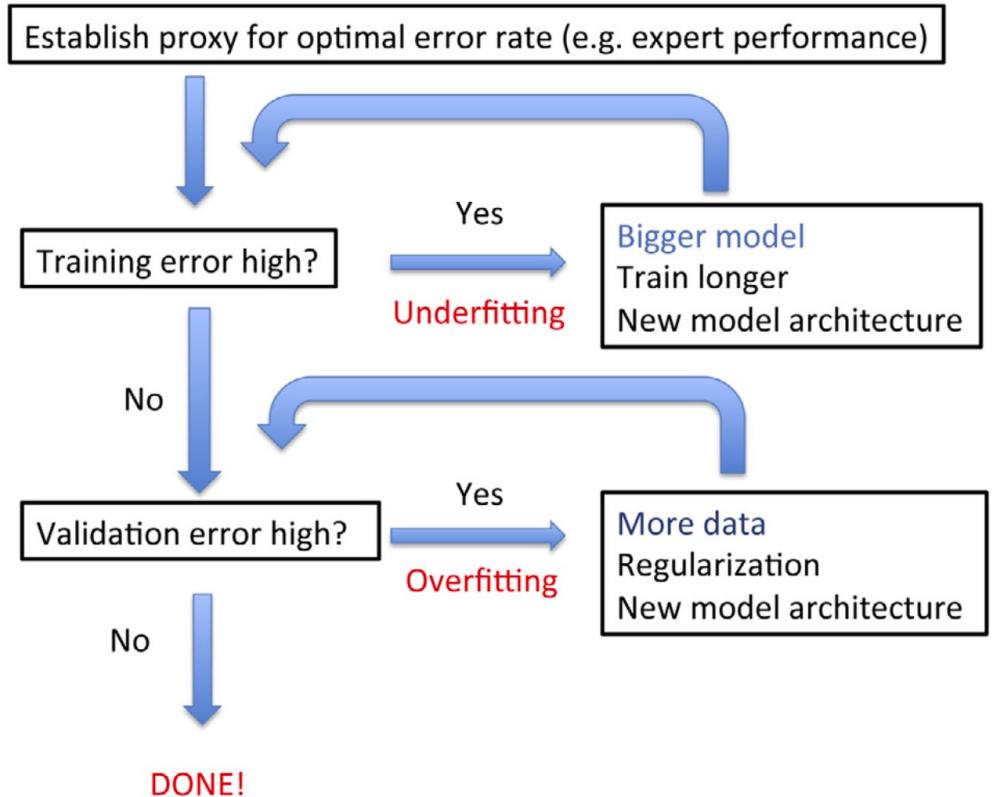
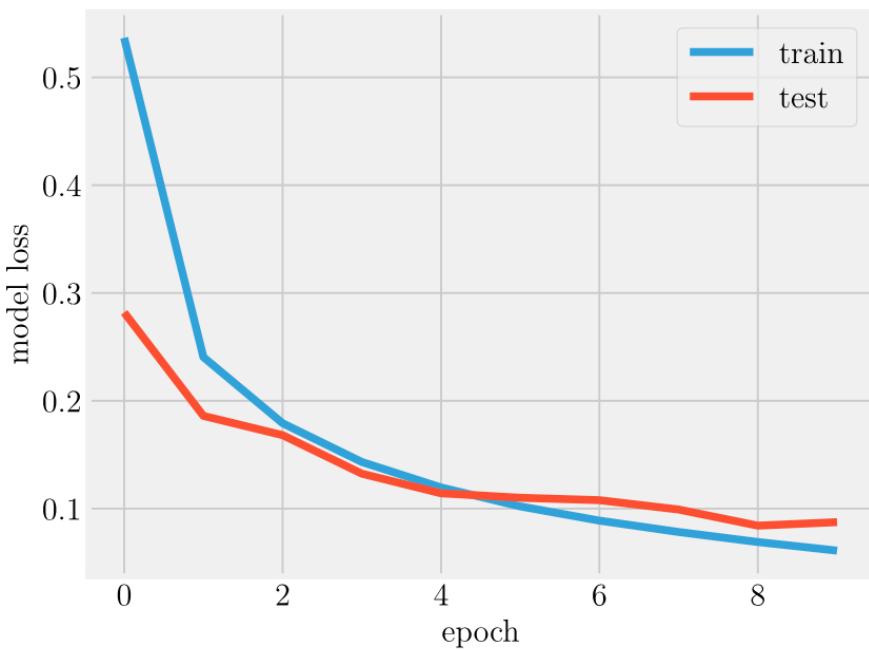
My model on training data



