What Is AI?

Module 2 of a course on *Ethical Issues in AI*

Prepared by

John Hooker

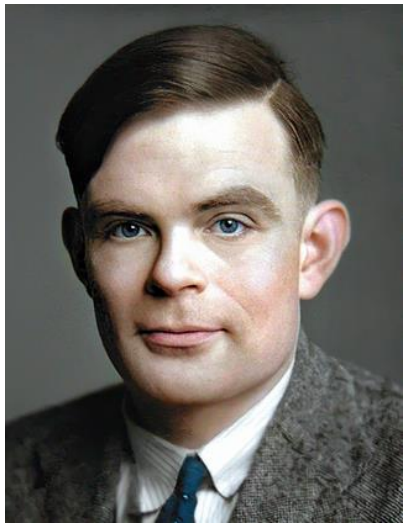
Emeritus Professor, Carnegie Mellon University

CMU Osher, January 2025

What is the essence of AI?

- Technology that enables machines to **simulate human intelligence**.
  - *Idea behind the **Turing test**.*
    - But this doesn't tell us what human intelligence **is**.

Alan Turing
1912-1954



What is the essence of AI?

- Technology that can solve **unstructured problems**.
  - *According to **Herbert Simon**, one of the founders of AI.*
    - Few, if any, AI applications achieve this
    - We now speak of **AGI** (artificial general intelligence) as the **next** goal.

Herbert Simon, 1916-2001

AI hype

- Decades of overpromising
  - ***1960**: Herbert Simon predicts:*
    - "Machines will be capable, within **20 years**, of **doing any work a man can do**."
  - ***1970s**: Marvin Minsky predicts:*
    - "In from **3 to 8 years** we will have a machine with the **general intelligence of an average human being**."
  - ***1993: Vernor Vinge predicts:***
    - "Within **30 years**, we will have the technological means to create **superhuman intelligence**. Shortly thereafter, the human era will be ended."

AI hype

- Claiming to use AI can **boost stock price**
  - *But it may be existing technology in a new wrapper.*

SC Media
A CRA Resource

What is 'AI washing?' Companies pay $400K to SEC for inflated claims

Laura French March 19, 2024

AI hype

- Tech fads hyped to MBAs often fizzle financially:

What happened to these?

Virtual reality
Wearables
Crypto
NFTs
Web3
Big data
IoT
Virtual assistants

AI hype

- Tech fads hyped to MBAs often fizzle financially:

What happened to these?

Virtual reality
Wearables
Crypto
NFTs
Web3
Big data
IoT
Virtual assistants

These took off

Email
World Wide Web
B2C
Smart phones
Social media*

***But in the wrong direction?**

Beneath the hype

- AI has achieved some remarkable successes:

Image processing
Language translation
Speech recognition
Pattern recognition
(e.g., medicine)
Recommender systems
Fraud detection
Robotics
(certain applications)

Beneath the hype

- No one can predict future technology

The Flying Automobile to Be *the* Car of the Future

We were supposed to have these long ago

AI technologies

- AI is fundamentally a combination of **statistics** and **optimization**.
  - *Implemented in code (e.g., Python)*
  - *A variety of technologies*

AI technologies

- AI is fundamentally a combination of **statistics** and **optimization**.
  - *Implemented in code (e.g., Python)*
  - *A variety of technologies*
- Best known is **machine learning**
  - *Neural networks ("deep learning")*
  - *Large language models (ChatGPT, Claude, CoPilot, etc.)*
- But there are many others…

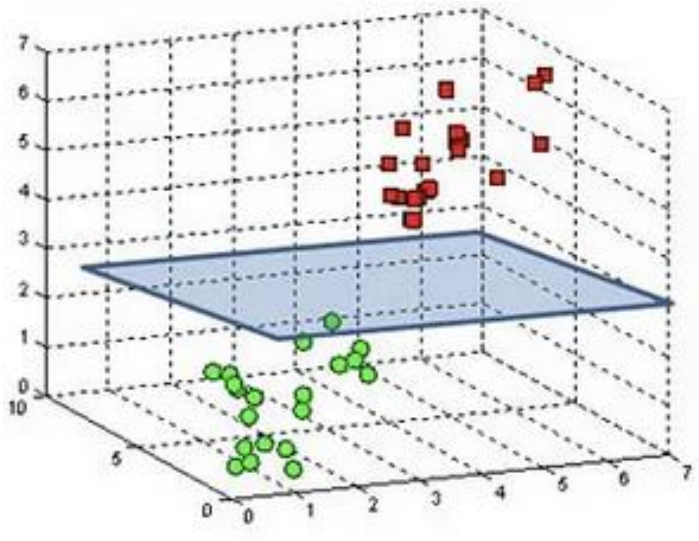
AI technologies

- There are many others…

Example:
Support vector machines

Used for **cancer diagnosis** etc.

