Types of ML and Data

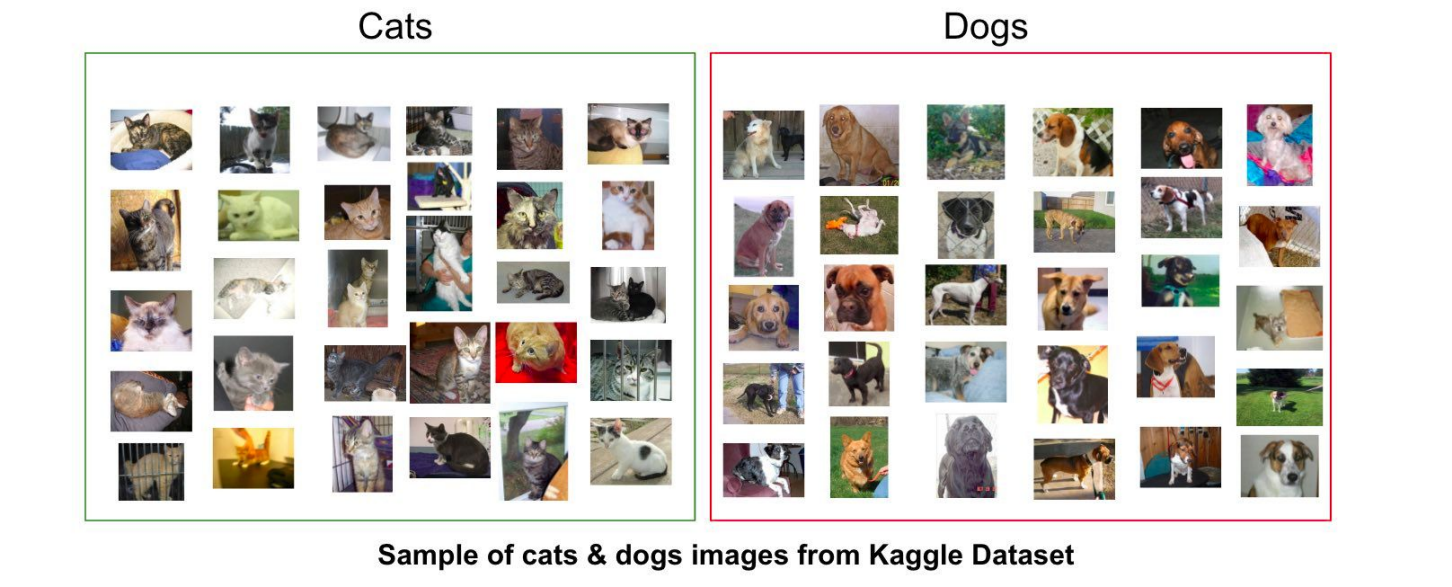
Broad Classification

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

Supervised Learning

Supervised Learning

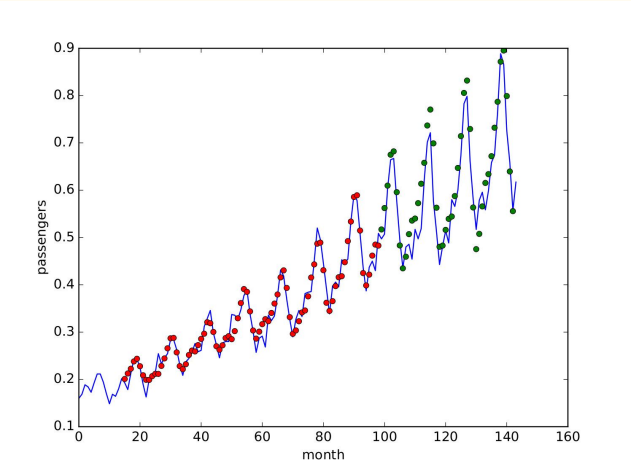
- Given labeled examples, learn to make predictions for unlabeled examples. (Ex. Image Classification)



