Special Topic: Machine Learning (ML) and Artificial Intelligence (AI)

ENGR 103: Engineering Computation and Algorithmic Thinking

Alex Ulbrich

Today's Lecture

- What is AI, ML, Deep Learning?
- A Typical ML Workflow
- Data Collection and Preparation
- Different Models

How many of you have used AI *today*?

Applications

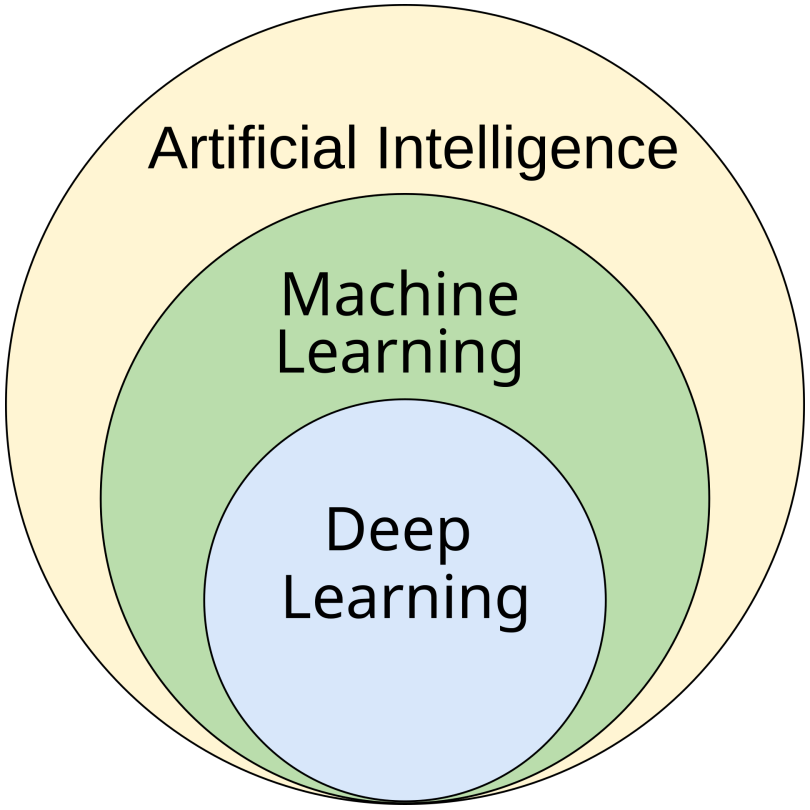
- Agriculture: optimal plant growth
- Banking: fraud detection
- Cybersecurity: attack detection and handling
- Civil: structural health monitoring
- Mechanical: predictive maintenance in machines
- Healthcare: medical diagnosis of patients
- Logistics: routing and optimization

AI, ML, and Deep Learning

Artificial Intelligence (AI) refers to the capability of computational systems to perform tasks typically associated with human *intelligence* (learning, reasoning, problem-solving, perception, and decision-making).

Machine Learning (ML) is s subset of AI that learns patterns from data.

Deep learning is a subset of machine learning that focuses on utilizing *neural networks*.



Basic ML Workflow

1. Collect and prepare data (images, sensor readings, text, ...)
2. Train a model (feeding examples)
3. Make predictions (on new data)
4. Improve over time (learning from mistakes)

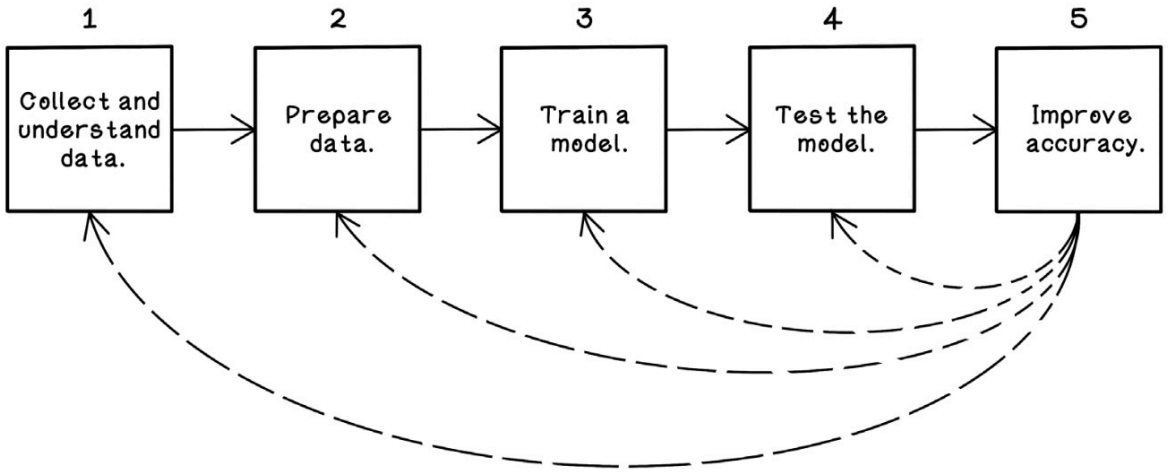


Image from R. Hurbans, Grokking Artificial Intelligence Algorithms.

Data Collection and Processing

Data scientists often spend **60-80% of their time on data-related tasks**: collecting, cleaning, transforming, and exploring data.

Raw data is **messy**: missing values, outliers, inconsistent formats, or irrelevant features can tank a model's performance.

