Welcome!

Machine Learning Decal

Hosted by Machine Learning at Berkeley

ML@B

Agenda

Who are we?

What is Machine Learning?

Class Logistics

General Overview and Context

Machine Learning Pipeline

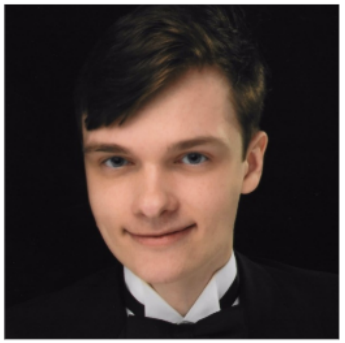
Python/Numpy

Scikit-Learn

Questions

Who are we?

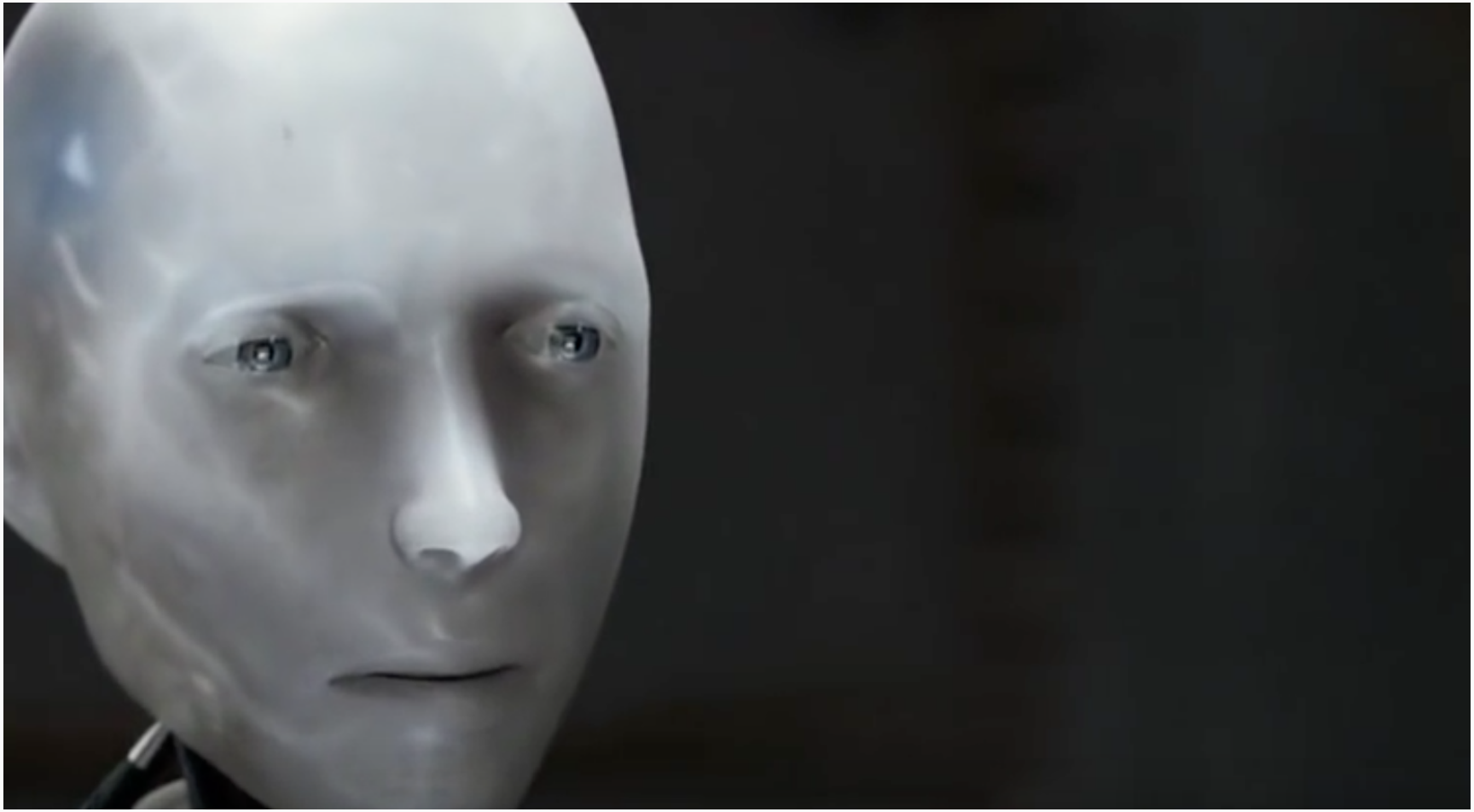
Machine Learning @ Berkeley Education Team



What is Machine Learning?

Age Old Question



Can AI compose music?

ML@B

Can AI paint a canvas?

Can AI paint a canvas?

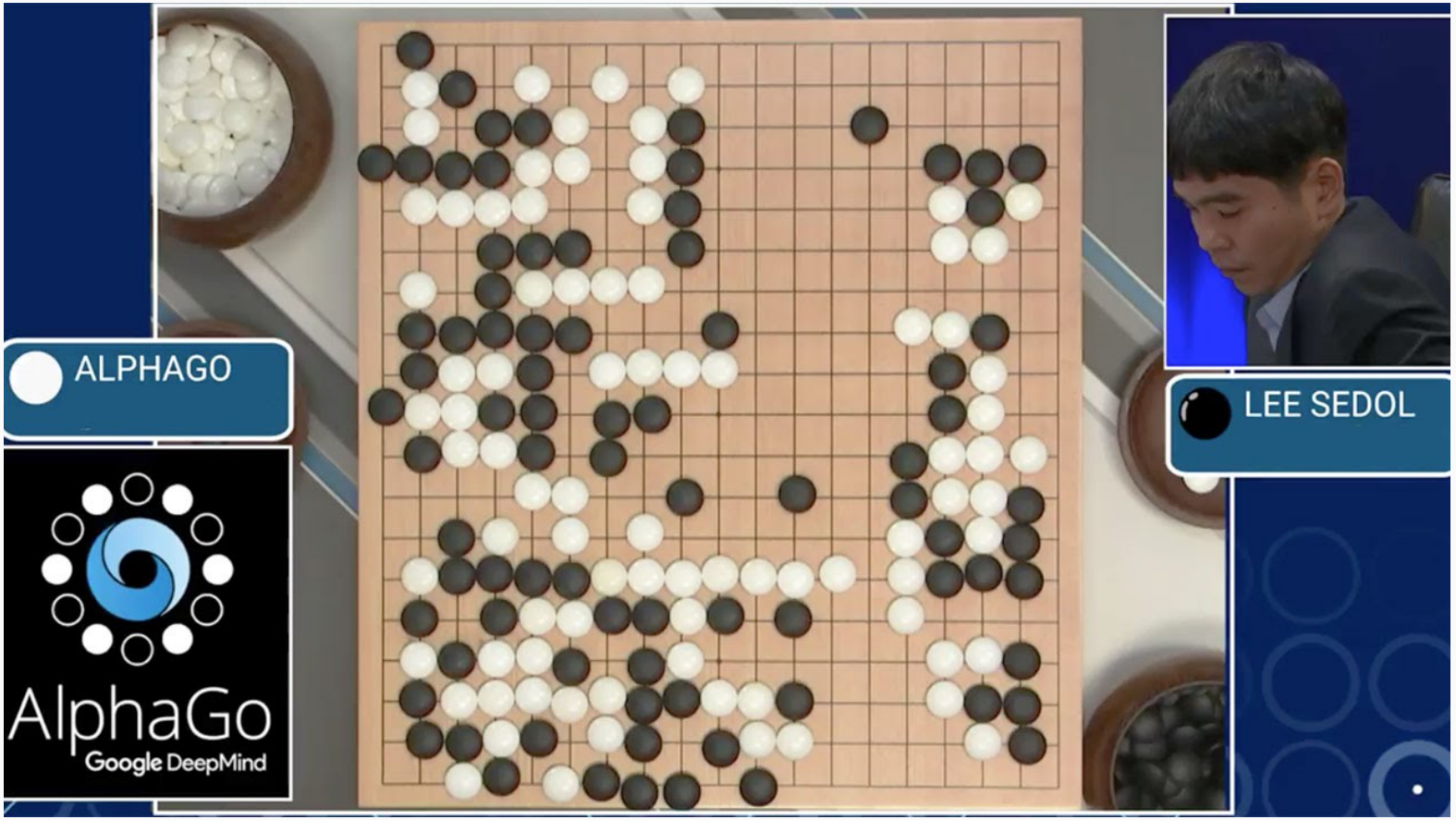
Dancing!


