What Is AI?

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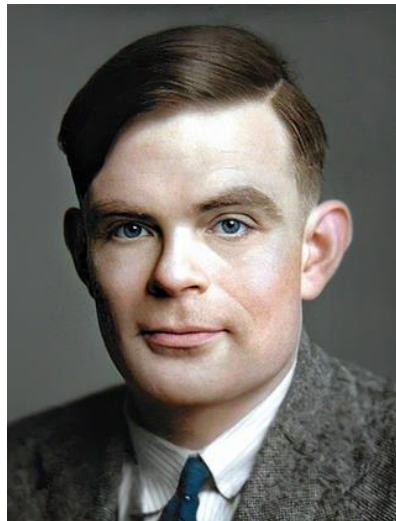
Prepared by

John Hooker
Emeritus Professor, Carnegie Mellon University

Chautauqua, June 2024

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

- Some fake or exaggerated claims.

The Register

Tesla faked self-driving demo, Autopilot engineer testifies

Brandon Vigliarolo   Wed 18 Jan 2023 // 20:29 UTC

TechCrunch

Google's best Gemini demo was faked

Devin Coldewey / 2:06 PM PST • December 7, 2023

The New York Times

Google's A.I. Search Errors Cause a Furor Online

The company's latest A.I. search feature has erroneously told users to eat glue and rocks, provoking a backlash among users.

May 24, 2024

LIVESCIENCE

GPT-4 didn't ace the bar exam after all, MIT research suggests — it didn't even break the 70th percentile

News   By Ben Turner published May 31, 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?

These took off

Virtual reality
Wearables
Crypto
NFTs
Web3
Big data
IoT
Virtual assistants

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

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.
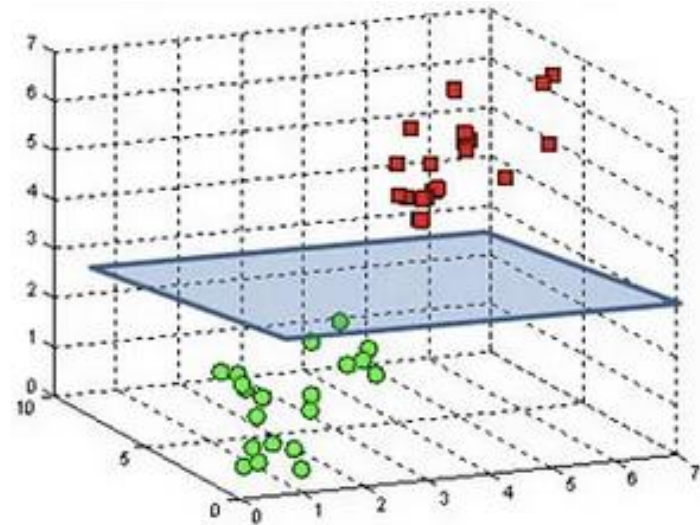
AI technologies

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

ML is not magic

- Machine "learning" is **statistical data fitting**

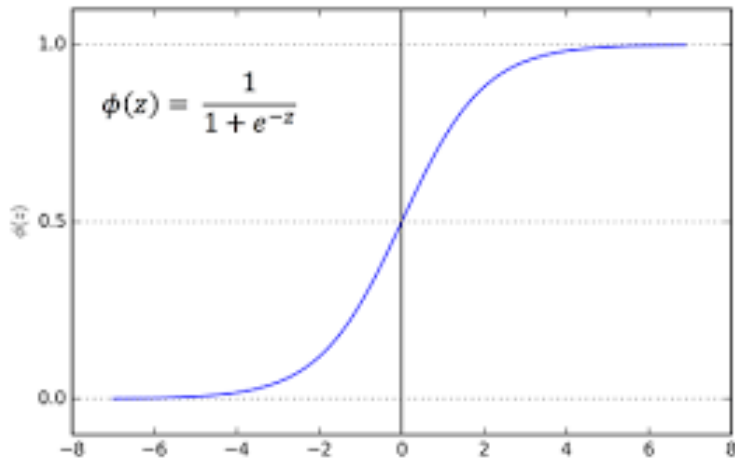This is a **neural network**… with **one neuron**.

