Introduction to machine learning in Hydrology

Lazaro J. Perez & Marc Berghouse

contact: lazaro.perez@dri.edu

Outline

- Introduction
- What is machine learning?
- Matlab

Introduction

- **Ice-breaker**

  Name

  Research topic

  What you hope to get out of the class

Class structure

- **Location**

  Computer lab in the CRVB @ DRI Campus

- **Plan**

  Tuesday: Lecture/Theory/Principles/Research examples

  Thursday: Coding/Work on example

- **Contact**

  Lazaro - lazaro.perez@dri.edu - Response time less than a day

  Office hours - Monday 10-14

  Marc -marc.berghouse@dri.edu - ?

Class structure

- **Tools**
  - Python (Jupiter notebooks)
    - Install Anaconda. If you need help work with Marc
    - Use Colab or Kaggle as workaround
  - Matlab (Regression Learner App // Code)
- **Grading**
  - You will grade yourself
  - Homework - Five models applied to specific cases - (60 %)
  - Project - Develop a predictive model to specific data - (35 %)
    - Project presentations start May 4th
  - Participation - 5%

Homework

- **Objectives**

  Assignments designed to reinforce the methods shown in class

  Develop a workflow useful for your research

  Learn by failure: Reproduce wrong predictions and detect error

  Investigate the performance of different models for single/multiple tasks

Plan

Date	Topic	Assignment
01/24	Intro & ML Basics (Laz)	
01/31	Applications of ML in Hydrology (Laz)	
02/02	Matlab & Random Forests (Laz)	H1
02/09	Python (Marc)	H2
02/14	K-Nearest Neighbor (Marc)	
02/21	Stochastic Gradient Descent (Marc)	H3
02/28	Support Vector Machine (Laz)	

ML Basics

- **What is machine learning?**

Humans learn from experience

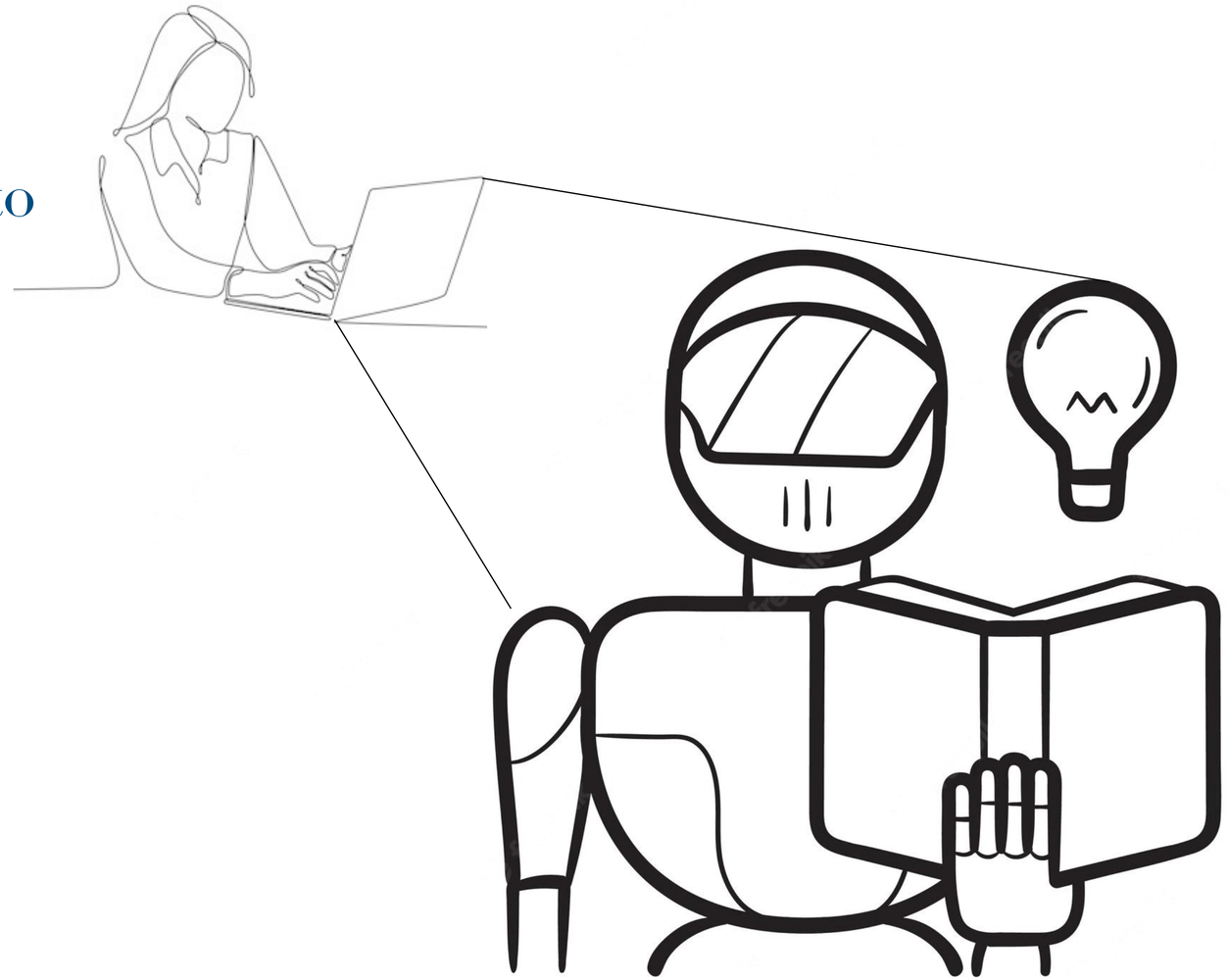
Machines follow orders

ML Basics

- **What is machine learning?**

What if humans can train machines to learn from past… data

Machine Learning!



ML Basics

- **What is machine learning?**

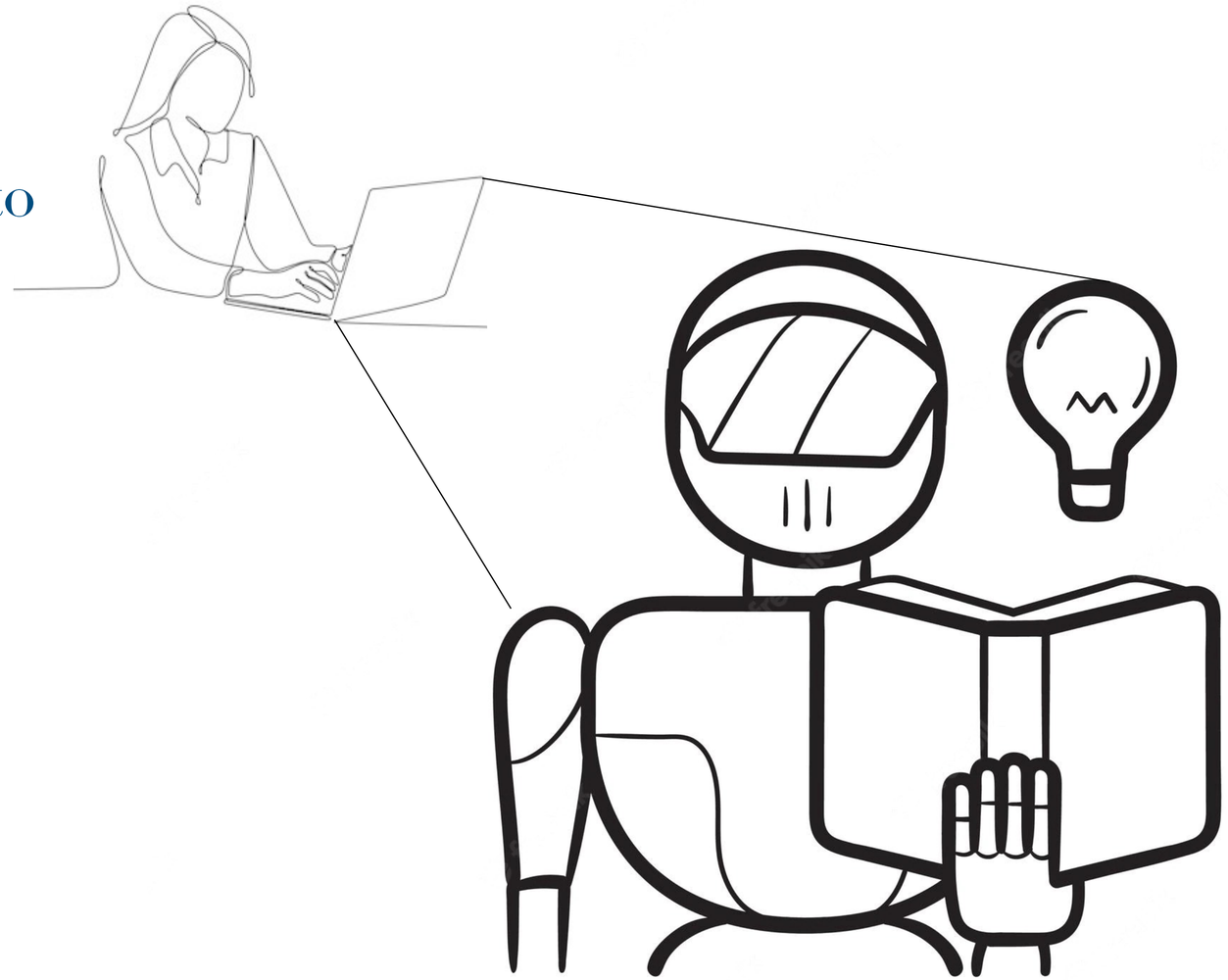
What if humans can train machines to learn from past… data

