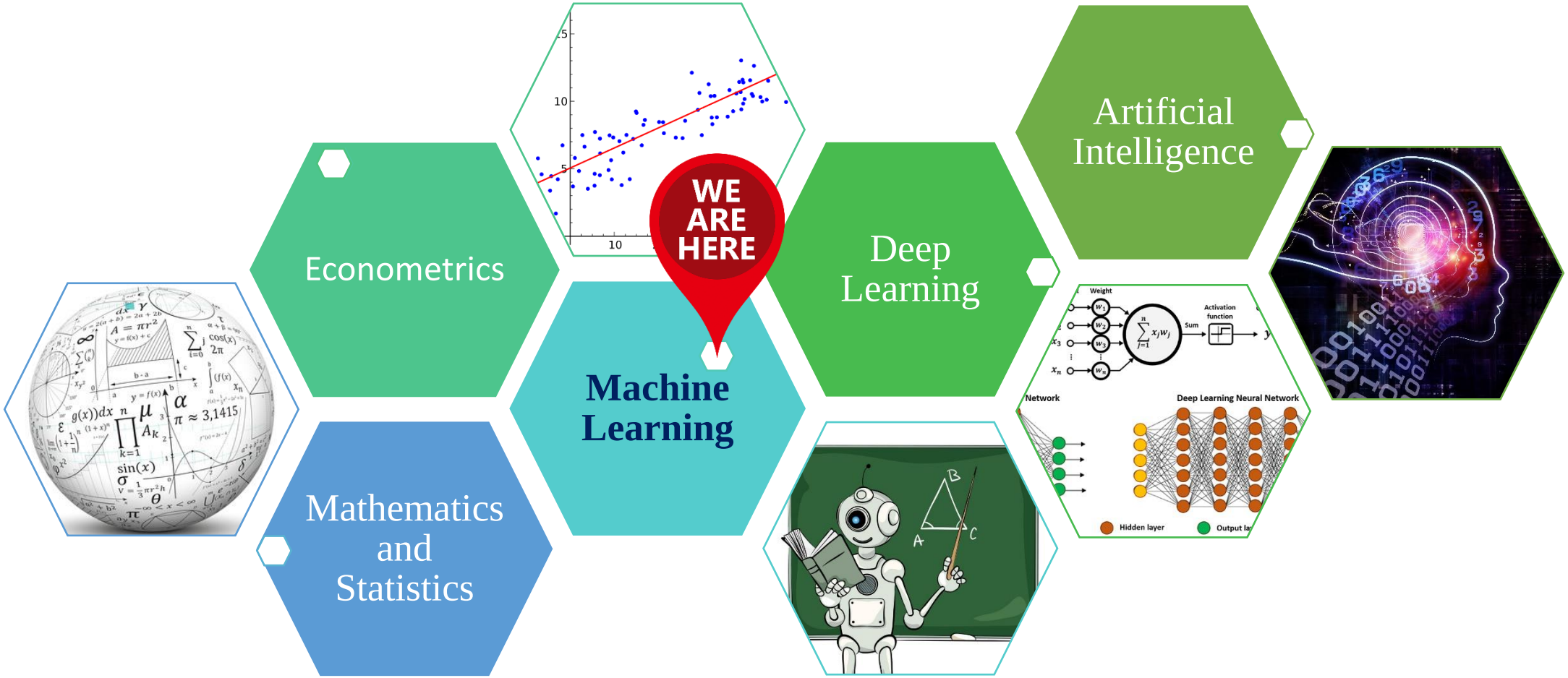
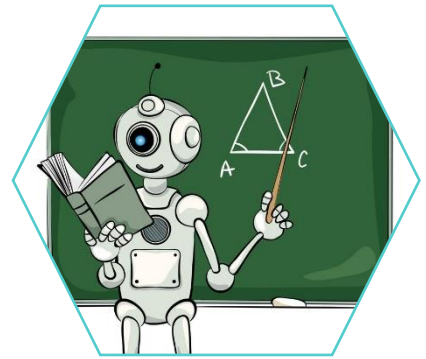
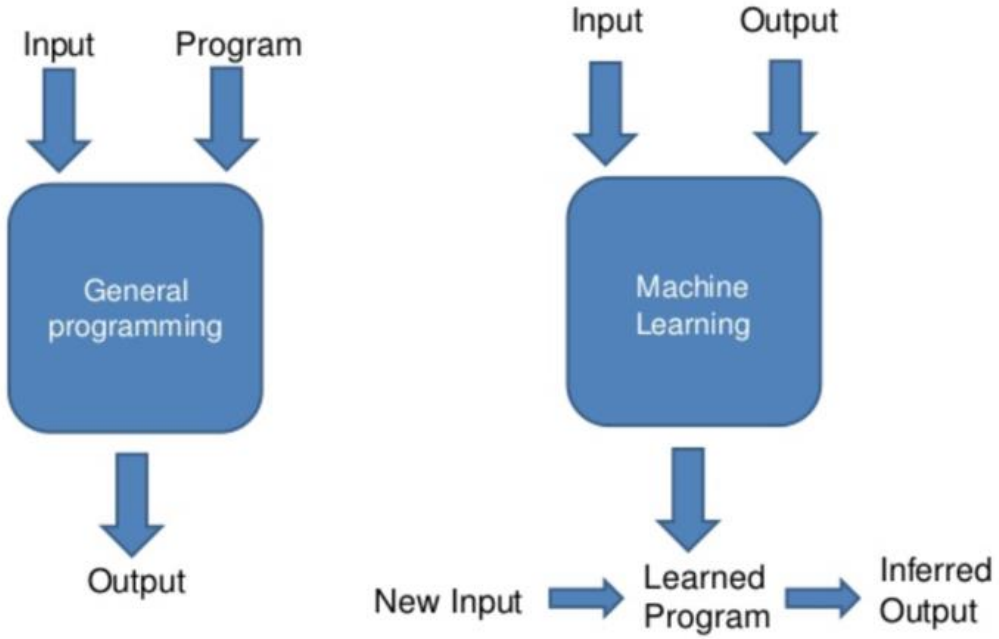
Class 2- What is Machine Learning?