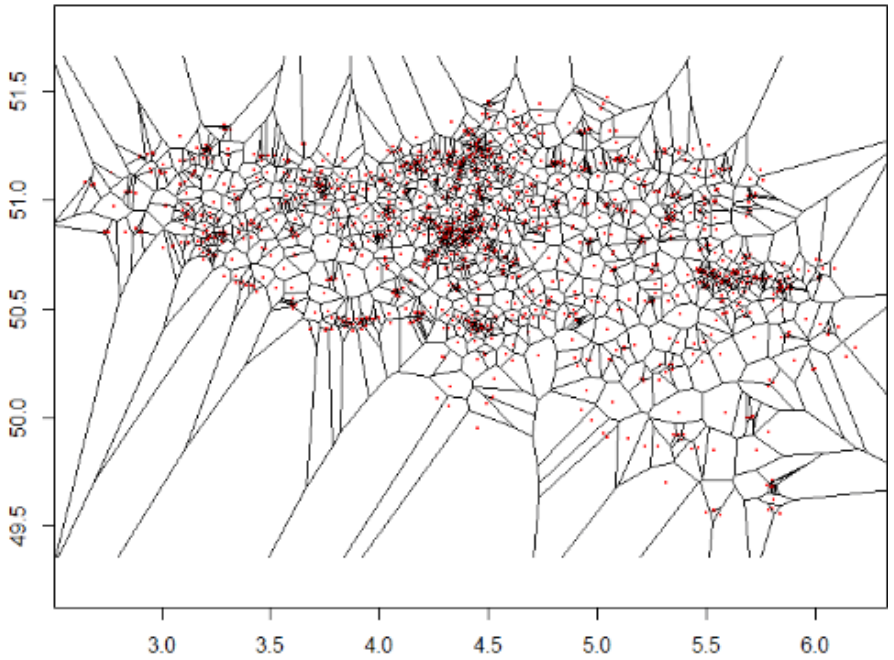
Garbage in, garbage out. A model's only as good as its data. Clean, well-prepared data ensures the model learns meaningful patterns, not noise.

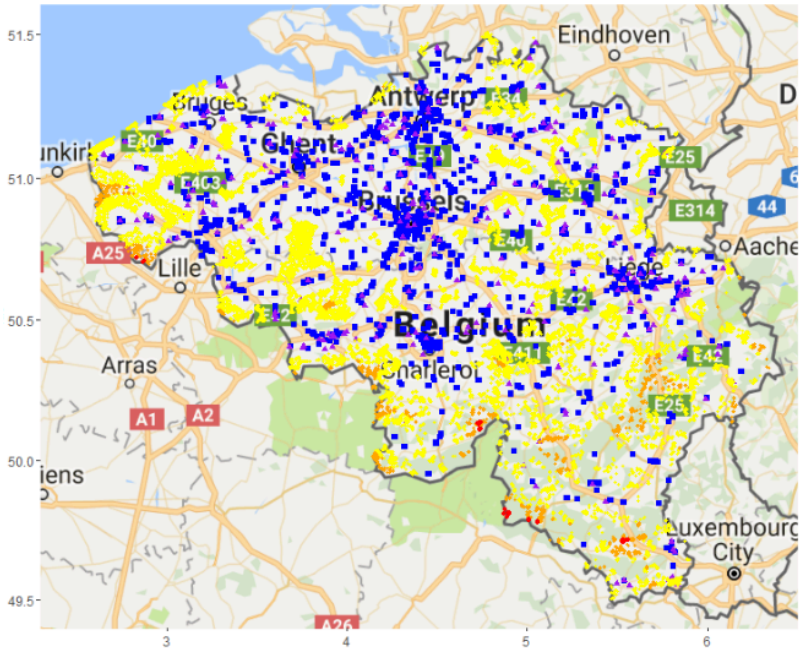
Not all datasets are terrible.

When doing supervised learning, we want to have *train* and *test* data.

Data Processing Example

Voronoi tesselation to compute distances between ~700k people and ~ 1.2k repair shops