Machine Learning!

But it is not as simple as learning…

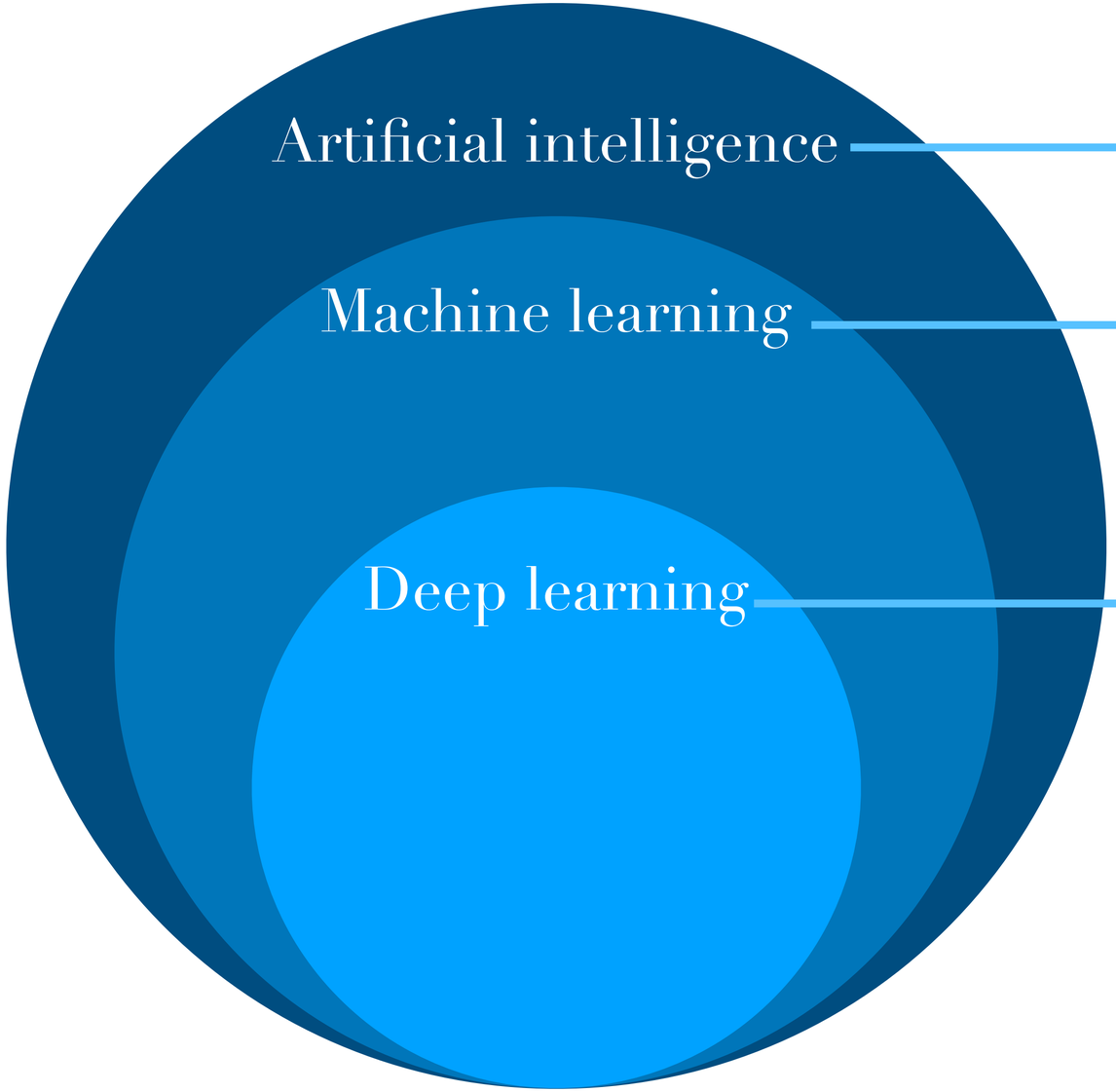
Understanding

Reasoning



ML Basics

- **What is machine learning?**



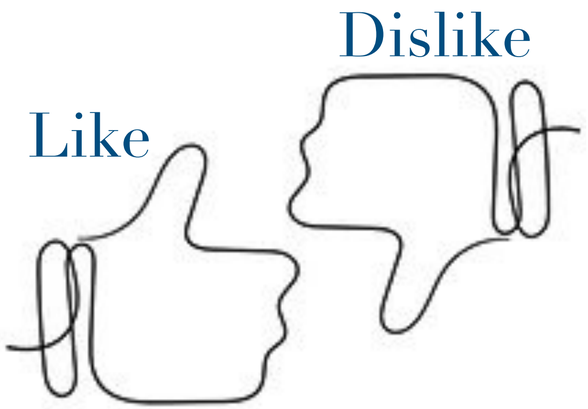
Technique that enables machines to mimic human behavior

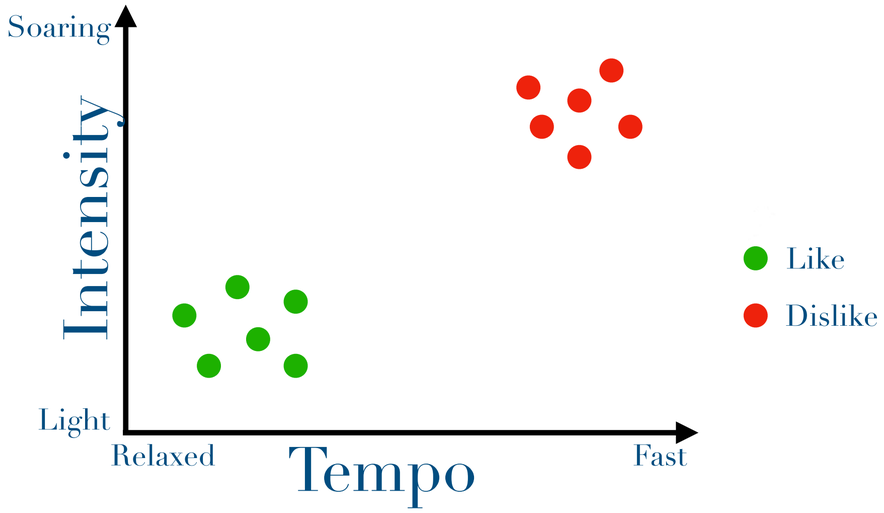
Subset of AI: Use statistical methods to enable machines to improve with experience