An optimization method based on hyperplane separation.



AI technologies

- There are many others…
  - *Major AI conferences receive 8000-12,000 paper submissions each year.*

Neural networks	**Transformer models**	**Image recognition**
Convolutional NNs	**Principal component analysis**	**Facial recognition**
Recurrent NNs		**Computer vision**
k-means clustering	**Singular value decomposition**	**Speech recognition**
Decision trees		**Recommender systems**
Q-learning	**Generative AI**	**Automated planning**
Support vector machines	**Generative adversarial networks**	**Robotics**
Knowledge representation		**Virtual agents**
Optimization	**Large language models**	**Internet of things**
Evolutionary computation	**Natural language processing**	**Inverse reinforcement learning**
Markov decision processes		
Causal networks	**Speech synthesis**	**Autoregression**
Bayesian inference	**Formal logic**	**Generative pre-trained transformers (GPTs)**
Reverse Bayesian inference	**AI-optimized hardware**	

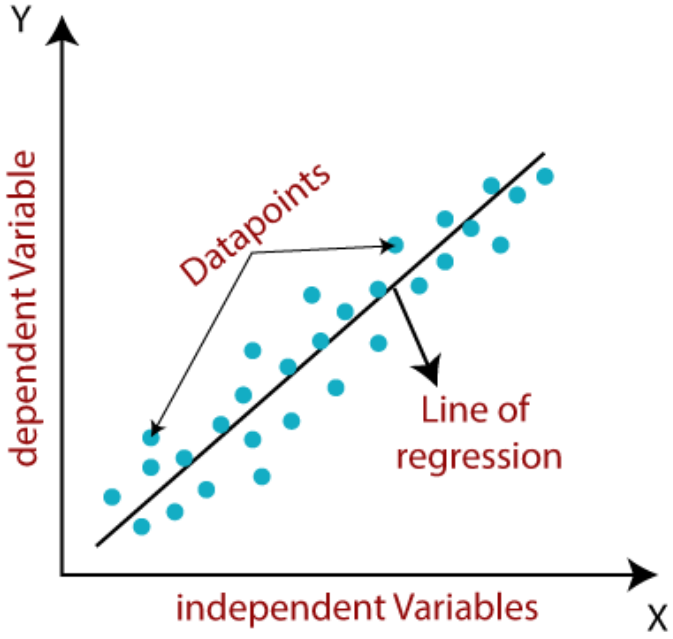
ML is not magic

- Machine “learning” is **statistical data fitting**

This is a **neural network**…
with **one neuron**.

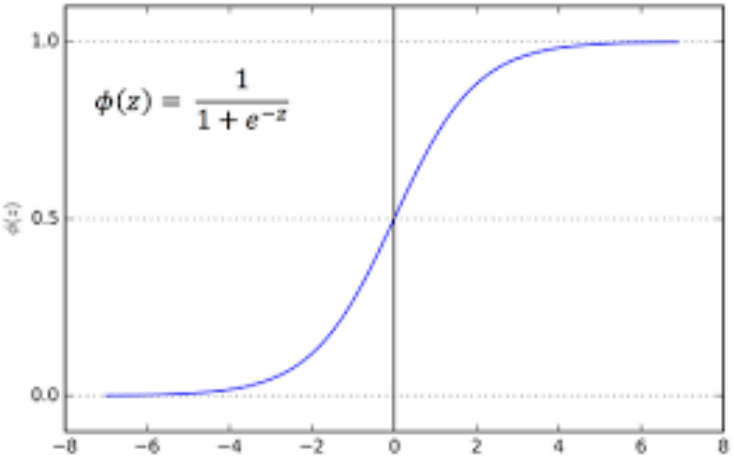
Fit a line $y = ax + b$ to data.

“**Learn**” *a* and *b* by solving
an **optimization problem**
(least squared error)

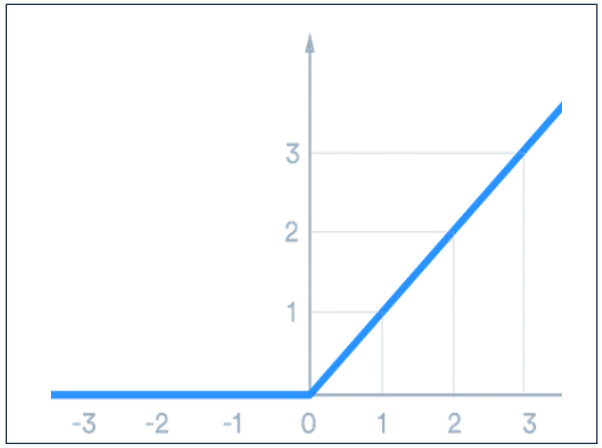


ML is not magic

- Machine “learning” is **statistical data fitting**
  - *ML normally prefers a nonlinear fit:*