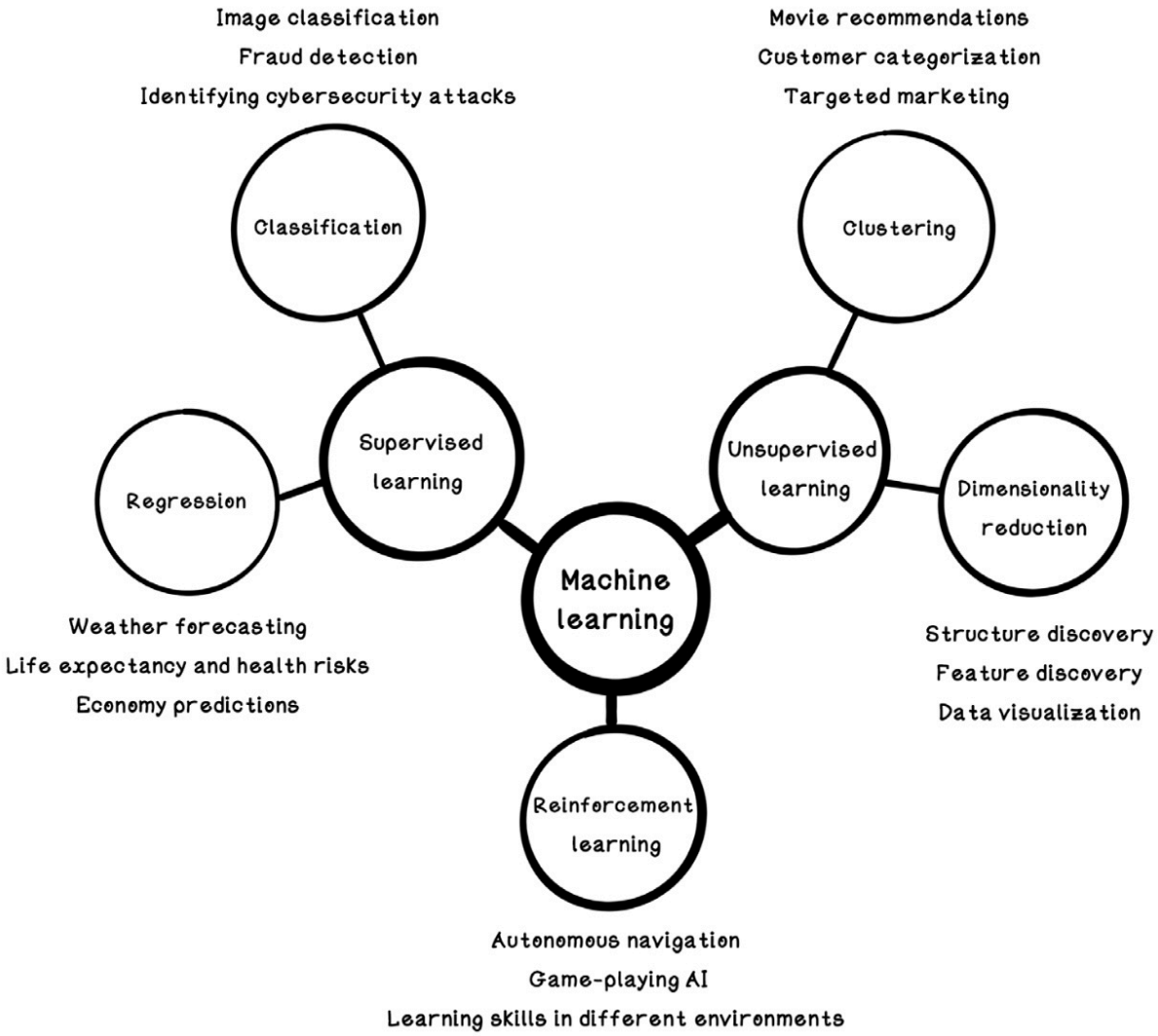
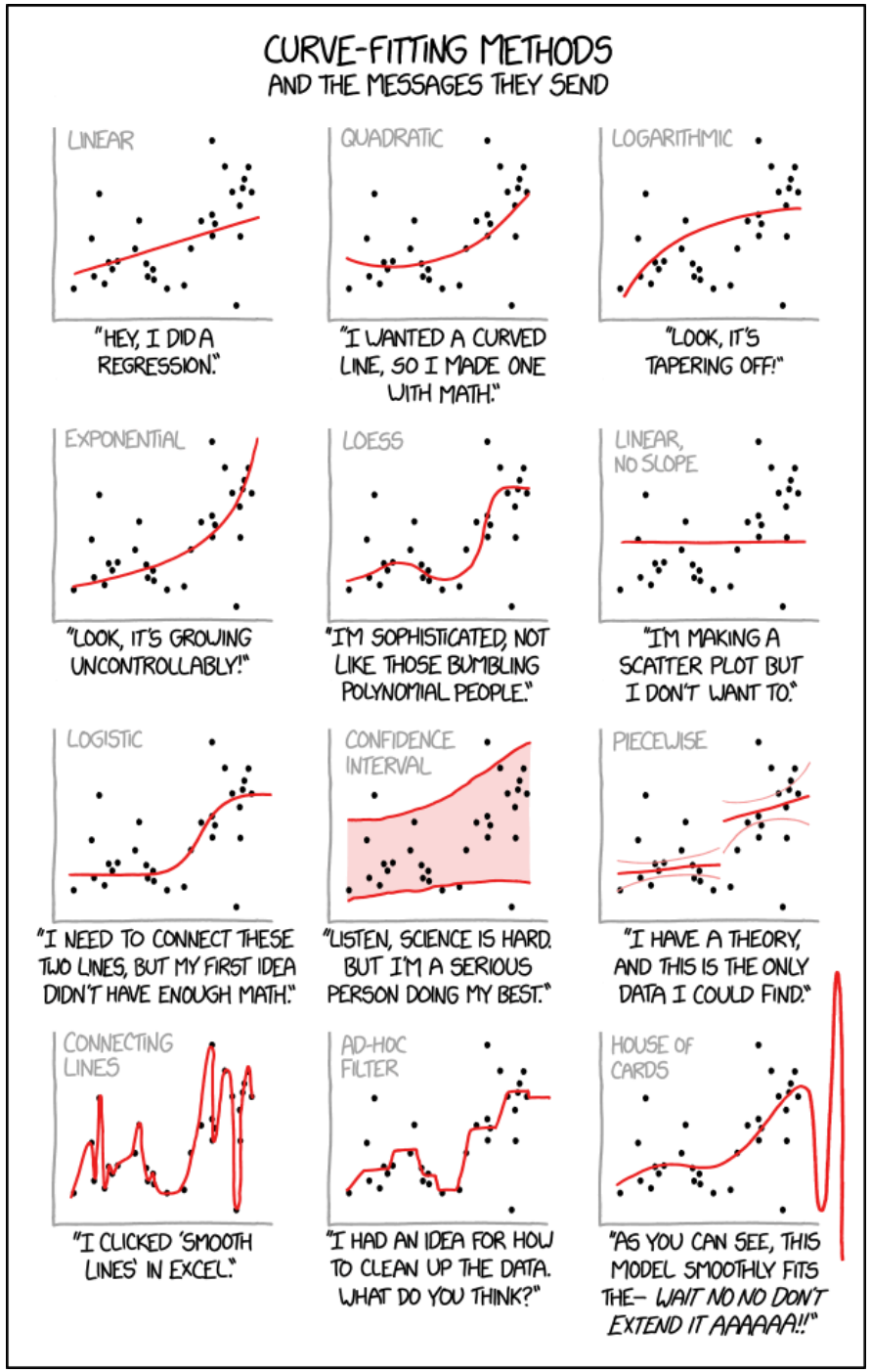


