Machine-Learning Mathematical Structures

YANG-HUI HE

Dept of Mathematics, City, **University of London**

Merton College, **University of Oxford**

School of Physics, **NanKai University**

String Data 2020

How does one *DO* mathematics, I ?

Russell-Whitehead *Principia Mathematica* [1910s] programme (since at least Frege, even Leibniz) to axiomatize mathematics, but . . .

Gödel [1931] Incompleteness ; Church-Turing [1930s] Undecidability

Automated Theorem Proving (ATP) The practicing mathematician hardly ever worries about Gödel

- Newell-Simon-Shaw [1956] Logical Theory Machine: proved subset of *Principia* theorems
- Type Theory [1970s] Martin-Löf, Coquand, . . . Coq interactive proving system: 4-color (2005); Feit-Thompson Thm (2012); Lean (2013)
- Univalent Foundation / Homotopy Type Theory [2006-] Voevodsky

We can call this Bottom-up Mathematics

How does one *DO* mathematics, I ?

Russell-Whitehead *Principia Mathematica* [1910s] programme (since at least Frege, even Leibniz) to axiomatize mathematics, but ...

Gödel [1931] Incompleteness ; Church-Turing [1930s] Undecidability

Automated Theorem Proving (ATP) The practicing mathematician hardly ever worries about Gödel

- Newell-Simon-Shaw [1956] Logical Theory Machine: proved subset of *Principia* theorems
- Type Theory [1970s] Martin-Löf, Coquand, ... Coq interactive proving system: 4-color (2005); Feit-Thompson Thm (2012); Lean (2013)
- Univalent Foundation / Homotopy Type Theory [2006-] Voevodsky

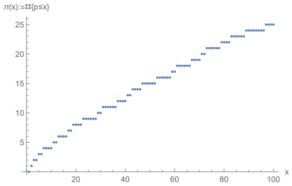
We can call this Bottom-up Mathematics

How does one do mathematics, II ?

- Late C20th - increasing rôle of computers: 4-color [Appel-Haken-Koch 1976]; Classif. Finite Simple Groups [Galois 1832 - Gorenstein et al. 2008] ...
- Buzzard: "Future of Maths" 2019: already plenty of proofs unchecked (incorrect?) in the literature, MUST use computers for proof-checking; XenaProject, Lean establish database of mathematical statements
- Davenport: ICM 2018 "Computer Assisted Proofs".
- Hale & Buzzard: Foresee within 10 years AI will help prove "early PhD" level lemmas, all of undergrad-level maths formalized;
- Szegedy: more extreme view, computers > humans @ chess (1990s); @ Go (2018); @ Proving theorems (2030)

How does one *DO* mathematics, III ?

- Historically, Maths perhaps more Top-Down: practice before foundation
  - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra . . .
  - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16

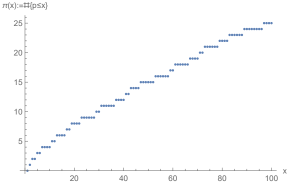


(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

- BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

How does one *DO* mathematics, III ?

- Historically, Maths perhaps more Top-Down: practice before foundation
  - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
  - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16

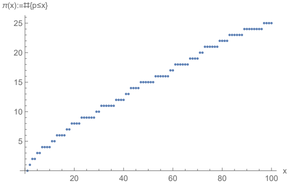


(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

- BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

How does one *DO* mathematics, III ?

- Historically, Maths perhaps more Top-Down: practice before foundation
  - Countless examples: calculus before analysis; algebraic geometry before Bourbaki, permutation groups / Galois theory before abstract algebra ...
  - A lot of mathematics starts with intuition, experience, and experimentation
- The best neural network of C18-19th? brain of Gauß ; e.g., age 16



(w/o computer and before complex analysis [50 years before Hadamard-de la Vallée-Poussin's proof]): PNT $\pi(x) \sim x/\log(x)$

- BSD computer experiment of Birch & Swinnerton-Dyer [1960's] on plots of rank r & N_p on elliptic curves

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena ... how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena ... how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena ... how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Question

- To extend the analogy: AlphaGo is top-down (need to see human games); even AlphaZero is not bottom-up (need to generate samples of games)
- In tandem with the bottom-up approach of Coq, Lean, Xena ... how to put in a little intuition and human results? If I gave you 100,000 cases of

$$\text{e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix}, \quad \text{or, labeled data e.g.} \begin{pmatrix} 5&3&4&3&5&1&4&4&1&2\\ 5&0&4&5&2&4&4&2&2&4\\ 1&1&2&2&0&4&1&4&5&0\\ 5&0&1&1&0&2&0&5&0&1\\ 2&5&0&1&1&3&2&3&0&3\\ 3&2&2&3&0&0&2&2&1&0\\ 2&2&5&1&4&4&0&0&1&2\\ 5&0&0&0&4&5&0&4&1&1\\ 4&3&4&3&3&1&0&0&2&5\\ 2&0&5&0&3&0&4&4&1&5 \end{pmatrix} \longrightarrow 3$$

- Q: Is there a pattern? Can one conjecture & then prove a formula?
- Q: What branch of mathematics does it come from?
- Perfect for (unsupervised & supervised) machine-learning; focus on labeled case because it encodes WHAT is interesting to calculate (if not how).

