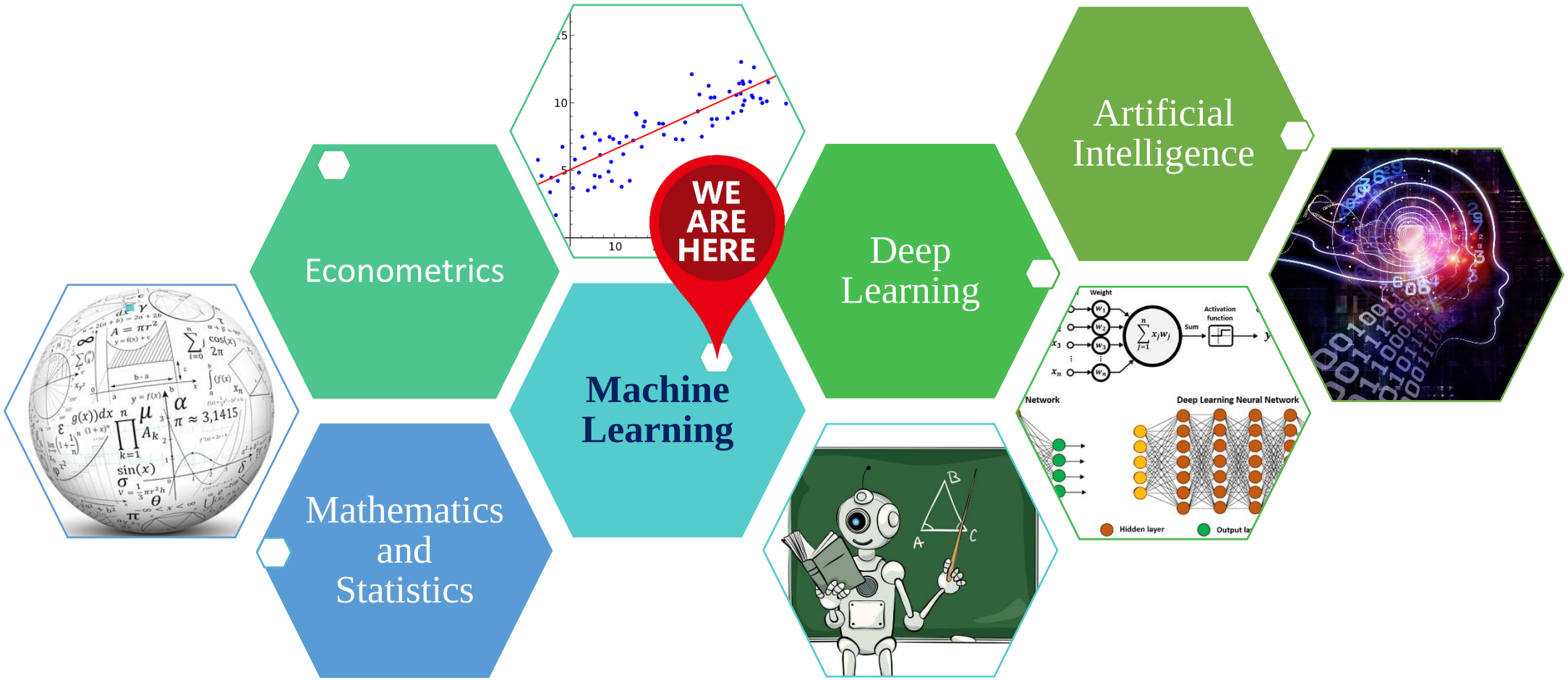
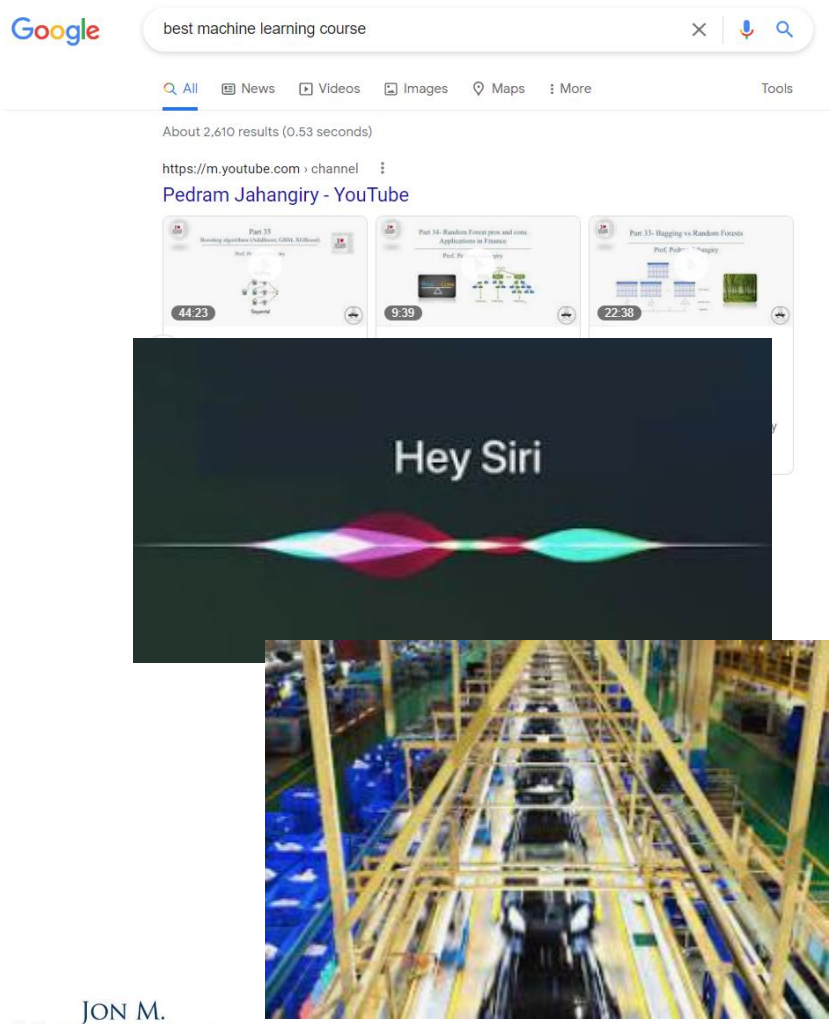