Source Video
Detected Pose
Source to Target 1 Result
Source to Target 2 Result

FAKE NEWS!

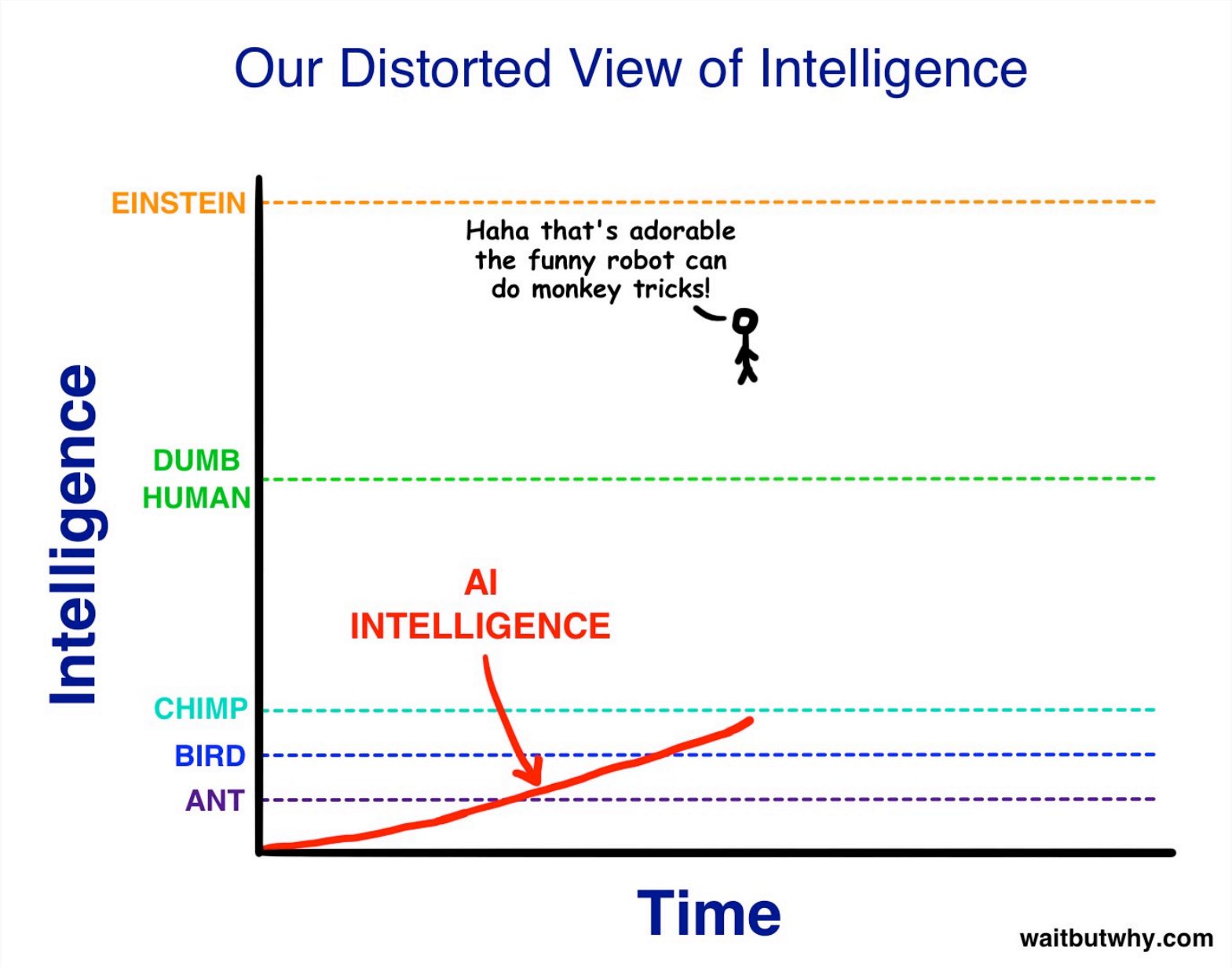
Pose tracking!

