Data Science: Machine Learning: Introduction

CPSC 501: Advanced Programming Techniques
Fall 2022

Jonathan Hudson, Ph.D
Assistant Professor (Teaching)
Department of Computer Science
University of Calgary

Monday, October 3, 2022

UNIVERSITY OF CALGARY

What is it?

Skynet?

UNIVERSITY OF
CALGARY

Artificial Intelligence

- **My definition**: take something you think only humans can do and make something computational that does it

- **Moving Target!**

- **Lots of things you take for granted were AI once.**

- Alexa/Siri/etc., google maps, biometrics, google search, automatic translation, natural language understanding, handwriting recognition, …

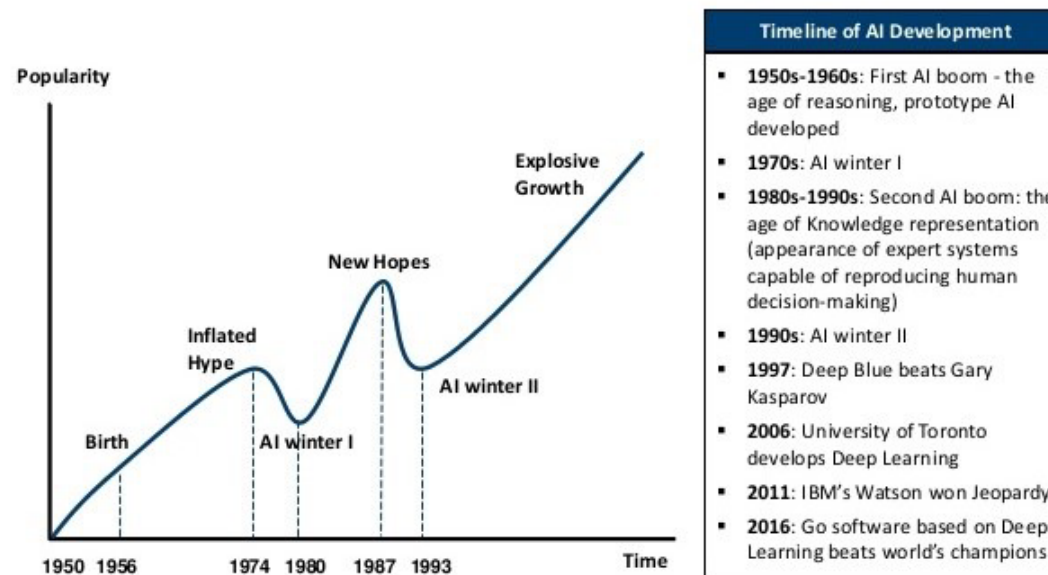- We'll talk about the history of AI, but in short the **trough of disillusionment** is historic key
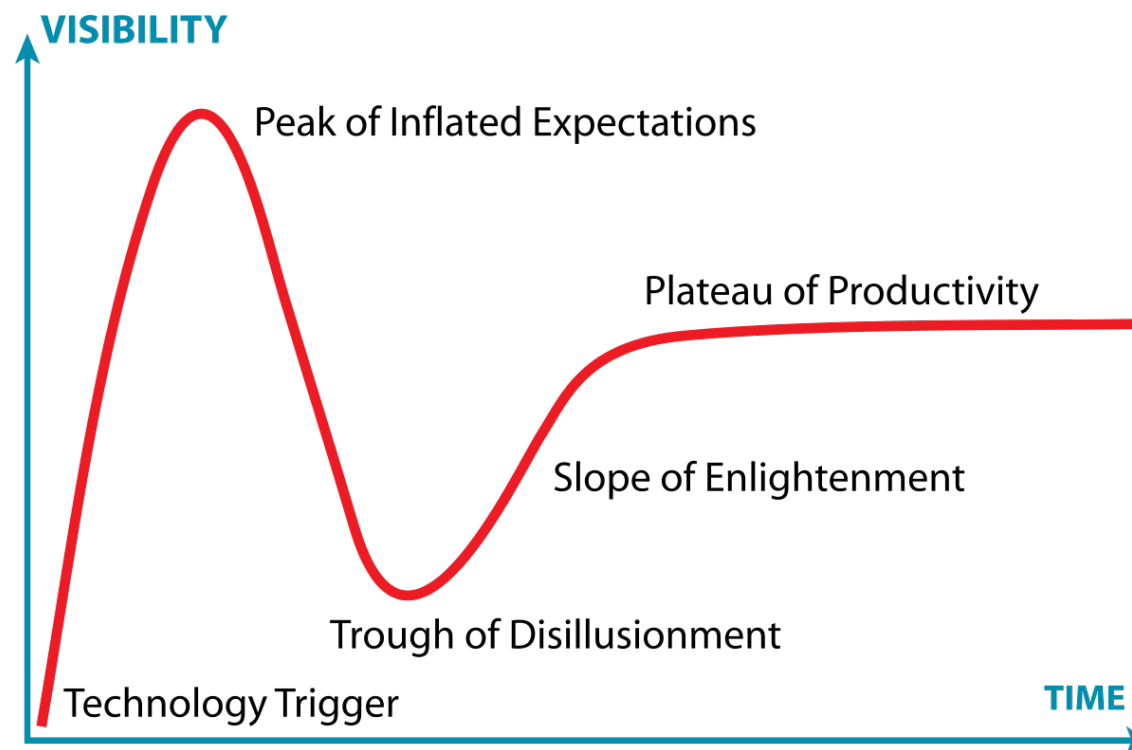
Artificial Intelligence (AI) in Computer Science

