Machine Learning: A Multidisciplinary View

SCIE1200 Guest Lecture
18 Sep 2024

Nan Ye

School of Mathematics and Physics
The University of Queensland

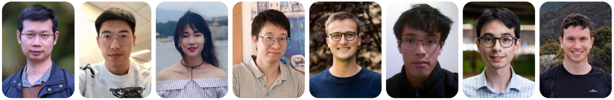
an academic...

teaching

proving

hacking

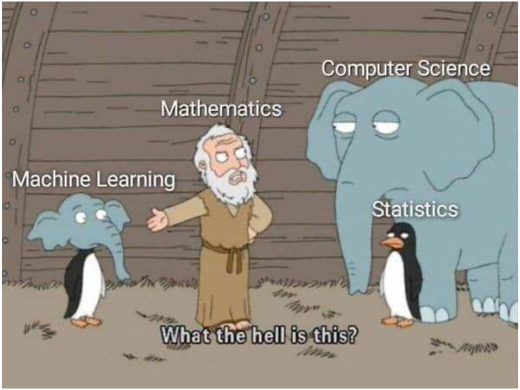
babysitting



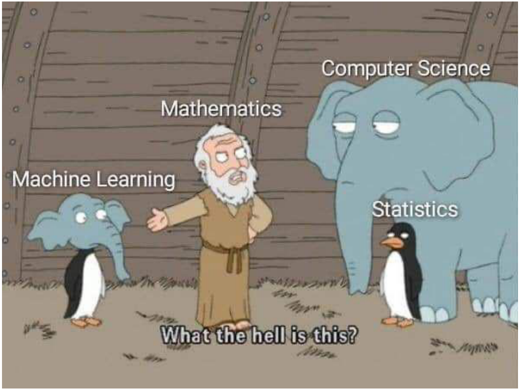
supervising some very intelligent people

cartoons generated by AI

big picture #1



big picture #1



Looks like the person who made this wanted to say machine learning is just taking something from computer science and something from statistics, and ended up being not recognised by either community and scorned by mathematics. . .

This might be the view held by quite a few people in the past, but perhaps not any more - the picture would be something like machine learning is a rocket, pulling mathematics, computer science, statistics, biology, physics, ..., up to the sky :D

email to a colleague

big picture #2

an attempt by AI to draw my picture

big picture #3 (my work)

autonomous driving

routing behavior analysis

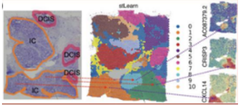
fishery stock assessment

MRI reconstruction

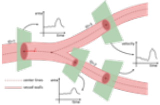
theoretically grounded practical algorithms
for
learning & decision-making

AAGI
Analytics for the Australian Grains Industry

agriculture analytics



spatial transcriptomics



haemodynamics modeling

big picture #4 (synergy)

philosophy
physics
linguistics
optimization
information theory
probability theory
mathematics

machine learning

operations research
computer science
psychology
statistics
formal logic
neuroscience

many ideas inspired by other fields & many applications in other fields

Roadmap

a multidisciplinary big picture

machine learning as a transformer in many fields

a multidisciplinary view on machine learning research

ending remarks

a transformer...

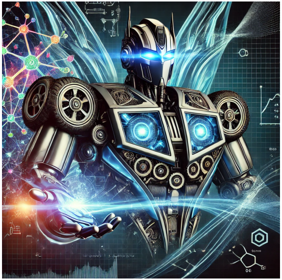
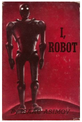


image generated by AI

Imaginations...


Golem
Judaism's Talmud (around 400)


Frankenstein
Mary Shelley's 1818 novel



I, Robot
Issac Asimov (1950)

The Iron Giant (1999)

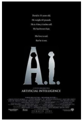
Robots (2005)

WALL-E (2008)

Astro Boy (2009)


Big Hero 6 (2014)

Next Gen (2018)



A.I. Artificial Intelligence (2001)

Iron Man (2008)

X-Men

Ex Machina (2014)

Star Wars

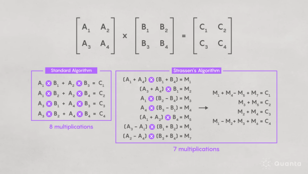
Avengers

The Wandering Earth (2019)

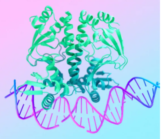
and realities...