Superhuman reasoning

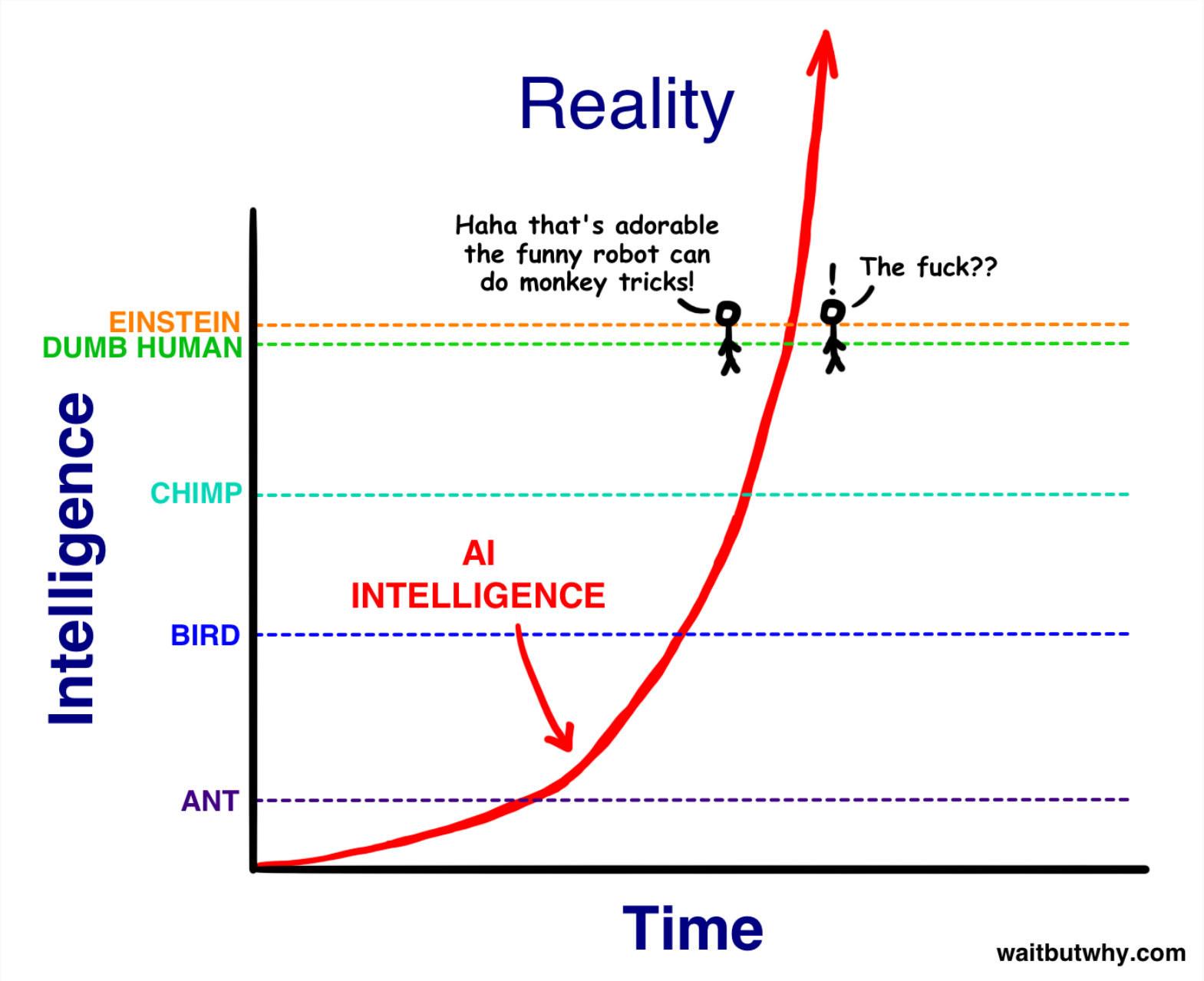


Selfdriving Cars

Our view of intelligence



What intelligence is actually like



Class Logistics

Goal

Goals:

- Understand the major concepts in machine learning
- Understand the tradeoffs between different approaches (what do I use when and why?)
- Gain familiarity with code to solve ML problems

How we accomplish this:

- Lecture (2 hrs/week) - theoretical intro to various techniques, along with demos
- Homeworks (3-6 hrs/week) - practice implementing different models on different datasets

Enrollment

ML@B

- Some of you here are in the class, others on the waitlist
- Must show up to first lecture if you wish to be enrolled
- We will give out course enrollment codes tonight - first to all enrolled students who showed up to this class, then waitlist students who showed up

Logistical

Join Piazza: piazza.com/berkeley/fall2018/cs198082
All communication will be through Piazza.

Join Gradescope:

Clone the github:
https://github.com/mlberkeley/Machine-Learning-Decal-Fall-2018

- All homeworks, lecture slides, and lecture recordings will be posted here
- Also check out the calendar which has all important dates (homework deadlines, office hours, etc.) for the class

Attendance

ML@B

- Attendance is mandatory and will be taken at every lecture
- You may miss up to 3 classes [not including first class]
- After your 3rd missed day of class, you will automatically be assigned a no pass

Grades

- 70 % average on all homework assignments
- Submit final project
- Automatic "no pass" for insufficient attendance

Assignments

ML@B

- 7 Homework Assignments
  - Work in groups up to 3
  - 1-2 week timeline
- Final Project/Hackathon
  - Can either submit a final project on own time or can attend and submit a project through the decal hackathon at the end of the semester
  - All the instructors will be there to answer questions and guide you through projects

Late Assignments

ML@B

- Each assignment will have 1 week late turn in period
- Small penalty for turning in late
- No submissions allowed after late turn in period over!

Homework 1 is out NOW, due next Tuesday (in one week)!

Office Hours

Date/Time/Location of office hours for the next two weeks will be posted on Piazza and on the calendar

Got Questions?

ML@B

During lecture:

- Raise your hand!
- Ask on the lecture questions thread on Piazza!

Outside lecture / about anything else:

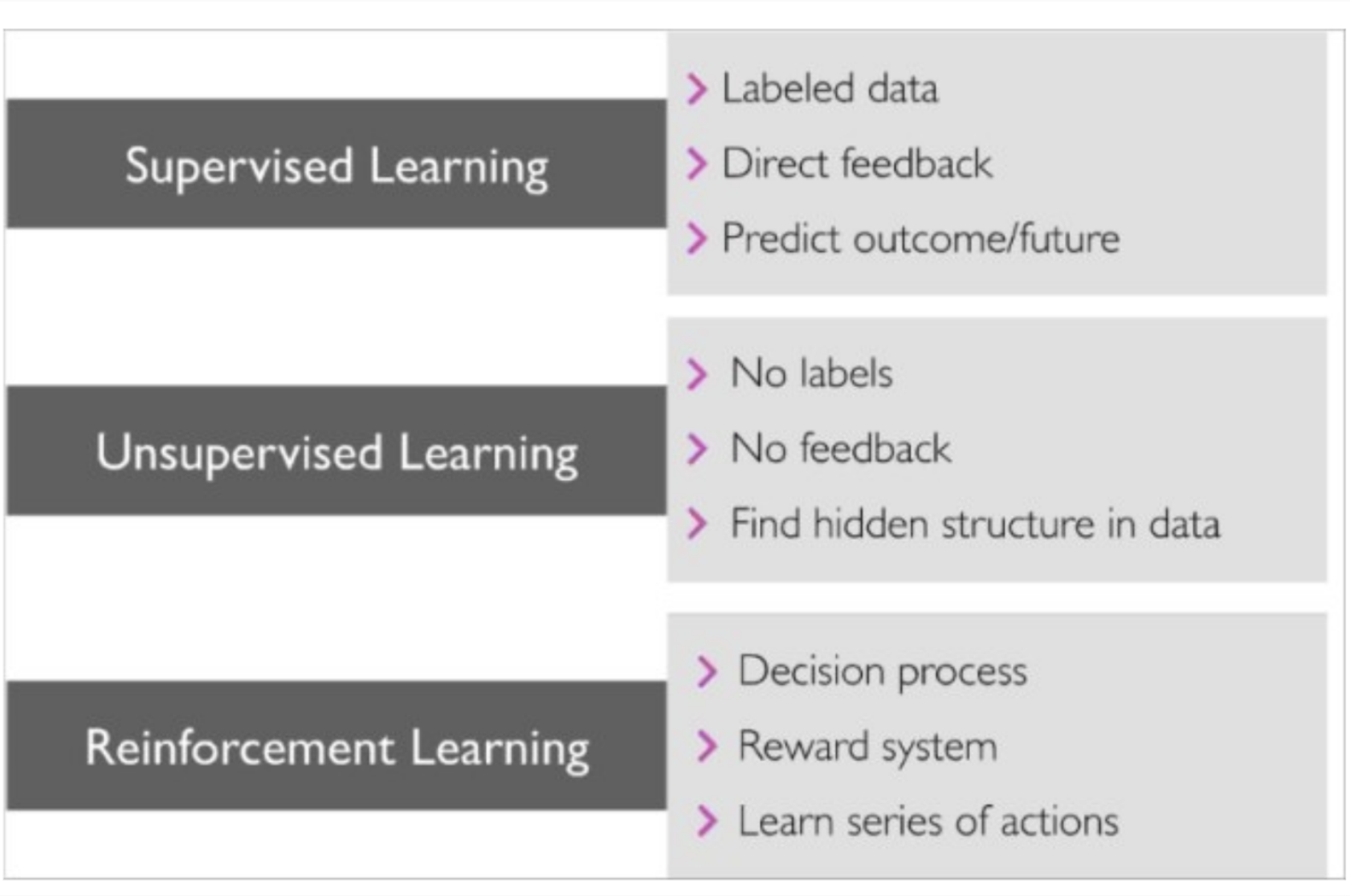
- Please don't email
- PIAZZA IS YOUR FRIEND!!!