Sample of cats & dogs images from Kaggle Dataset

Regression vs. Classification

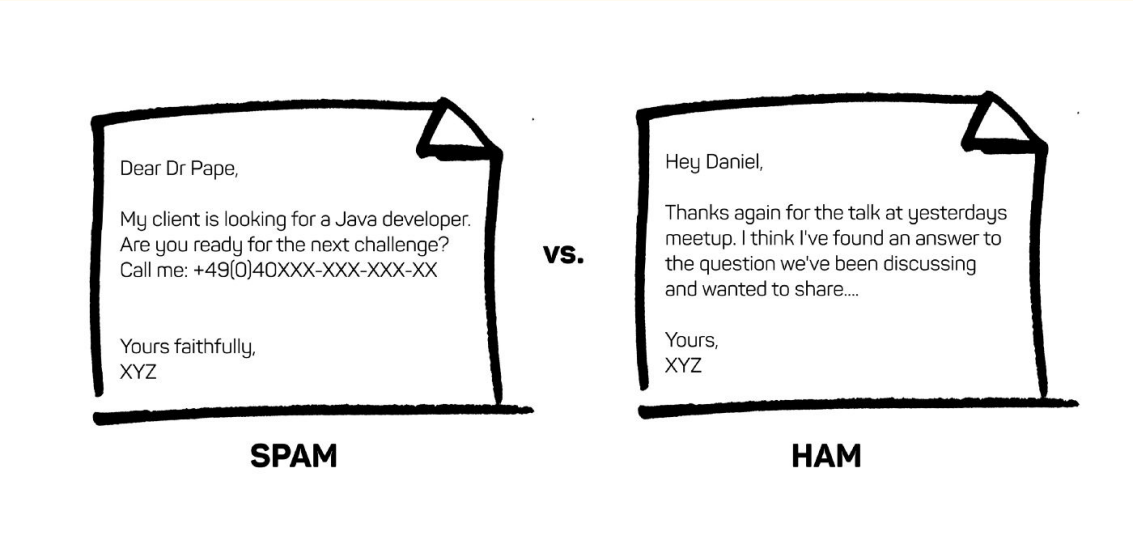
- The output variable takes continuous values
- Estimating or predicting a continuous numerical value



- The output variable takes class labels
- Identifying group membership

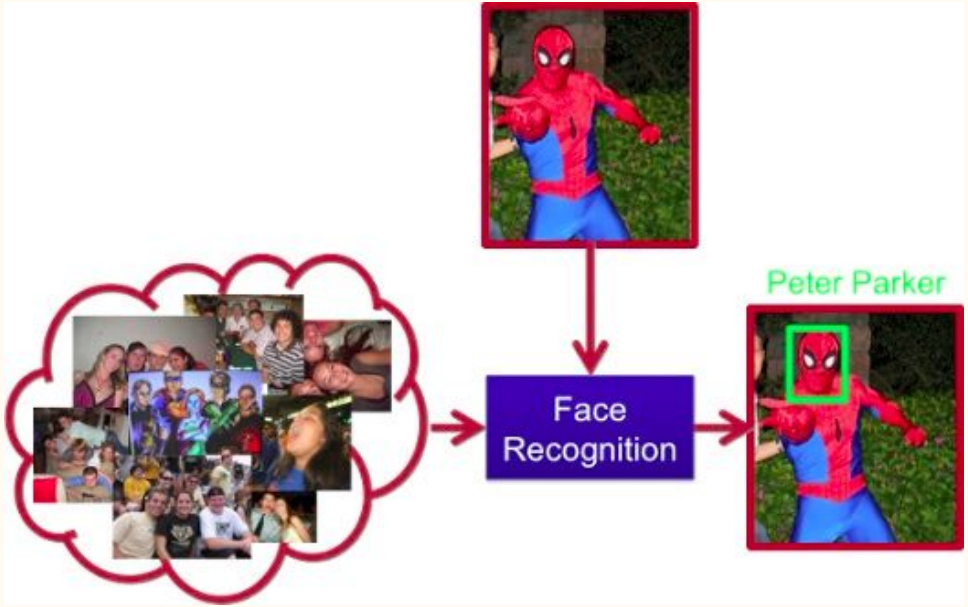
Example - Spam Detection

- Problem: classify each e-mail message as SPAM or non-SPAM (binary classification problem).
- Data: large collection of SPAM and non-SPAM messages (labeled examples).