- **First attempt:** Make 'intelligent' human-like machines.
  - Example: the GPS program (General Problem Solver) -> solve formal logic problems
  - **Result: Complete failure and enormous negative reactions**
  - **First AI Winter**
  - Problems: One knowledge representation scheme with one knowledge processing mechanism not enough + search spaces are enormous.
- **Second attempt:** Take application area and make something that works
  - Example: Natural language, diagnosis, board game playing
  - Result: Worked within expectations but very specialized
  - From First AI Winter up until around 'deep learning'

AI HAS A LONG HISTORY OF BEING "THE NEXT BIG THING"...

Timeline of AI Development

- **1950s-1960s:** First AI boom - the age of reasoning, prototype AI developed
- **1970s:** AI winter I
- **1980s-1990s:** Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s:** AI winter II
- **1997:** Deep Blue beats Gary Kasparov
- **2006:** University of Toronto develops Deep Learning
- **2011:** IBM's Watson won Jeopardy
- **2016:** Go software based on Deep Learning beats world's champions

UNIVERSITY OF CALGARY

What was that Turing Test at the start?

- **'Does a machine exhibit behaviour which is indistinguishable from that of a human?'**
- Place a human interrogator is in a room with two computer terminals
  - One terminal is connected to another person
  - The other terminal is connected to an artificial intelligence
- The interrogator asks questions by typing them into the terminals
- If the interrogator cannot reliably determine which terminal connects to the human and which connects to the AI then the AI passes the test
  - The goal of the AI is not correct answers –it's answers that resemble those of a person (Eugene Goostman!)
- Based on work by Alan Turing in the 1940s and 1950s
  - 'The Imitation Game', 'Codebreaker'
  - **Father of Computer Science** (mathematician)

UNIVERSITY OF
CALGARY

Connectionist vs Symbolic AI

- Connectionist → Make something simple, like a digital brain cell, connect them together with some simple network structures, intelligence falls out? Profit

- Symbolic AI → examine physical/animal/human world, find/create abstract structures and processes, apply to application problem that fits

- Differences:
  - rely on emergence of capability (conn.) or bake it in (symb)
  - flexible to a degree (conn.) or often can't think beyond what is baked in (symb)
  - great with learning patterns (conn.) or better at less common (symb.)

