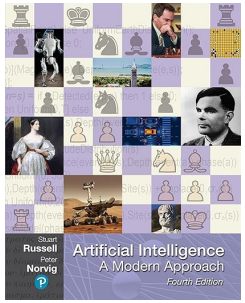
UNIVERSITY OF MARYLAND

MSML610: Advanced Machine Learning

Lesson 01.1: AI and Machine Learning

Instructor: Dr. GP Saggese - gsaggese@umd.edu

References: - AIMA (Artificial Intelligence: a Modern Approach), Chap 1



- ***AI and Machine Learning***
  - What is AI?
  - What is Machine Learning?

- AI and Machine Learning
  - ***What is AI?***
  - What is Machine Learning?

ML, AI, and Intelligence

- Machine Learning is a subset of AI
  - All of it confused with deep learning, large-language models, predictive analytics, ...
- What is artificial intelligence?
- What is intelligence?

Human Intelligence

- We call ourselves "homo sapiens" because intelligence sets us apart from animals
- For thousands of years, we tried to understand how we think
- One of the biggest mysteries
  - Brain is a small mass of matter
  - Our brain can understand nature secrets, e.g., theory of relativity, quantum mechanics, black holes in the universe
  - How can brain understand, predict, and manipulate a world more complicated than itself?

Artificial Intelligence

- The term "Artificial Intelligence" was coined in 1956
- **AI aims to:**
  - Understand human intelligence
  - Create intelligent entities
  - *"What I cannot create, I do not understand"* (Feynman, 1988)
- **AI is a technology**
  - Universal and applicable to any human activity and task
  - Its impact greater than any previous historical event
  - Currently generates trillions of dollars annually in revenue
  - Presents many unresolved problems
    - E.g., major concepts in physics might be established

AI Formal Definition

- AI is defined around **two axes**:
  - Thinking vs. Acting
  - Human vs. Rational (ideal performance)
- Four possible definitions of AI as a machine that can:
  1. Think humanly
  2. Think rationally
  3. Act humanly
  4. Act rationally
- **Q**: Which one do you think is the best definition?
- We will see that building machines that can **"act rationally"** should be ultimate goal of AI

1. AI as Thinking Humanly

- To build machines that think like humans we need to **determine how humans think**
- **Pros**
  - Express precise theory of the human mind as a computer program
- **Cons**
  - Unknown workings of the human mind
  - Anthropocentric definition

2. AI as Thinking Rationally

- What are the rules of **correct thinking**?
  - Given correct premises, yield correct conclusions
- **Logic** studies the "laws of thought"
  - Formalize statements about objects and their relations
- **Automatic theorem proving**
  - Programs solve problems in logical notation
  - Run indefinitely if no solution exists (related to the halting problem)

Thinking Rationally: Cons

1. **Formalizing informal knowledge is difficult**
   - Example: *"A handshake occurs when two people extend, grip, shake hands, then release."*
   - Formal logic representation:

$$\begin{aligned}\exists x, y\, (&\text{Person}(x) \land \text{Person}(y) \land x \neq y \land \\ &\text{Hand}(x, h_x) \land \text{Hand}(y, h_y) \land \\ &\text{MoveToward}(h_x, h_y) \land \text{Contact}(h_x, h_y) \land \\ &\text{Shake}(h_x, h_y) \land \\ &\text{Release}(h_x, h_y))\end{aligned}$$

2. **Probabilistic nature of knowledge**
   - Example in medicine: *"Fever, cough, and fatigue could indicate flu, COVID-19, or another illness."*
3. **Scalability challenges**
   - Large problems may need heuristics for practical solutions
4. **Intelligence requires more than rational thinking**
   - Importance of agent interaction with the world
   - Problem of the "body"

3. AI as Acting Humanly

- **Agent** is something that perceives and acts to reach a goal
- **Definition**: AI designs **agents that can act like humans**
- **Turing test**
  - *"A computer passes the Turing test if a human cannot tell whether the answers to questions came from a person or a computer"*
- Passing the (embodied) Turing test requires:
  1. Natural language processing to communicate
  2. Knowledge representation to store information
  3. Automated reasoning to use stored knowledge and answer questions
  4. Machine learning to detect patterns
  5. Computer vision and speech recognition to perceive objects and understand speech
  6. Robotics to manipulate objects and move

Turing Test: Pros and Cons

- **Pros**
  - Operational definition of intelligence
  - Sidestep philosophical vagueness
    - "What is consciousness?"
    - "Can a machine think?"
    - ...
- **Cons**
  - **Anthropomorphic** criteria define intelligence in human terms
    - Multiple forms of non-human intelligence exist
  - Intelligence in terms of Turing test is **fooling humans** into thinking it's human
  - E.g., aeronautical engineering is about:
    - Yes: Focus on wind tunnels and aerodynamics
    - No: Designing machines that imitate birds