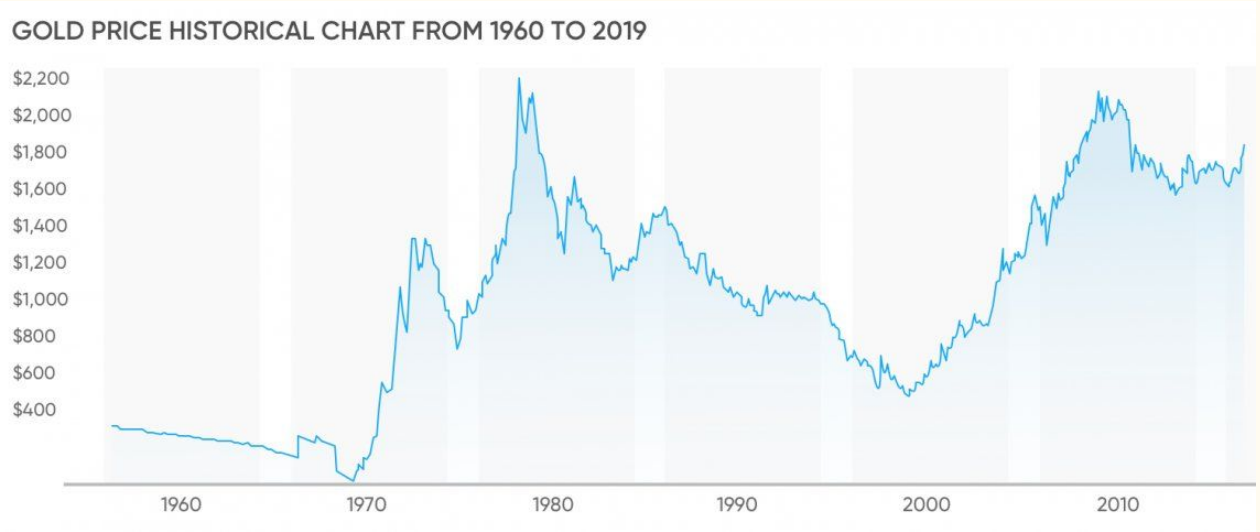
Example - Face Recognition

- Problem: Identify and label a face from images (classification)
- Data: Large collection of images with faces labeled (labeled data)



Example - Price prediction

- Problem: Predict the price of an object in the future on a given date (regression problem)
- Data: Past prices with date (labeled data)