General programming vs Machine learning





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

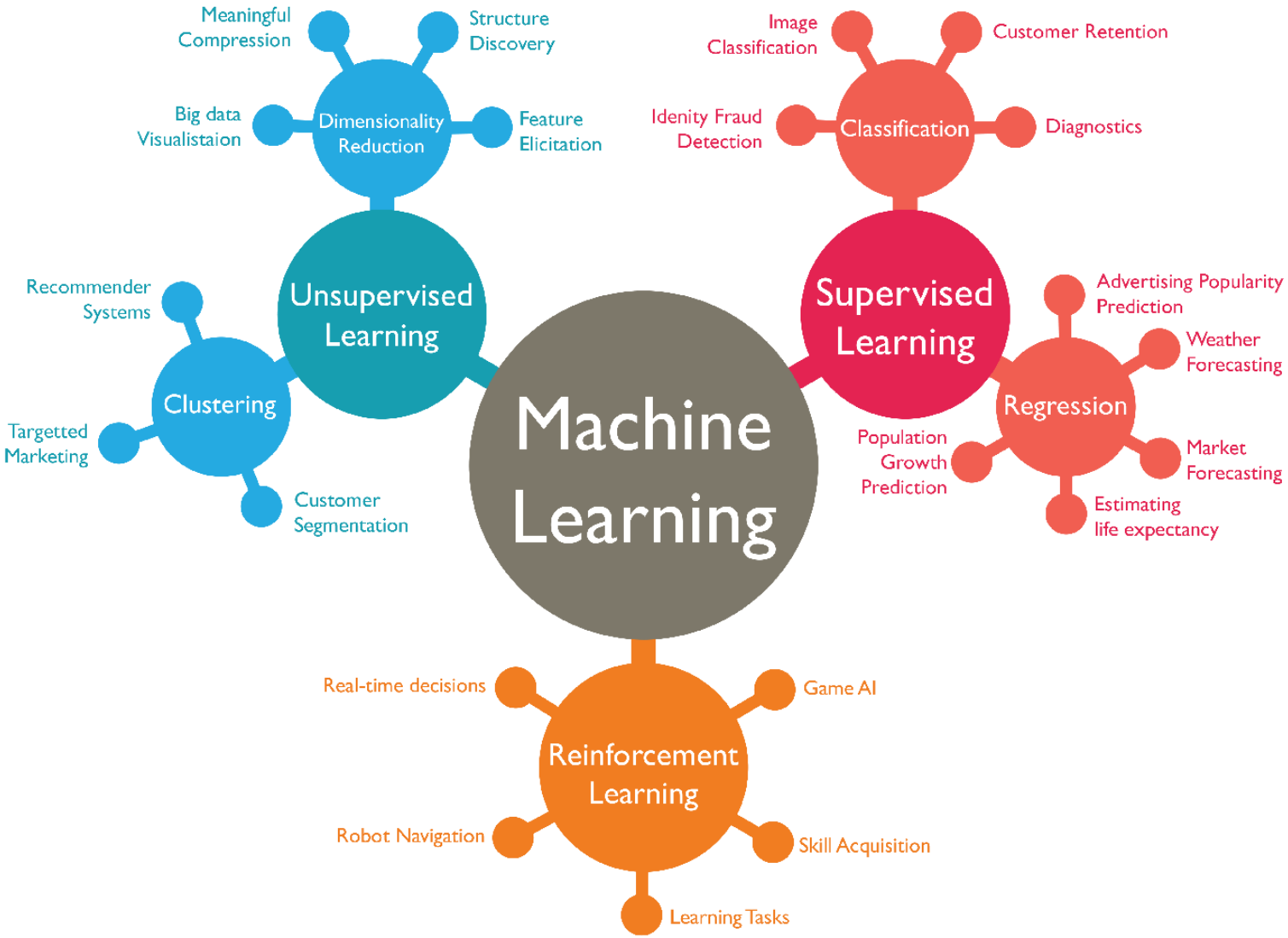
What is Machine Learning?