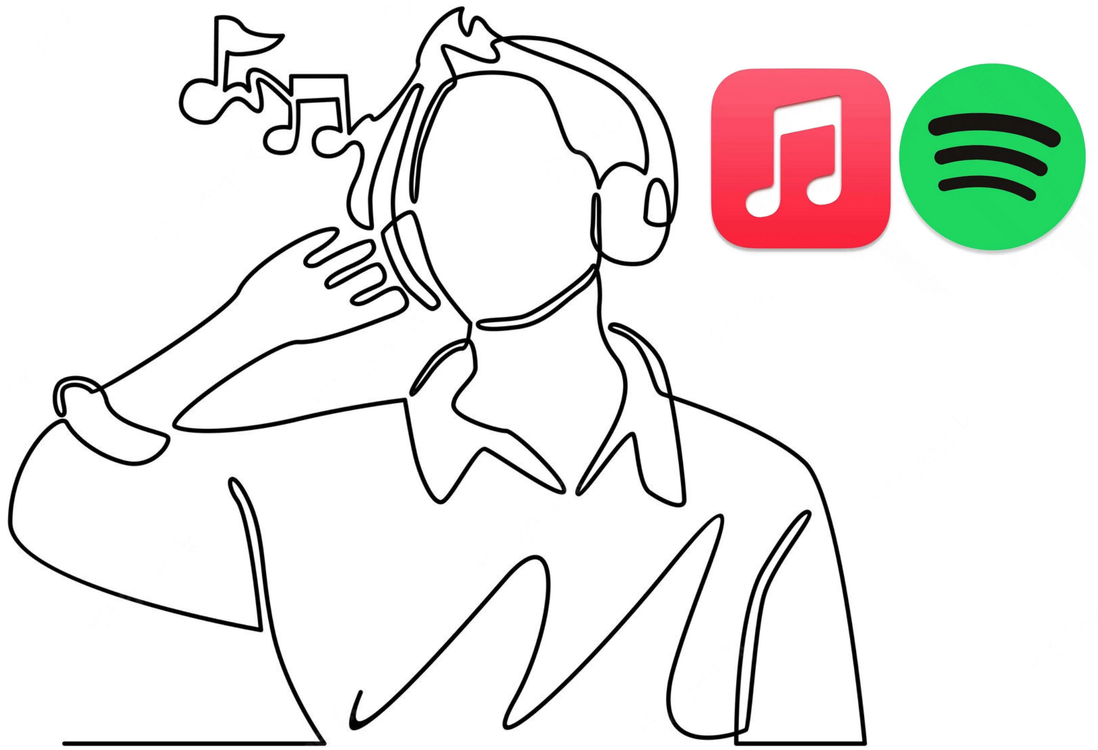
Subset of ML: Make the computation of multi-layer neural network feasible

ML Basics: Examples in daily life

Classification

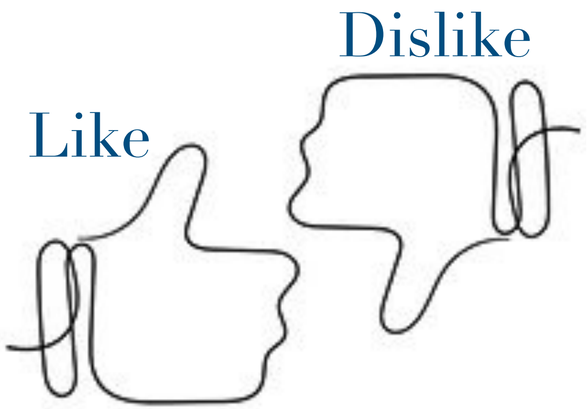




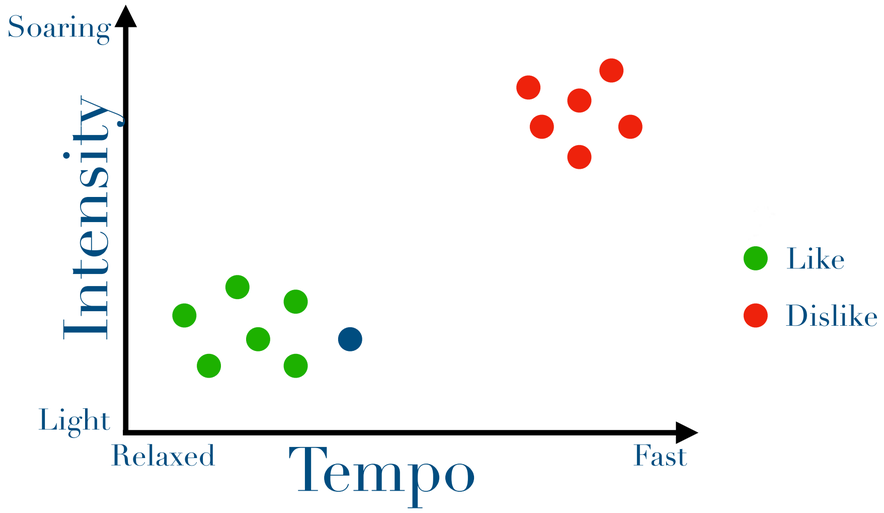


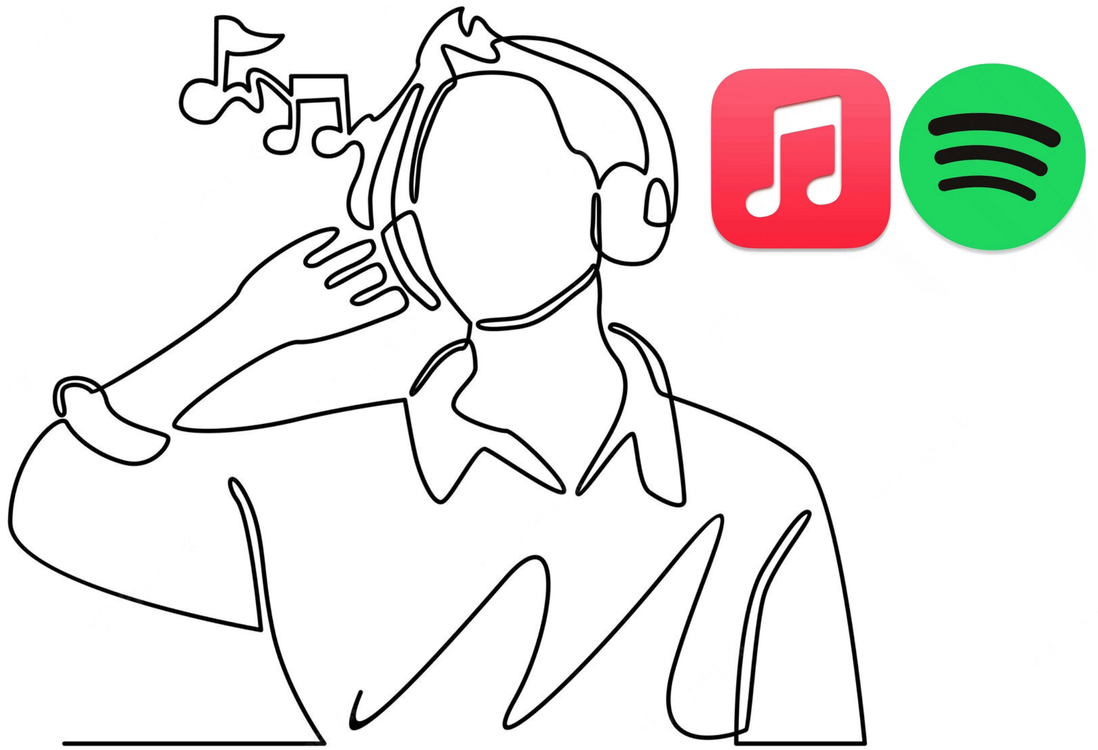
ML Basics: Examples in daily life

Classification



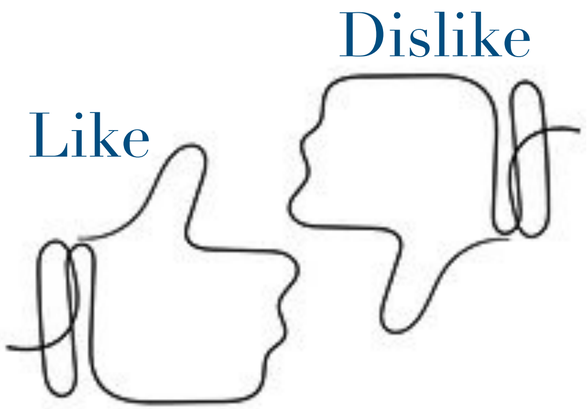
?