4. AI as Acting Rationally

- **Rational agents**: agents that do the "right thing" given what they know
- Agents that **act rationally** should:
  1. Operate autonomously
  2. Perceive environment
  3. Persist over a prolonged time period
  4. Adapt to change
  5. Create and pursue goals

Acting Rationally as Ultimate Goal of AI

- Which definition of AI to use?
  - Acting vs. Thinking
  - Rational vs. Human
- **Acting > Thinking**
  - Acting rationally is broader than just thinking rationally
- **Rational > Human**
  - Rationality can be mathematically defined
  - Human behavior is shaped by evolutionary conditions
- AI focuses on **agents acting rationally**

Rationally is Not Absolute

- AI wants to build agents that **do the right thing**
  - What is the right thing?
- E.g., you leave the house and a branch strikes you
  - **Q**: Did you act rationally?
  - Probably
- E.g., you cross the street and a car knocks you over
  - **Q**: Did you act rationally?
  - It depends, but probably no
- E.g., moral issues with self-driving car
  - Swerve and hit a pedestrian to avoid a frontal crash that would kill 2 people

Problems of a Rational Agent

- **Probabilistic environment**
  - A rational agent aims for:
    - The best outcome in a deterministic setup
    - The best expected outcome under uncertainty
- **Best** is determined by the objective function:
  - E.g., cost function, sum of rewards, loss function, utility
- Omniscience vs **no-regrets**
  - Best based on available information
- Sometimes **no provably correct action** exists
  - Yet, an action must be taken
- Even **with perfect information** rationality can't be feasible due to:
  - Cost of acquiring all data (e.g., in medicine)
  - Computational demands
- Perfect good enough vs perfect
  - Acting appropriately ("satisficing")

- AI and Machine Learning
  - What is AI?
  - ***What is Machine Learning?***

Machine Learning: Definitions

- How to define machine learning?
- *"Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed"* (Samuel, 1959)
- **Machine learning** is about building machines to do **useful things** without being **explicitly programmed**
  - E.g. a computer learns to play checkers by playing against itself, memorizing positions that lead to winning
- *"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if $P(T)$ improves with experience E"* (Mitchell, 1998)
- E.g.,
  - Computer vision
  - Speech recognition
  - Natural language processing

Limits of ML Compared to Human Intelligence

- **AI differs from human intelligence**
  - Machines don't learn like humans (e.g., LLMs)
- **Fragility to input variations**
  - ML models fail with slight input distortions
    - Adversarial attacks cause misclassification by altering one pixel
    - A model trained for a video game may fail if the screen is slightly rotated; humans continue effortlessly
- **Lack of transfer learning**
  - ML systems cannot apply knowledge across domains without retraining
- **Massive data and compute requirements**
  - ML requires enormous datasets and computational resources
    - A teenager learns to drive in hours
    - Self-driving systems need billions of compute hours and extensive data
- **Poor common sense and reasoning**
  - ML lacks built-in world knowledge and intuitive logic

Limits of ML Compared to Human Intelligence

- **Opaque decision-making**
  - Many ML models offer little transparency into decision processes
    - Limits trust, interpretability, and accountability in critical applications
- **Dependence on narrow objectives**
  - ML systems excel at optimizing narrow tasks but fail with ambiguous goals
    - E.g., an algorithm maximizing user engagement may promote harmful content
- **Susceptibility to bias and data quality**
  - Models inherit and amplify biases in training data
- **Lack of embodiment and physical interaction**
  - Human cognition is grounded in physical and sensory experience

The 3 Machine Learning Assumptions

- In practice, ML involves solving a practical problem by:
  - Gathering a dataset
  - Building a statistical model from the dataset algorithmically
- The **three assumptions** of machine learning
  - A **pattern exists**
  - Pattern cannot be **precisely defined mathematically**
  - **Data is available**
- Which ML assumption is **essential**?
  - A pattern exists
    - If no pattern, try learning, measure effectiveness, conclude it doesn't work
  - Pattern cannot be precisely defined mathematically
    - If solution is direct, ML not recommended, but may still apply
  - Data is available
    - Without data, no progress can be made
    - **Data is crucial**

AI vs ML vs Deep-Learning

- **AI**
  - Machines programmed to reason, learn, and act in a rational way
- **ML**
  - Machines capable of performing tasks without being explicitly programmed
- **AI without ML**:
  - Example: Rule-based systems (e.g., IBM Deep Blue playing chess)
- **Deep Learning (DL)**
  - ML using neural networks with many layers
- **Large Language Models (LLM)**
  - Neural networks trained on massive text datasets and RLHS