UNIVERSITY OF
CALGARY

Connectionist vs Symbolic AI
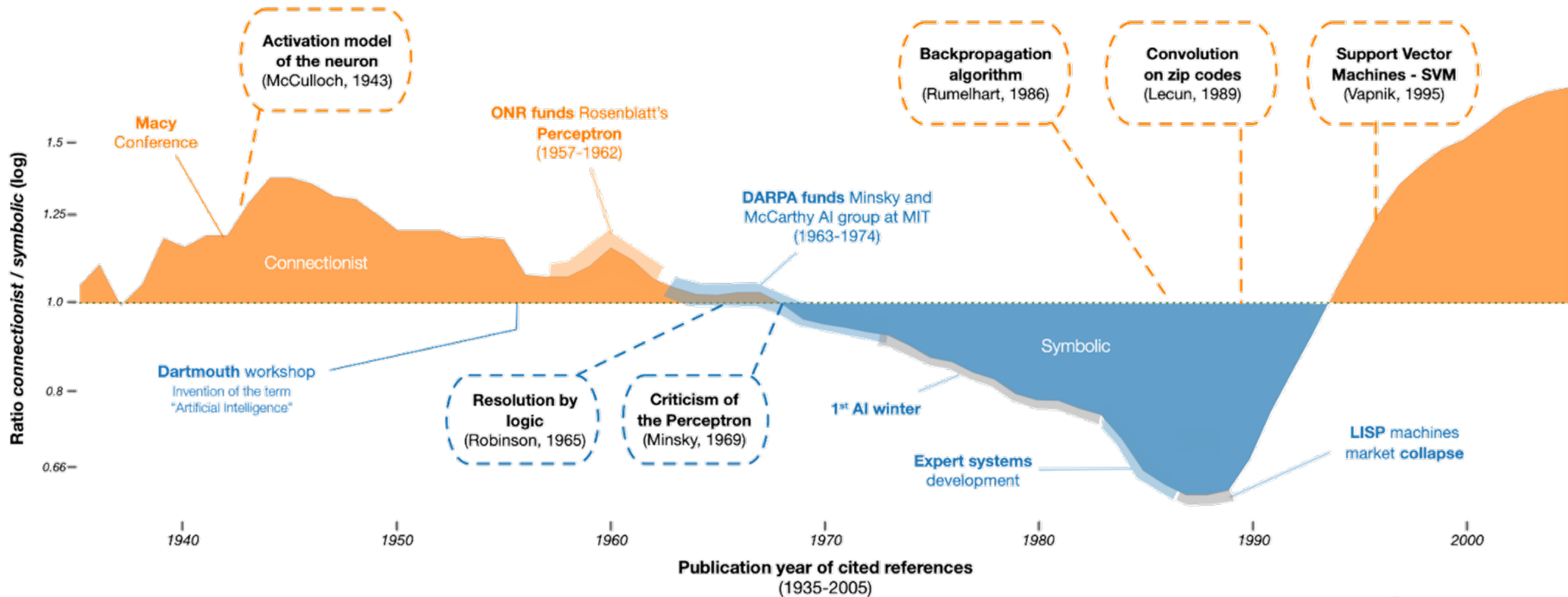
,Cardon, D., Cointet, J.-P. & Mazieres, A. (2018). Neurons spike back. The invention of inductive machines and the artificial intelligence controversy. *Réseaux*, 36(211), 173-220.

UNIVERSITY OF CALGARY

Current
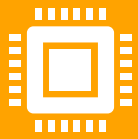
After second AI Winter computer science made a lot of progress on taking the specialized system on the symbolic side and generalizing their ideas

On the other side our computing power reached the point that connectionist became powerful again

Current

- Popular areas: Deep learning (neural networks), machine learning (AI for finding patterns), digital assistants (connect symbolic systems), service architectures, internet of things, self-automation, etc.

- Things likely less close then you are sold: universal self-driving cars (visual identification has a lots of weaknesses), machine intelligence (?), true generic systems (we've generalized but there's usually a lot more specificity under the hood than you think)

- Machine Learning: In many ways is used as a modern term for AI by many. Really just means machine changes/adds knowledge with less human influence.
  - in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data.