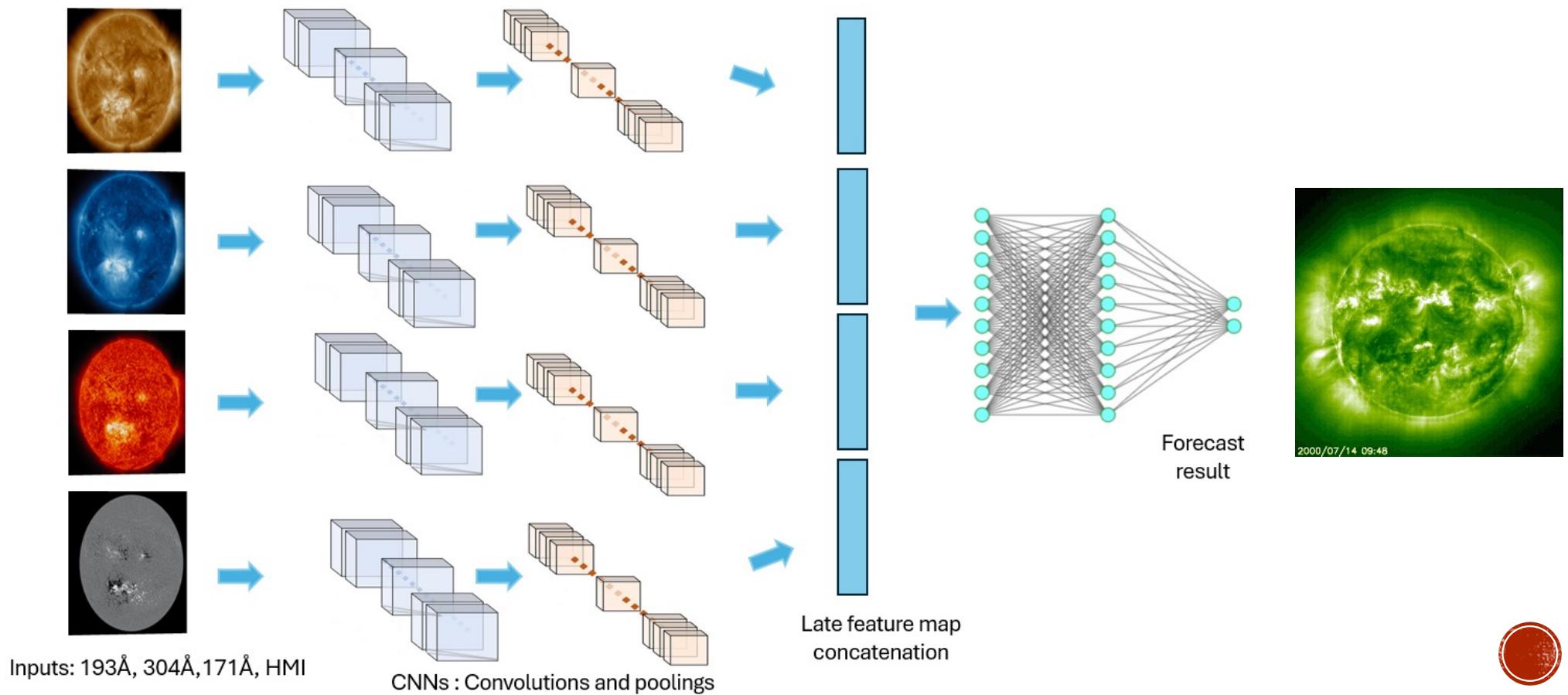
My model on test dataset



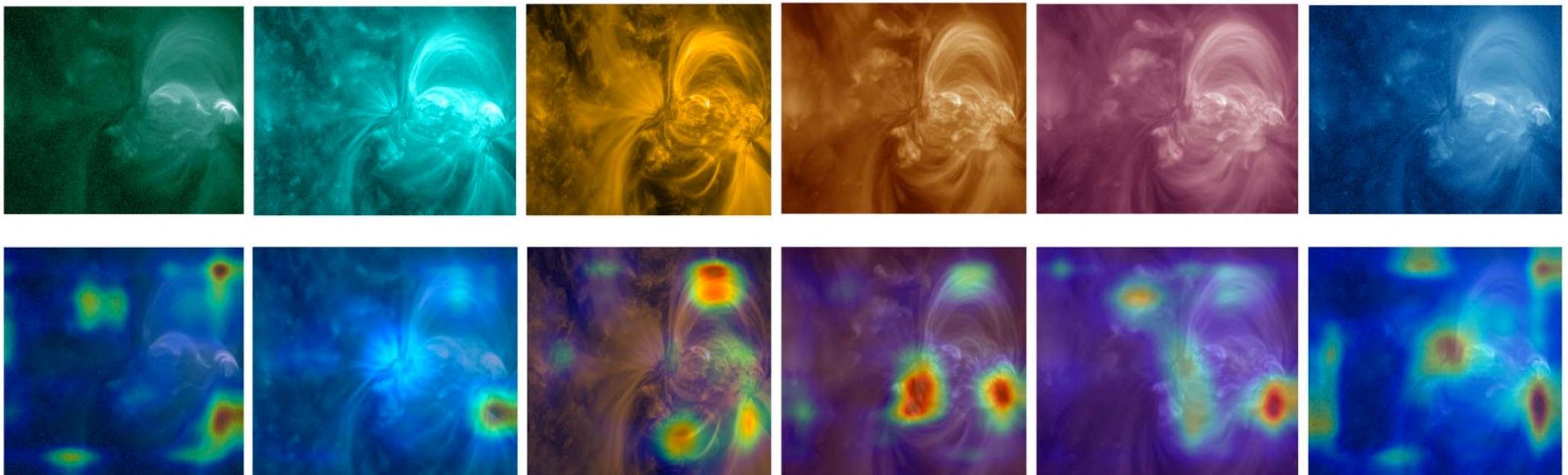
TRAINING: WHEN TO STOP?



FORECASTING FLARES



FLARES: GRADIENTS TO CREATE HEATMAPS



SOMETHING IS WRONG WITH THE PANDA



$+ .007 \times$



$=$



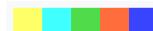
\mathbf{x}

$y = \text{"panda"}$
w/ 57.7%
confidence

$\text{sign}(\nabla_{\mathbf{x}} J(\boldsymbol{\theta}, \mathbf{x}, y))$
“nematode”
w/ 8.2%
confidence

$\mathbf{x} +$
 $\epsilon \text{ sign}(\nabla_{\mathbf{x}} J(\boldsymbol{\theta}, \mathbf{x}, y))$
“gibbon”
w/ 99.3 %
confidence