Sigmoid function
Used in early days

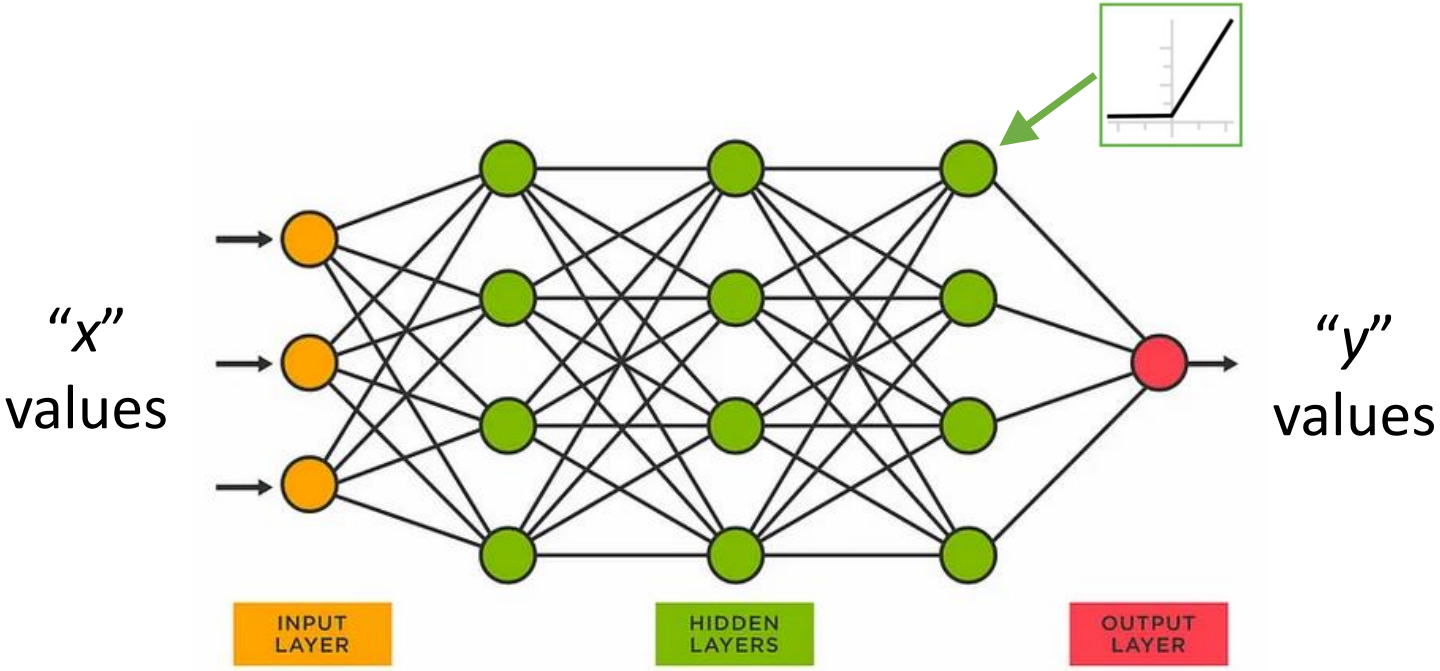


ReLU function*
Popular today

*Rectified Linear Unit

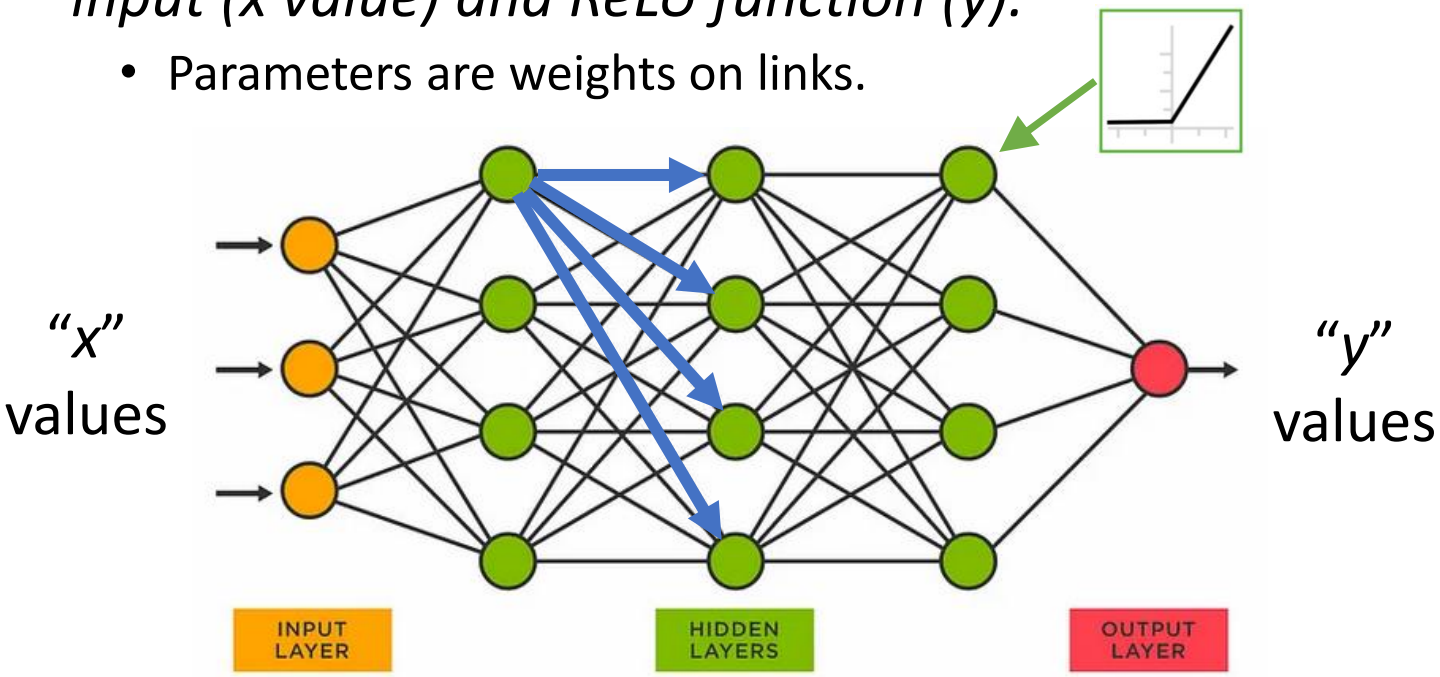
ML is not magic

- Machine “learning” is **statistical data fitting**