Mathematical Data: perfect for mining

- Mathematical Data is more structured than "real world" data, much less susceptible to noise; Outliers even more interesting, e.g. Sporadics, Exceptionals, . . .
- Last 10-20 years: large collaborations of computational mathematicians, physicists, CS (cf. SageMATH, GAP, Bertini, MAGMA, Macaulay2, Singular, Pari, Wolfram, . . .) computed and compiled vast data
  - links
  - Generic computation *HARD*
  - mining provides some level of "intuition" & is based on "experience"

Methodology

Bag of Tricks Hilbert's Programme of *Finitary Methods*, Landau's *theoretical minimum*, Migdal's *Mathmagics* ...

IMO Grand Challenge (2020-) Good set of concrete problems to try on AI

Standard Supervised ML Methods Regressor & Classifiers

- NN: MLPs; CNNs; RNNs, ... (gentle tuning of architecture and hyper-parameters)
- SVM, Bayes, Decision Trees, PCA, Clustering, ...
- ML: emergence of complexity via connectivity $\rightsquigarrow$ Intution (?)

This Talk: Status Report of Experiments in the last couple of years

- all standard methods $\simeq$ same performance
- $\sim$ 20-80 split; training on 20 (precision, Matthews' ϕ or R^2)

String/Algebraic Geometry

- CICY configuration $\to$ Hodge Numbers: YHH (1706.02714) Bull-YHH-Jejjala-Mishra (1806.03121, 1903.03113), Krippendorf-Syvaeri [2003.13679] Erbin-Finotello (2007.13379; : $(0.99, 0.9)$ YHH-Lukas [2009.02544] CICY4: $(0.98, 0.9)$
- Elliptic fibrations (from CICYs): YHH-SJ Lee (1904.08530) $(0.99, 0.9)$
- Distinguishing Heterotic SMs from the sum-line-bundle database (Anderson-Constantin-Gray-Lukas-Palti) and extrapolating beyond Deen-YHH-Lee-Lukas (2003.13339): $(0.98, 0.99)$
- Calabi-Yau metric: improves Donaldson alg. for numerical CY metric by 10-100 times Ashmore–YHH–Ovrut '19, q.v. Anderson, Gray, Krippendorf, Raghuram, Ruehle; Douglas–Lakshminarasimhan–Qi, '20

String/Algebraic Geometry: 2018-

- q.v., Bundle Cohomology (Ruehle, Brodie-Constantin-Lukas, Larfors-Schneider, Otsuka-Takemoto, Klaewer-Schlechter)
- q.v., Kreuzer-Skarke Dataset (Halverson, Long, Nelson; McCallister-Stillman)
- q.v., Calabi-Yau volumes in AdS/CFT (Krefl-Seong)
- q.v., MSSM from orbifold models (Parr-Vaudrevange-Wimmer)
- q.v. Particle Masses Gal-Jejjala-Pena-Mishra ...
- q.v. Knot invariants: Jejjala-Kar-Parrikar, Craven-Jejjala-Kar Gukov-Halverson-Ruehle-Sułkowski, using NLP
- YHH-Jejjala-Nelson NLP on ArXiv sections
- q.v. DEEP CONNECTIONS K. Hashimoto: AdS/CFT = Boltzmann Machine; Halverson-Maiti-Stoner: QFT = NN; de Mello-Koch: NN = RG; Vanchurin 2008: Universe = NN.

Representation/Group Theory

- ML Algebraic Structures (GAP DB) [YHH-MH. Kim 1905.02263]
  - When is a Latin Square (Sudoku) the Cayley (multiplication) table of a finite group? Bypass quadrangle thm (0.95, 0.9)
  - Can one look at the Cayley table and recognize a finite simple group?
    - bypass Sylow and Noether Thm; (0.97, 0.95) rmk: can do it via character-table T, but getting T not trivial
    - **SVM:** space of finite-groups (point-cloud of Cayley tables) seems to exist a hypersurface separating simple/non-simple
- ML Lie Structure Chen-YHH-Lal-Majumder [2011.00871] Weight vector → length of irrep decomp / tensor product: (0.97, 0.93); (train on small dim, predict high dim: (0.9, 0.8))
- [Chen-YHH-Lal-Zas 2006.16114]: even/odd/reflection sym (>0.99); distinguishing CFT 3pt functions (>0.99); Fourier coefficients / conformal block presence (>0.97) . . .