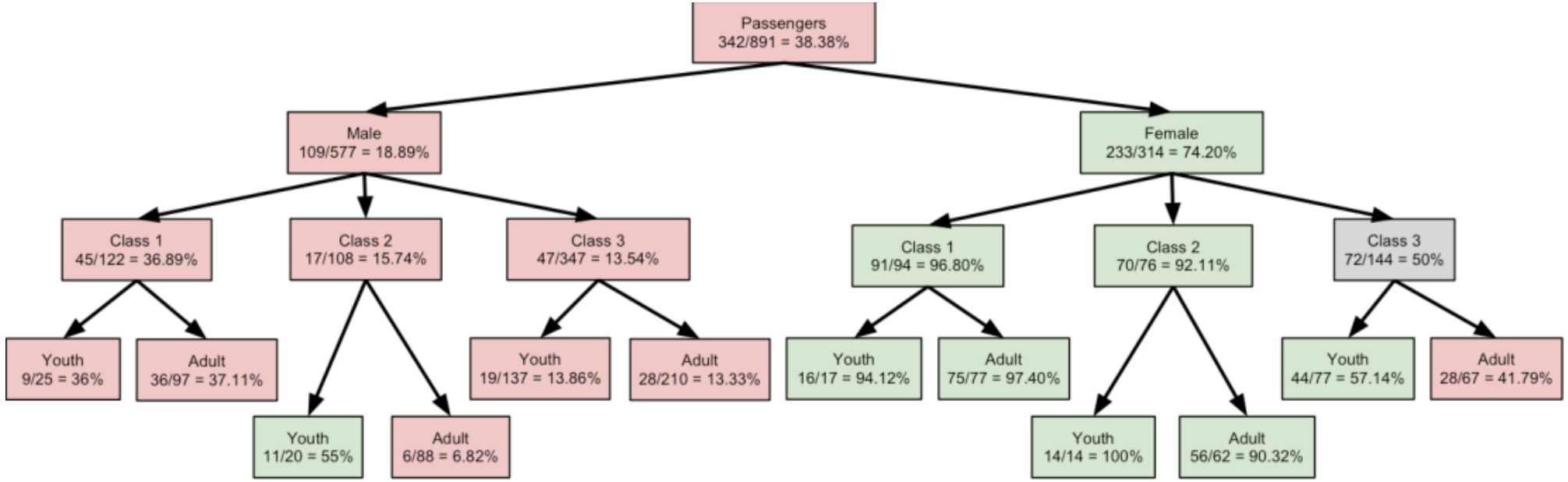
Image from R. Hurbans, Grokking Artificial Intelligence Algorithms.

Regression

https://xkcd.com/2048/



Classification



Can we predict Titanic survivors? **Green** mostly survived. **Red** mostly died.

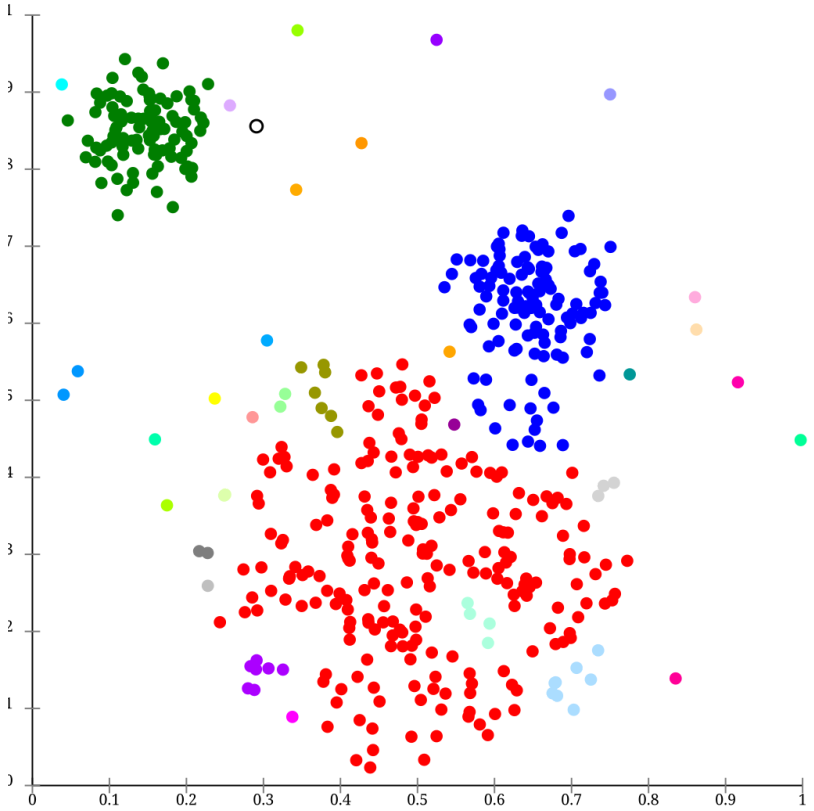
CS229 Titanic – Machine Learning From Disaster