  - *ReLUs, etc., are linked together in a “neural”* ***network***



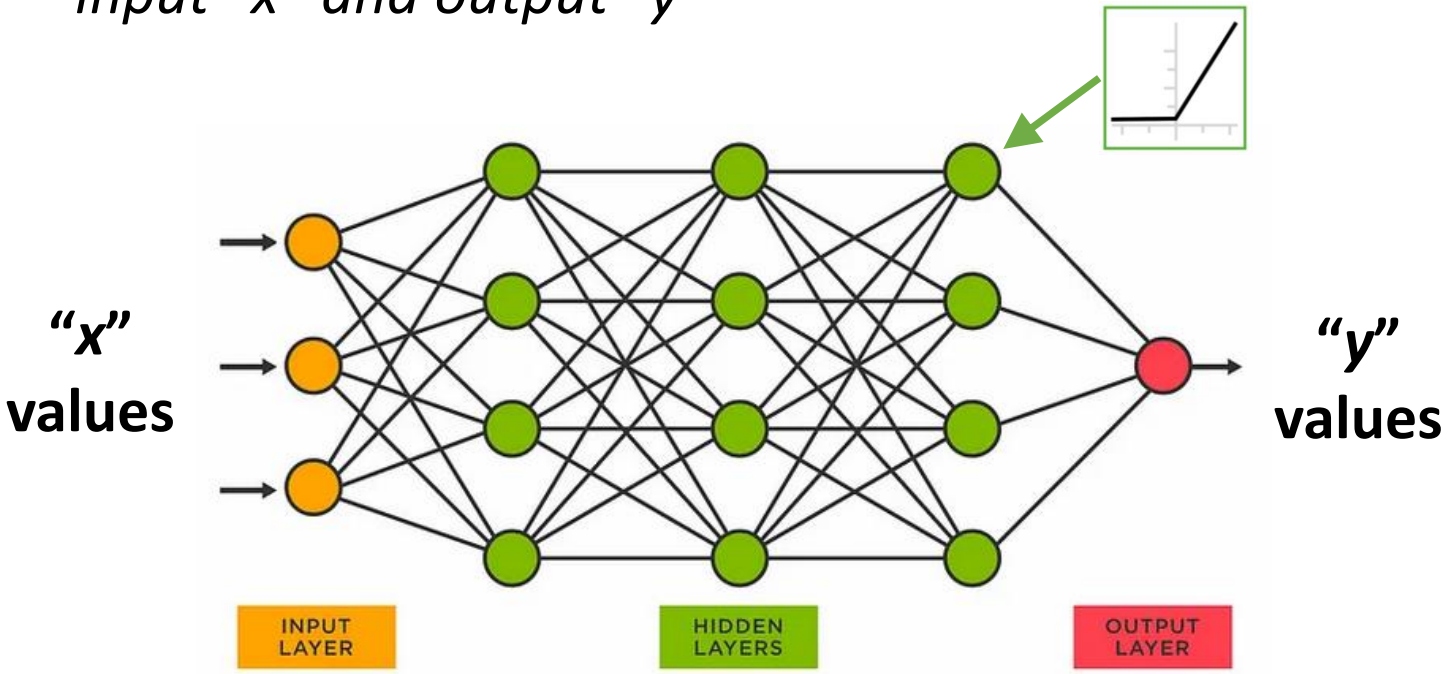
ML is not magic

- Machine "learning" is **statistical data fitting**
  - *Each neuron sends a signal determined by its input (x value) and ReLU function (y).*
    - Parameters are weights on links.