"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)

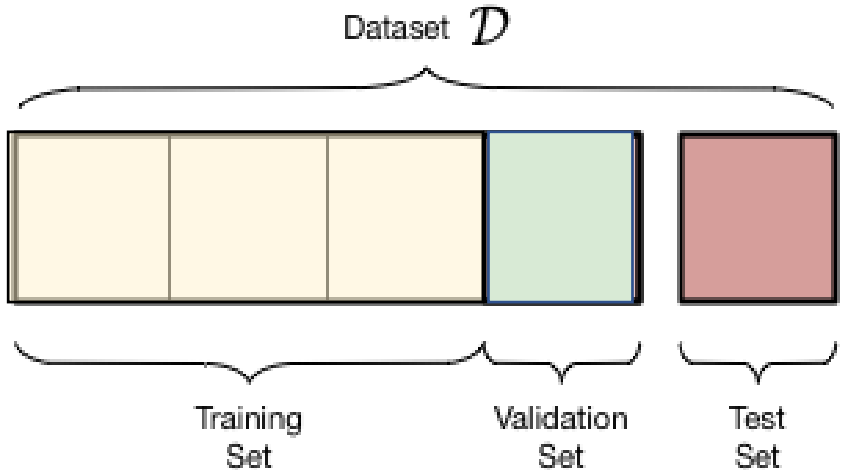
Let's break this statement into its components:

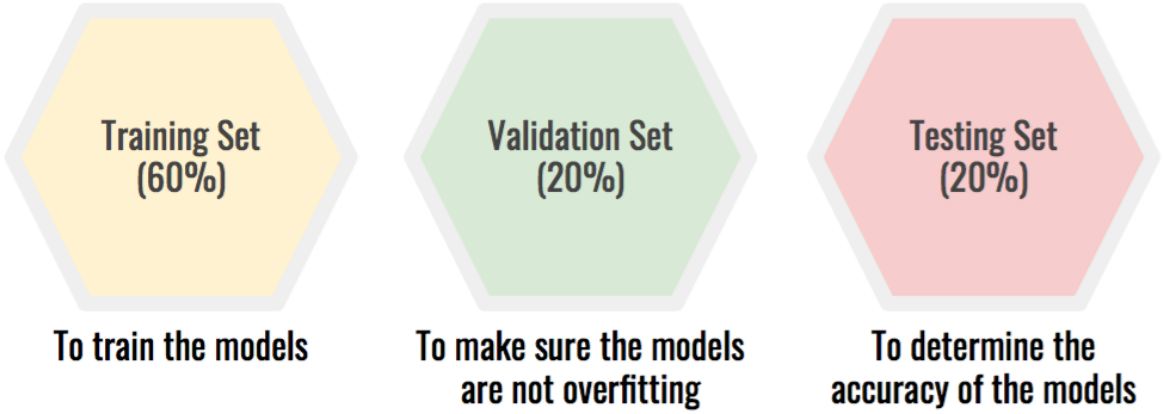
- ✓ **learns complex patterns**: The ML algorithm may find a pattern that cannot be easily represented with a finite set of equations.
- ✓ **learns … in a high-dimensional space**: Solutions often involve many variables and the interactions between them.
- ✓ **learns … without being specifically directed**: 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

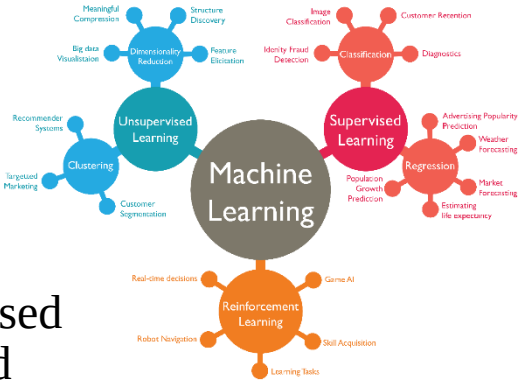


Train, validation and test data





Supervised Learning



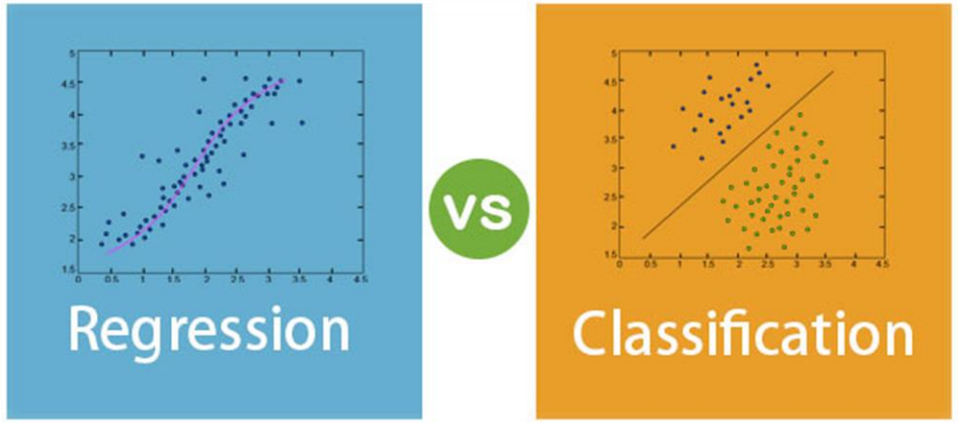
- In **supervised learning**, computers learn to model relationships based on train data. In supervised learning, inputs and outputs are labeled for the algorithm. After learning the pattern, the trained algorithms are used to predict outcomes for test data.