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

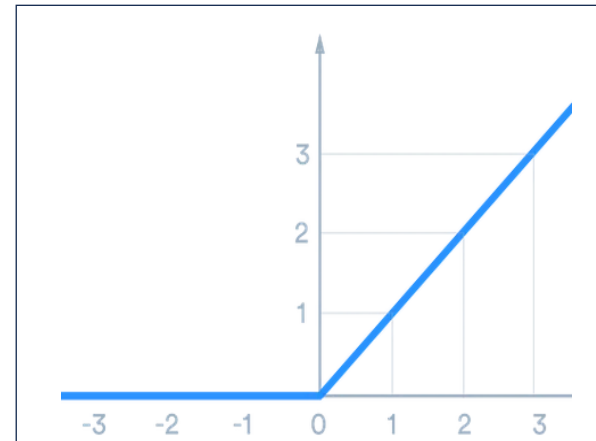"**Learn**" a and b by solving an **optimization problem** (least squared error)
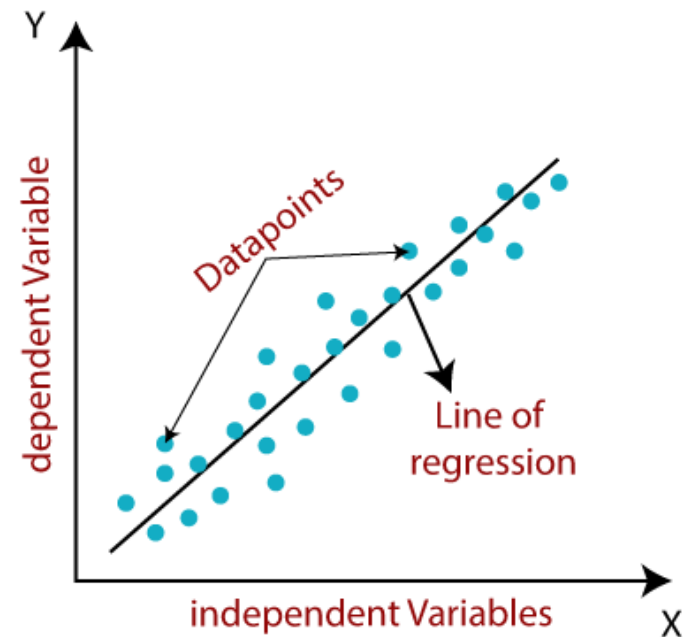
ML is not magic