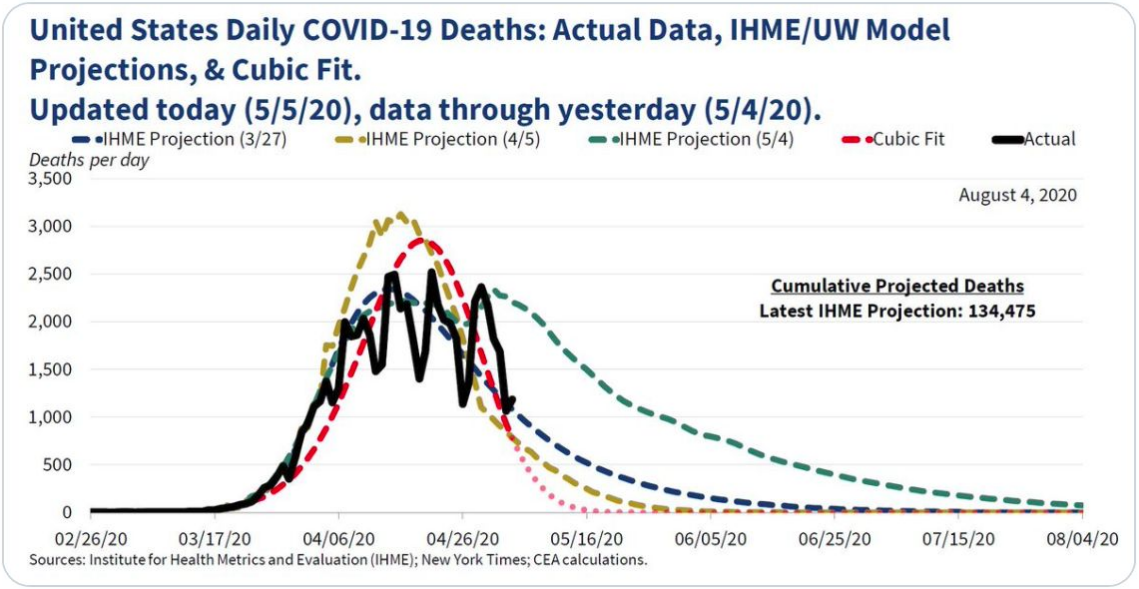


Example - COVID-19 cases prediction

- Problem: Predict the number of cases on a future date (regression)
- Data: Historical record of number of cases

CEA 
@WhiteHouseCEA

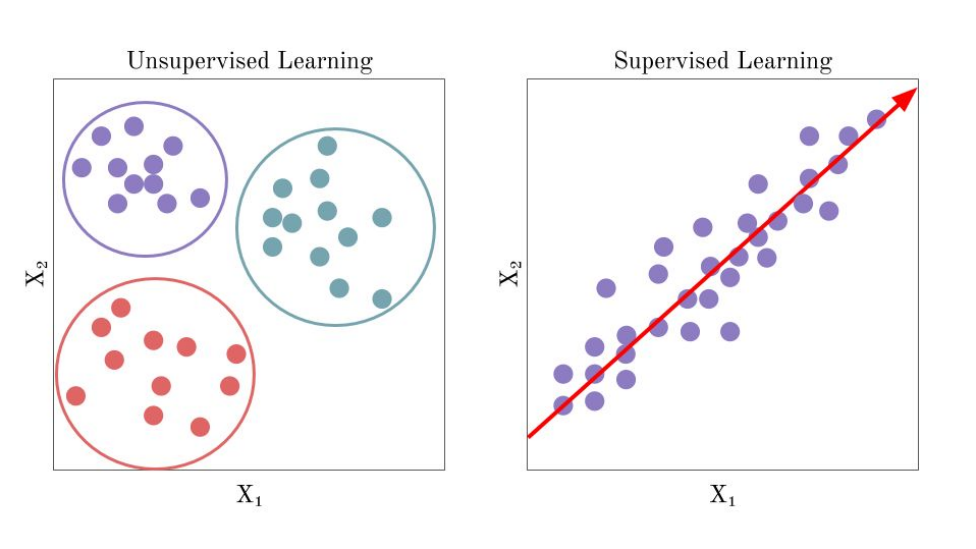
To better visualize observed data, we also continually update a curve-fitting exercise to summarize COVID-19's observed trajectory. Particularly with irregular data, curve fitting can improve data visualization. As shown, IHME's mortality curves have matched the data fairly well.