  (q.v. [Krippendorf-Syvaeri 2003.13679])

Combinatorics, Graph/Quivers

- [YHH-ST. Yau 2006.16619] Wolfram Finite simple graphs DB
  - ML standard graph properties:

    ?acyclic (0.95, 0.96); ?planar (0.8, 0.6); ?genus $>, =, < 0$ (0.8, 0.7); ?∃ Hamilton cycles (0.8, 0.6); ?∃ Euler cycles (0.8, 0.6)

    (Rmk: NB. Only "solving" the likes of traveling salesman stochastically)
  - spectral bounds ($R^2 \sim 0.9$) ...
  - Recognition of Ricci-Flatness (0.9, 0.9) (todo: find new Ricci-flat graphs);
- [Bao-Franco-YHH-Hirst-Musiker-Xiao 2006.10783]: categorizing different quiver mutation (Seiberg-dual) classes (0.9 - 1.0, 0.9)

Number Theory: A Classical Reprobate?

Arithmetic (prime numbers are Difficult!)

- [YHH 1706.02714, 1812.02893:]
    - Predicting primes $2 \to 3,\ 2,3 \to 5,\ 2,3,5 \to 7$; no way
    - fixed window of $(\text{yes/no})_{1,2,\ldots,k}$ to $(\text{yes/no})_{k+i}$ for some i; ML PRIMES problem (0.7, 0.8) NOT random! (prehaps related to AKS algorithm [2002], PRIMES is in P)
    - Sarnak's challenge: same window $\to$ Liouville Lambda (0.5, 0.001) Truly random (no simple algorithm for Lambda)
- [Alessandretti-Baronchelli-YHH 1911.02008]

  ML/TDA@Birch-Swinnerton-Dyer Ш and Ω ok with regression & decision trees: RMS < 0.1; Weierstrass $\to$ rank: random

Number Theory: A Modern Hope?

Arithmetic Geometry (Surprisingly Good)

- [Hirst-YHH-Peterken 2004.05218]: adjacency+permutation triple of dessin d'enfants (Grothendieck's Esquisse for $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$) ; predicting transcendental degree (0.92, 0.9)
- YHH-KH Lee-Oliver arithmetic of curves
  - 2010.01213: Complex Multiplication, Sato-Tate $(0.99 \sim 1.0, 0.99 \sim 1.0)$
  - 2011.08958: Number Fields: rank and Galois group (0.97, 0.9)
  - 2012.04084: BSD from Euler coeffs, integer points, torsion (0.99, 0.9); Tate-Shafarevich Ш (0.6, 0.8) [Hardest quantity of BSD]

Clearly useful for maths and physics

looking for new conjectures e.g.,

- '19 YHH-Kim: separating hyperplane - simple/non-simple groups; open
- '19 Brodie-Constantin-Lukas: exact formulae for cohomo surf.; proved.
- '20 YHH-Lee-Oliver: L-coefs and integer pt./torsion on ell; proved.
- '20 Craven-Jejjala-Par: Jones poly best-fit function; open
- ...

speed up computations and accuracies e.g.,

- computing/estimating (top.inv., charges, etc) MUCH FASTER
- '19 Ashmore-YHH-Ovrut: speed up Donaldson alg@CY metric 10-100
- '20 Douglas et al., Anderson et al. accuracy improvement on Donaldson 10-100 times
- ...

The other Extreme (?) View-Point

On the other hand, what is analyticity?

- n-th pime $= \left\lfloor \frac{n! \bmod (n+1)}{n} \right\rfloor (n-1) + 2$ (not efficient)
- bundle-cohomology:

Bott for Projective space:

$$h^q(\mathbb{P}^n, (\wedge^p T\mathbb{P}^n) \otimes \mathcal{O}(k)) = \begin{cases} \binom{k+n+p+1}{p}\binom{k+n}{n-p} & q=0 \quad k > -p-1, \\ 1 & q = n-p \quad k = -n-1, \\ \binom{-k-p-1}{-k-n-1}\binom{-k-n-2}{p} & q = n \quad k < -n-p-1, \\ 0 & \text{otherwise} \end{cases}$$

e.g. $(2,4)$-CY3 hypersurface:

$$h^q(X, \mathcal{O}_X(-k,m)) = \begin{cases} (k+1)\binom{m}{3} - (k-1)\binom{m+3}{3} & q=0 \quad k < \frac{(1+2m)(6+m+m^2)}{3(2+3m(1-m))} \\ (k-1)\binom{m+3}{3} - (k+1)\binom{m}{3} & q=1 \quad k > \frac{(1+2m)(6+m+m^2)}{3(2+3m(1-m))} \\ 0 & \text{otherwise} \end{cases}$$

- ...
- better suited for a computer programme any way

An Inherent Hierarchy?