- **Regression:**

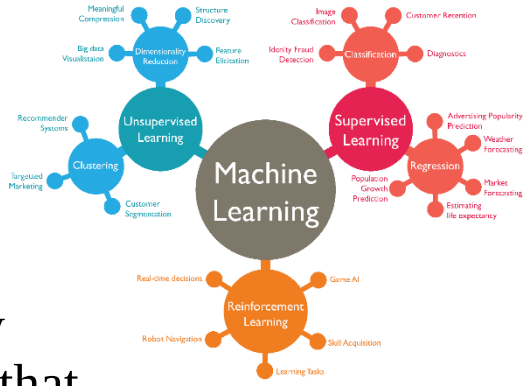
1. Predicting stock market returns
2. Predicting revenue growth
3. ….

- **Classification:**

1. Generating buy, sell, hold signals.
2. Likelihood of a successful M&A or IPO
3. Enhancing detection of fraud in financial statements
4. Classification on winning and losing funds or ETFs
5. …



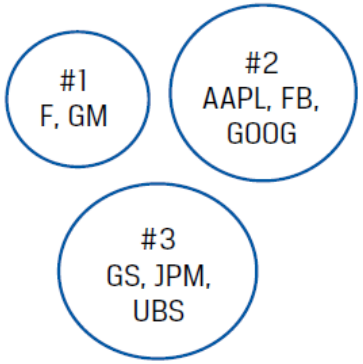
Unsupervised Learning



- 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. Examples:

- **Clustering:**

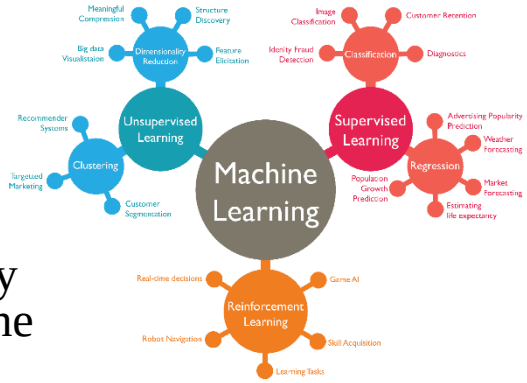
1. Grouping companies into peer groups based on some non-standard characteristics like financial statement data or corporate characteristics rather than sectors or countries.
2. Client profiling and asset allocation
3. Portfolio diversification and stock selection based on co-movements similarities



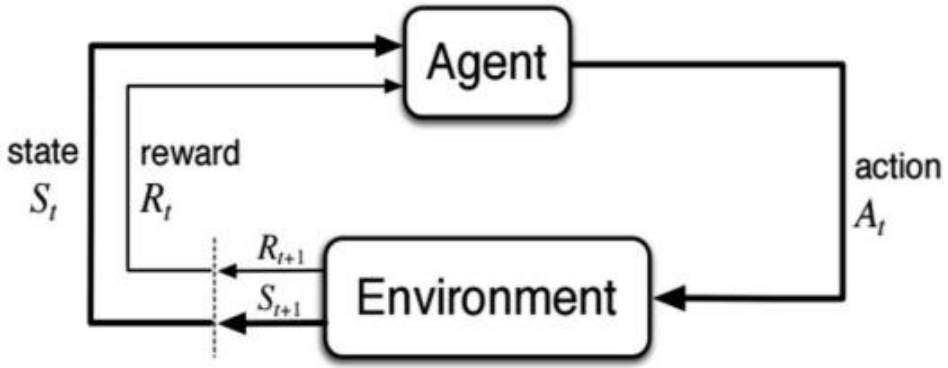
- **Dimensionality Reduction:**

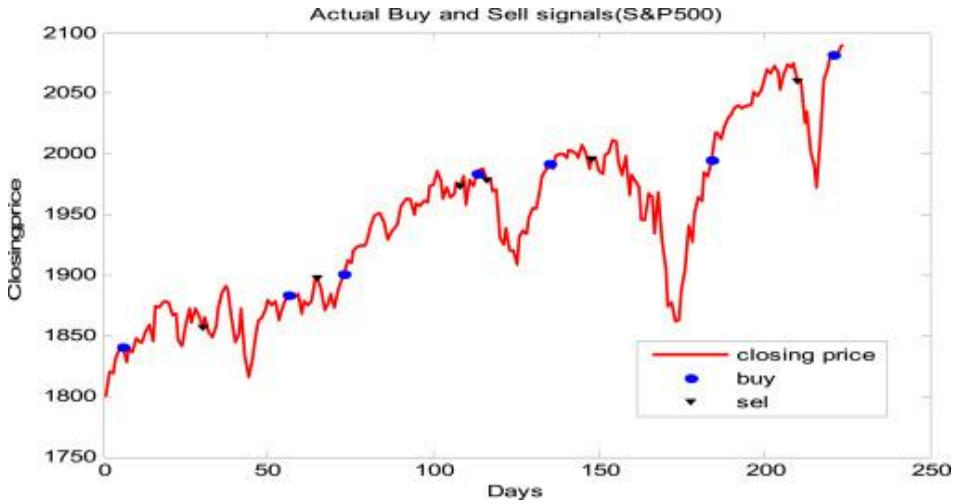
1. Identify the most predictive factors underlying asset price movements (to avoid factor zoo)