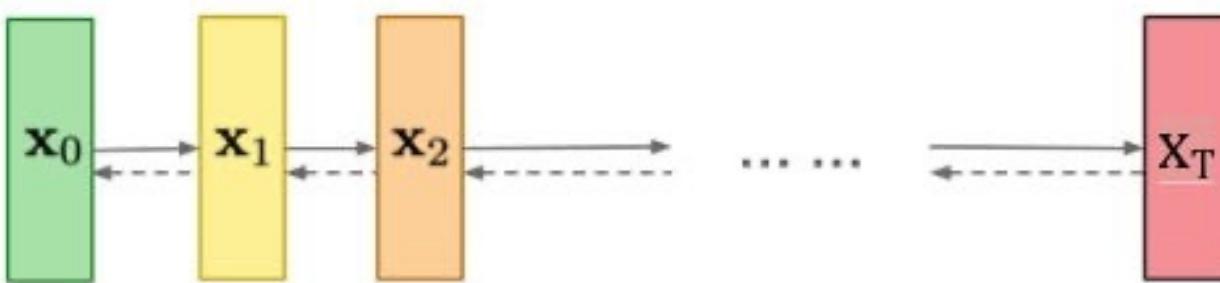
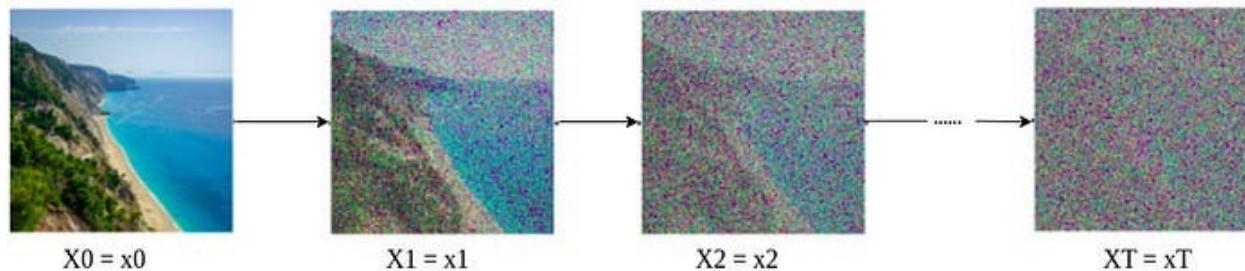
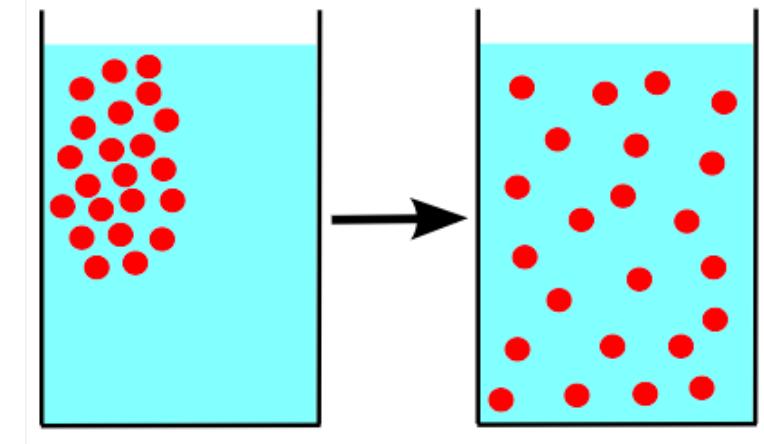
GENERATIVE AI