UNIVERSITY OF
CALGARY

Machine Learning

- Around since Optical Character recognition research

- Prominence in 1990s with spam filters

- Good at seeing patterns in images, classifying things with many parts (text), filtering by pattern, chat systems, modeling complex things like, voice recognition, etc.

UNIVERSITY OF
CALGARY

Machine Learning

- Categories

1. Supervised Learning : Machine is given guidance (often based on pre-labeling training data)
   - Common examples are classifying something under labels, or finding parameters for a model that best matches data

2. Unsupervised Learning: Machine is left unguided and has to decide most everything (often minor guidance like # of things to find)
   - Common example is clustering

- Of course there are hybrid variants (semi-supervised)

UNIVERSITY OF
CALGARY

Symbolic: Search

Brains?

UNIVERSITY OF
CALGARY

What is Search?

- The process of navigating from a start state to a goal state by transitioning through intermediate states
  - Determine states (**knowledge representation can be really hard, 'baked in'**)
  - The transitions can be limited
    - It may not be possible to move directly from any state to any other state
- A search model consists of
  - The set of states
  - The set of transitions between the states
- Search control
  - Determines which transition is traversed to move to the next state
  - Simple types: breadth first search, depth first search, hill climbing, A*, alpha-beta pruning, minimax, branch and bound, …

UNIVERSITY OF
CALGARY

Search Example

- Define a problem clearly
    - Zombie Apocalypse! I'm looting an abandoned house.
    - I see five items I, each item $i \in I$ has a weight w_i and a value v_i.
    - I have a backpack with a weight limit W
    - I can't fit all the items in my backpack but I can select some $S \subseteq I$.
    - I want to maximize the value of the items I take without breaking my backpack!
    - That is pick an S to maximize $\sum_{i \in S} w_i \leq W$.

- One way to view the problem is to think of all the possible solutions S.

- I could pick items 1,2,3 or items 4,7.

- How many of these solutions exist?

UNIVERSITY OF
CALGARY

Search Example (cont'd)

- One way to view the problem is to think of all the possible solutions S.

- I could pick items 1,2,3 or items 4,5.

- How many of these solutions exist?

- Binary notation. Items 1,2,3 can be 11100 and 4,5 can be 00011.

- We are using a 1 to indicate taken, and 0 to indicate not taken.

- So there are $2 \times 2 \times 2 \times 2 \times 2 = 2^5$ different solutions.

- This is known as our **search space** (set of all possible solutions).

- **Exponential!!!** (doubles in size each item more we have to consider.)

UNIVERSITY OF
CALGARY

Search Example (cont'd)

- Binary notation. Items 1,2,3 can be 11100 and 4,5 can be 00011.

- We are using a 1 to indicate taken, and 0 to indicate not taken.

- So there are $2 \times 2 \times 2 \times 2 \times 2 = 2^5$ different solutions.

- This is known as our **search space** (set of all possible solutions).

- Exponential!!! (doubles in size each item more we have to consider.)

- **Brute force $O\left(2^{|I|}\right)$**

- **Dynamic Programming $O\left(|I| \times W\right)$**

- Even DP doesn't scale great if W is large

UNIVERSITY OF
CALGARY

What if we were happy with a pretty good solution?

(Zombies are breaking down the door!)