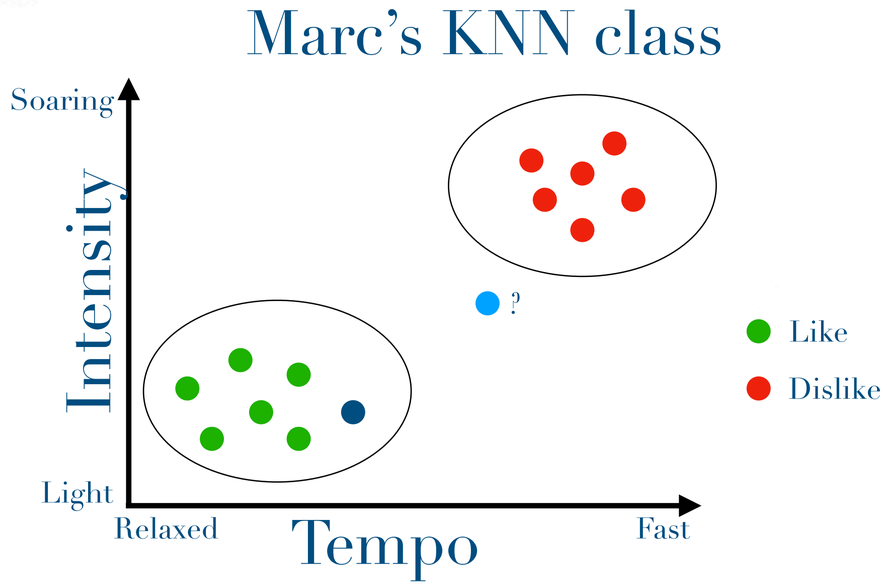


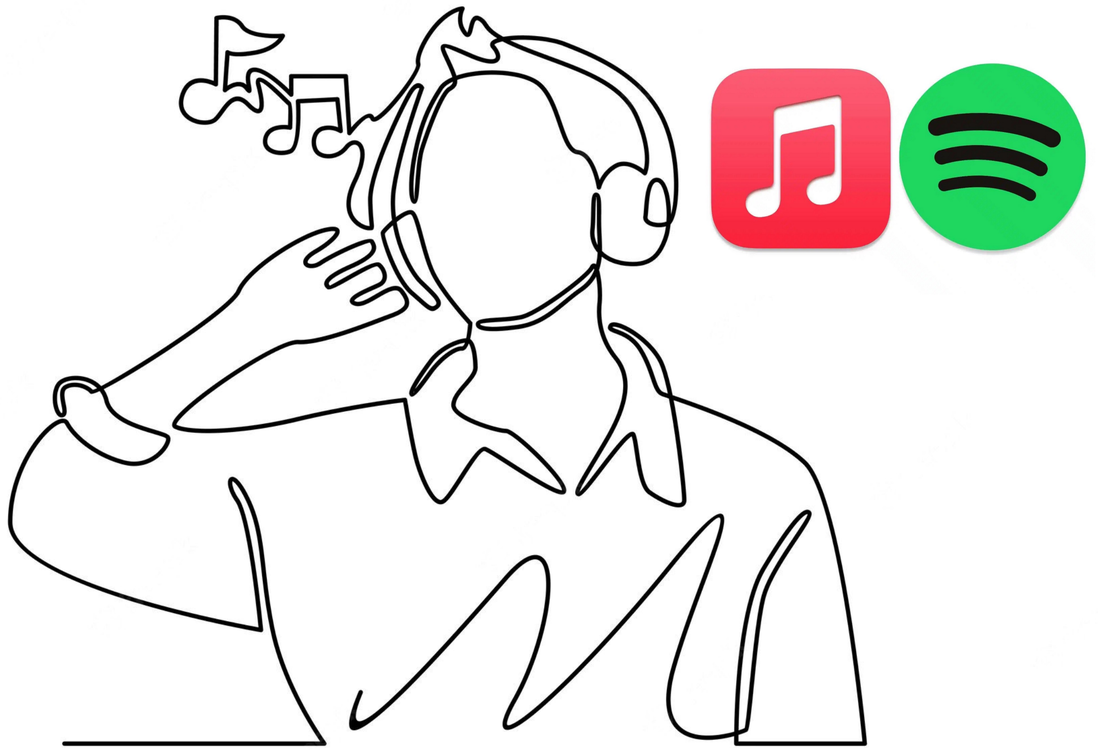


ML Basics: Examples in daily life

Classification

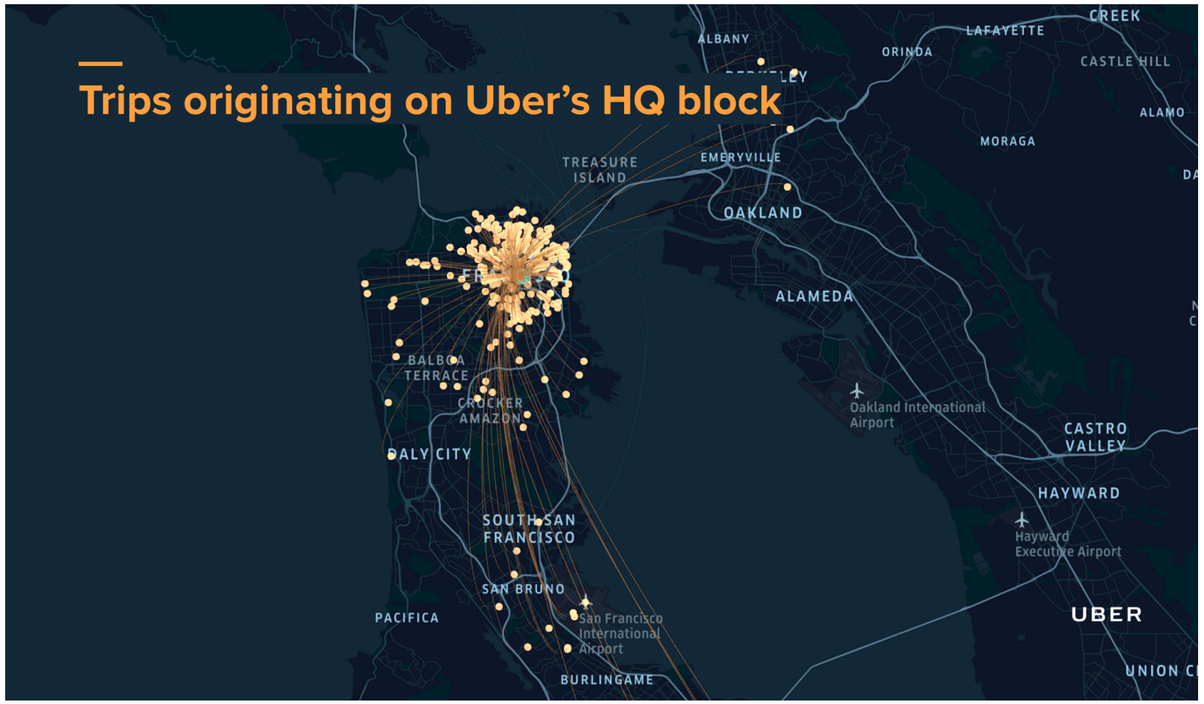


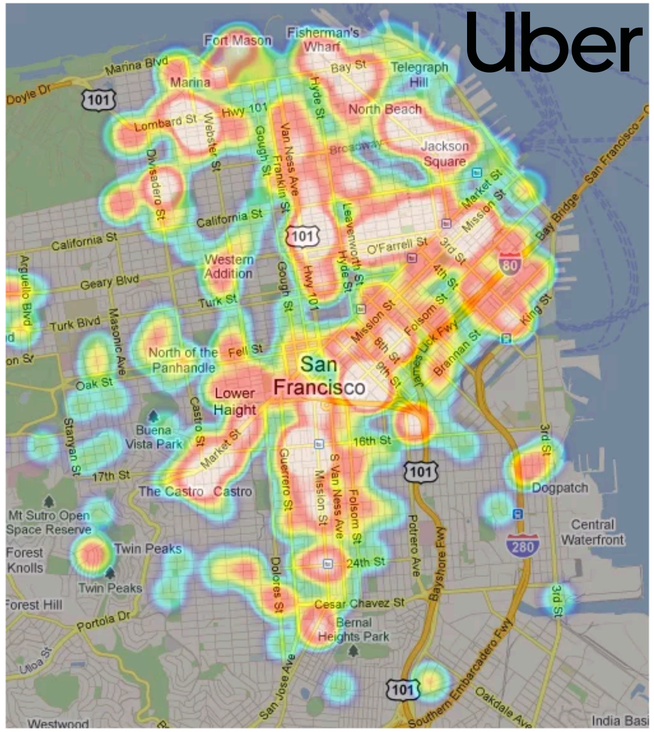




ML Basics: Examples in daily life

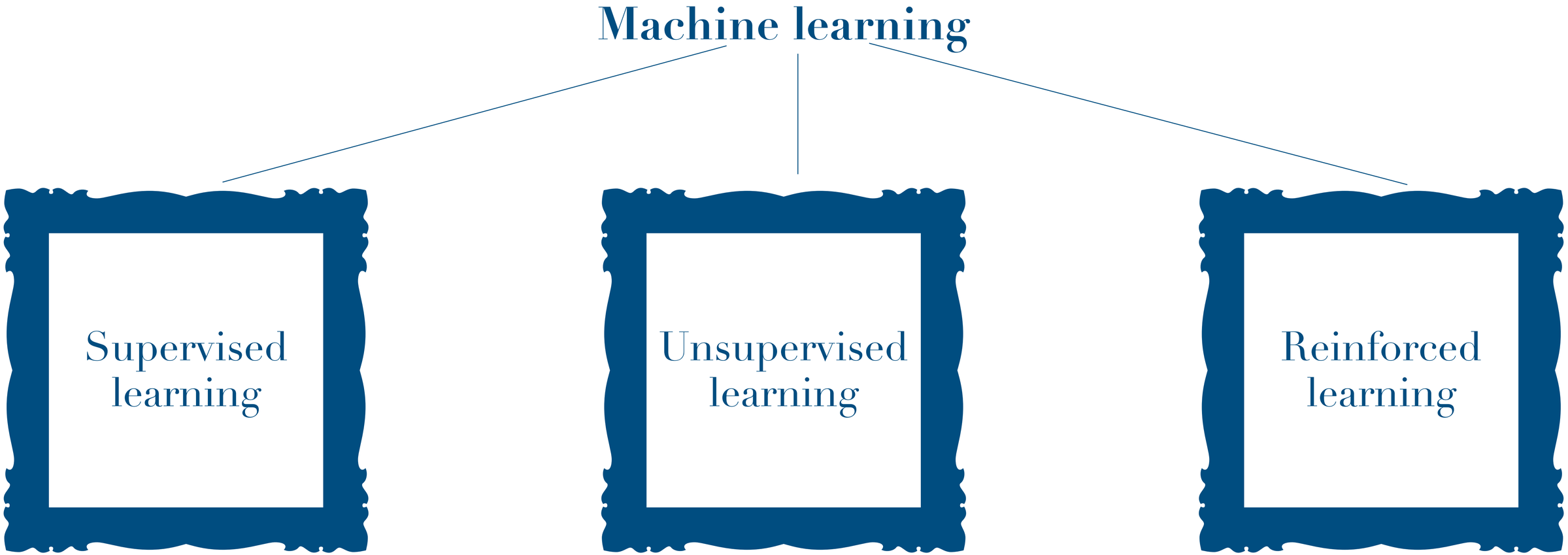
Regression





tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	RateCodeID	store_and_fwd_flag	dropoff_longitude	dropoff_latitude	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount