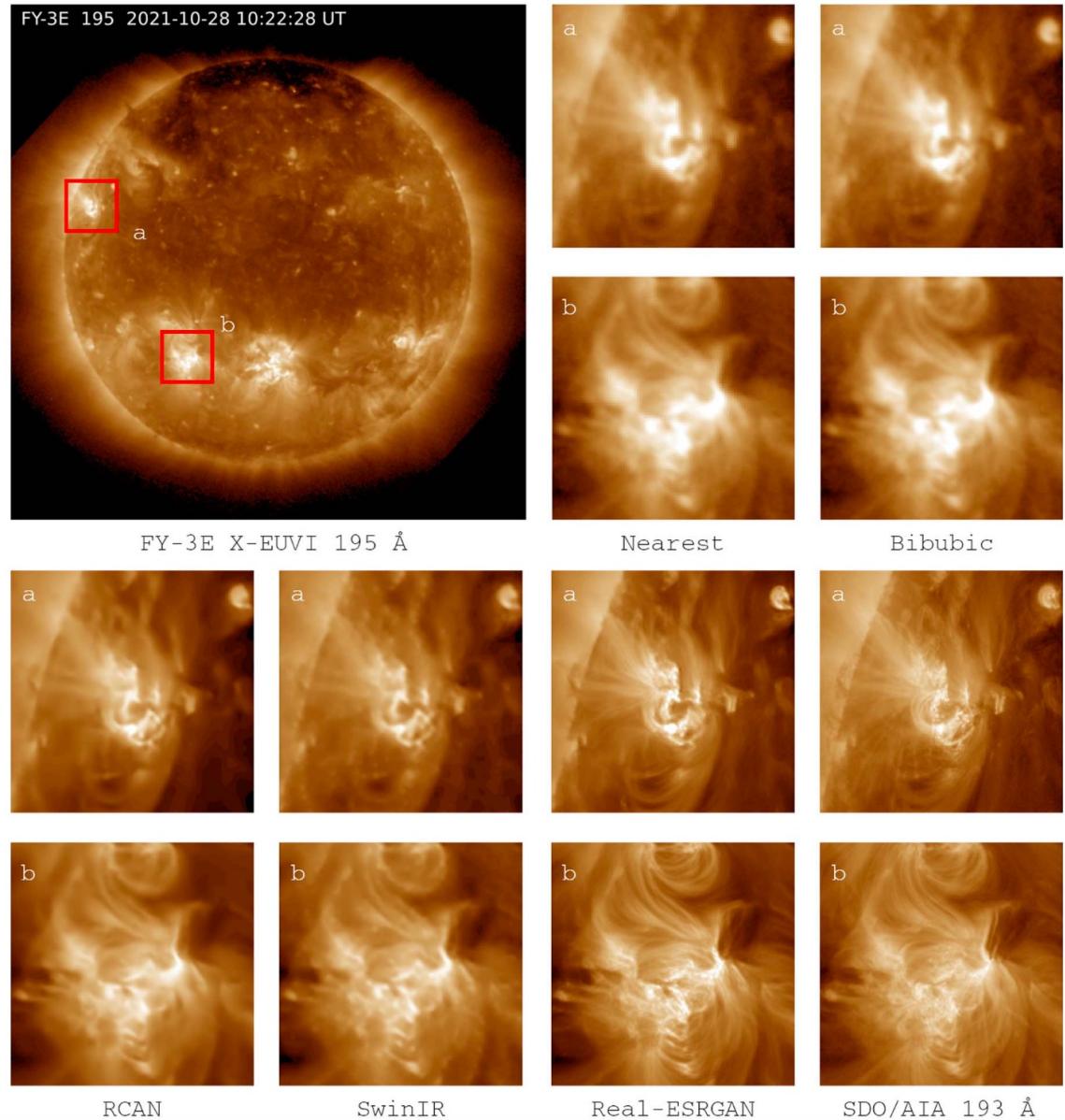
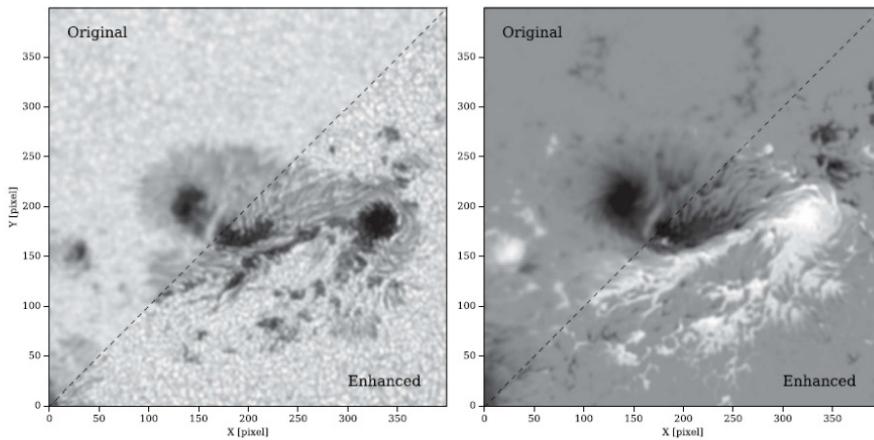
 Designer

Powered by DALL-E 3

HOW DALL-E WORKS?