UNIVERSITY OF CALGARY

Heuristic Symbolic Search

- Instead of guaranteed exploration there are a number of symbolic techniques that cannot guarantee us the best answer (ex. for that 0/1 knapsack problem example)

- But, we generally **get an rather good answer quickly**
  - Lots of broadly used AI falls into here

- Based on **heuristics** (rules that seem to work)
  - Specific rules like: If I want to go to Mountains from Calgary I likely want to choose roads that go West!
  - Or more abstract: using biological or physical inspirations for search algorithm design

UNIVERSITY OF
CALGARY

Heuristic Symbolic Search

- **Physical** Patterns examples:
    - Simulated Annealing : Cooling of metal slowly to find good state
    - Particle Swarm Optimization : gravity and energy like behaviour of particles

- **Biological** Patterns examples:
    - Genetic Algorithms : DNA and Survival of the fittest
    - Ant Colony Optimization : colony exploratory properties

UNIVERSITY OF CALGARY

Symbolic: Hill-Climbing

Zombie brain?
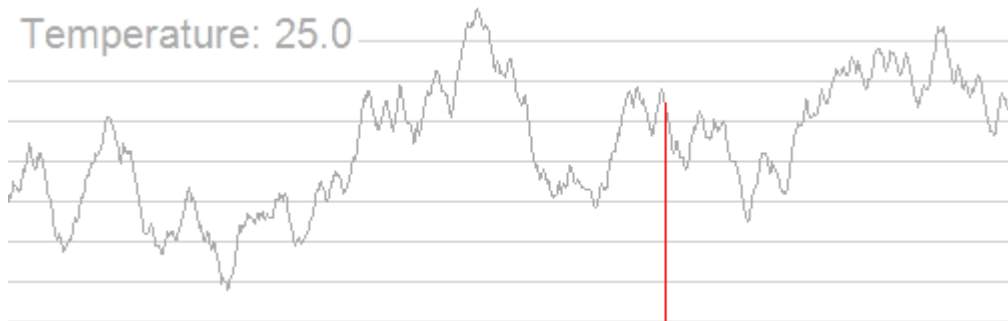
UNIVERSITY OF CALGARY

Hill-Climbing
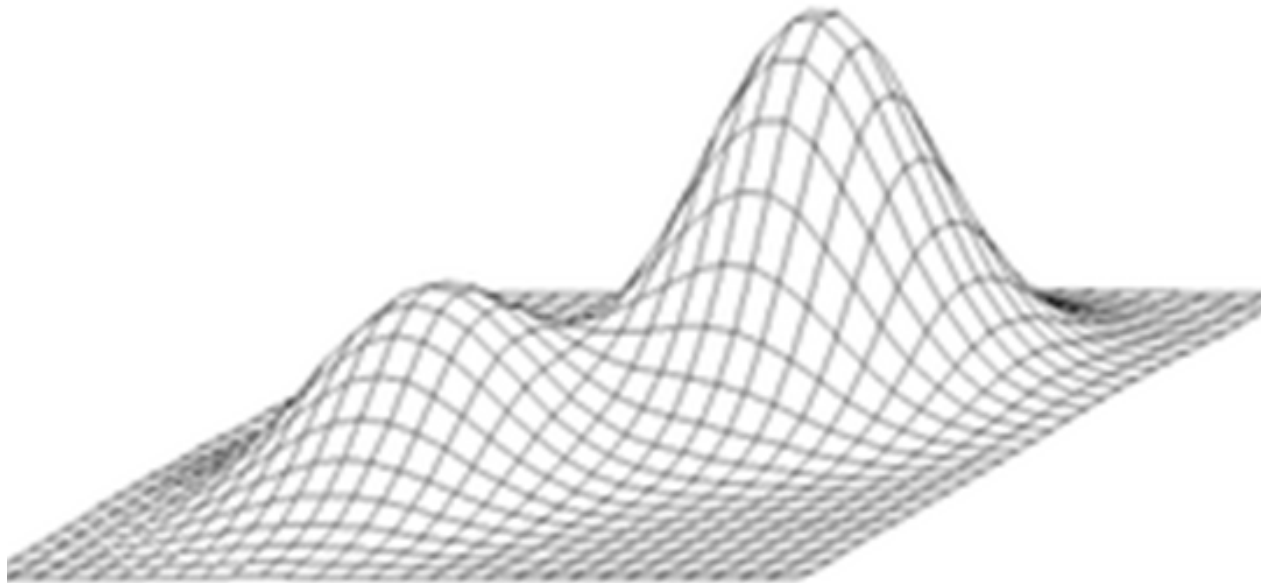
- Simple search
- Greedy style algorithm (CPSC 413)
- Have one solution
- Make a simple change (if things get better, then keep change)
  - Put/swap something in backpack

- Challenges with local optima

Symbolic: Simulated Annealing

Smarter brain?