Reinforcement Learning



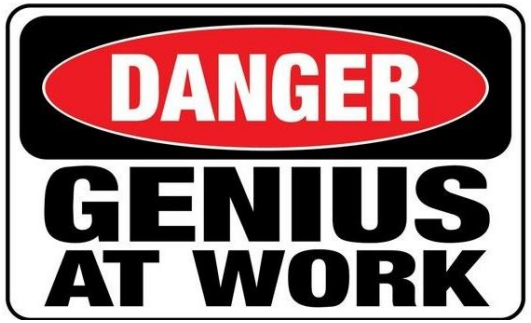
- In **reinforcement learning**, a computer (agent) learns from interacting with its environment by producing actions and discovering rewards. You need to define the environment, actions and the reward system. The machine will then explore and exploit to maximize the reward. The new actions may not be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



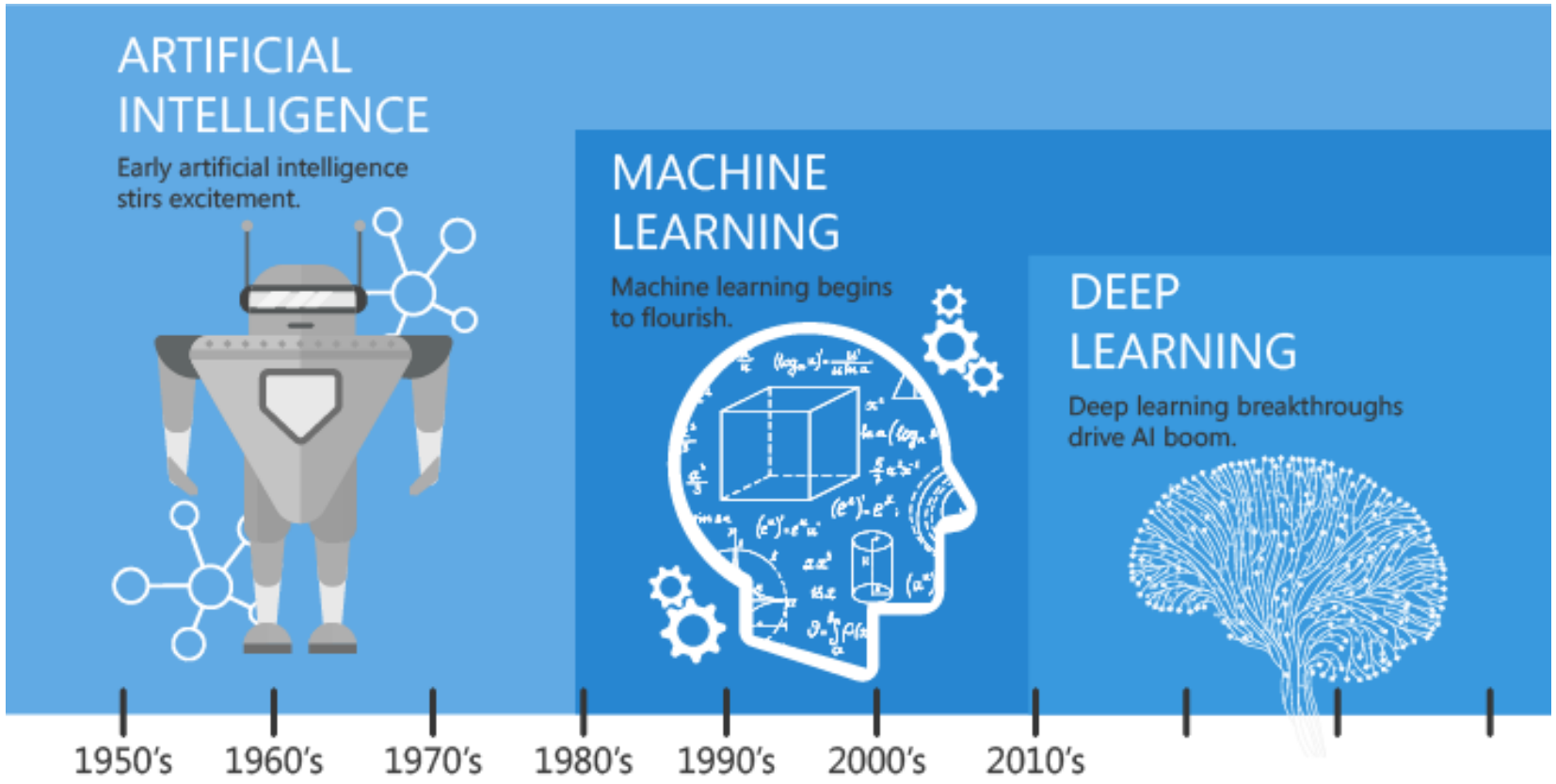


- Example: a virtual trader (agent) who follows certain trading rules (the actions) in a specific market (the environment) to maximize its profits (its reward).

How are you doing so far?