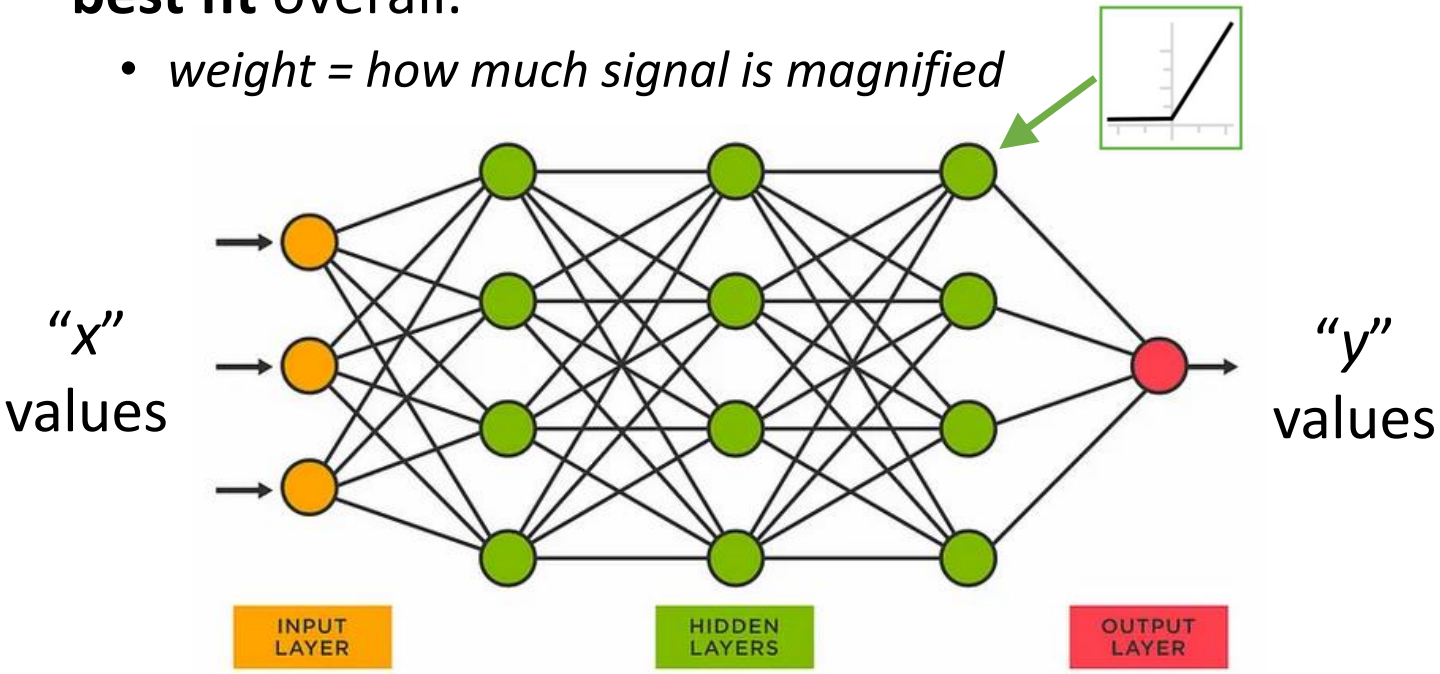
ML is not magic

- Machine “learning” is **statistical data fitting**
  - *We want to “learn” the relationship between input “x” and output “y”*