UNIVERSITY OF
CALGARY

Simulated Annealing

- Have one solution through the whole process

- Solution starts out random and you track a Temperature value

- Each iteration
  - Propose a new solution
  - If it is better, then keep it
  - If it is worse, then keep it only if its loss of value is below temperature (sometimes)
    - In practice pick random probability
    - Calculate a difference in value term (that decreases with temp)
    - Only accept worse if first less than second
  - Temperature decreases (cools)

- Idea, we know local optimum can trap us, so we get worse sometimes

- Parameters touchy. How much to update? Not a ton of exploration.

Temperature: 25.0

Symbolic: Particle Swarm Optimization

Everyone go looking for brains?

UNIVERSITY OF
CALGARY
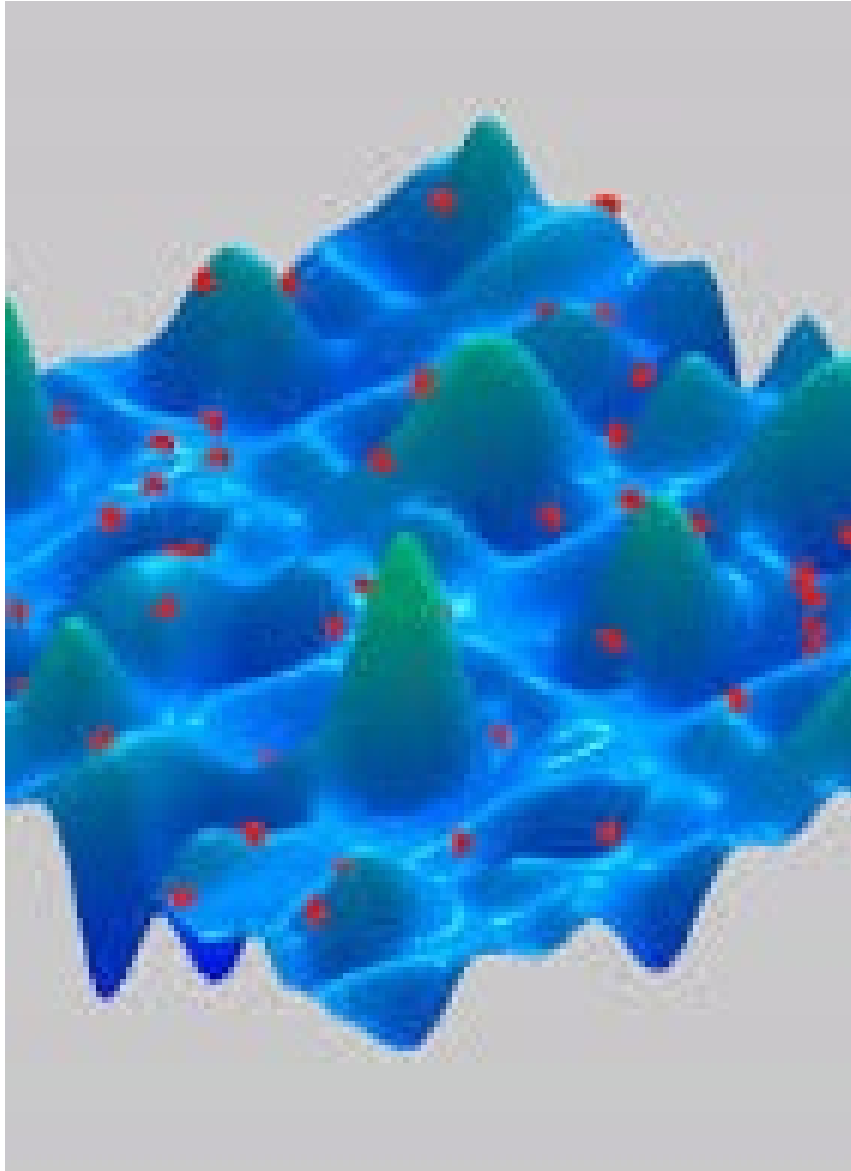
Particle Swarm Optimization

- Design complexity grows.