Unsupervised Learning

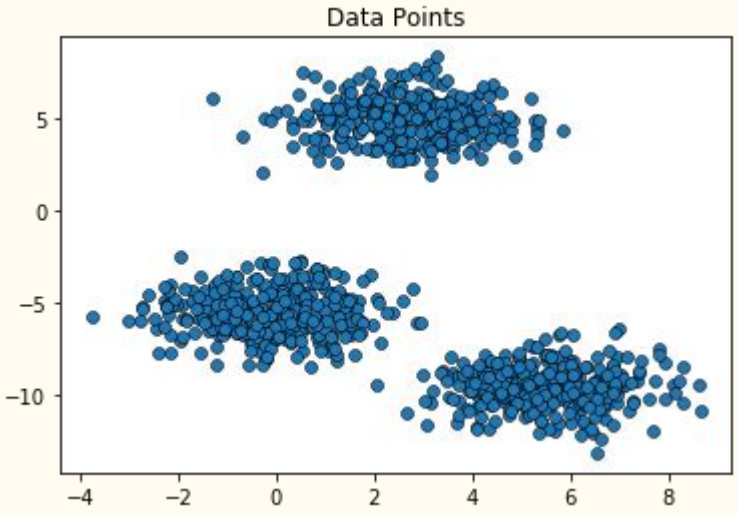
Unsupervised Learning

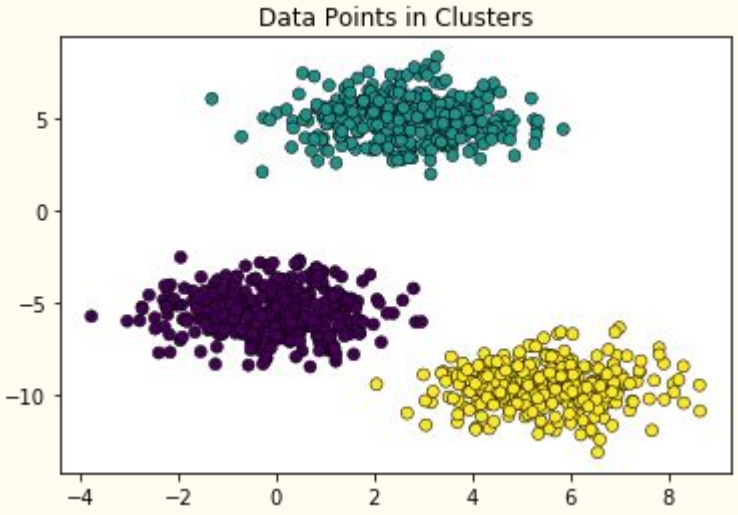
- Learn the underlying patterns in data in the absence of labels
- The goal of unsupervised learning algorithms is to analyze data and find important features.



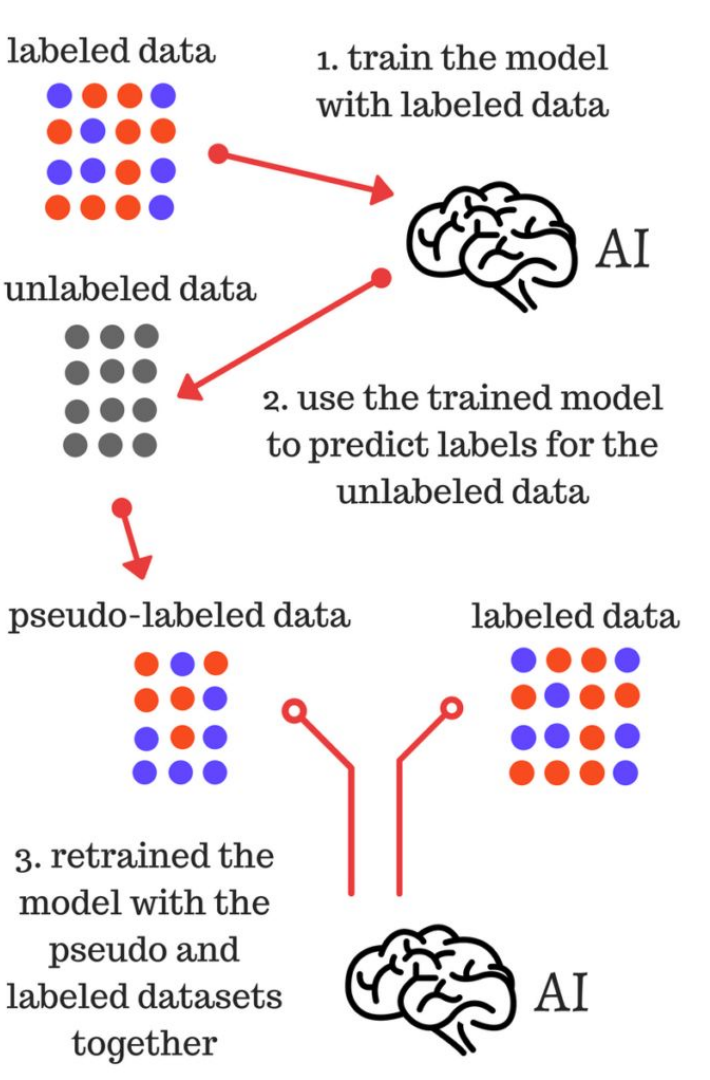
Clustering

- Clustering aims to discover “clusters”, or subgroups within unlabeled data.