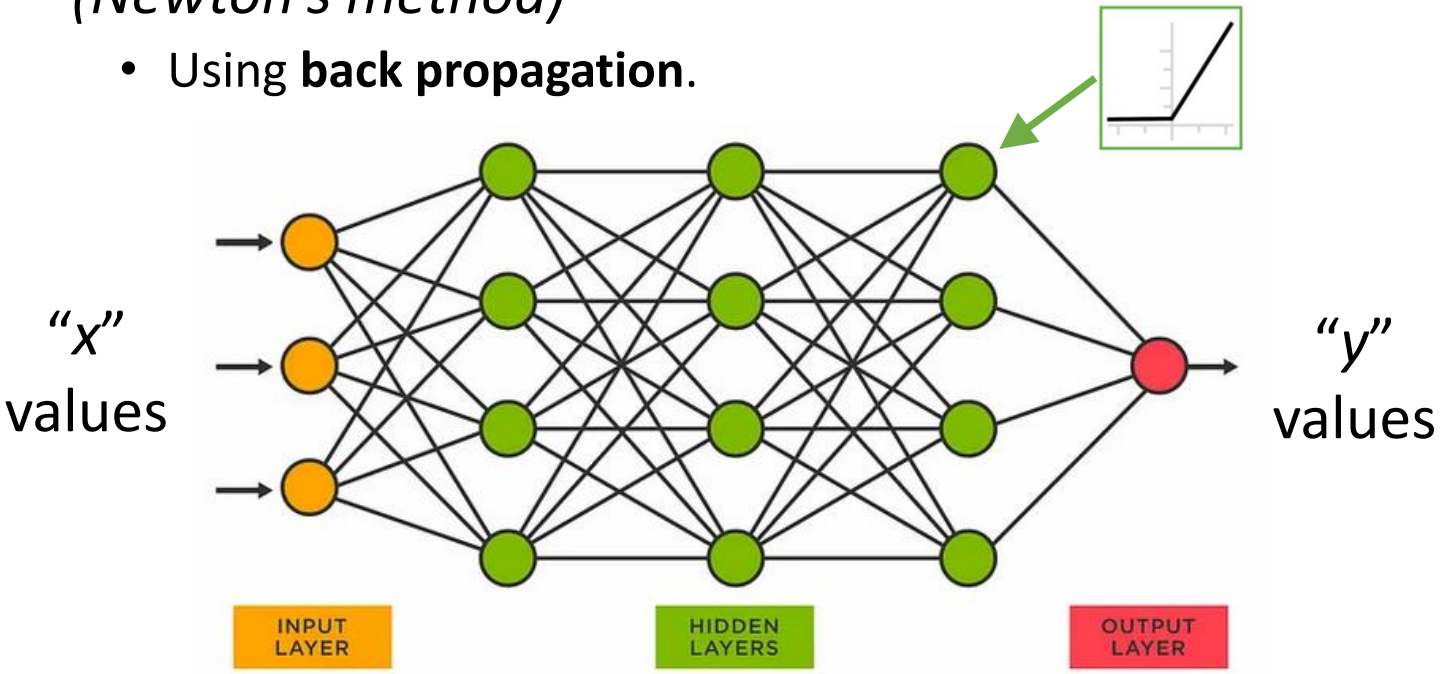
ML is not magic

- Machine “learning” is **statistical data fitting**
  - So we find the weights on links that give the **best fit** overall.
    - *weight = how much signal is magnified*