1/15/15 14:00	1/15/15 14:13	1	3	-73.964272	40.7730217	1	N	-73.965775	40.8046684	2	12	0	0.5	0	0	0.3	12.8
1/15/15 14:00	1/15/15 14:05	1	0.67	-73.970932	40.7959213	1	N	-73.970169	40.7891235	1	5	0	0.5	1	0	0.3	6.8
1/7/15 14:58	1/7/15 15:06	1	0.98	-73.948692	40.7777824	1	N	-73.955284	40.786869	1	7	0	0.5	1.4	0	0.3	9.2
1/7/15 14:58	1/7/15 15:12	3	4.39	-73.988663	40.7227058	1	N	-73.987221	40.6944084	2	15.5	0	0.5	0	0	0.3	16.3

ML Basics: How do machines learn?



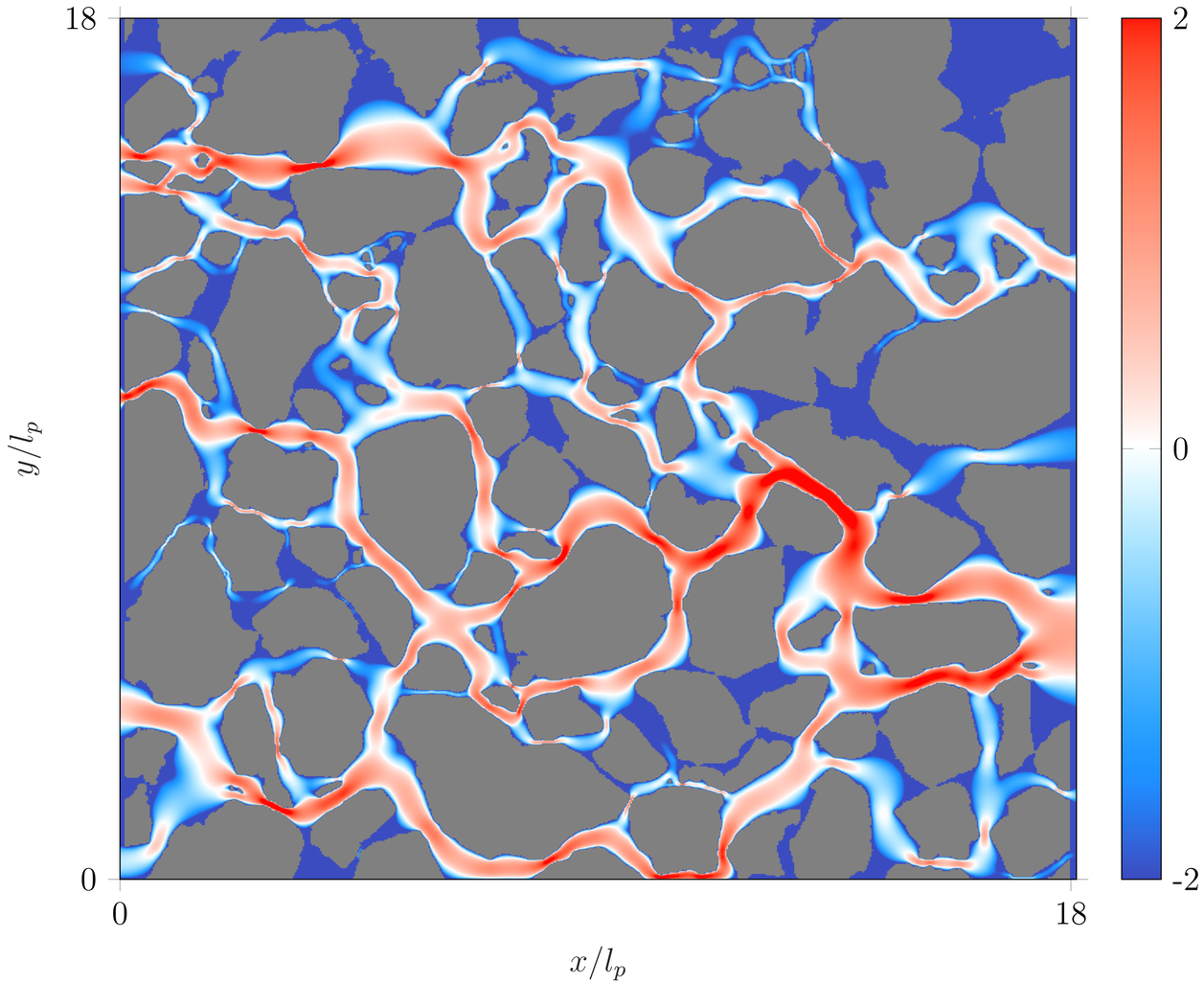
ML Basics: How do machines learn?