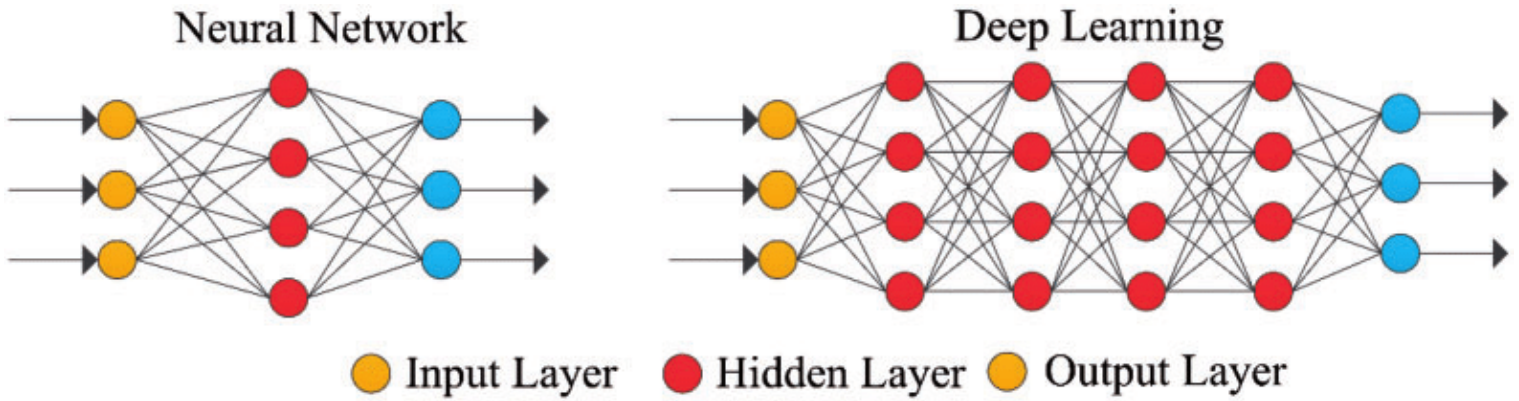


AI vs ML vs DL


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

Deep Learning

Deep learning is a subset of machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data.



- Examples:

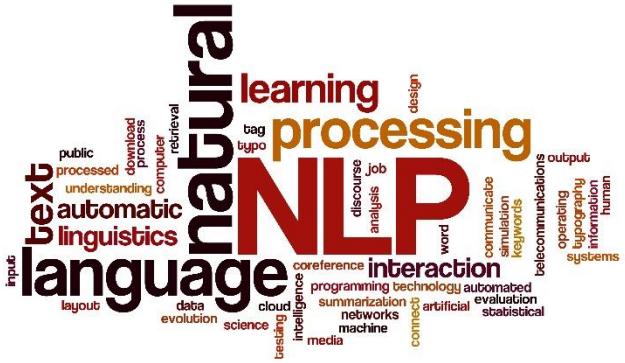
1. Image recognition algorithms can now analyze data from satellite-imaging systems to provide intelligence on the number of consumers in retail store parking lots,
2. Shipping activity and manufacturing facilities, and
3. Yields on agricultural crops.

Natural Language Processing (NLP)

NLP is a field of research at the intersection of **computer science**, **artificial intelligence**, and **linguistics** that focuses on developing computer programs to analyze and interpret **human language**.

- Automated tasks using NLP include translation, speech recognition, text mining, sentiment analysis, and topic analysis.

- Examples:
  - Reading millions of pages of annual reports, thousands of hours of earning calls, transcripts, news articles and social media posts to identify trend in shorter timespans!
  - Analyzing communications and transcripts from policymakers (FED, ECB, …) to provide insights around trending topics like interest rate policy, GDP, inflation expectation and etc
  - Chatbots answer basic retirement savings questions, learning from their interactions with investors.