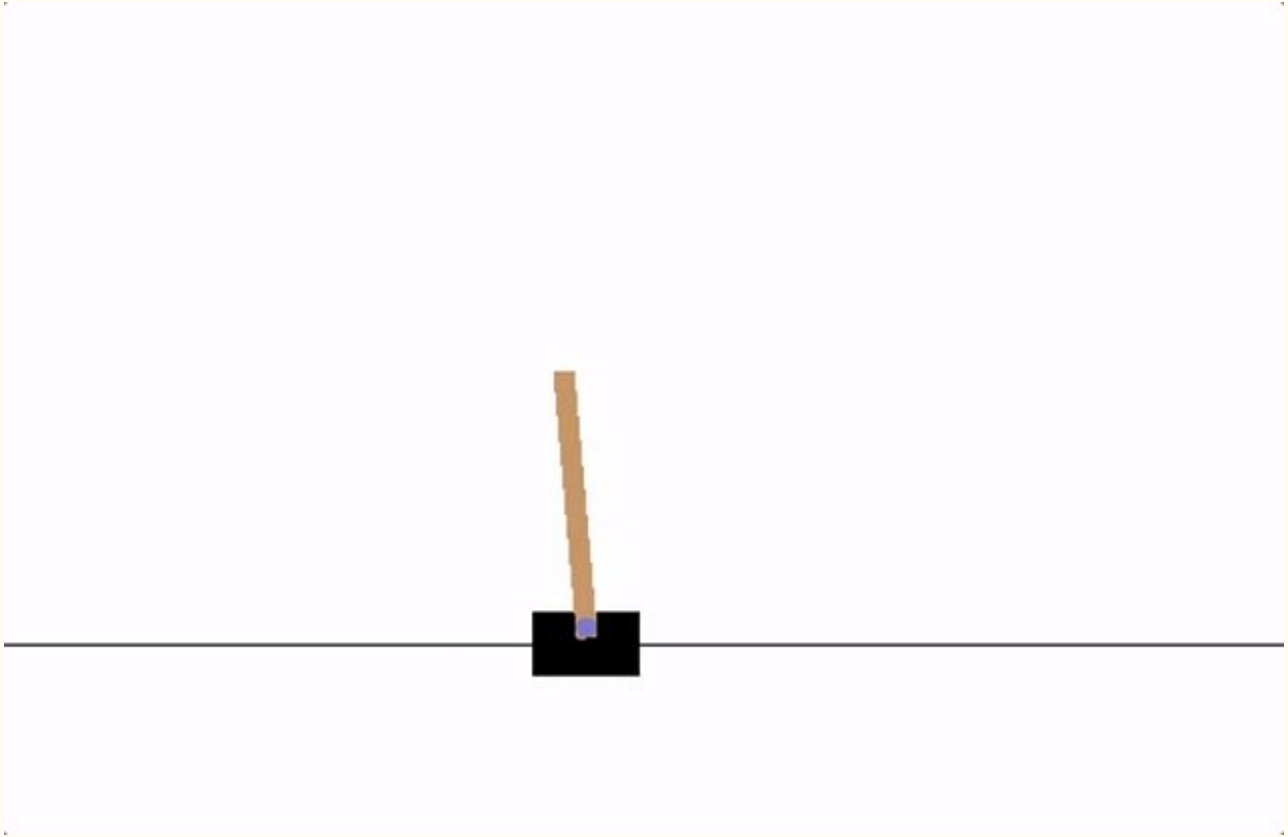


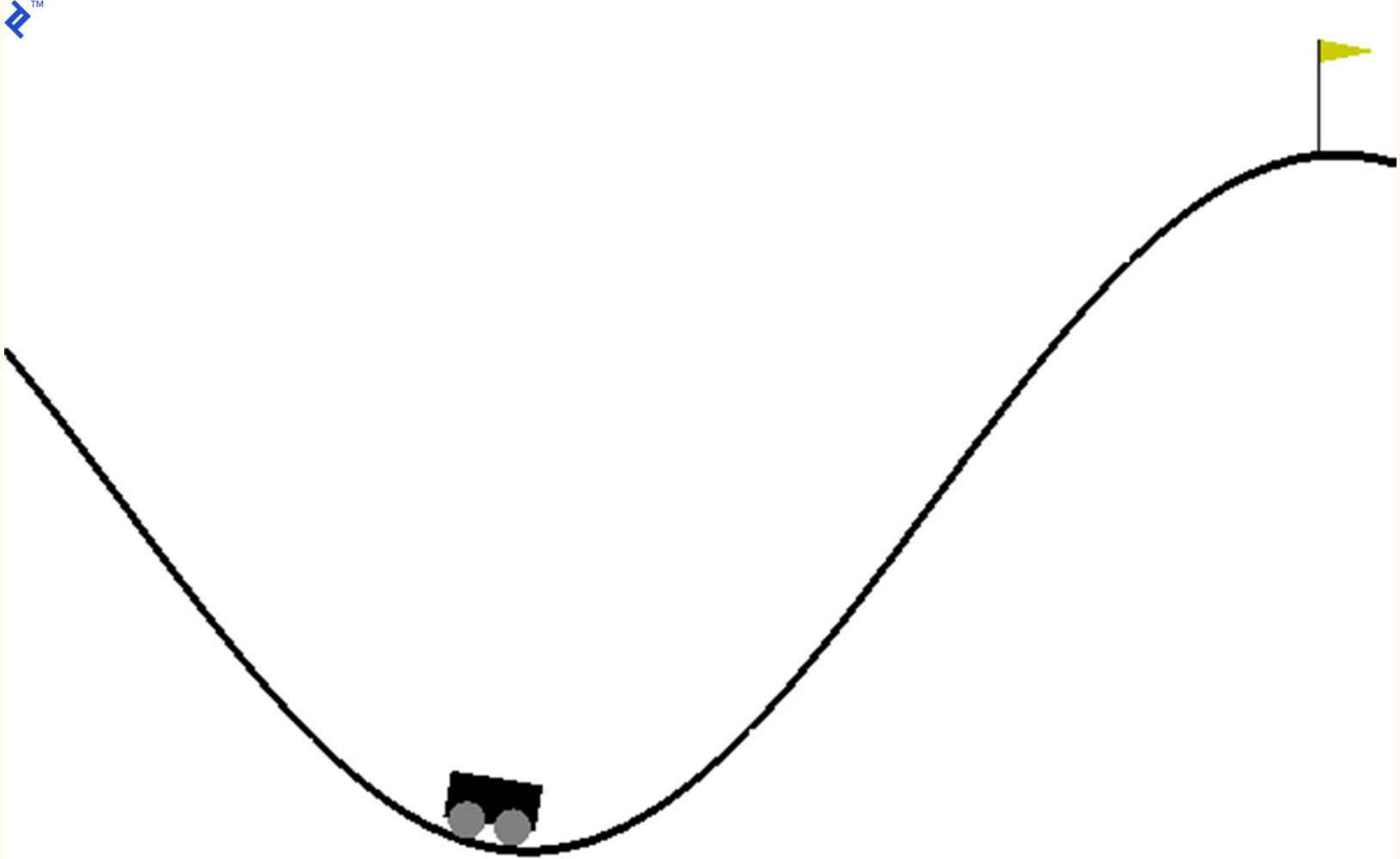
Semi-supervised Learning


labeled data
1. train the model with labeled data
AI
unlabeled data
2. use the trained model to predict labels for the unlabeled data
pseudo-labeled data
labeled data
3. retrained the model with the pseudo and labeled datasets together
AI