Machine learning

Reinforced learning

ML Basics: How do machines learn?

Machine learning



Reinforced learning

Find the optimal flow rate to displace contaminant

ML Basics: How do machines learn?

Machine learning

Unsupervised learning

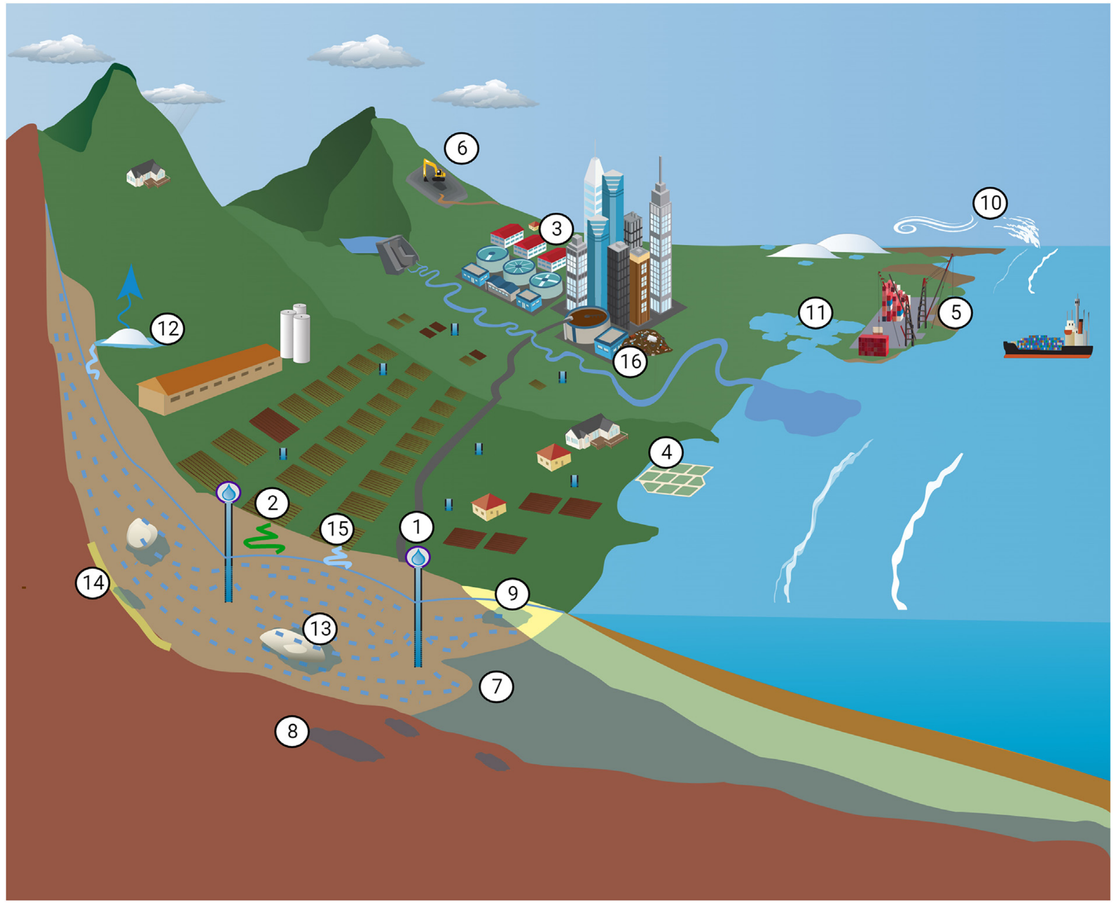
Customer	N items	Food	Clothes
Gata	2	1	0
Alejandra	5	1	4
Maria	4	2	2
Elvira	2	1	1

- Clustering
- Association
- Dimension reduction

ML Basics: How do machines learn?

Machine learning

Unsupervised learning



Narvaez et al., 2023

Seawater intrusion

Association: between well pumping and snow melt

Clustering: Water treatment plant and city consumption
vs.
Water discharge from streams

ML Basics: How do machines learn?

Machine learning

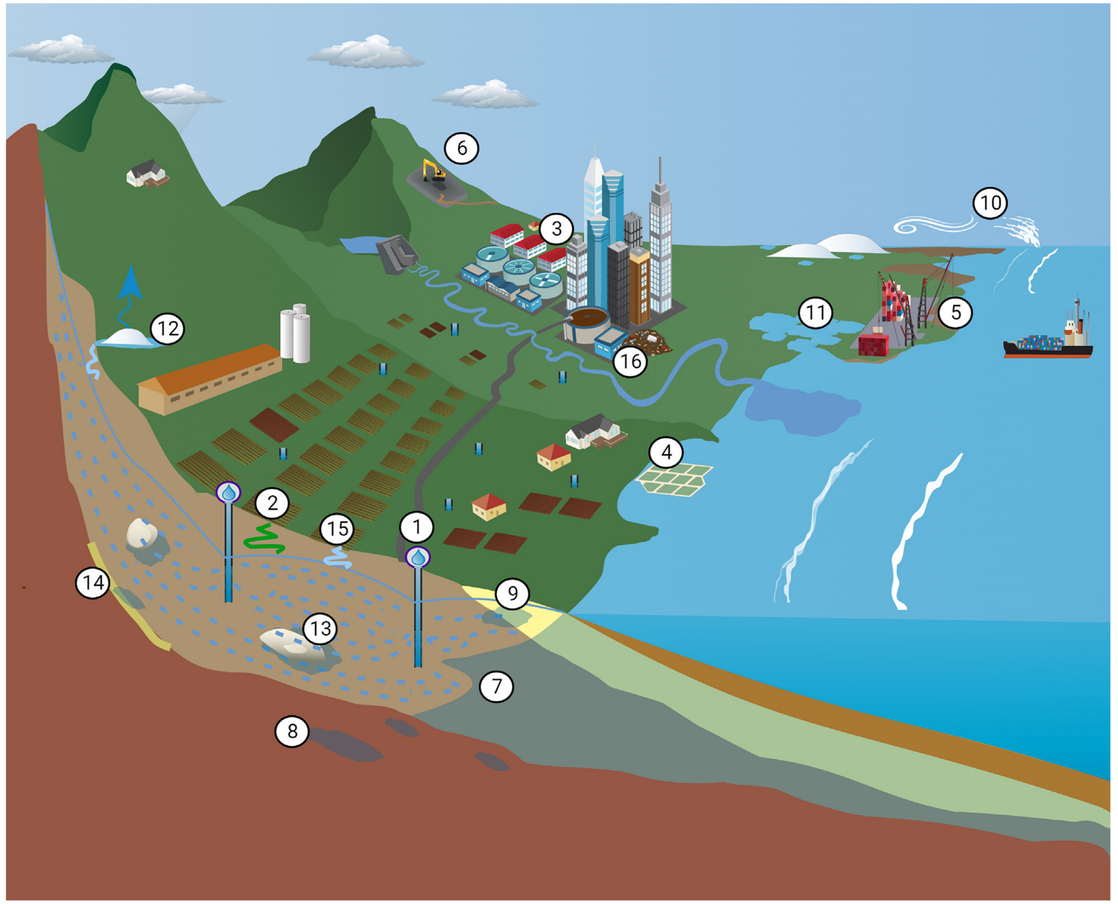
Supervised learning

Value	Weight (g)	Diameter (mm)
0.01	2.3	16.25
0.02	3.05	18.75
0.05	3.9	21.25
0.1	4.1	19.75
0.2	5.75	22.25
0.5	7.8	24.25
1	7.5	23.25
2	8.5	25.75

ML Basics: How do machines learn?