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

$$\phi(z) = \frac{1}{1 + e^{-z}}$$

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***

"*x*" values

"*y*" values

INPUT LAYER

HIDDEN LAYERS

OUTPUT LAYER
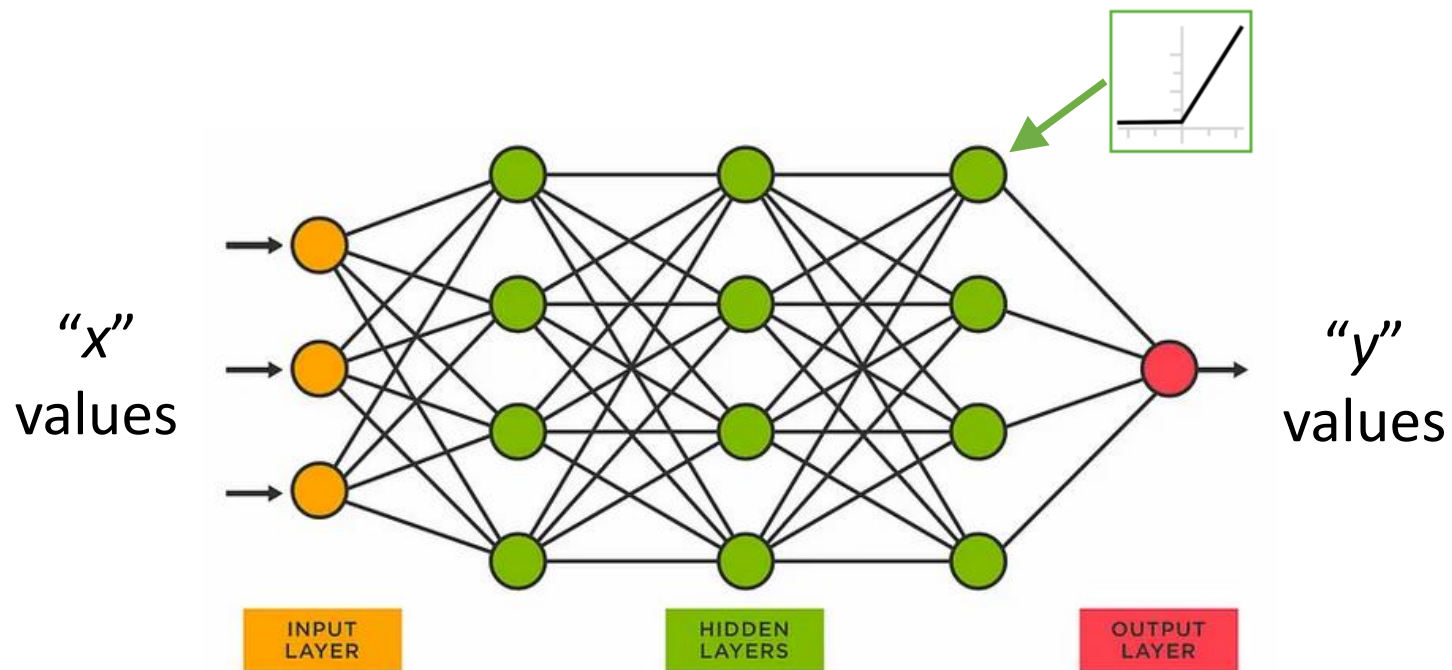
ML is not magic

- 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.