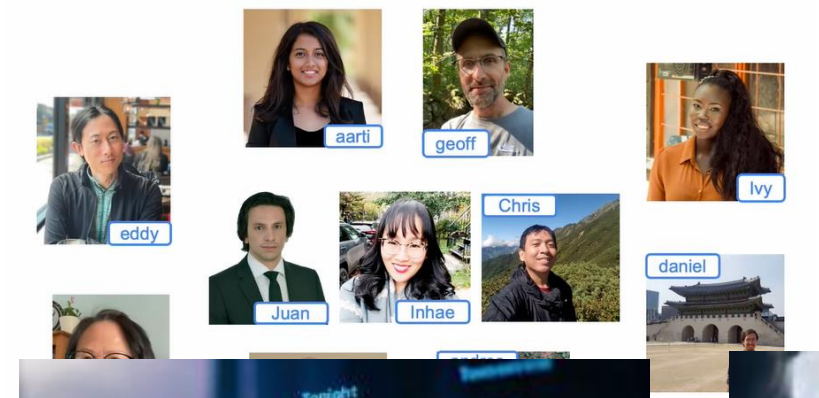
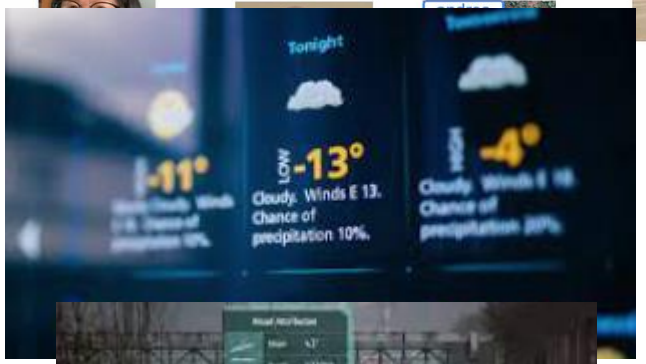
Class 2- What is Machine Learning?