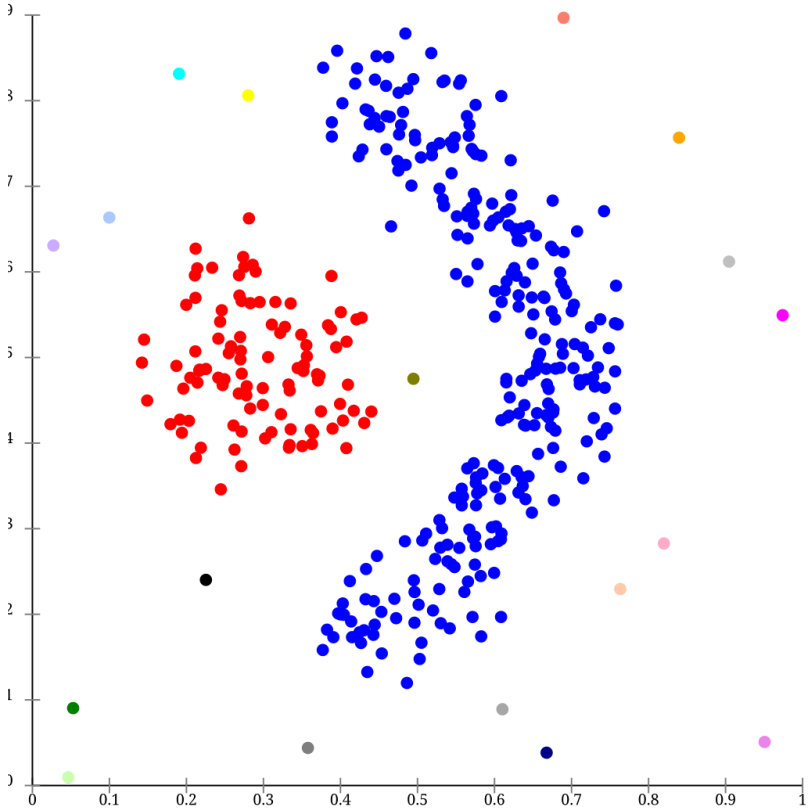
Clustering

Centroid vs

Density

Source



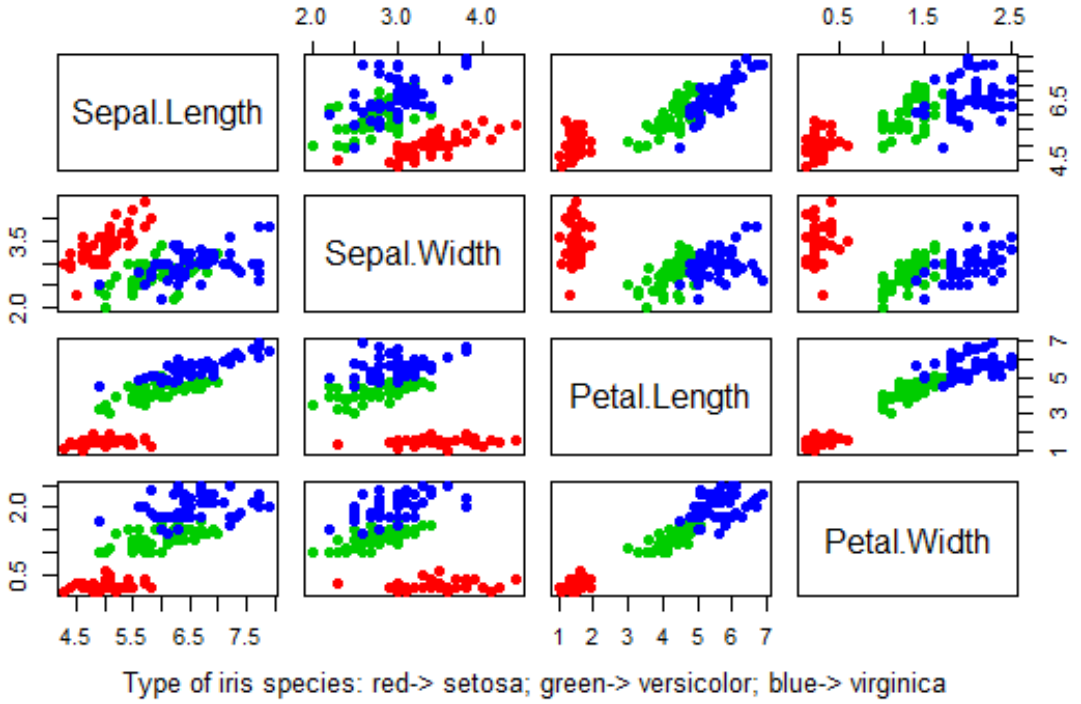


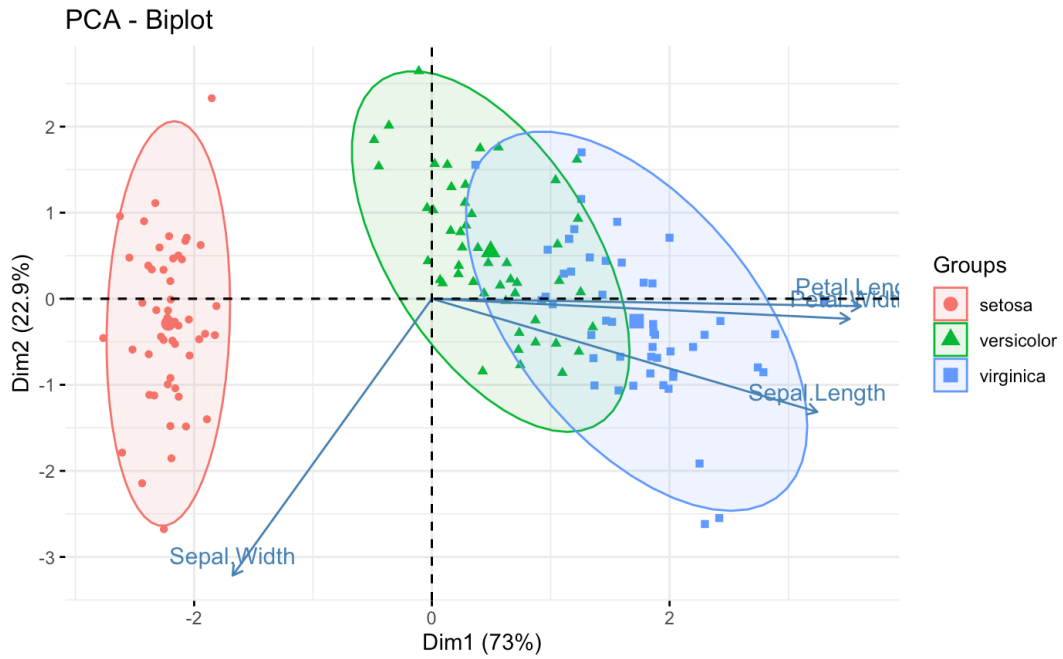
Dimensionality Reduction

Scatter Plots

Principal Component Analyis (PCA)