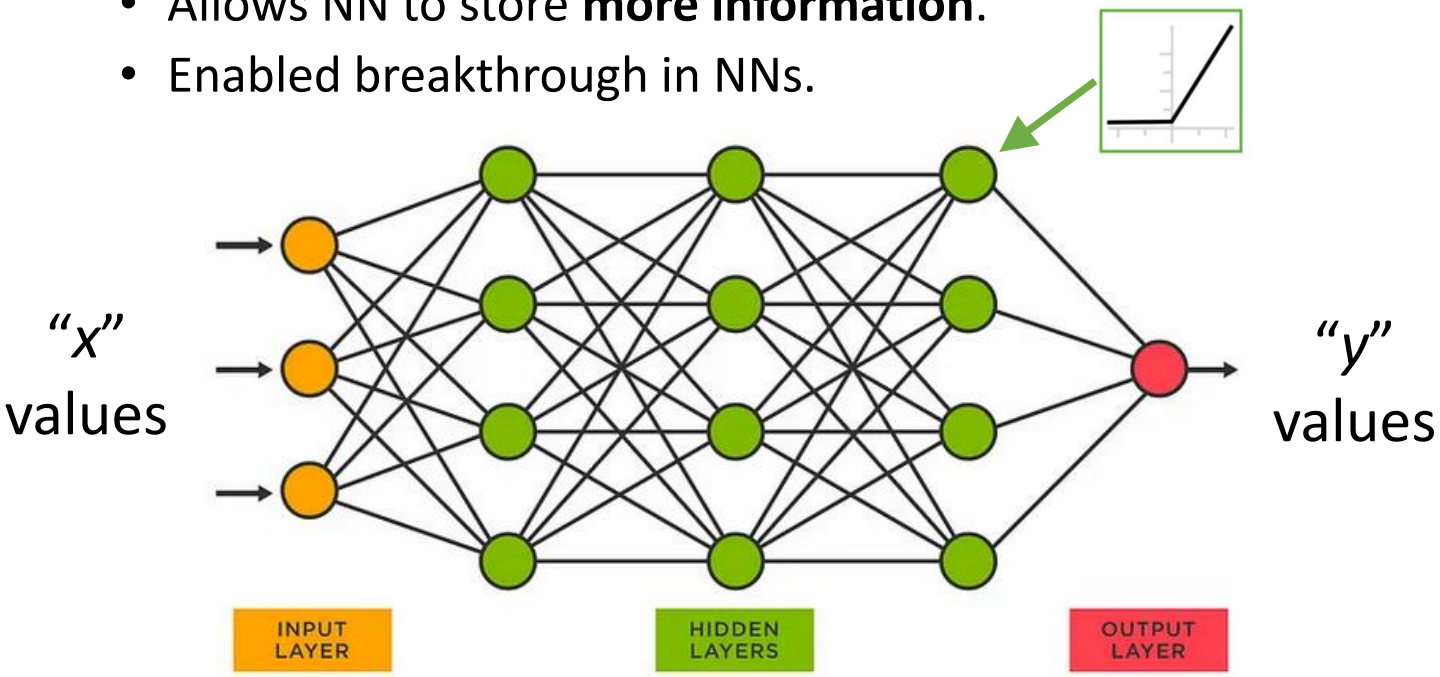
ML is not magic

- Machine “learning” is **statistical data fitting**
  - *We find best fit with a **gradient descent** algorithm (Newton’s method)*
    - Using **back propagation**.



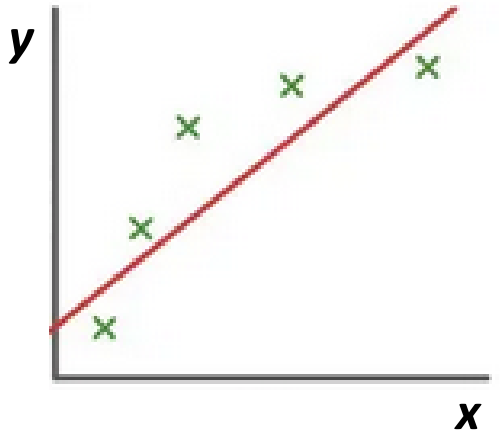
ML is not magic

- Machine “learning” is **statistical data fitting**
  - *“**Deep learning**” = many layers*
    - Allows NN to store **more information**.
    - Enabled breakthrough in NNs.



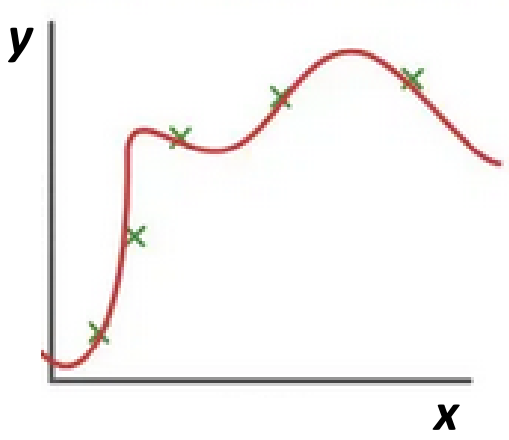
Overfitting

- This is a **no-no** in classical statistics.
  - *Too many parameters capture **random** variations*
    - and miss the **overall pattern**.



$y = ax + b$

2 parameters (too few)

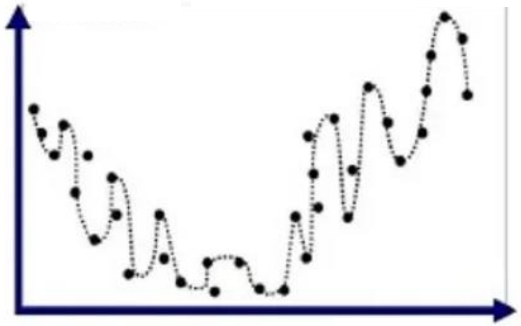


$y = ax + bx^2 + cx^3 + dx^4 + ex^5$

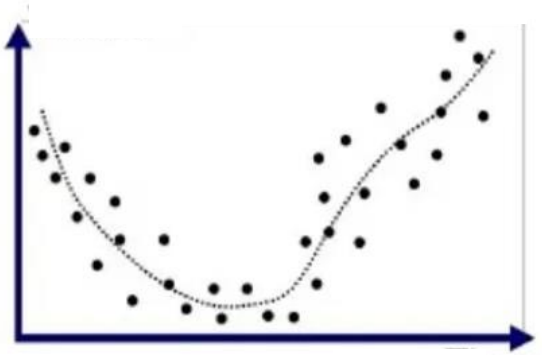
5 parameters (too many)

Overfitting

- NNs may use **billions** of parameters
  - **More** than the number of data points.
  - Because we don't know in advance which are important.
- *This often **smooths out** the fit.*
  - Why? **Unknown** (but one can guess)