"*x*" values

"*y*" values

INPUT LAYER

HIDDEN LAYERS

OUTPUT LAYER
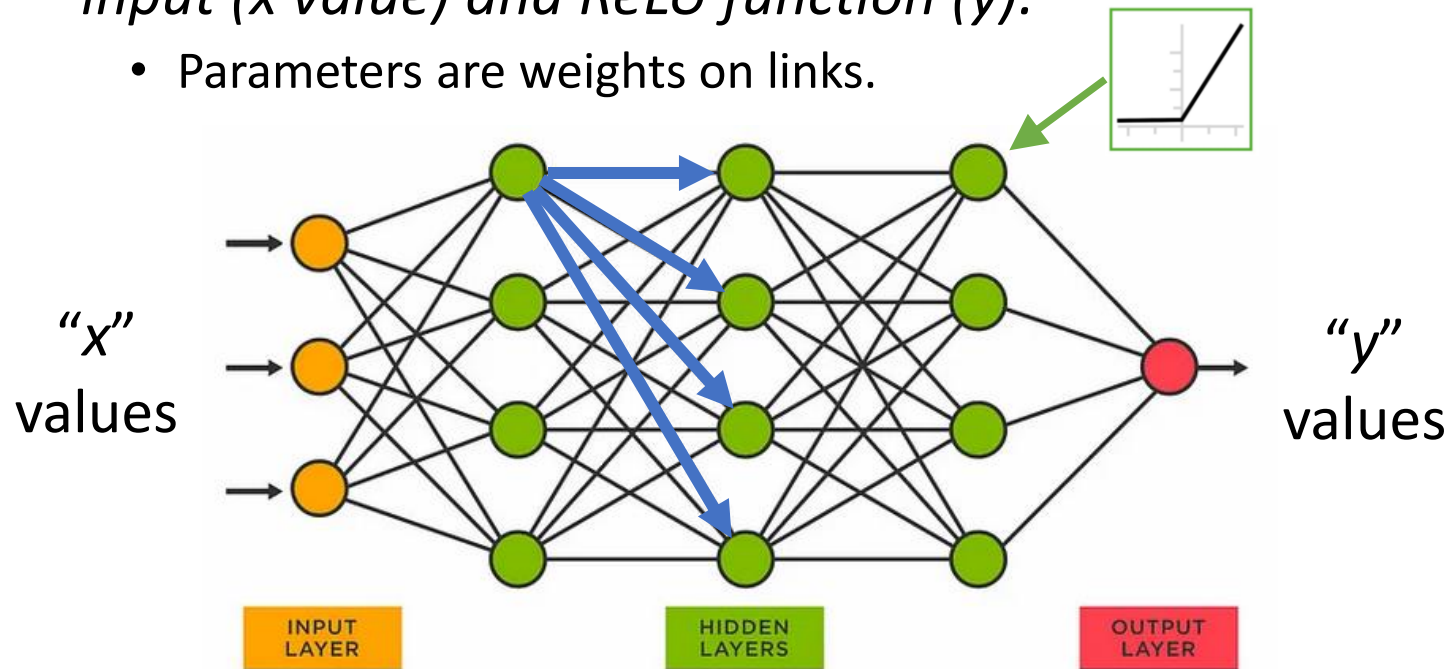
ML is not magic

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

"**x**" values

"**y**" values

INPUT LAYER

HIDDEN LAYERS

OUTPUT LAYER
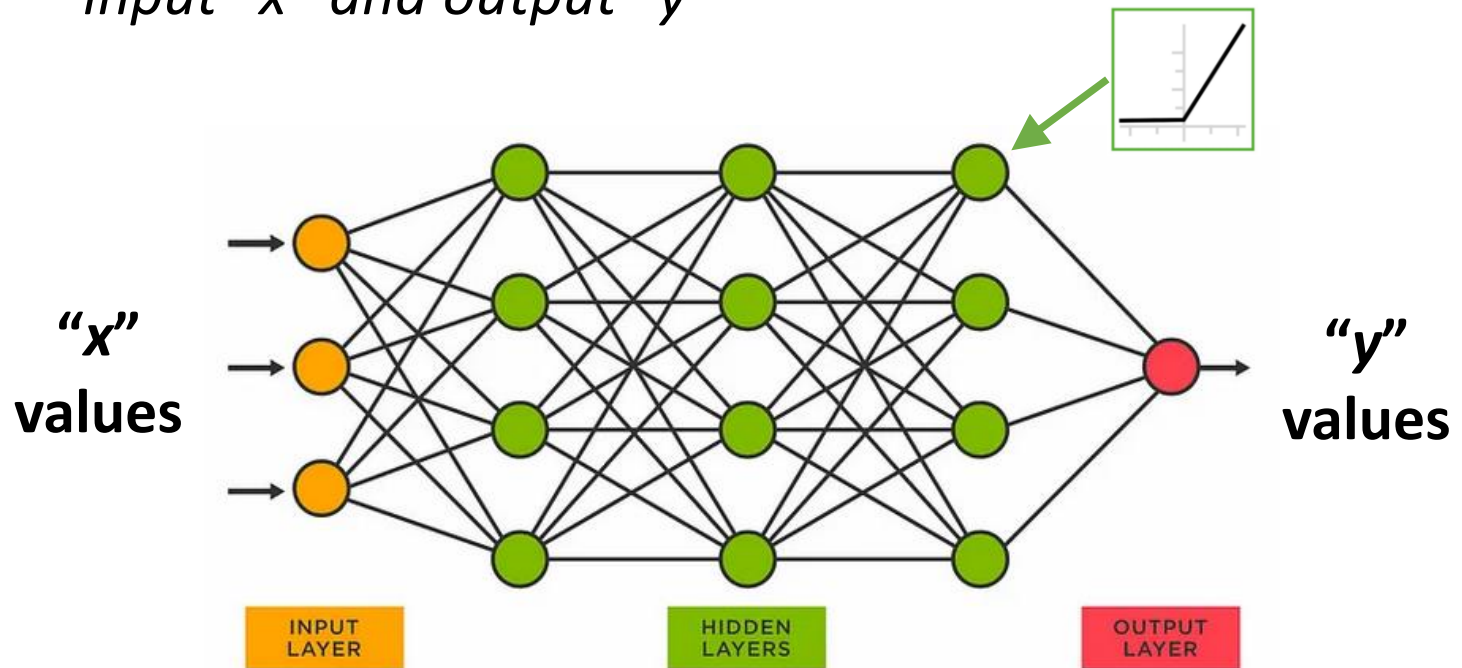
ML is not magic