Course Content

Syllabus

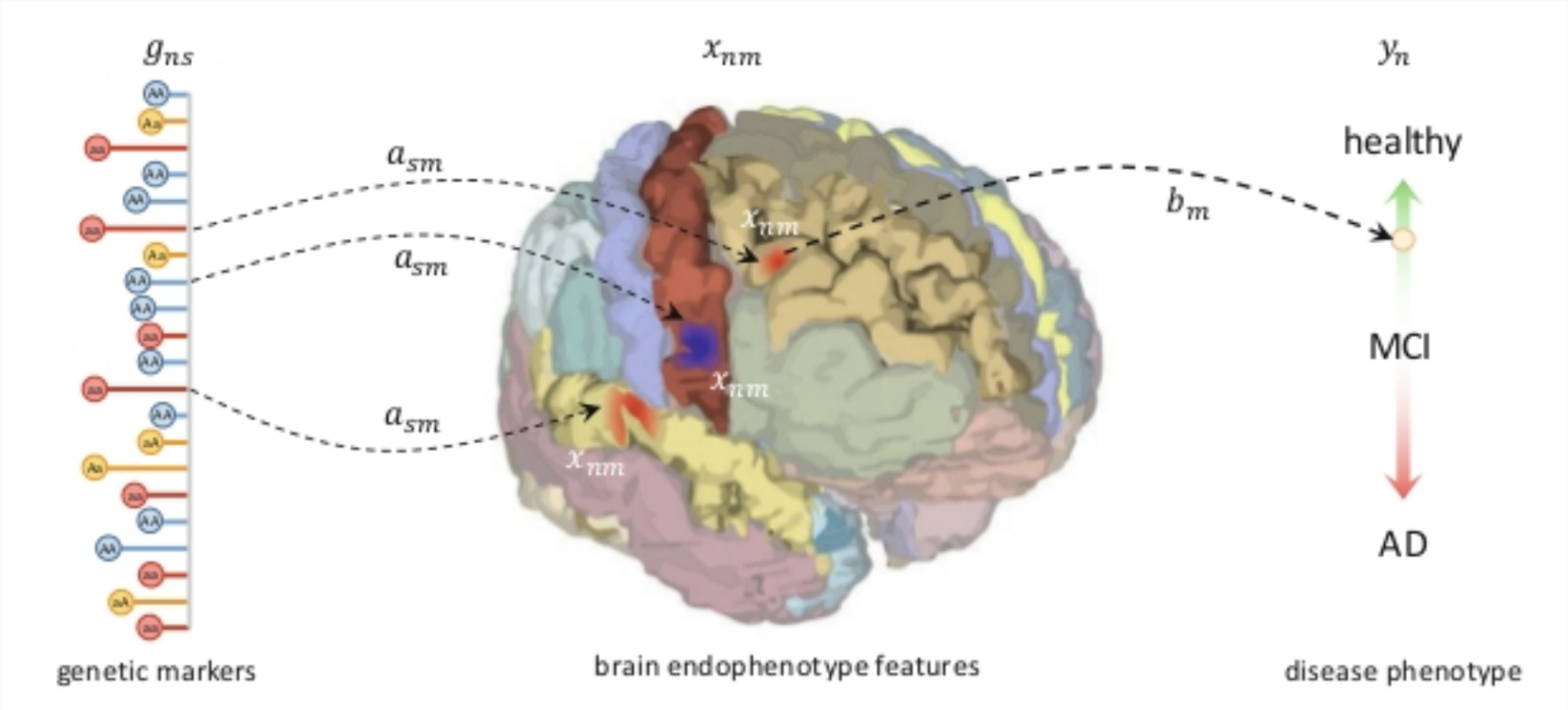
General Overview and Context

3 Different Classes of Machine Learning Problems



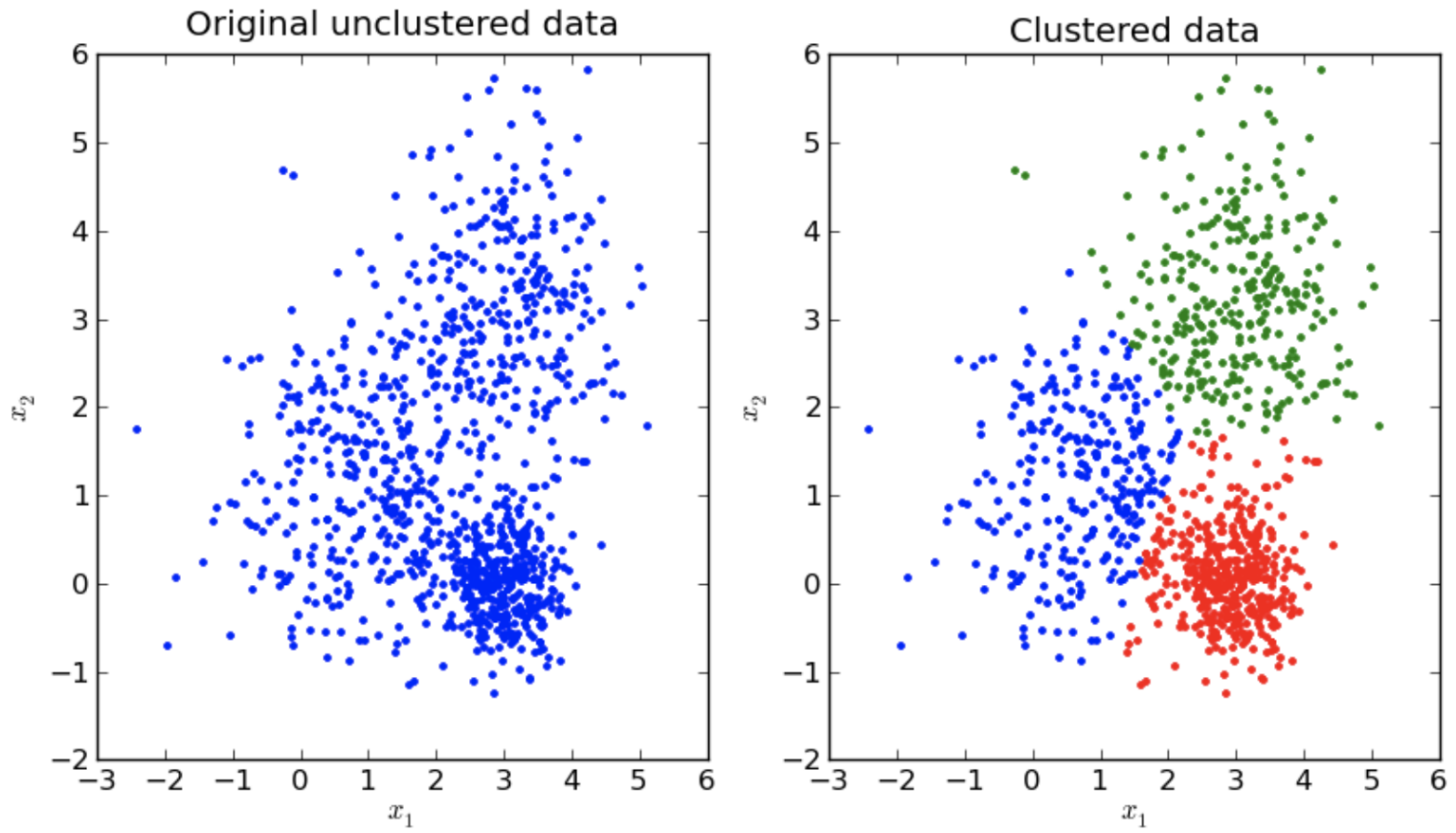
Supervised Learning (Classification) Ex: ImageNet (10M+ images), 1000 classes

Supervised Learning (Regression) Ex: Imaging Genetics and Genome-Wide Association Studies

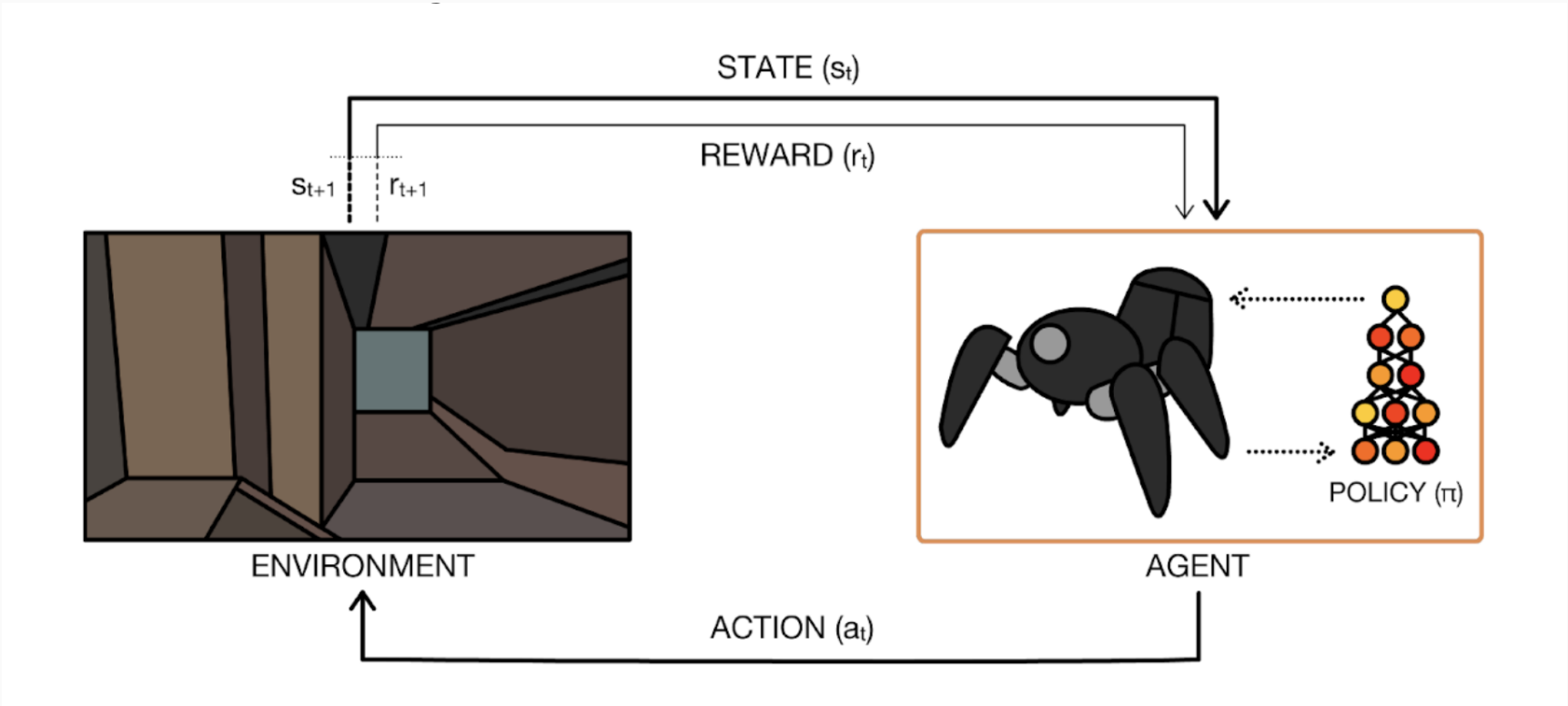


ML@B

Unsupervised Learning (Clustering)