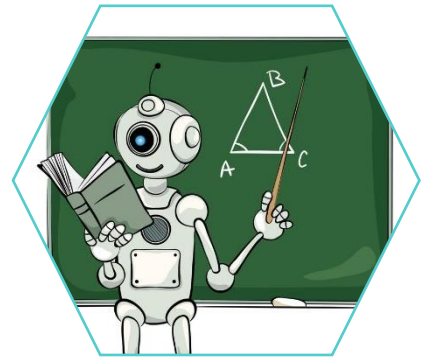
Machine Learning is everywhere!

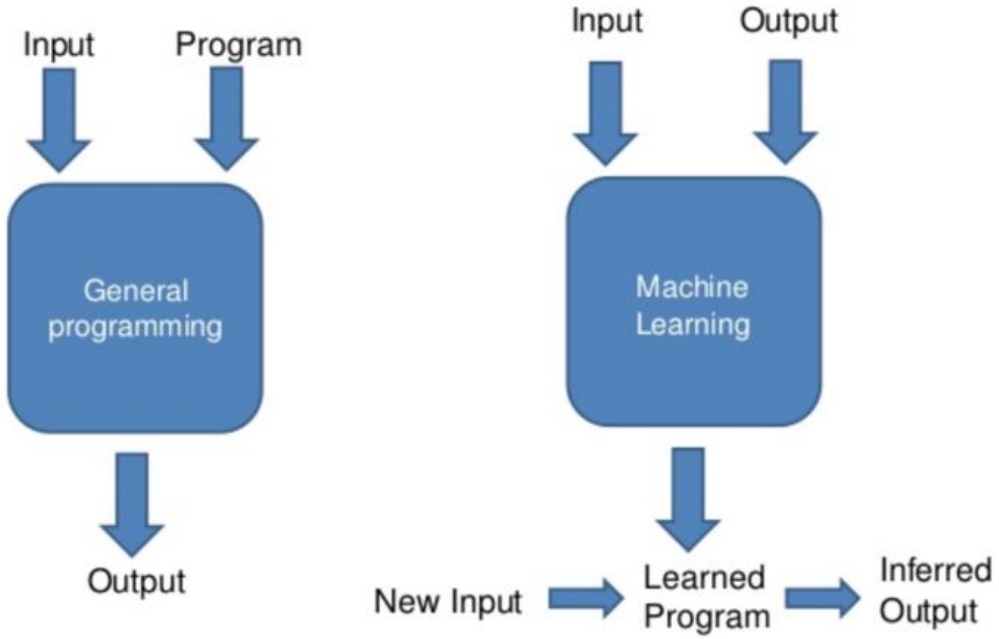






General programming vs Machine learning





- Machine Learning: Involves automated detection of meaningful patterns in data and apply the pattern

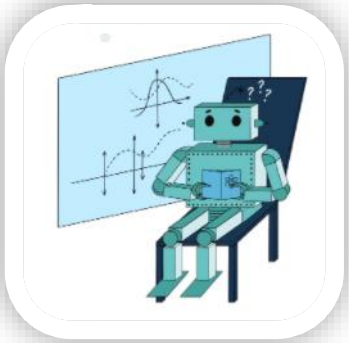
What is Machine Learning?