Road map

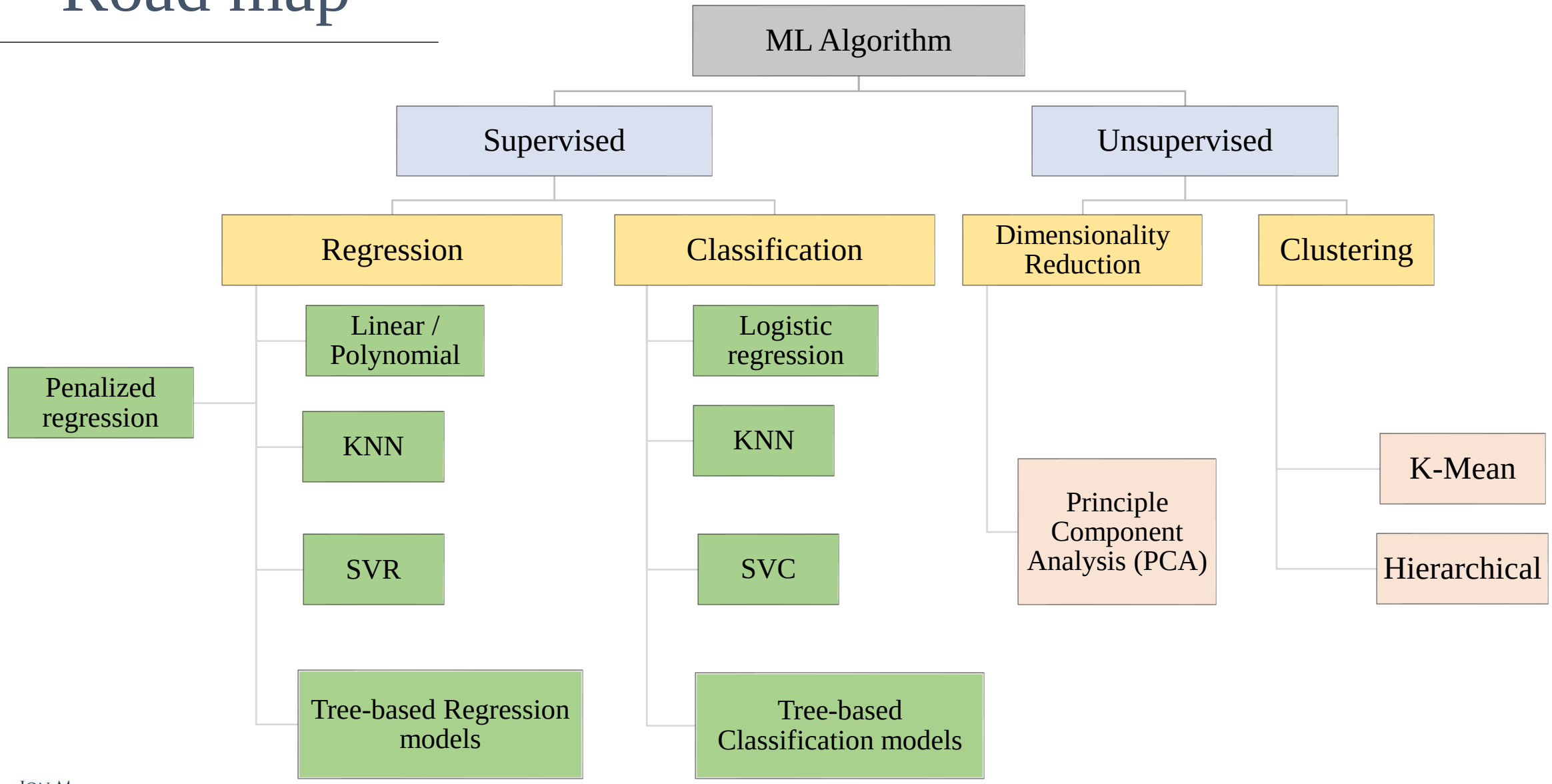
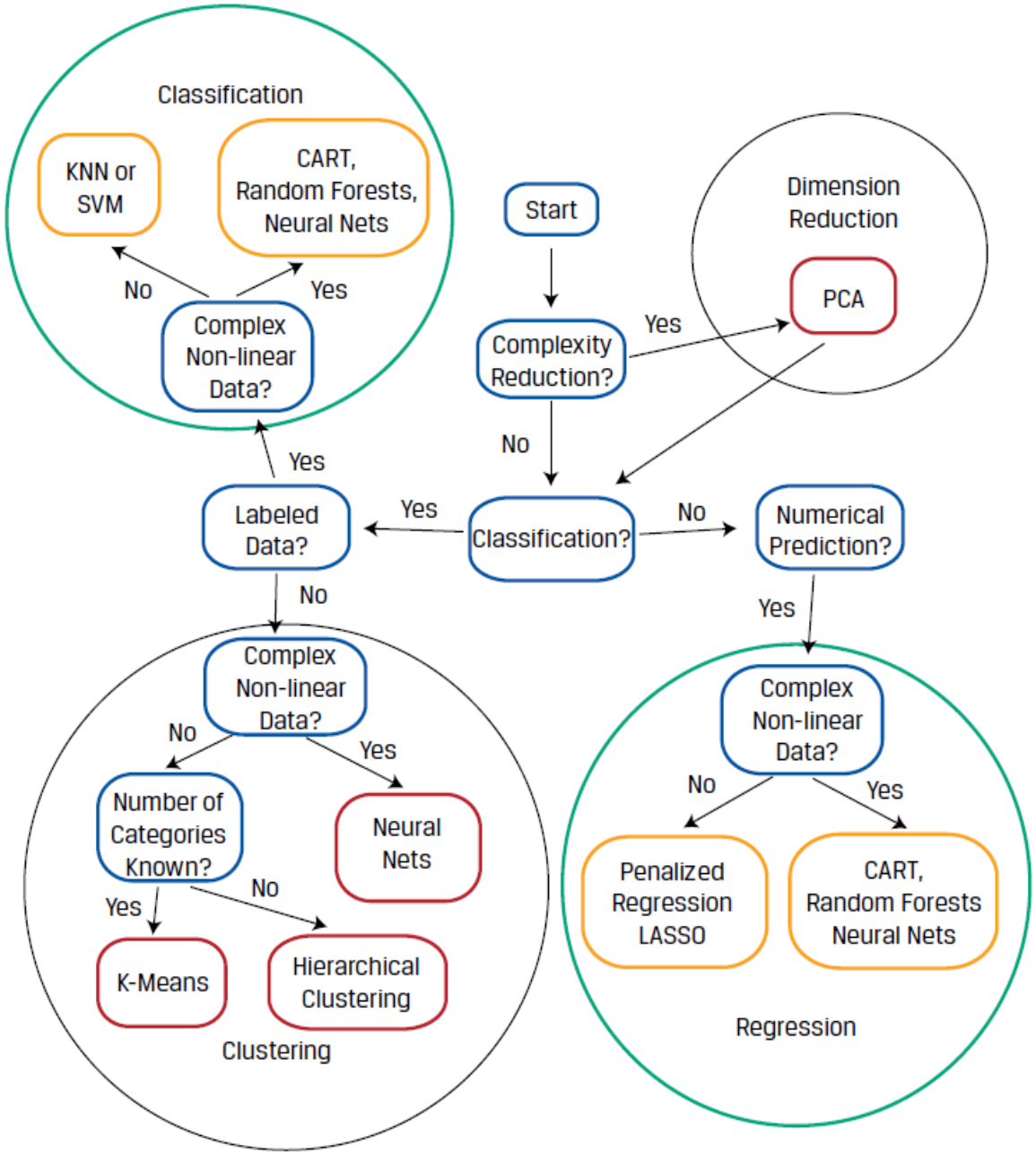
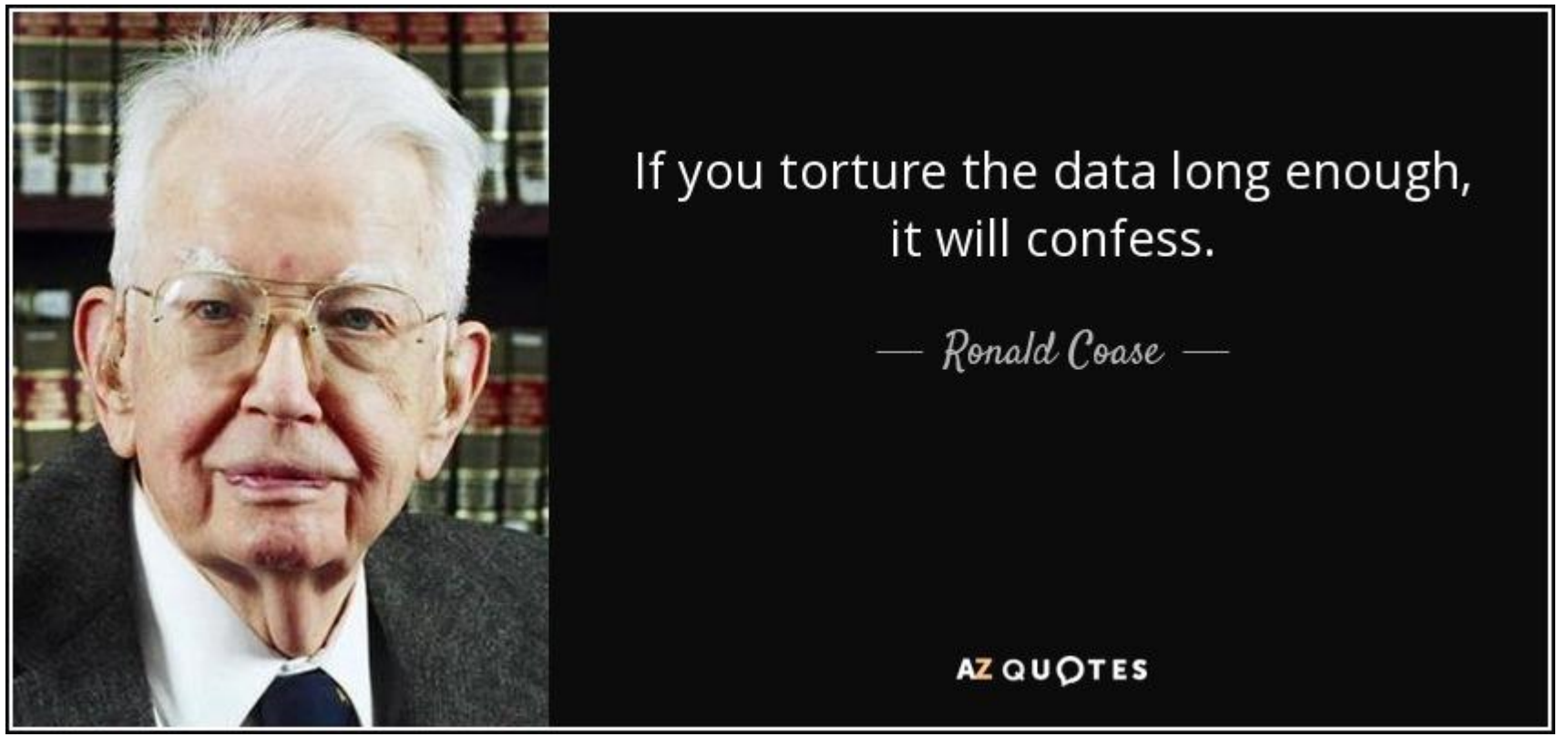

ML Algorithm
Supervised
Unsupervised
Regression
Classification
Dimensionality Reduction
Clustering
Linear / Polynomial
Penalized regression
KNN
SVR
Tree-based Regression models
Logistic regression
KNN
SVC
Tree-based Classification models
Principle Component Analysis (PCA)
K-Mean
Hierarchical

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.



Students' questions

1. Is there a difference between statistical linear regression and machine learning linear regression?
2. How exactly the machine learning algorithms are built. could anyone create a machine learning algorithm as long as they have a python environment?
3. I'm having a difficult time coming to grip with the "it" of machine learning. If it is always this "agent" that I am teaching or if that only applies to reinforcement learning. Thanks!
4. If the algorithms used for each type of learning are widely available or if each entity that collects data via machine learning has to write its own algorithm to put into use?
5. whether the machine in reinforcement learning is given a specific goal or not. Or maybe I should say, I can't picture what a reward looks like to a computer.
6. Can data clustering be used outside of python? What real world applications does it have?


thank you all
CLASS DISMISSED