Reinforcement Learning



Machine Learning Pipeline

Typical Pipeline

ML@B

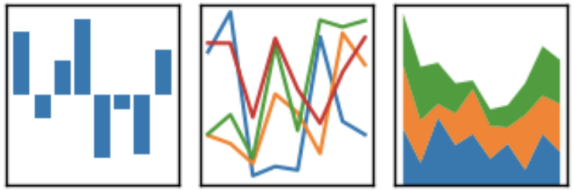
- Acquire the data
- Prepare/Visualize Data
- Choose a Model
- Train a Model on the Training Set
- Evaluate Model Performance
- Tune Model Hyperparameters
- Prediction!

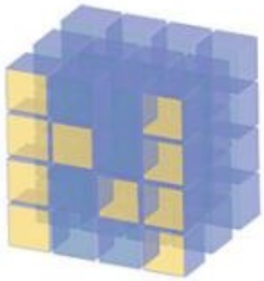
STEP ONE: Acquire the Data

STEP TWO: Prepare and Visualize Data

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$





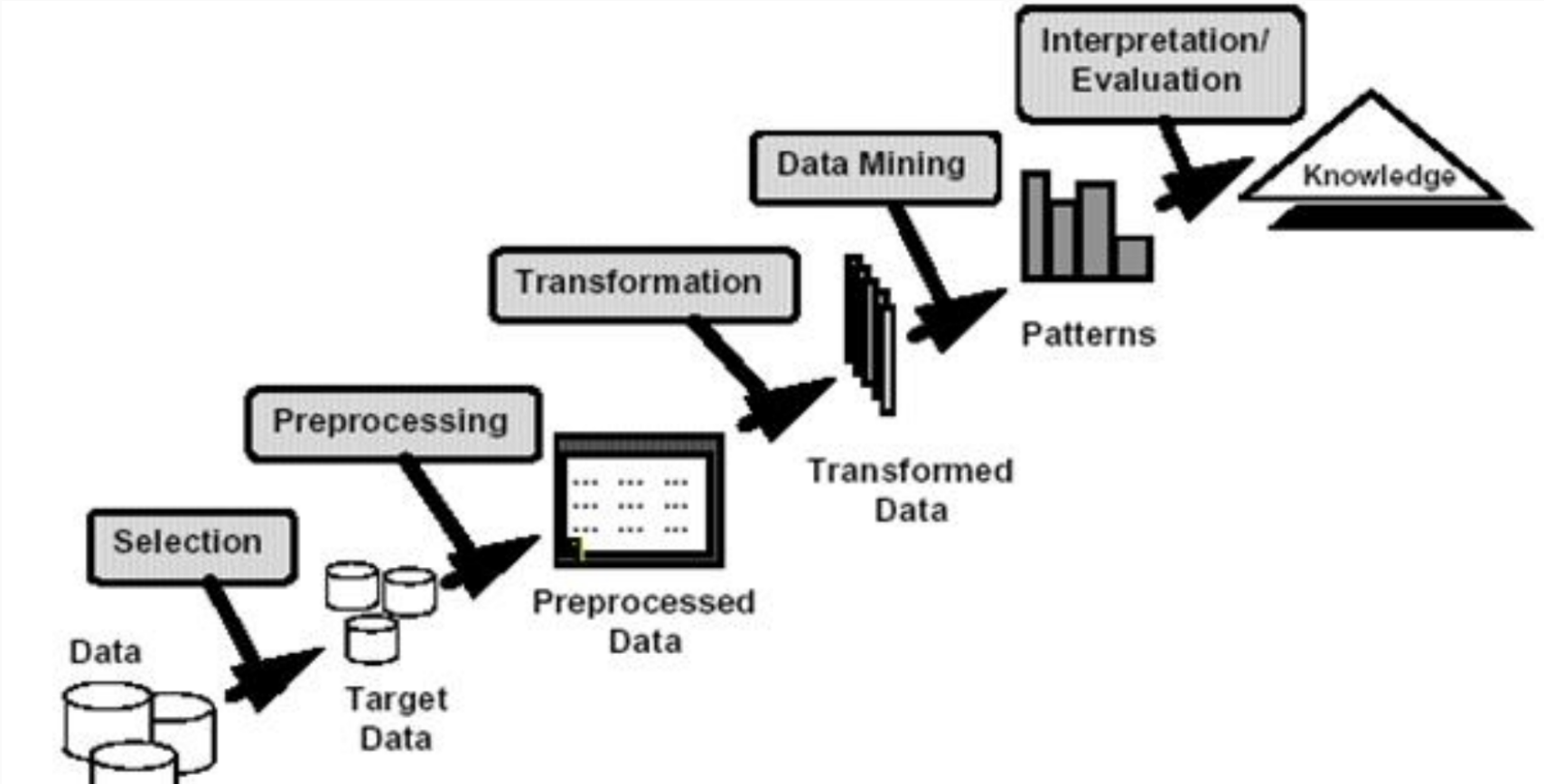
NumPy

scikits learn

machine learning in Python

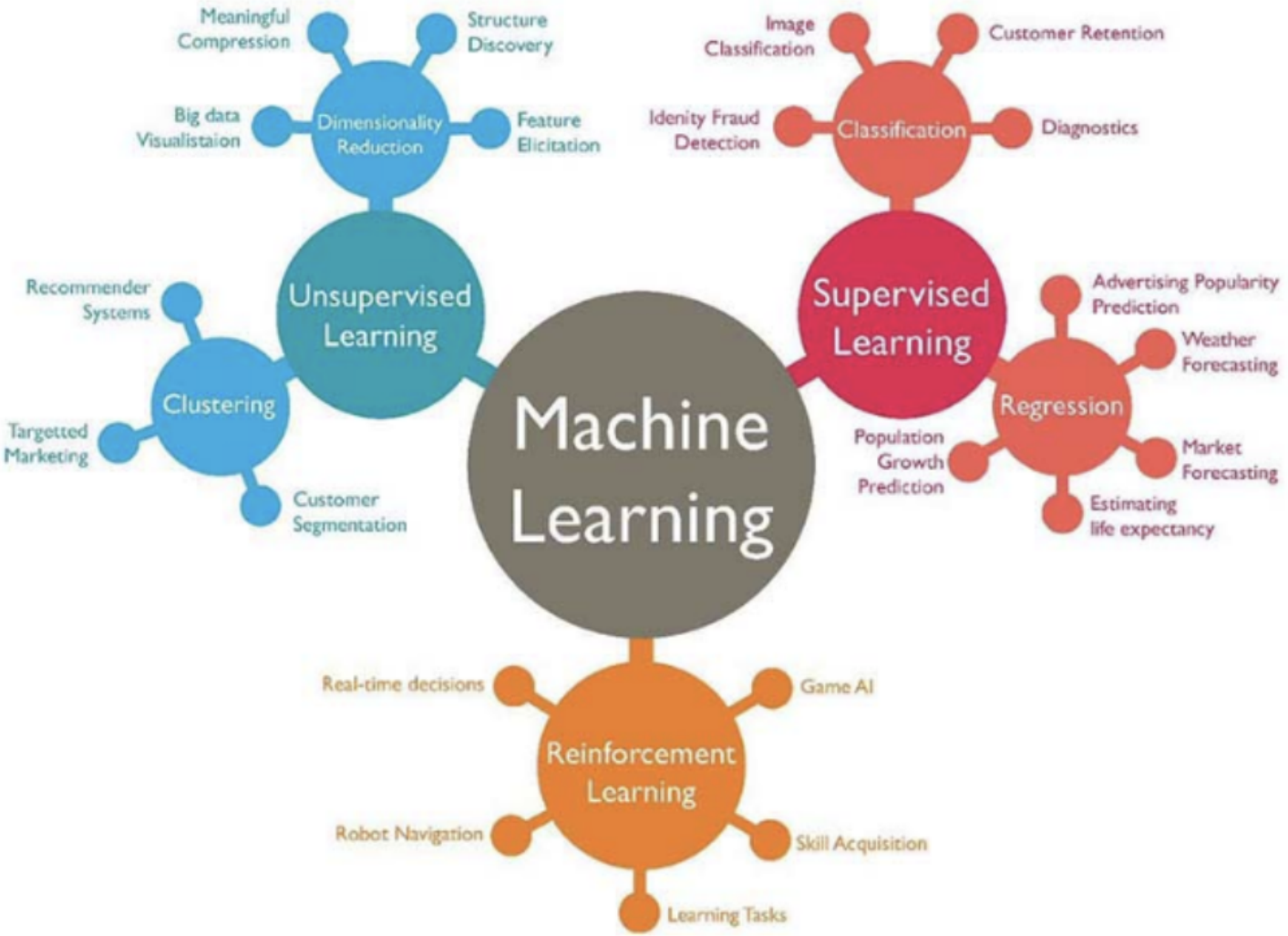
Importance of Feature Selection and Data Preprocessing in ML problems

- Data Preprocessing and Feature Selection must be performed before proceeding to actual data mining and ML analyses.

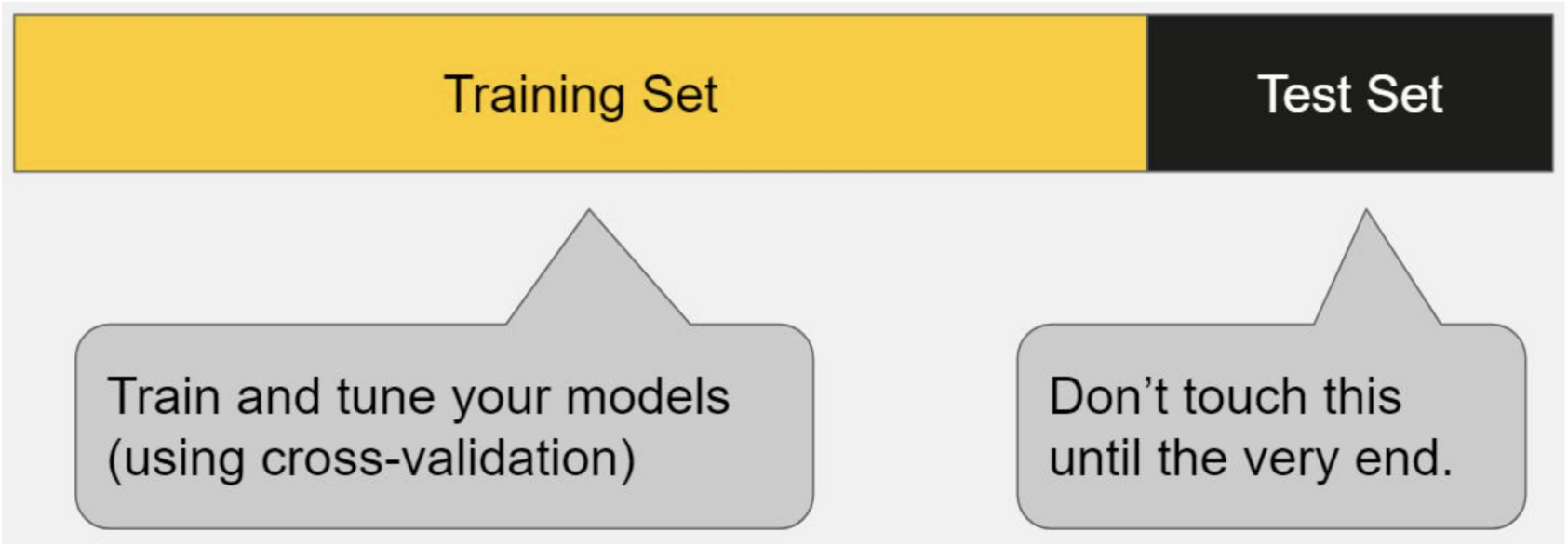


STEPs THREE and FOUR: Choose A Model (3) and Train It! (4)