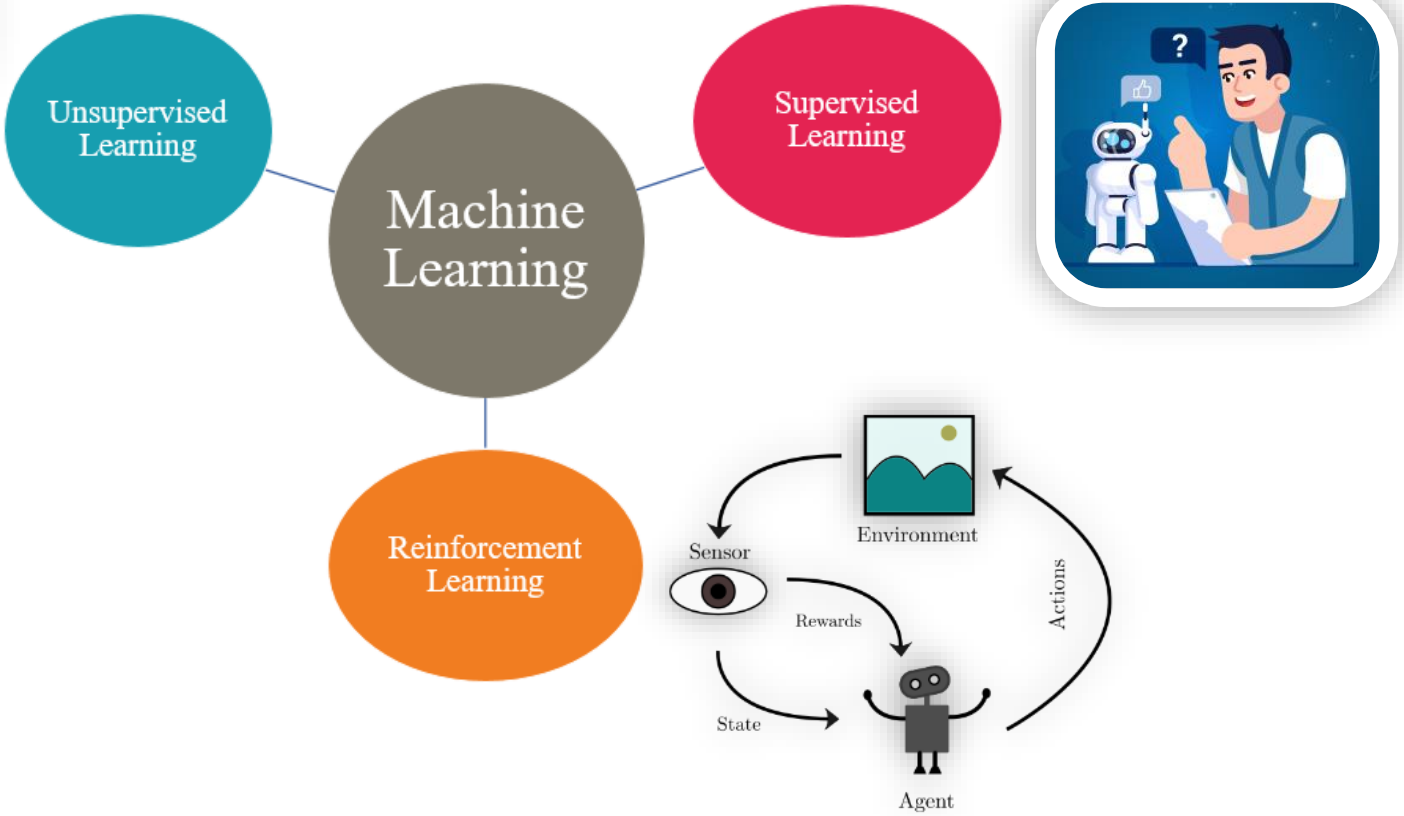
"A ML algorithm learns complex patterns in a high dimensional space without being specifically directed." *Marcos Lopez De Prado: Advances in Financial Machine Learning* (2018, p.15)

Let's break this statement into its components:

- ✓ **learns complex patterns**: The ML algorithm may find a pattern that cannot be easily represented with a finite set of equations.
- ✓ **learns … in a high-dimensional space**: Solutions often involve many variables and the interactions between them.
- ✓ **learns … without being specifically directed**: Unlike with other empirical tools, researchers do not impose a particular structure on the data. Instead, researchers let the data speak.

Types of Machine Learning





Supervised Learning

- In **supervised learning**, computers learn to model relationships based on train data. In supervised learning, inputs and outputs are labeled (we know the right answers) for the algorithm. After learning the pattern, the trained algorithms are used to predict outcomes for test data.

- **Regression:**

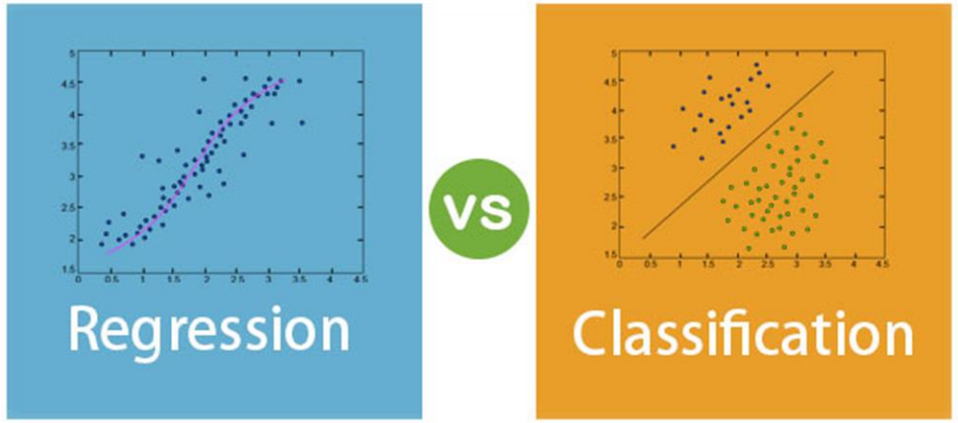
1. Predicting housing price
2. Predicting stock market returns
3. Predicting revenue growth

….