- 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*

"*x*"
values

"*y*"
values

INPUT LAYER

HIDDEN LAYERS

OUTPUT LAYER
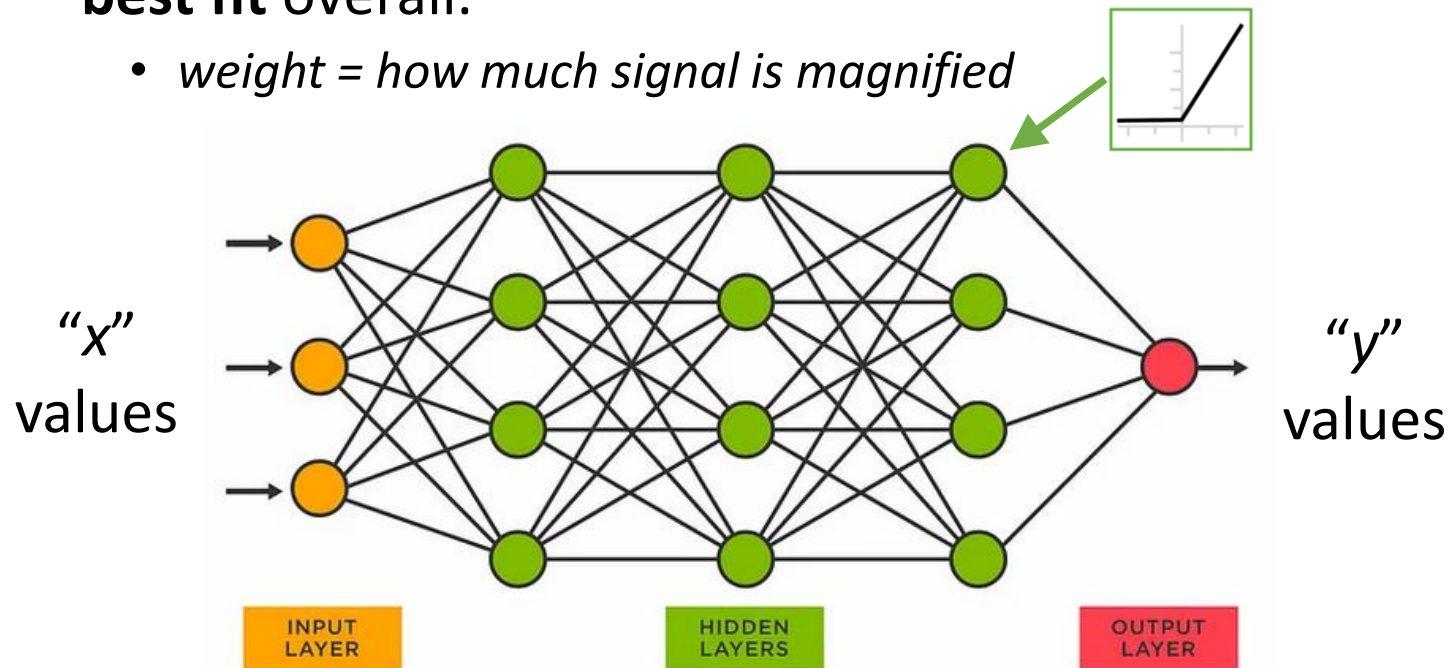
ML is not magic

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

"*x*"
values

"*y*"
values

INPUT LAYER

HIDDEN LAYERS

OUTPUT LAYER