Iris Dataset



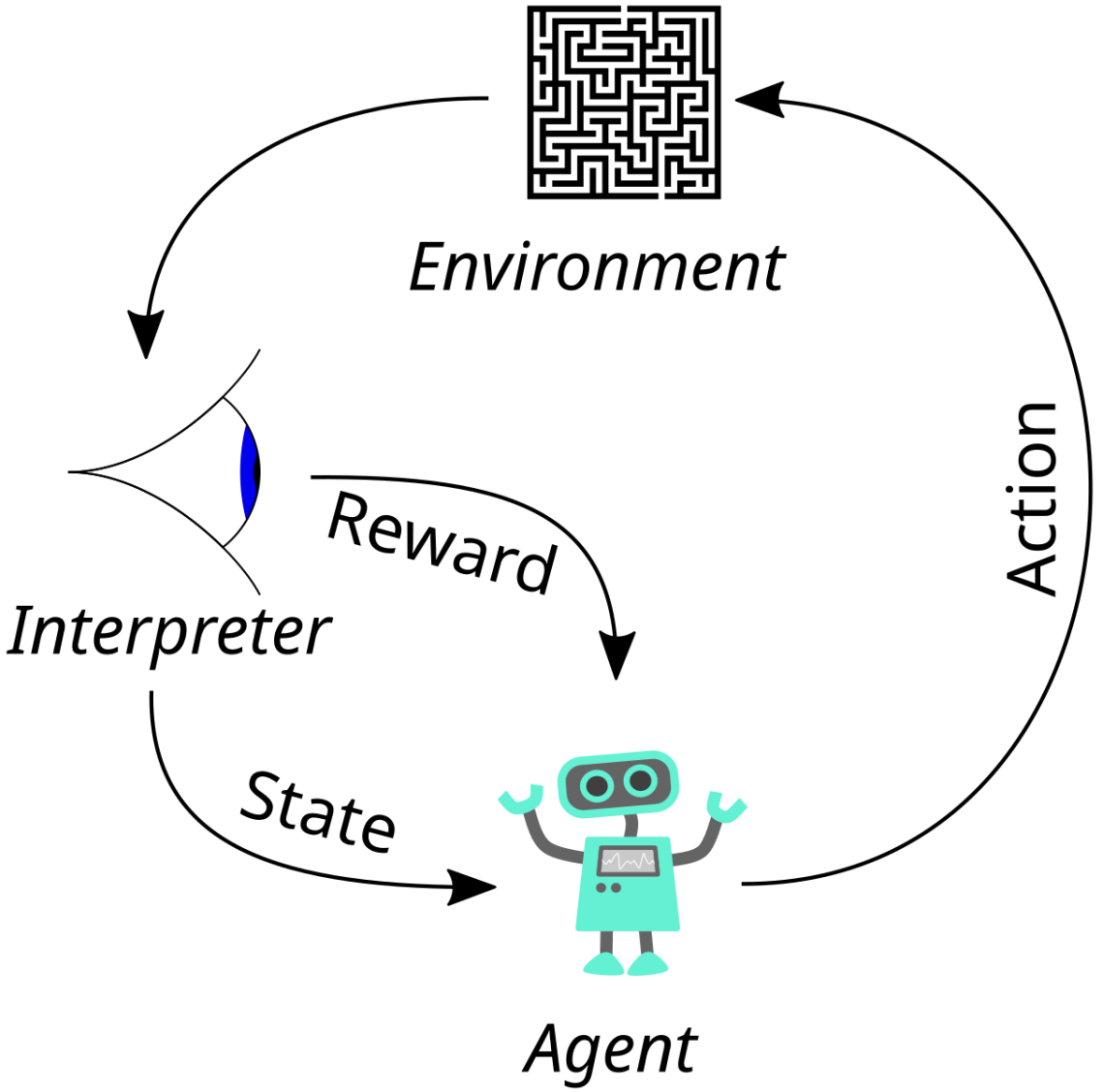


Reinforcement Learning

Imagine trying to teach a pup how to sit by giving them a reward if they succeed.

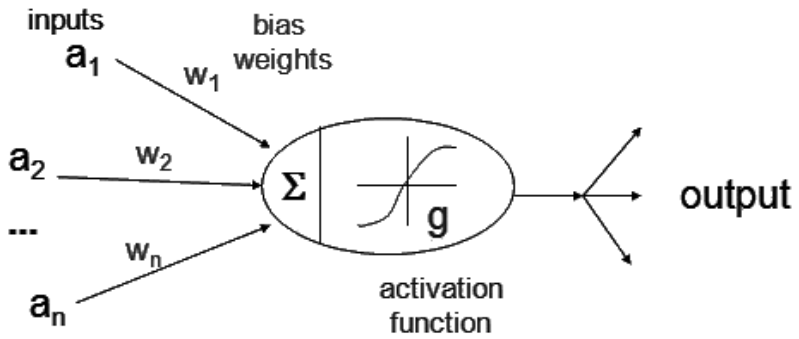
In this case, the model/agent improves when making successful attempts.