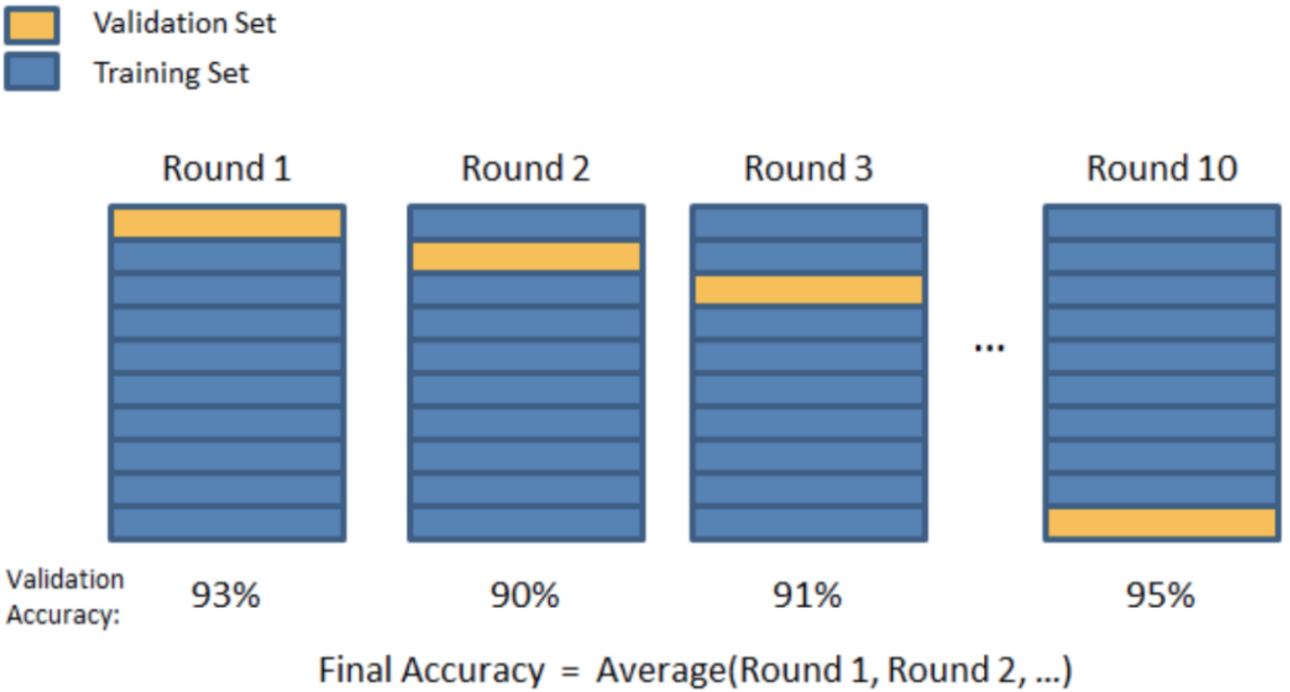
ML@B



Dealing with Datasets in Machine Learning



STEP FIVE: Evaluate the Models Performance in the Validation Step (K-fold Cross-Validation)



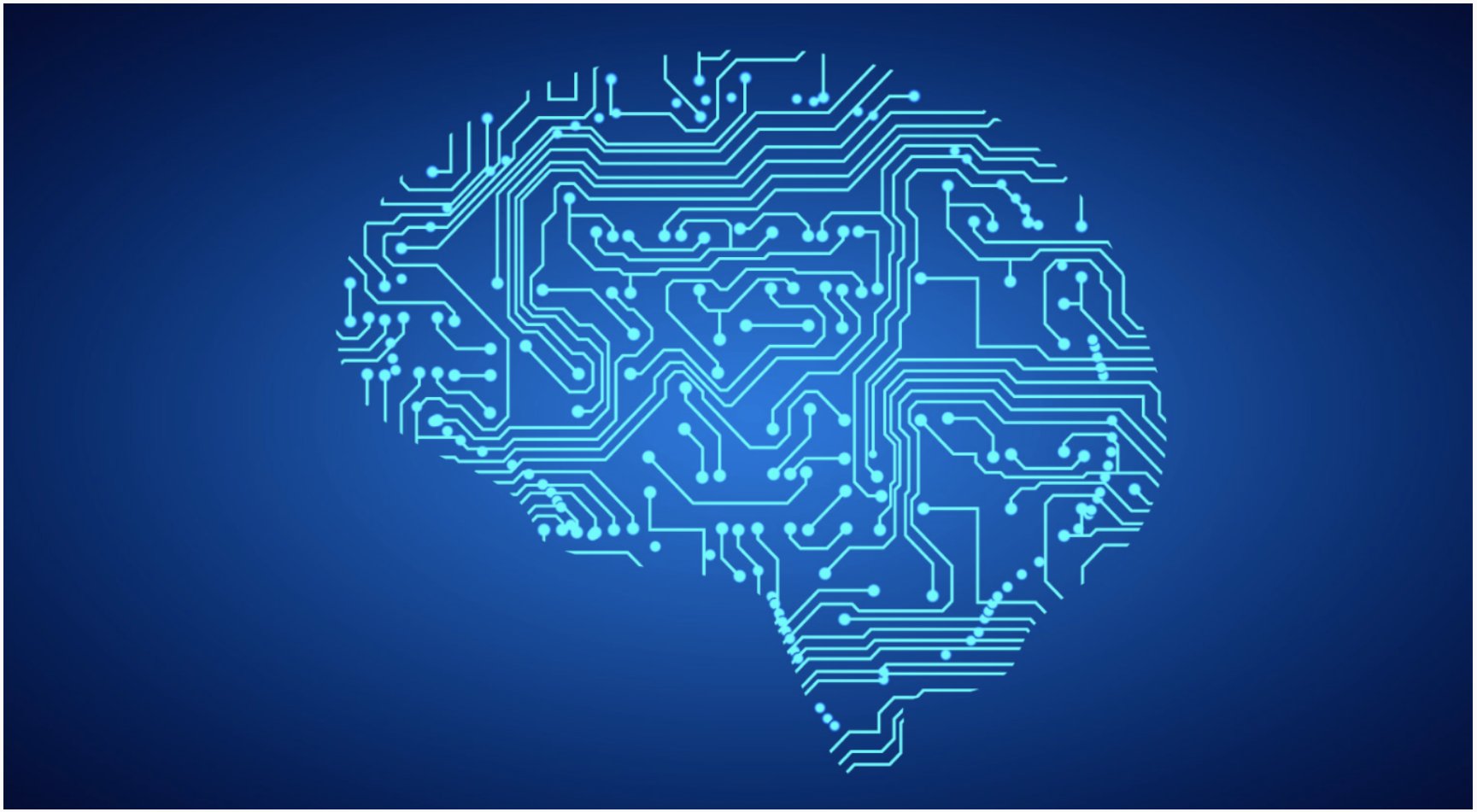
STEP SIX: Tune Model Hyperparameters Via K-Fold Cross-Validation

ML@B

- What happens if our cross-validation accuracy is rather low, and we would like to improve it?
- Choose different Hyperparameters to our model!
- While the ML model finds optimal values of the parameters to minimize some type of loss function, the ML practitioner gets to choose values for the hyperparameters.

STEP SEVEN: Characterize Your Models Performance on a Test Dataset and Predict using Your Model!

ML@B



Python/Numpy

DEMO

Scikit-Learn

DEMO

Questions

Questions?