Machine learning

Supervised learning



Narvaez et al., 2023

Seawater intrusion

Classify: salinity increase from seawater intrusion & salinity increase from salt dissolution

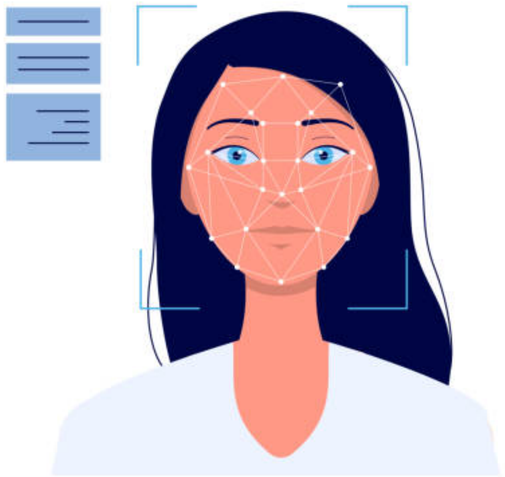
Regression: Predict groundwater levels at specific location

Quiz: Supervised vs. Unsupervised vs. Reinforced?

Daily life

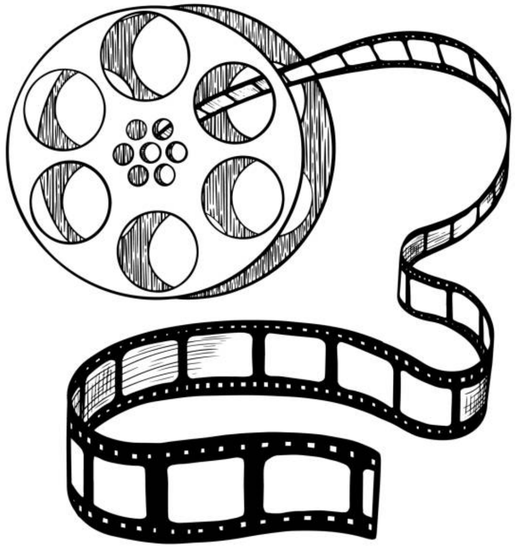
Scenario 1: Facebook

Face recognition



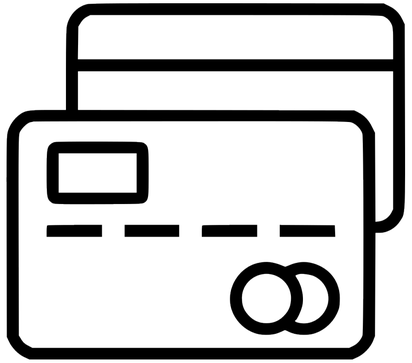
Scenario 2: Netflix

Movie recommendation



Scenario 3: Fraud detection

Credit card transactions

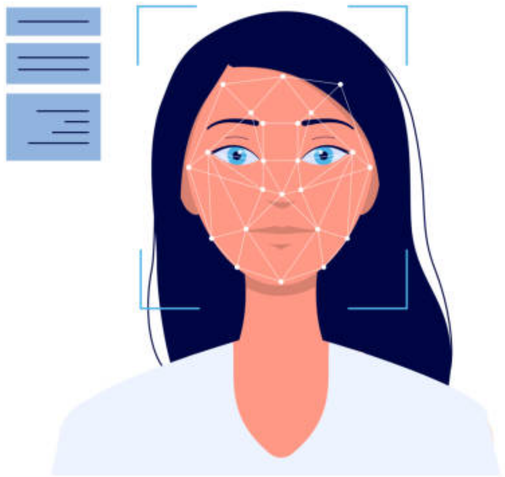


Quiz: Supervised vs. Unsupervised vs. Reinforced?

Daily life

Scenario 1: Facebook

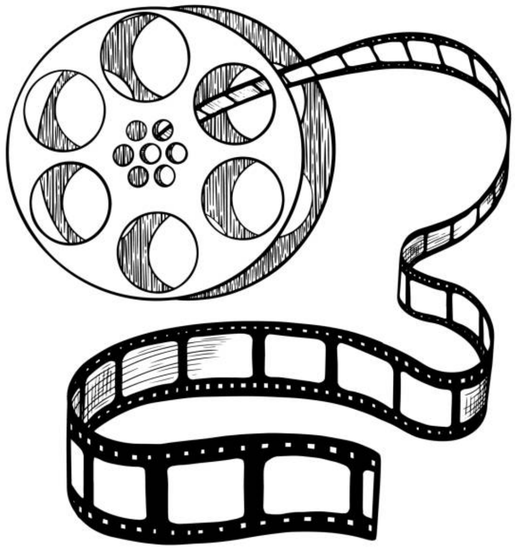
Face recognition



It is supervised learning!
Here Facebook is using tagged photos to recognize the person.
Therefore, the tagged photos become the labels of the pictures.

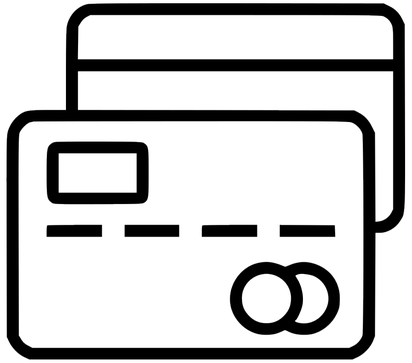
Scenario 2: Netflix

Movie recommendation



Scenario 3: Fraud detection

Credit card transactions

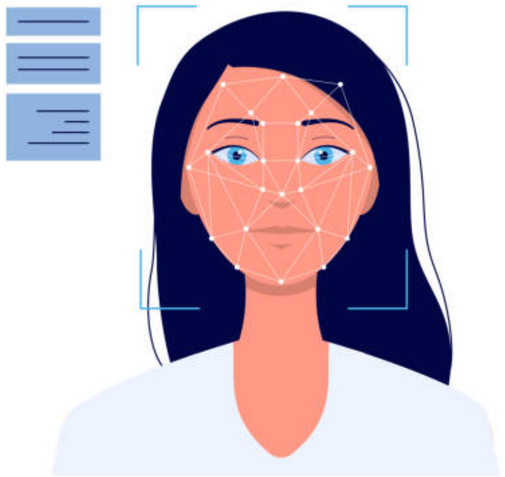


Quiz: Supervised vs. Unsupervised vs. Reinforced?