Source

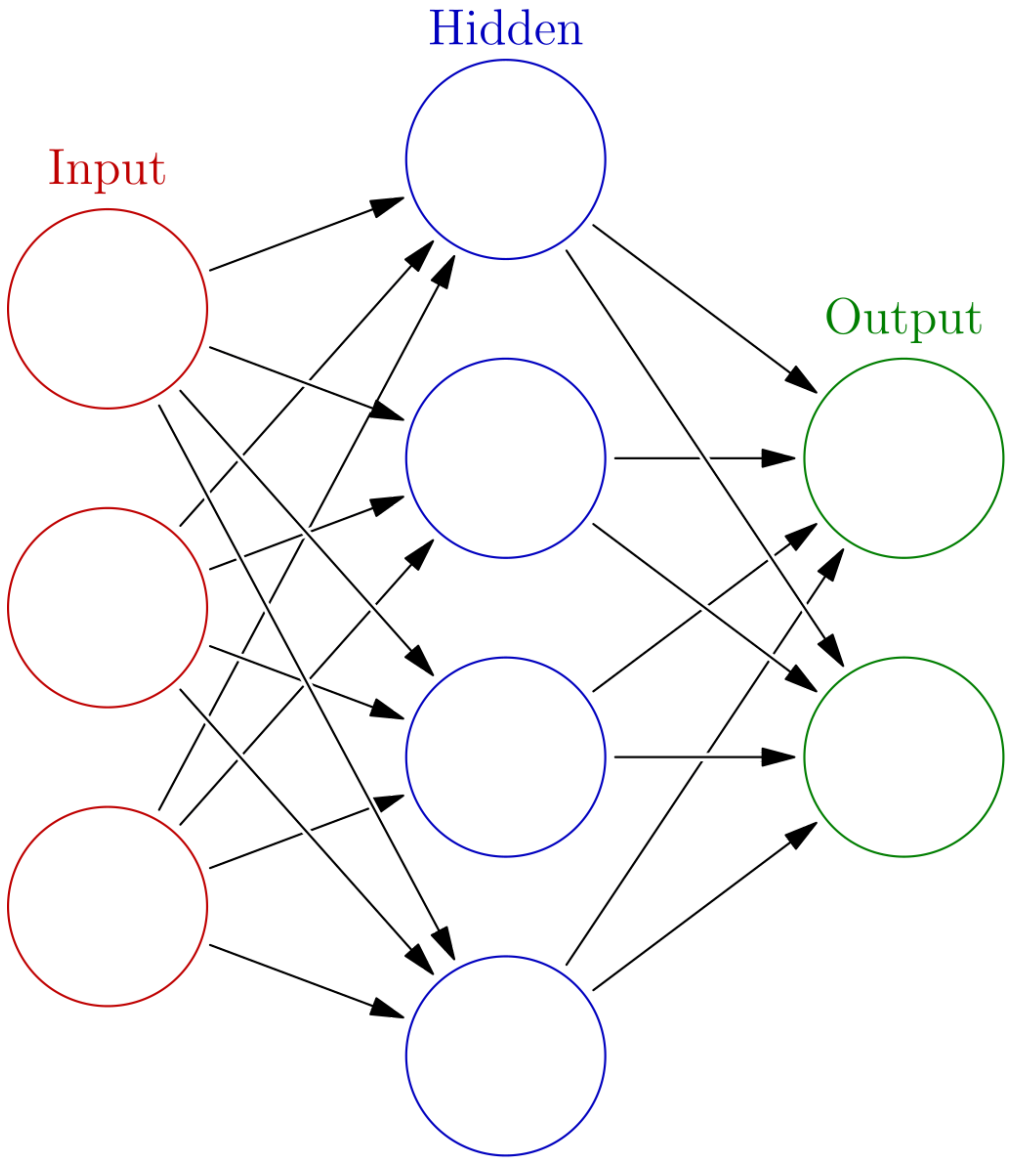


Neural Networks

Nodes (neurons) connected by edges (synapses). Each node receives information from connected nodes, processes it, then sends it to other nodes.



Source



Evaluating Model Quality

It depends on the type of model:

- classification: **confusion matrix** or similar
- regression: some measure of the "error"
- reinforcement learning: some measure of the "reward"

All metrics have some trade-offs.

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