ENHANCING SOLAR IMAGES



Prototypical Inversion Prompt in Midjourney Version 5.1

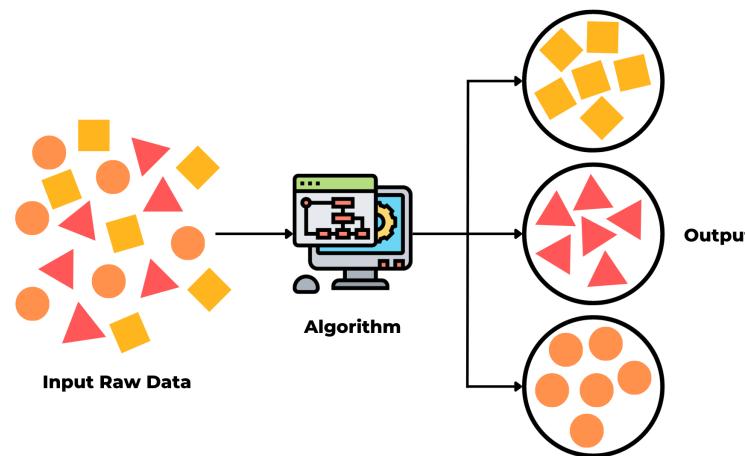
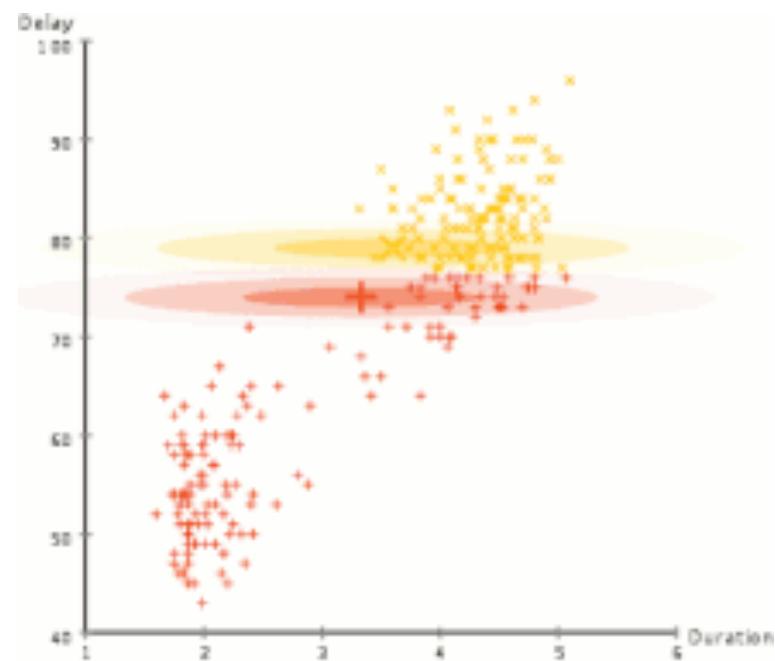
“An astronaut riding a horse”



“A horse riding an astronaut”