- Think of particles as having 'gravity'. The better the solution the more 'gravity'.

- Particles also have momentum.

- Have many particles.

- Each step, particles follow their current direction of change with influence of the nearby local optima and global optima.

- Less touchy to parameters and good at exploration. Often cooling principle included to help find best at end.

- Challenges with discrete problems.

Symbolic: Genetic Algorithms

People brains?

UNIVERSITY OF
CALGARY

Genetic Algorithms

- Designed around the concept of DNA, biological populations, and survival if the fittest.

- Good for discrete as DNA is more discrete form.

- Also the idea of adding, removing, swapping like DNA is a more natural operation.

- Make a population of solutions.

- Over iterations make more solutions by selecting 'parents' based on fitness (solution value).

- Then make 'children' based on sharing chunks of DNA (crossover) or by change single DNA spots (mutation).
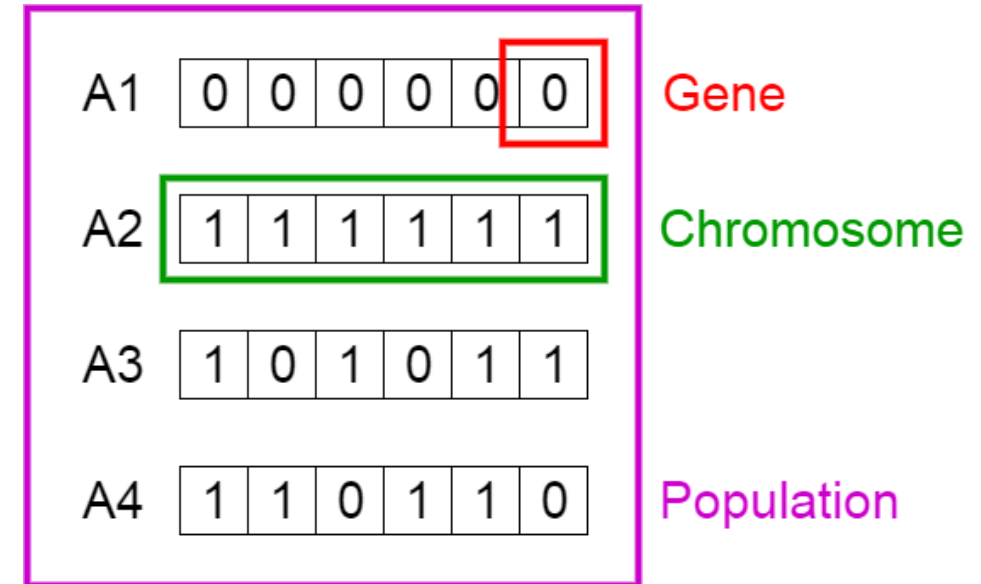
Connectionist: Neural Networks

Actually brains?

UNIVERSITY OF CALGARY

What are neural networks?

- Inspired by the Human Brain.

- The human brain has about 86 Billion neurons and requires 20% of your body's energy to function.

- These neurons are connected to between 100 Trillion to 1 Quadrillion synapses!