- In decreasing precision/increasing difficulty:

	numerical
string theory $\to$	algebraic geometry over $\mathbb{C}$ $\sim$ arithmetic geometry
	algebra
string theory $\to$	combinatorics
	analytic number theory

- **Categorical Theory**
  - suggested by & in prog. w/ B. Zilber, Merton Prof. of Logic, Ox
  - major part of Model Theory: Morley-Shelah Categoricity Thm
  - Hart-Hruskovski-Laskowski Thm: 13 classes (levels) of iso-classes $I(T, k)$ of a theory T in first order logic over some cardinality k.

Thank you!

Syntax		Semantics
Alpha Go	→	Alpha Zero
ML Patterns	→	Auto Thm Pf&Chk
- Renner et al., PRL/Nature News, 2019: ML (*SciNet*, *autoencoder*)
- Lample-Charton, 2019: ML Symolic manipulations in mathematics
- Tegmark et al., 2019 AI Feynman, symb regressor
- Raayoni et al. 2020 Ramanujan-Machine
- Barbaresco-Nielson 2021 Infor Geom/ML

Sophia (Hanson Robotics, HK)

1st non-human citizen (2017, Saudi)

1st non-human with UN title (2017)

1st String Data Conference (2017)

Various Databases

- Kreuzer-Skarke: `http://hep.itp.tuwien.ac.at/~kreuzer/CY/`
  - new PALP: Braun-Walliser: ArXiv 1106.4529
  - Triang: Altmann-YHH-Jejjala-Nelson: `http://www.rossealtman.com/`
- CICYs: resurrected Anderson-Gray-YHH-Lukas, `http://www-thphys.physics.ox.ac.uk/projects/CalabiYau/cicylist/index.html`
- q.v. other databases of interesting to the math/physics community:

  Graded Rings/Varieties: Brown, Kasprzyk, et al. `http://www.grdb.co.uk/`

  Finite Groups/Rings: GAP `https://www.gap-system.org/`

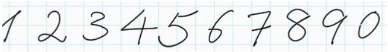
  Modular Forms: Sutherland, Cremona et al. `https://www.lmfdb.org/`

  Knots & Invariants: KnotAtlas `http://katlas.org/` Return

  . . .

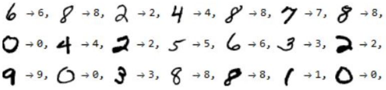
A Prototypical Question

- Hand-writing Recognition, e.g., my 0 to 9 is different from yours:



- How to set up a bijection that takes these to $\{1, 2, \ldots, 9, 0\}$? Find a clever Morse function? Compute persistent homology? Find topological invariants? ALL are inefficient and too sensitive to variation.

- What does your iPhone/tablet do? What does Google do? Machine-Learn
    - Take large sample, take a few hundred thousand (e.g. NIST database)



$\ldots$

$28 \times 28 \times (RGB)$

A Single Neuron: The Perceptron

- began in 1957 (!!) in early AI experiments (using CdS photo-cells)
- DEF: Imitates a neuron: activates upon certain inputs, so define
    - Activation Function $f(z_i)$ for input tensor z_i for some multi-index i;
    - consider: $f(w_i z_i + b)$ with w_i weights and b bias/off-set;
    - typically, $f(z)$ is sigmoid, Tanh, etc.
- Given training data: $D = \{(x_i^{(j)}, d^{(j)}\}$ with input x_i and **known output** $d^{(j)}$, minimize

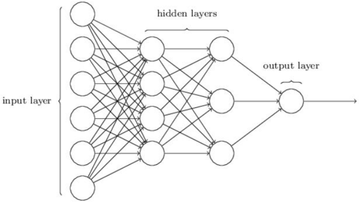
$$SD = \sum_j \left(f(\sum_i w_i x_i^{(j)} + b) - d^{(j)} \right)^2$$

to find optimal w_i and $b \rightsquigarrow$ "learning", then check against Validation Data

- Essentially (non-linear) regression

The Neural Network: network of neurons $\rightsquigarrow$ the "brain"

- DEF: a connected graph, each node is a perceptron (*Implemented on Mathematica $\geq$ 11.1 / TensorFlow-Keras on Python*)
  1. adjustable weights/bias;
  2. distinguished nodes: 1 set for input and 1 for output;
  3. iterated training rounds.



Simple case: forward directed only,

called multilayer perceptron

Many Layers : DEEP Learning

Connectivity $\rightsquigarrow$ Emergence of Complexity

- Essentially how brain learns complex tasks; **apply to our Landscape Data**

Back to Landscape