UNSUPERVISED LEARNING: CLUSTERING



How clustering works

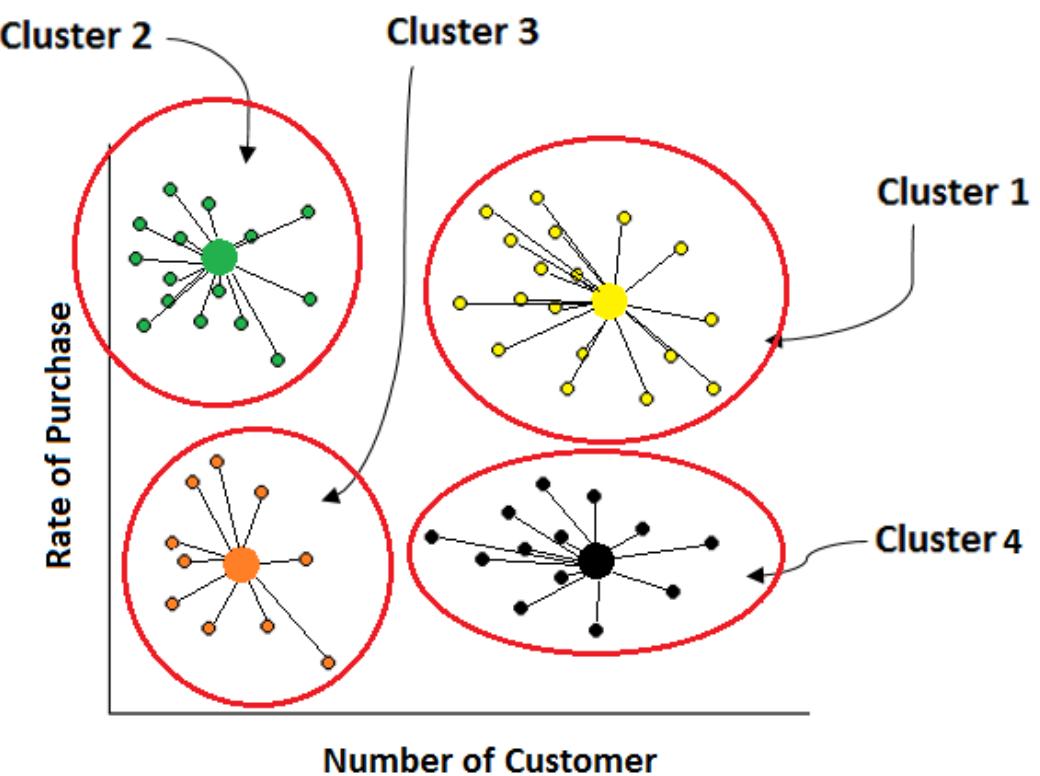


K-MEANS ALGORITHM

observations

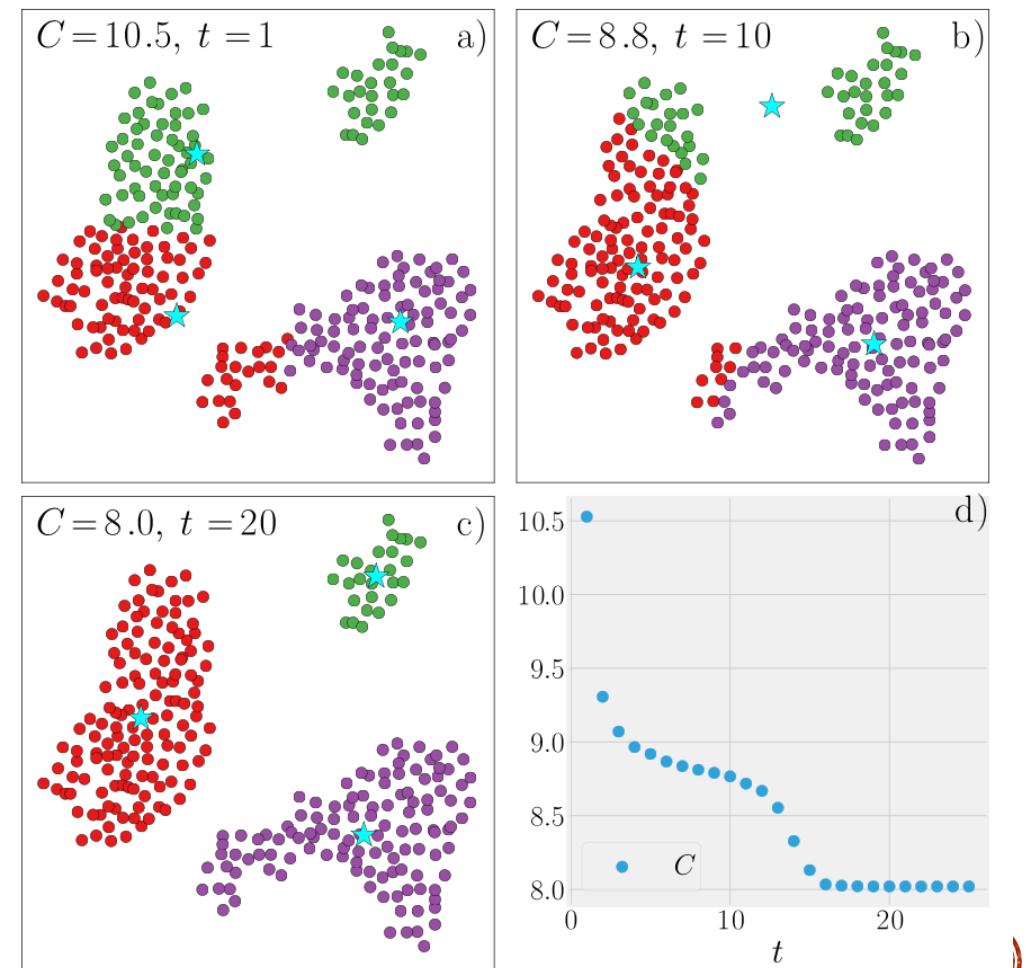
$$\text{Cluster means: } \mu_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} x_n$$