Daily life

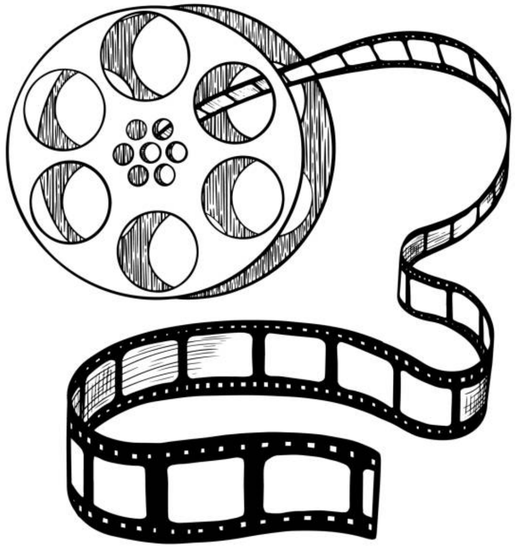
Scenario 1: Facebook

Face recognition



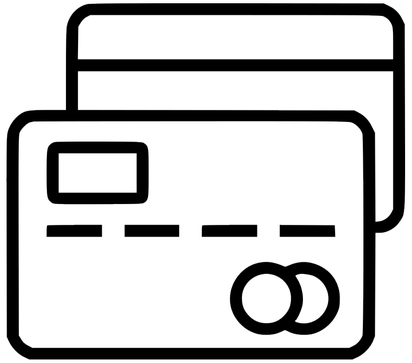
Scenario 2: Netflix

Movie recommendation



Scenario 3: Fraud detection

Credit card transactions



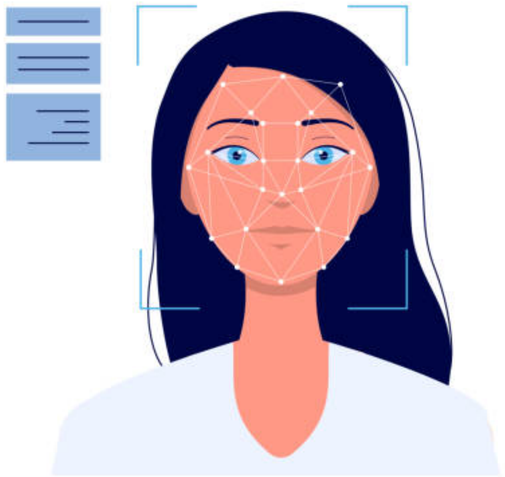
It is supervised learning!
The model trains a classifier on pre-existing labels (genres of movies).
Evaluate the features based on your likes/dislikes and then recommend new movies based on similar features.

Quiz: Supervised vs. Unsupervised vs. Reinforced?

Daily life

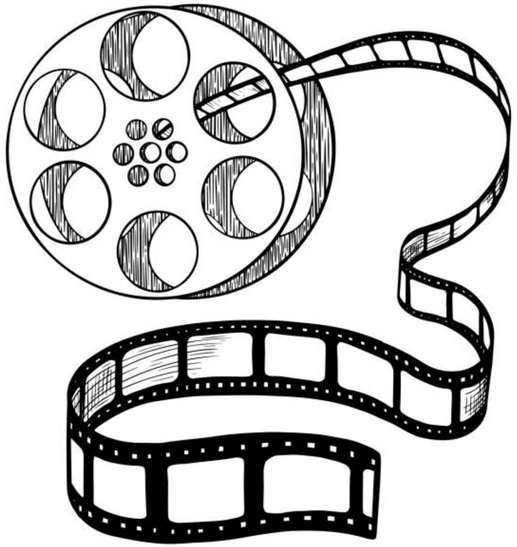
Scenario 1: Facebook

Face recognition



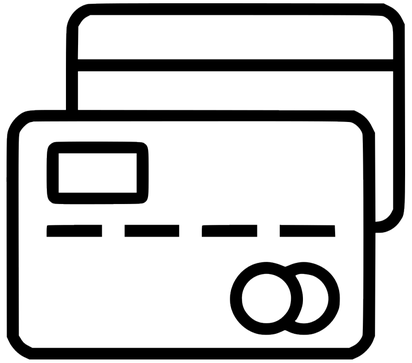
Scenario 2: Netflix

Movie recommendation



Scenario 3: Fraud detection

Credit card transactions



It is unsupervised learning!
Suspicious transactions are not defined, hence there are no labels of "fraud" and "not fraud". The model tries to identify outliers by looking at anomalous transactions and flags them as 'fraud'.

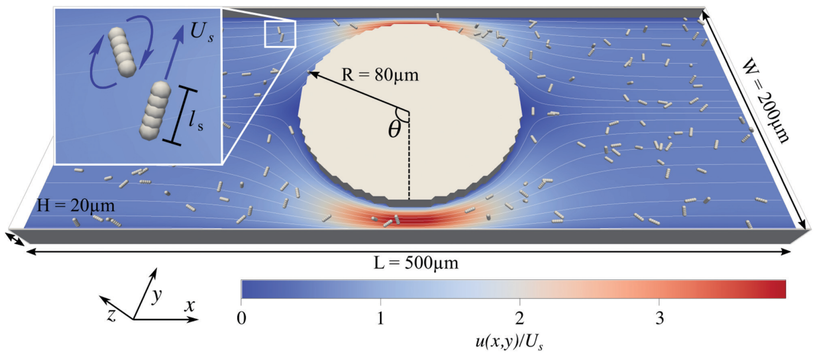
Quiz: Supervised vs. Unsupervised vs. Reinforced?

Hydrology

Scenario 1: Bioremediation

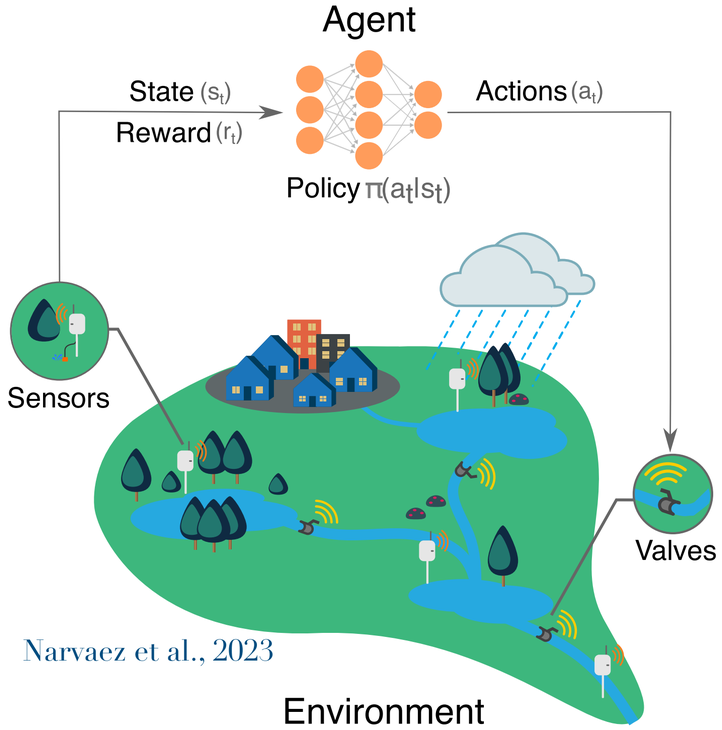
Determine best species

Leat et al., 2020



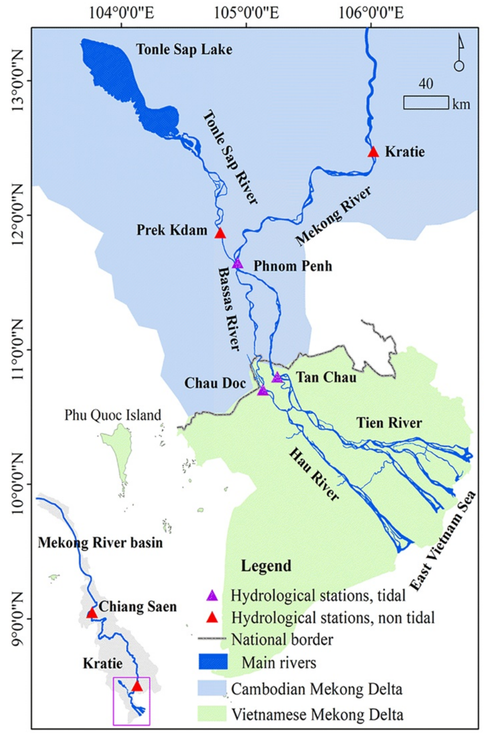
Scenario 2: Storms

Systems adapt to storms



Narvaez et al., 2023

Scenario 3: River discharge



Than et al., 2021

SB LAB

DRI