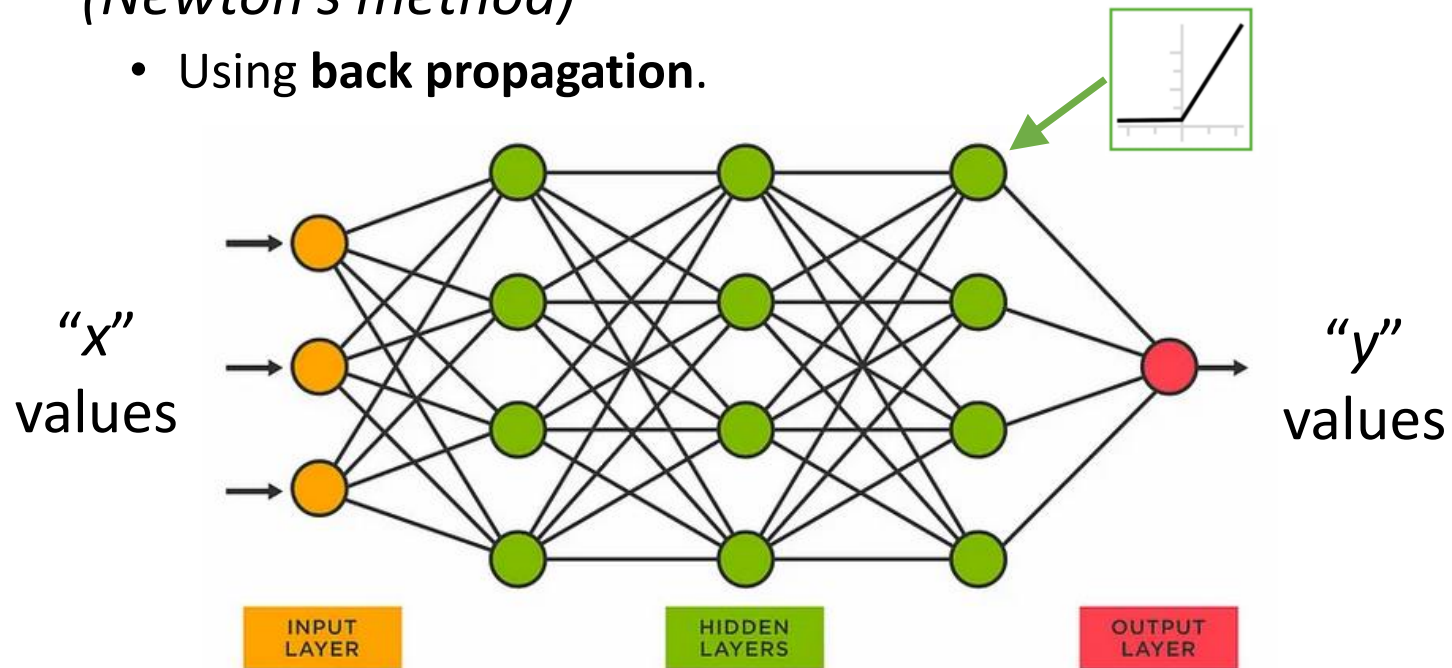
ML is not magic

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

"*x*"
values

"*y*"
values
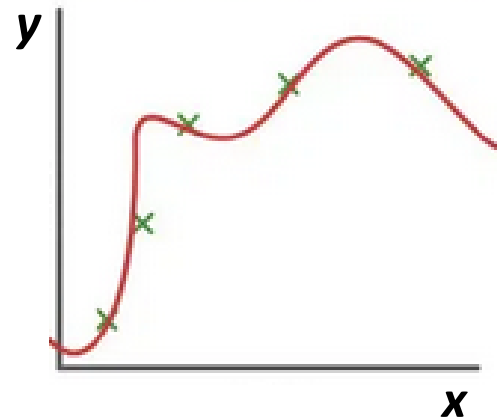
INPUT LAYER

HIDDEN LAYERS

OUTPUT LAYER
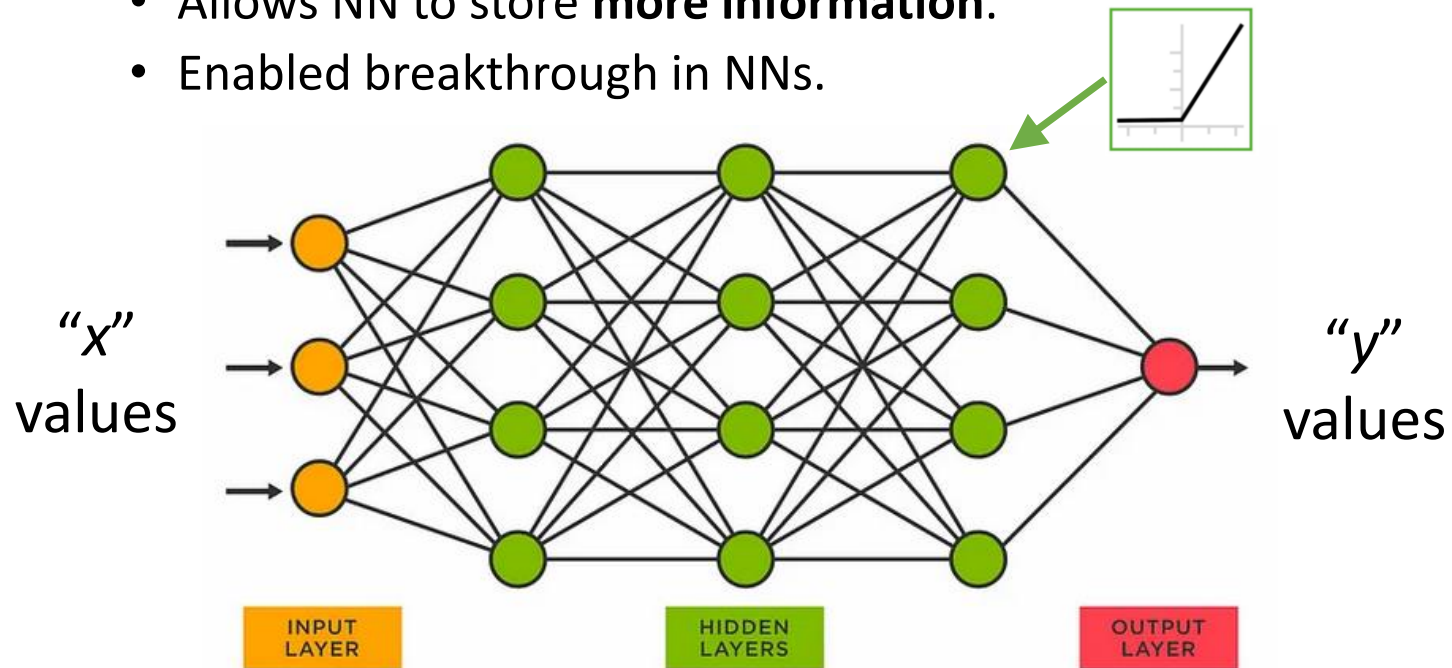
Overfitting

- This is a **no-no** in classical statistics.