AlphaGo


self-driving car



AlphaTensor


AlphaFold

Siri

Cortana

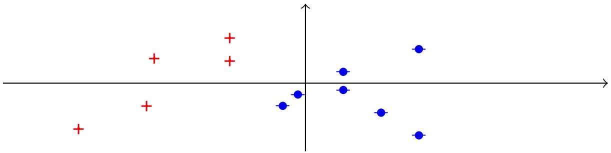

Sora

ChatGPT

and many others, powered by machine learning

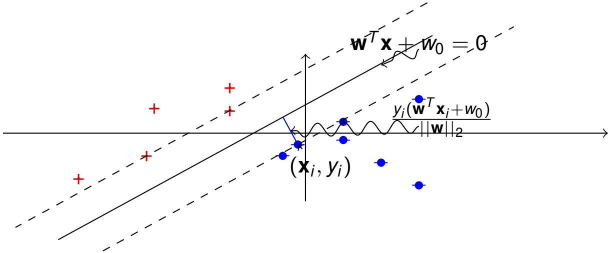
machine learning and ...

... and geometry!



Q. What's the best line to separate the red and blue points?

… and geometry!



Q. What's the best line to separate the red and blue points?

A. Support Vector Machines: the one farthest away from all points.

linearly separable data

$$\min_{\mathbf{w}, w_0} \frac{1}{2}||\mathbf{w}||_2^2$$
$$\text{s.t.} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + w_0) \geq 1, \quad i = 1, \ldots, n.$$

nonlinearly separable data

$$\min_{\mathbf{w}, w_0, \xi_1, \ldots, \xi_n} \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_i \xi_i$$
$$\text{s.t.} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + w_0) \geq 1 - \xi_i, i = 1, \ldots, n,$$
$$\xi_i \geq 0, i = 1, \ldots, n.$$

... and physics!

Maximum entropy principle (Jaynes, 1957): the most noncommittal probabilistic models satisfying given constraints should be preferred.

... and physics!

Maximum entropy principle (Jaynes, 1957): the most noncommittal probabilistic models satisfying given constraints should be preferred.

Q. A machine randomly outputs 1, 2, 3 with an observed mean of 2. what are the probabilities of these numbers?

... and physics!

Maximum entropy principle (Jaynes, 1957): the most noncommittal probabilistic models satisfying given constraints should be preferred.

Q. A machine randomly outputs 1, 2, 3 with an observed mean of 2. what are the probabilities of these numbers?

A. $p_1 = p_2 = p_3 = 1/3$, as it is the distribution that solves

$$\max - \sum_{i=1}^{3} p_i \ln p_i$$
$$s.t. \quad p_1 + 2p_2 + 3p_3 = 2$$

Jaynes,Information Theory and Statistical Mechanics,1957

Q. How to learn a model for labeling words with their parts-of-speech tags?

NNP	VBZ	PRP$	JJ	NN	.
SCIE1200	is	my	favourite	course	!

Q. How to learn a model for labeling words with their parts-of-speech tags?

NNP	VBZ	PRP$	JJ	NN	.
SCIE1200	is	my	favourite	course	!

A. Conditional Random Fields (Lafferty, McCallum, and Pereira, 2001): learn a distribution $p(\mathbf{y} \mid \mathbf{x})$ of the label sequence given the input sequence, using the maximum entropy principle

$$\max H(p) = \sum_{\mathbf{x}} \pi(\mathbf{x}) \sum_{\mathbf{y}} -p(\mathbf{y} \mid \mathbf{x}) \ln p(\mathbf{y} \mid \mathbf{x})$$

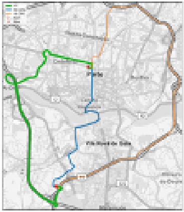
$$\begin{aligned} \text{s.t.} \quad \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) &= 1, \quad \forall \mathbf{x} \\ p(\mathbf{y}|\mathbf{x}) &\geq 0, \quad \forall \mathbf{x}, \mathbf{y} \\ \mathbb{E}(f_i) &= \tilde{f}_i, \quad \forall i. \quad \text{(expected feature = empirical feature)} \end{aligned}$$

The solution has a form similar to a random field in physics (e.g., Ising model).

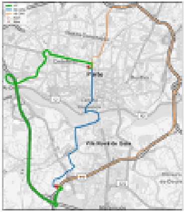
Ye et al. (2009) & Cuong et al. (2014): extend pairwise features to higher-order features.