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



K-MEANS ALGORITHM

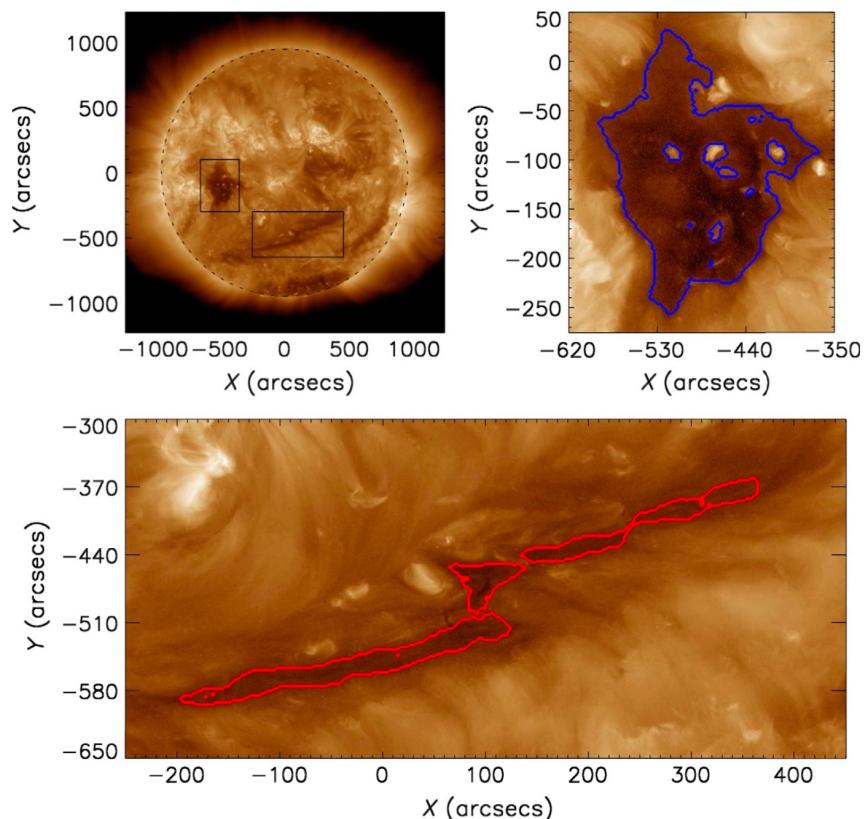
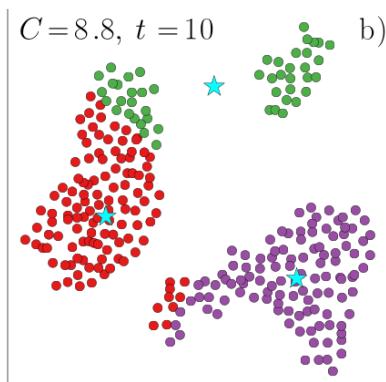
$$r_{nk} = \begin{cases} 1, & \text{if } k = \arg \min_{k'} (x_n - \mu_{k'})^2 \\ 0, & \text{otherwise} \end{cases}$$



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>