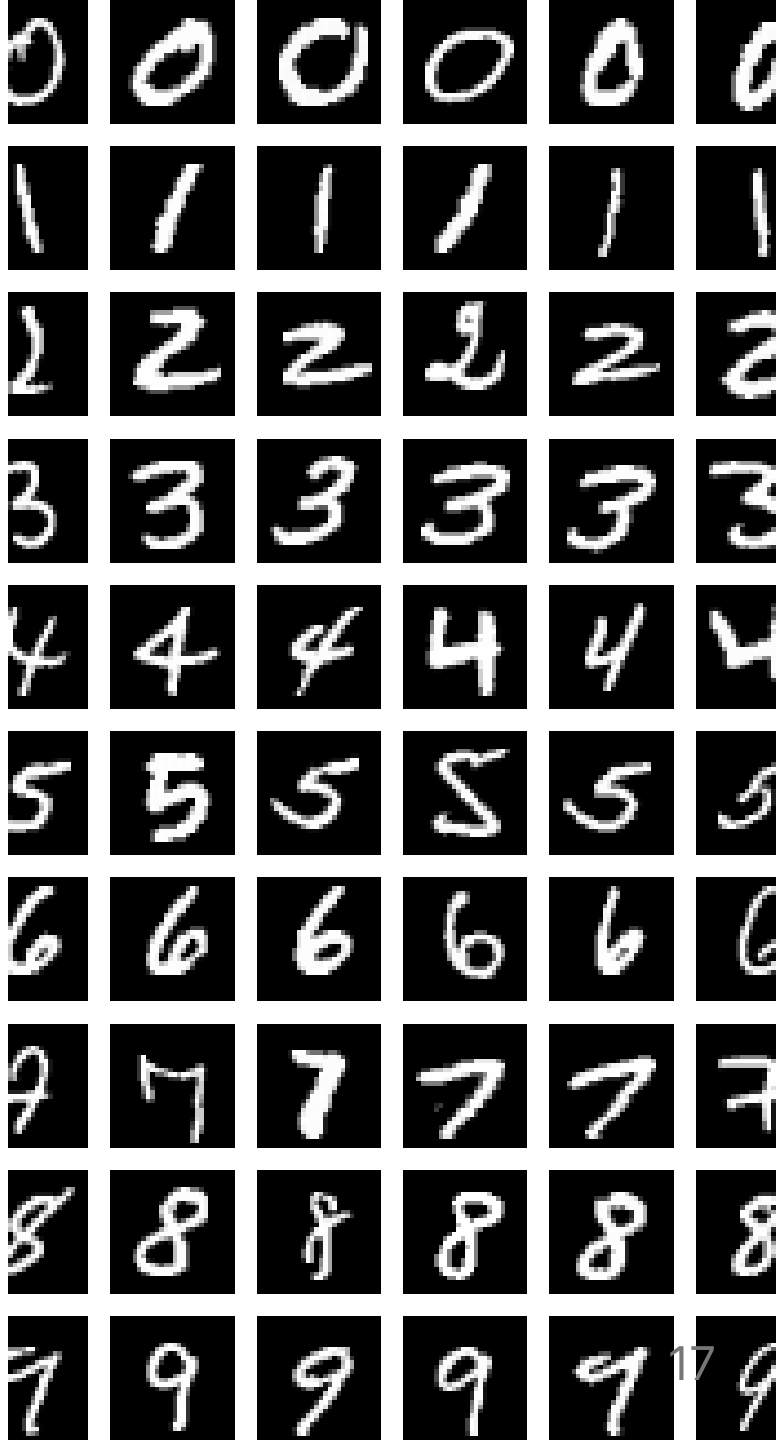
Google Colab Exercise

Classifying handwritten digits: MNIST Database

Using neural networks (simple to complex).

Open https://colab.research.google.com/ to get started.

Code is available on GitHub as well.



Worth Checking

- AlphaGo Documentary
- HuggingFace LeRobot
- Conversational AI and Search: Grok, ChatGPT, Gemini, LeChat, Copilot, Perplexity
- Teachable Machine
- TensorFlow Playground
- Kaggle's Intro to Machine Learning
- AI Dungeon
- Quick, Draw!