Lafferty, McCallum, and Pereira,Conditional random fields: Probabilistic models for segmenting and labeling sequence data,2001
Ye et al.,Conditional random fields with high-order features for sequence labeling,2009
Cuong et al.,Conditional Random Field with High-order Dependencies for Sequence Labeling and Segmentation,2014

Q. How to determine the reward being maximized by observed behaviors?



Q. How to determine the reward being maximized by observed behaviors?



A. Inverse reinforcement learning using maximum entropy principle (Ziebart et al., 2008)

$$\min_p D(p \parallel q) \quad \text{s.t.} \quad E_{\tau \sim p}\phi(\tau) = \tilde{\phi},$$

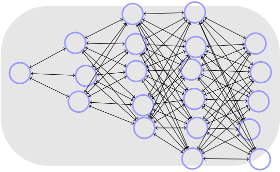
where D denotes the KL divergence, $\tau = (s_1, a_1, \ldots, s_T, a_T)$ a trajectory, $\phi(\tau)$ the feature vector for τ, and $q(\tau)$ a given reference trajectory distribution.
$\Rightarrow$ a reward linear in the features.

Snoswell, Singh, and Ye (2020): a generalized maximum entropy formulation and exact inference and learning algorithms.

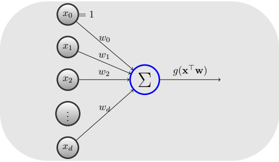
Ziebart et al.,Maximum entropy inverse reinforcement learning.2008
Snoswell, Singh, and Ye,Revisiting Maximum Entropy Inverse Reinforcement Learning: New Perspectives and Algorithms,2020

... and biology/neuroscience

artificial neural networks (ANNs)



ANN



artificial neuron

- ANNs
  - interconnected simple computational units (neurons)
  - universal approximators
  - often trained to minimize loss
- Neurons
  - input from incoming edges, output along outgoing edges
  - computes nonlinearly transformed weighted input sum $g(\mathbf{w}^\top \mathbf{x})$
  - nonlinearity g known as activation/transfer function

Q. How much fish is out there?

Area of the Northern Prawn Fishery

Q. How much fish is out there?

A. Lei, Zhou, and Ye (2024a,b): use structured neural networks to reliably "standardize" catch data to remove variations caused by factors other than abundance (e.g., fishing gear, skipper experience, weather).

Lei, Zhou, and Ye,Spatial-temporal neural networks for catch rate standardization and fish distribution modeling,2024
Lei, Zhou, and Ye,Structured neural networks for CPUE standardization: A case study of the blue endeavour prawn in Australia's Northern Prawn Fishery,2024

ending remarks

TensorFlow

theano

mxnet

Caffe2

GLUON

Sonnet

PyTorch

DL4J

Keras

ConvNetJS
Deep Learning in your browser

scikit learn

many powerful libraries for you to use

AI achieves silver-medal standard solving International Mathematical Olympiad problems

25 JULY 2024

AlphaProof and AlphaGeometry teams

Share

At present, these are great tools for mathematicians - they allow mathematicians to focus more on ideas and insights rather than the chores (not saying that an ingenious proof, such as one for an IMO question, is a chore), though the process of doing maths may be less fun without the strenuous effort to come up with ingenious solutions.

These tools don't ask good questions yet. They are still the mechanical proof assistant that exploits computers to efficiently search the proof space, except that they are now guided by AI to make more promising moves, so that the search completes in a reasonable amount of time.

One downside of these tools is that everything is done symbolically as in the traditional automatic theorem provers: the natural language theorem statements need to be converted to formal symbolic statements, and the proofs are written as formal symbolic statements, which are generally not fun to read. For example, AlphaProof uses the formal symbolic representation of Lean - see this `https://en.wikipedia.org/wiki/Lean_(proof_assistant)` for some examples.

Perhaps someone will build an interface to support human-style mathematical writing in the near future. If that happens, I'll be curious how the new generation of maths students will do maths - will they lose the ability to come up with new ideas/insights because they don't do enough chores and thus never become mathematically mature?

email to a colleague

use both artificial and human intelligence well :)

Sources of knowledge, diverse and wide,
Combining fields where insights reside,
Ideas from math, stats, and code,
Enhancing progress as data flows.

1 powerful tool that spans every domain,
2gether with art, science, and human gain,
0ld theories renewed with modern might,
0ngoing impact, reaching new heights.

ChatGPT