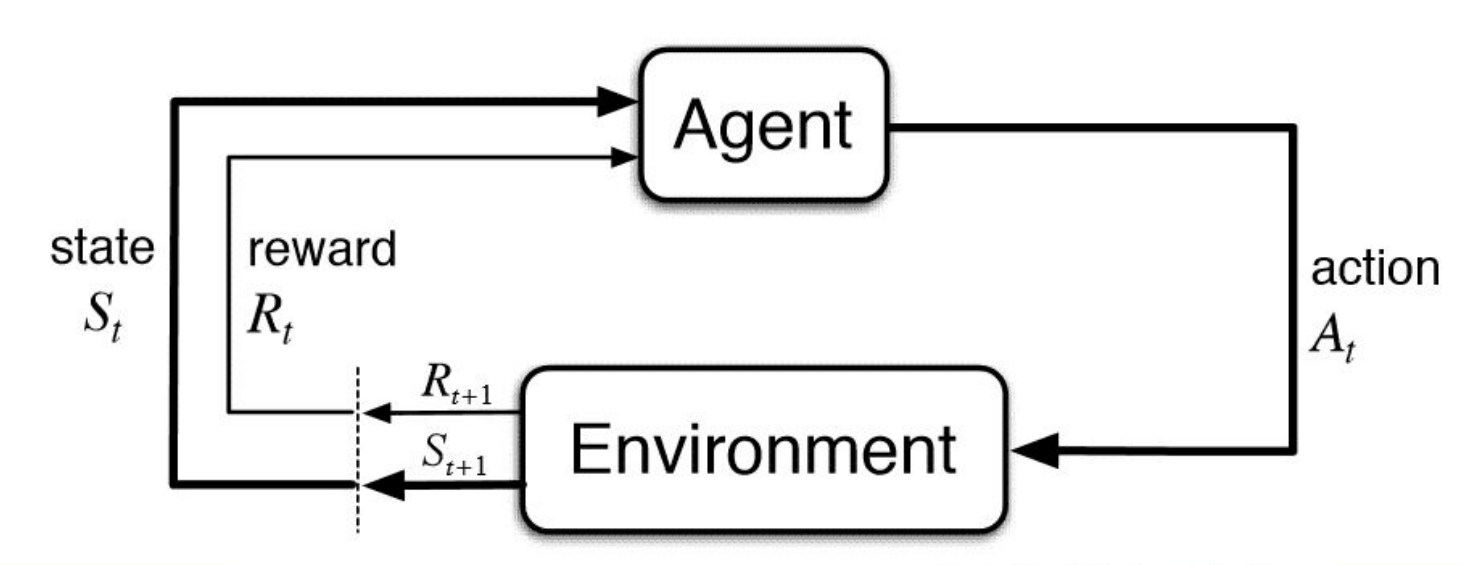
Reinforcement Learning

- Take action that maximizes future reward.





Source: OpenAI Gym


Agent
state
S_t
reward
R_t
action
A_t
R_{t+1}
S_{t+1}
Environment


agent
actions
rewards
observations

Examples

- Robotics
- Chess
- Go
- Dota2

Types of Data

Discrete vs. Continuous

- Discrete type of data is of categorical or ordinal type. For example, whether a component is faulty or is not faulty.
- Aka Quantitative

- Continuous (metric) of data represents amount or magnitude such as lines of source code.
- Aka Qualitative

Data unit	Numeric variable	= Quantitative data	Categorical variable	= Qualitative data
A person	"*How many* children do you have?"	**4** children	"In *which country* were your children born?"	**Australia**
	"*How much* do you earn?"	$**60,000** p.a.	"*What* is your occupation?"	**Photographer**
	"How many hours do you work?"	**38** hours per week	"Do you work *full-time or part-time*?"	**Full-time**
A house	"*How many* square metres is the house?"	**200** square metres	"In *which city or town* is the house located?"	**Brisbane**
A business	"*How many* workers are currently employed?"	**264** employees	"*What* is the *industry* of the business?"	**Retail**
A farm	"*How many* milk cows are located on the farm?	**36** cows	"*What* is the main *activity* of the farm?"	**Dairy**

Discrete data

- Nominal: No ranking between categories or classes.
- Ex: Dog, Cat

- Ordinal: The categories have ranking / ordering.
- Ex: High, Mid, Low