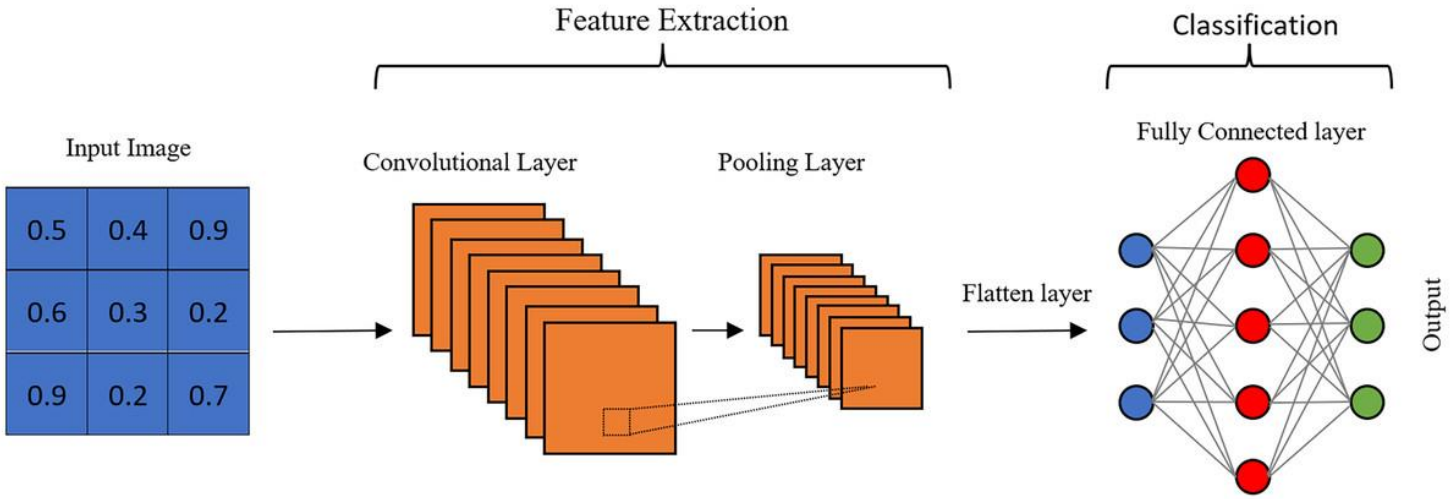
Overfit



Super overfit

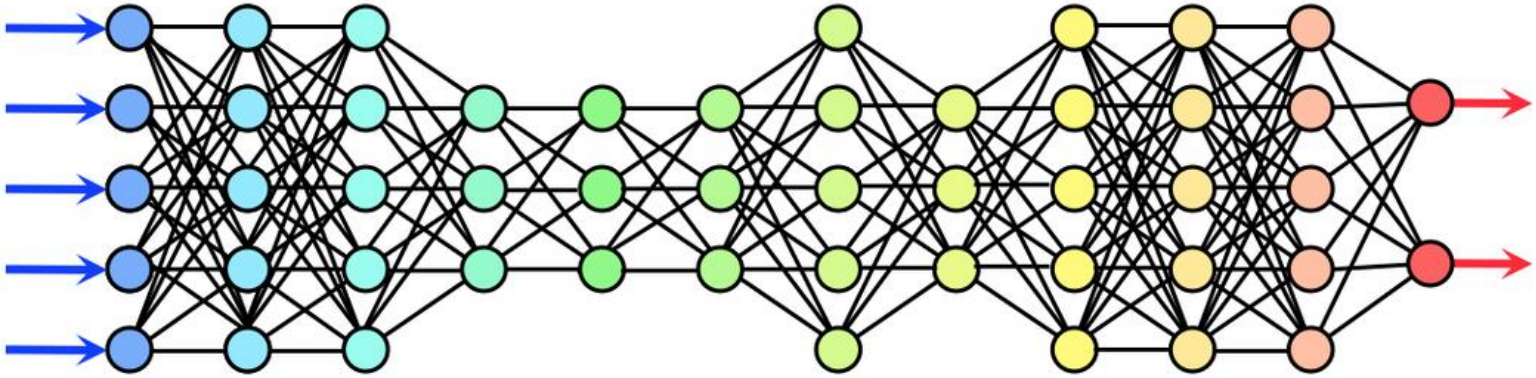
Trial and error

- The first NN you try probably **won't work**.
  - *It may require **hand coding**.*
    - As in **convolutional** NNs for image processing.
    - Some layers consist of fixed code, **no learning**.



Trial and error

- The first NN you try probably **won't work**.
  - *It may require **1000s of trials** to get the **right design***
    - Also to address **numerical problems** in gradient descent.
    - Requires **enormous computational power** ("compute")
    - Trials run **in parallel** on many computers.



How about ChatGPT, etc.?

- They aren’t magic, either.
  - *More on these later…*
    - **LLM**s (Large Language Models)
    - **Generative AI**
    - **GAN**s (Generative Adversarial Networks)
    - **Transformers**
    - **GPT**s (Generative Pre-trained Transformers)