  - *Too many parameters capture **random** variations*
    - and miss the **overall pattern**.
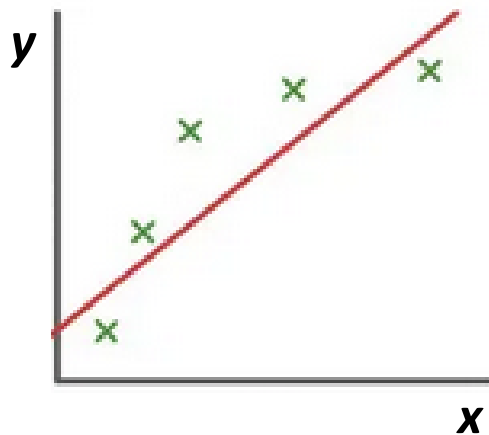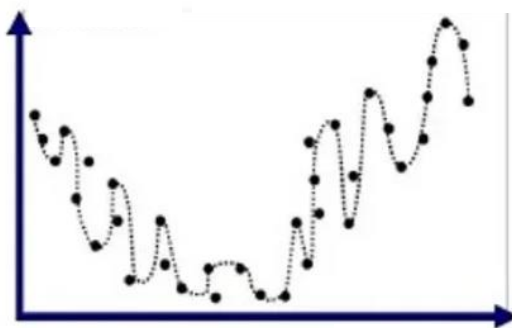
$y = ax + b$
2 parameters (too few)
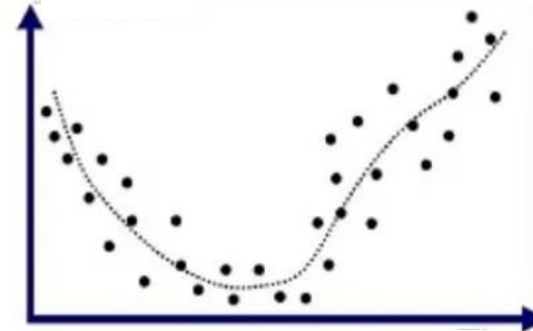
$y = ax + bx^2 + cx^3 + dx^4 + ex^5$
5 parameters (too many)