- **Classification:**

1. Generating buy, sell, hold signals.
2. Estimating the likelihood of a successful M&A or IPO
3. Predicting credit default rate.

…



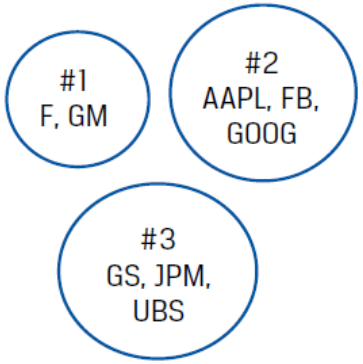
Class exercise

- Can you think of an example of Regression? Classification?

Unsupervised Learning

- In **unsupervised learning**, computers are trained on unlabeled train data without any guidance. The goal is to discover the underlying patterns and find groups of samples that behave similarly. Find something interesting!

- **Clustering:** group similar data points together
  - ✓ Grouping companies into peer groups based on some non-standard characteristics like financial statement data or co-movement similarities rather than sectors or countries.
  - ✓ Client profiling and asset allocation



- **Anomaly detection:** Find anomalies (unusual data points)
  - ✓ Fraud detection

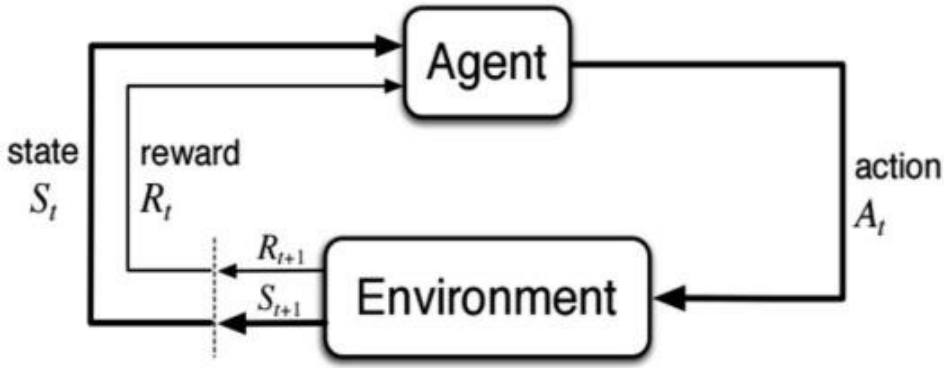
- **Dimensionality Reduction:** compress data in lower dimension
  - ✓ Identify the most predictive factors underlying asset price movements (to avoid factor zoo)

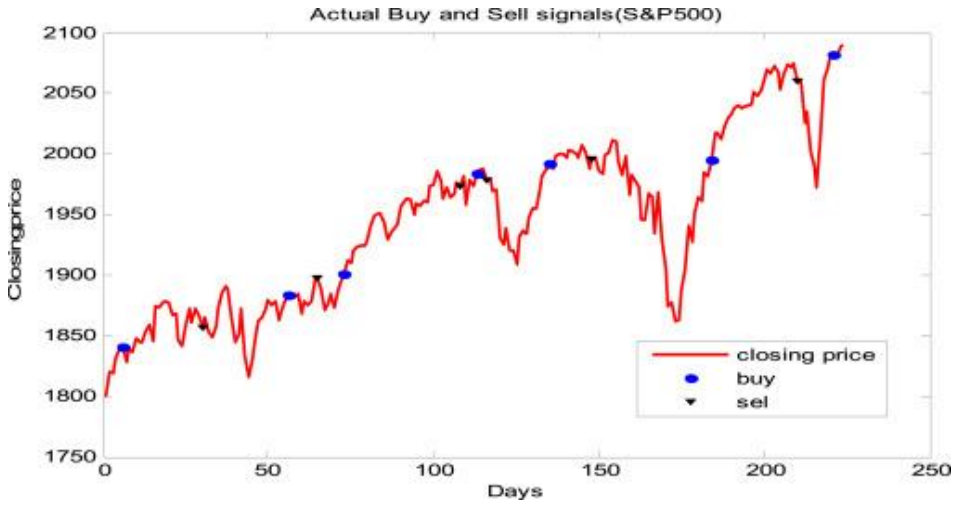
Class exercise

- Can you think of an example of Clustering?