- Deep learning neural networks really popular right now

- Connectionist method
  - Make network, train, hope result is useful

UNIVERSITY OF
CALGARY

Neuron Model of Connections
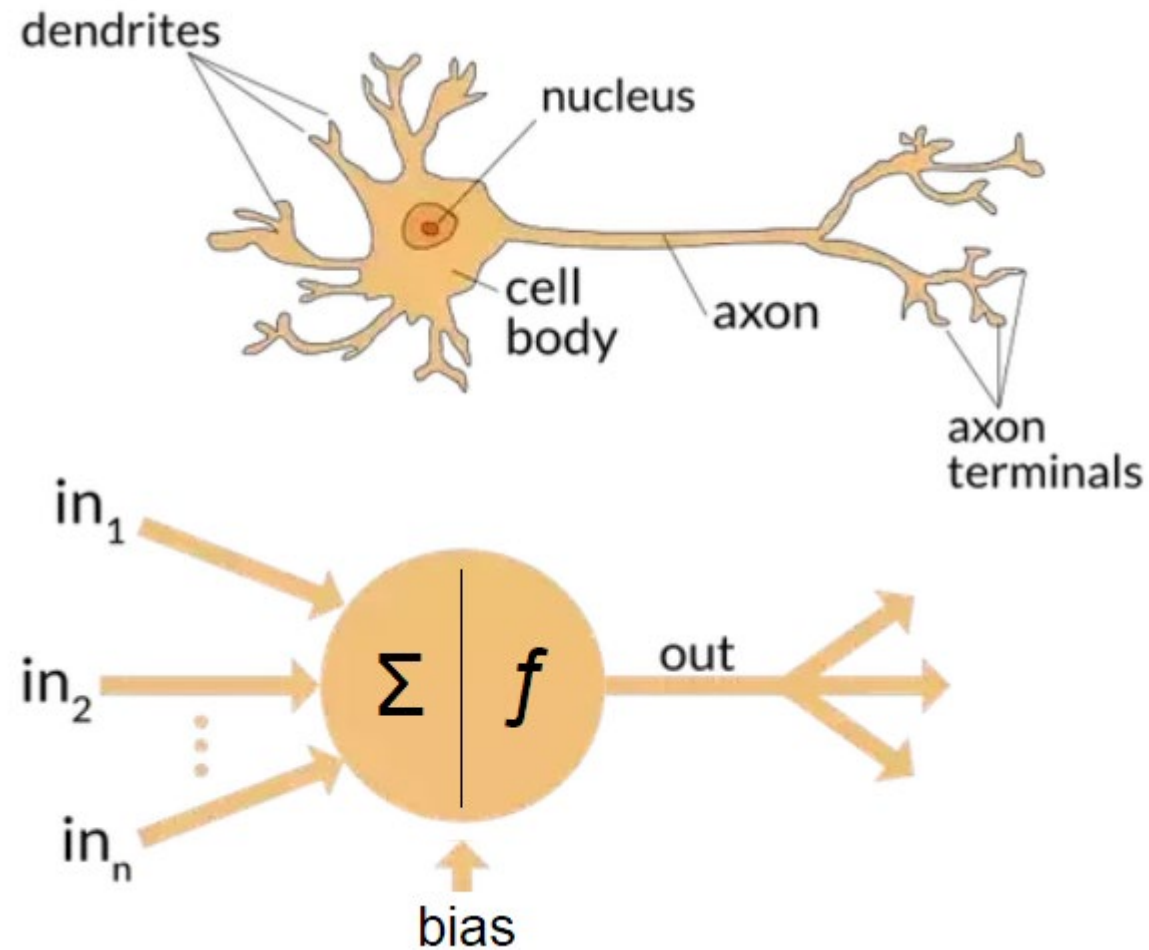
Summary

- AI is a moving target of making computer do previously thought to be human-like things.

- Machine learning is mostly a modern term for AI, that means AI learns by itself.

- Search spaces can get really big, so heuristic techniques are symbolic ways of exploring the space quicker but don't guarantee best solution.

- Connectionist AI like deep learning neural networks are currently really popular (although balance likely in the middle).
  - i.e. humans have brains but within our brains and between brains we've developed and communicated symbolic structures to help process the world

UNIVERSITY OF CALGARY

Onward to ...
scikit.

Jonathan Hudson
jwhudson@ucalgary.ca
https://pages.cpsc.ucalgary.ca/~jwhudson/

UNIVERSITY OF
CALGARY