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.*
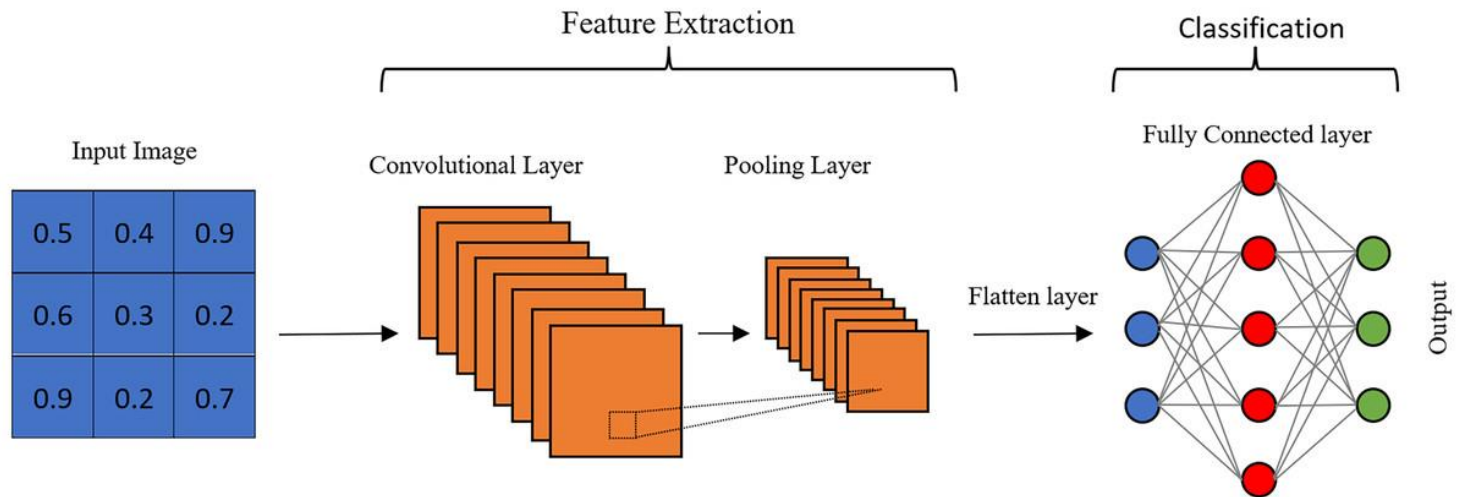   - Why? **Unknown**.


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.
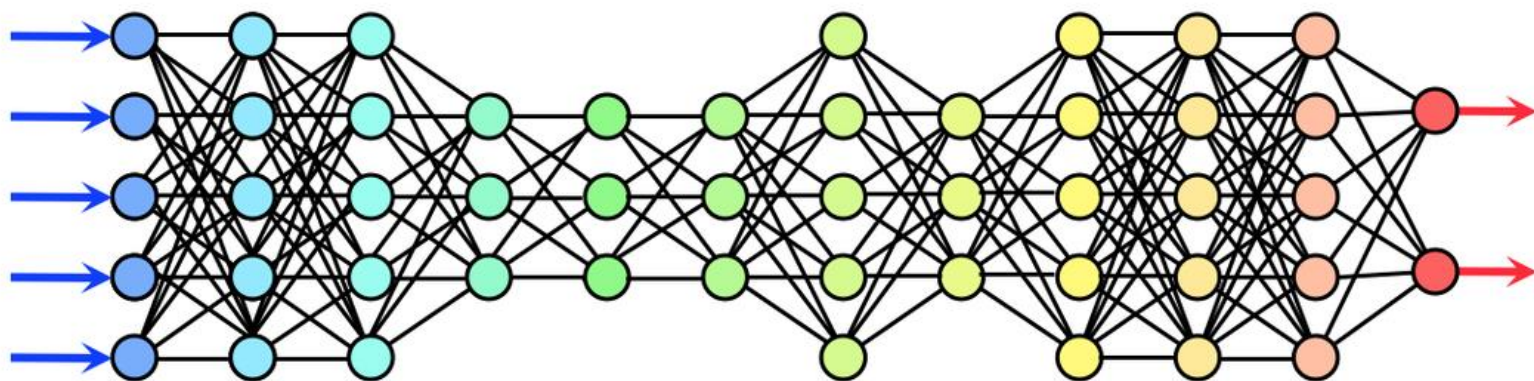
How about ChatGPT, etc.?

- 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)