Reinforcement Learning

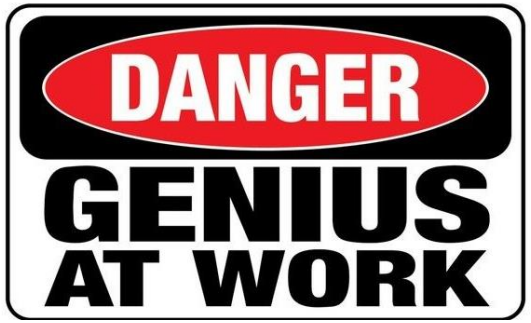
- In **reinforcement learning**, a computer (agent) learns from interacting with its environment by producing actions and discovering rewards. You need to define the environment, actions and the reward system. The machine will then explore and exploit to maximize the reward. The new actions may not be immediately optimal. The learning subsequently occurs through millions of trials and errors.





- Example: a virtual trader (agent) who follows certain trading rules (the actions) in a specific market (the environment) to maximize its profits (its reward).

How are you doing so far?



Road map

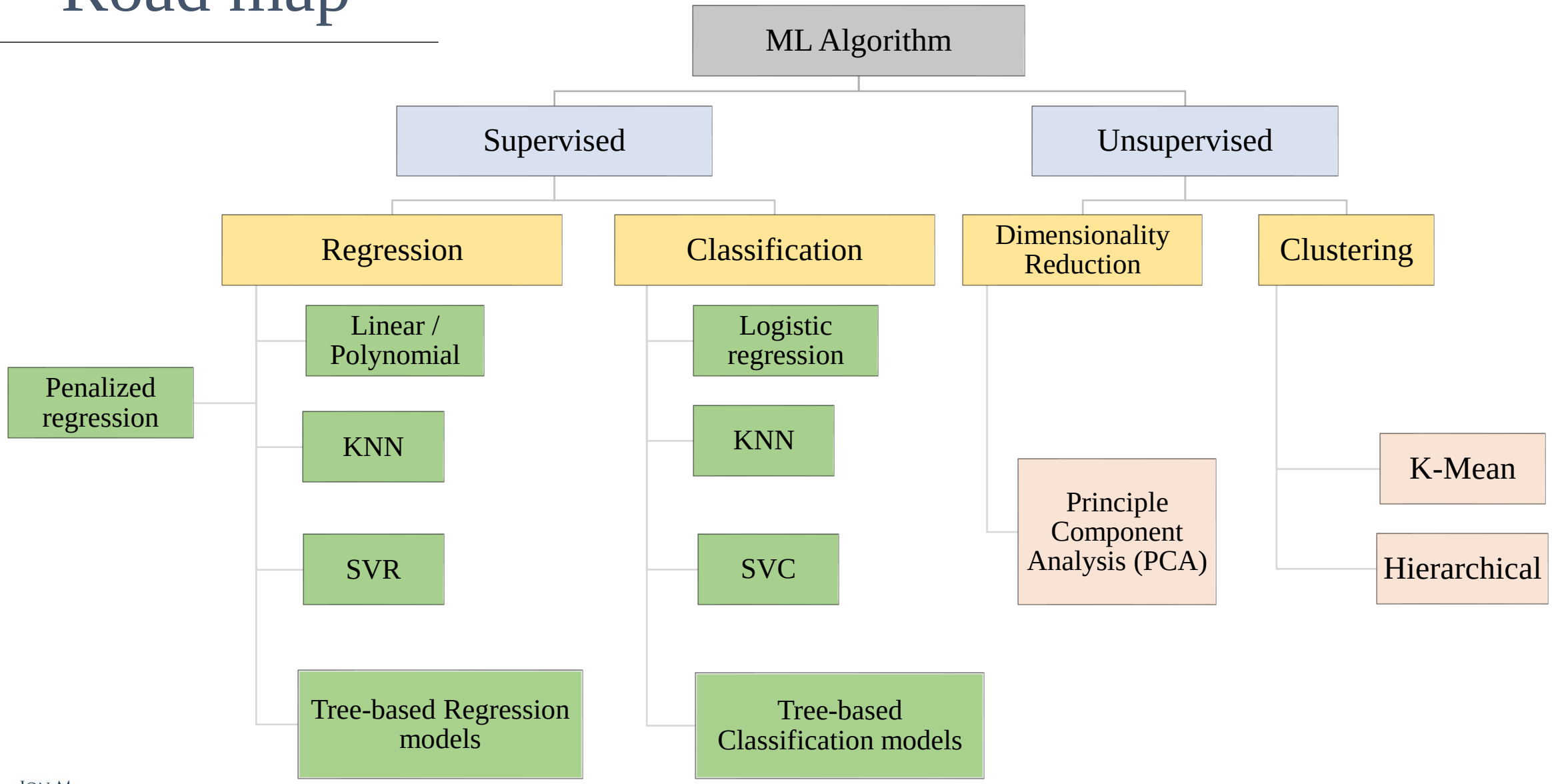
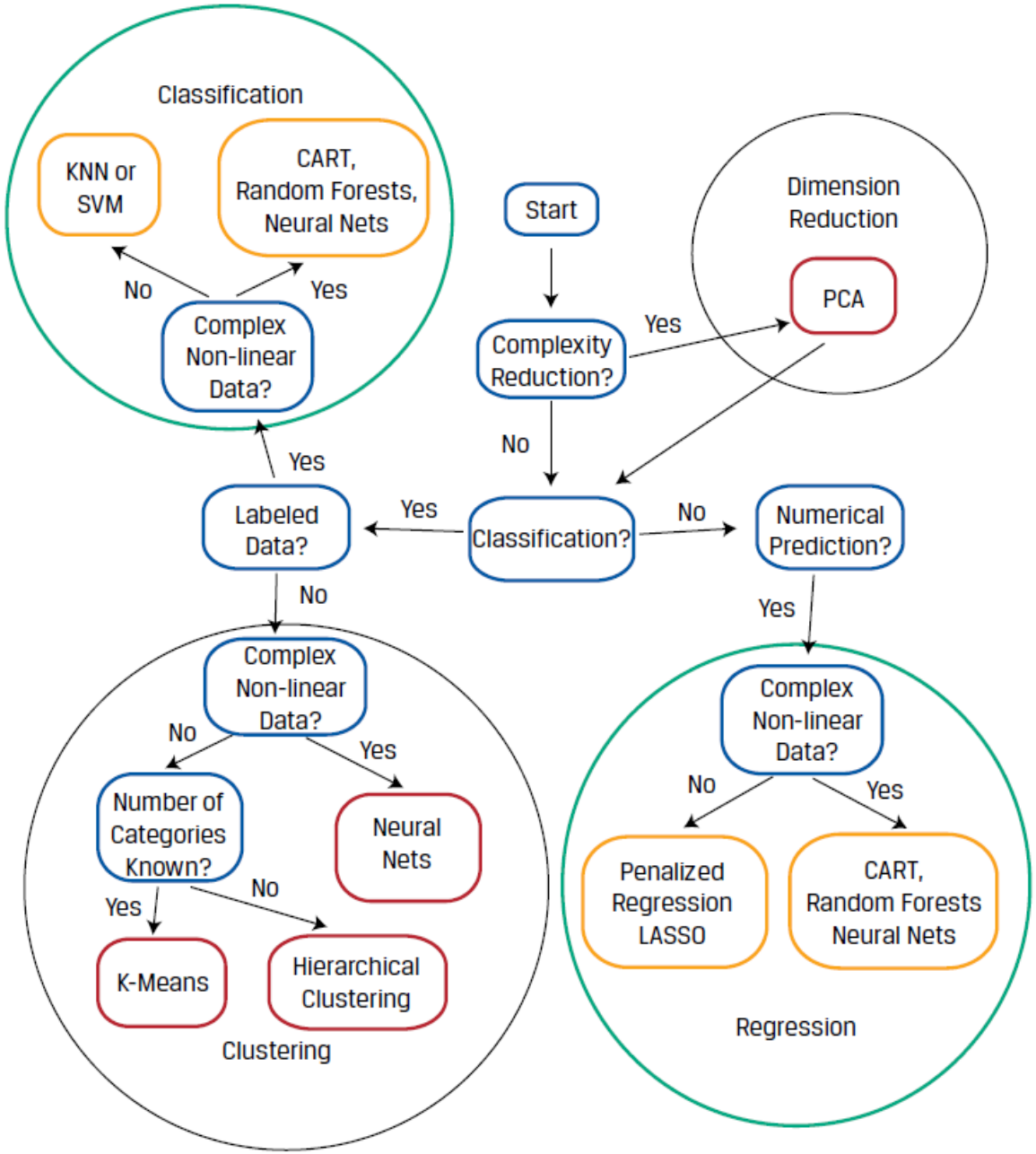
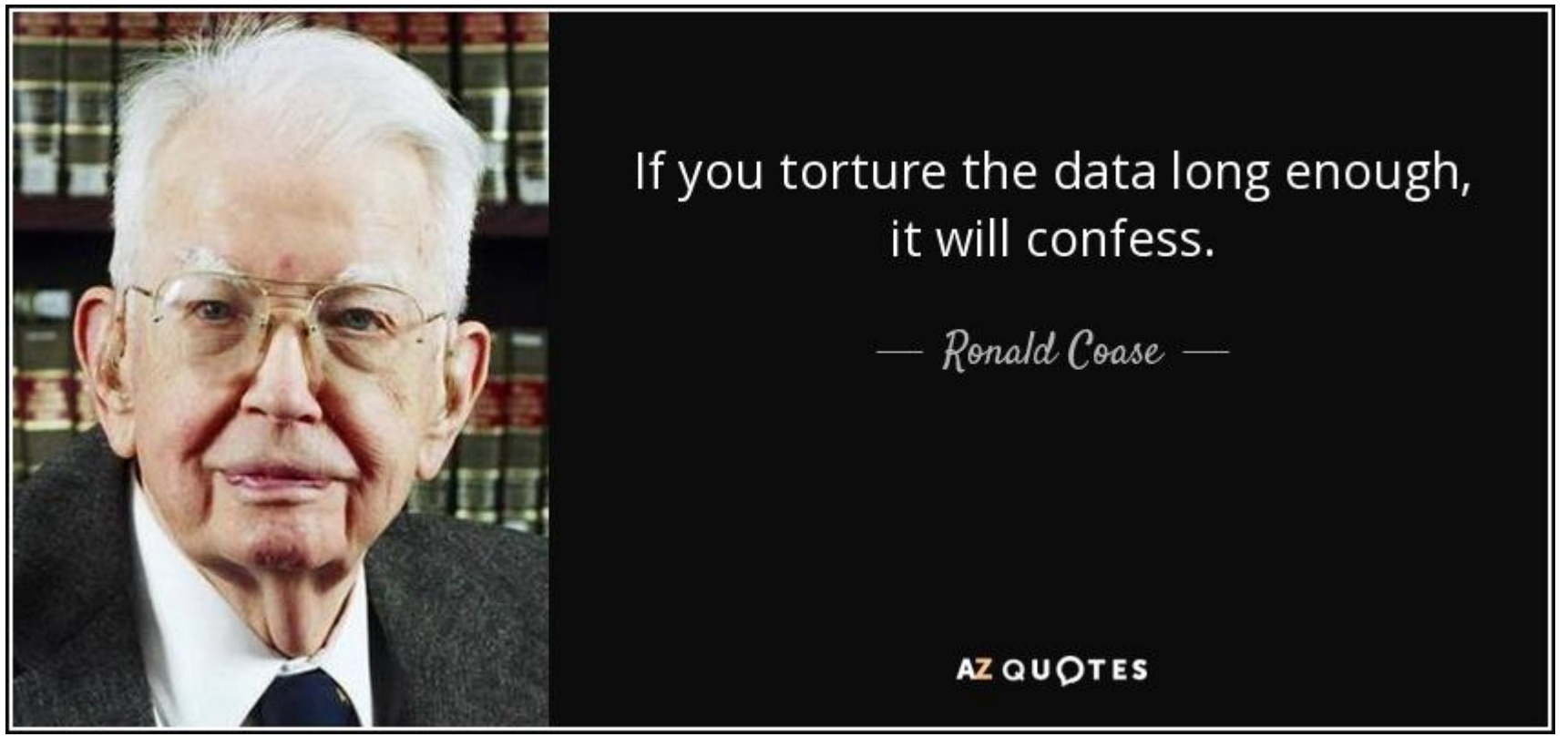

ML Algorithm
Supervised
Unsupervised
Regression
Classification
Dimensionality Reduction
Clustering
Linear / Polynomial
Penalized regression
KNN
SVR
Tree-based Regression models
Logistic regression
KNN
SVC
Tree-based Classification models
Principle Component Analysis (PCA)
K-Mean
Hierarchical

Exhibit 37 Stylized Decision Flowchart for Choosing ML Algorithms



Having said that…

- **Warning**: A ML algorithm will always find a pattern, even if there is none.




class
DISMISSED