"Machine Learning & String Theory"

Lectures DLAP 2019

Lecturer: Jim Halverson

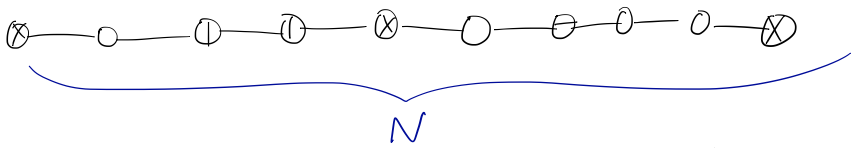
Lecture 1: Computational Complexity in Theoretical Physics

Introduction i) intro ii) complexity basics iii) NP-hardness iv) physical examples

TIME COMPLEXITY ASYMP T-SCALING OF PROCESS/ALG W/ INPUT SIZE N

CS EXAMPLE SORTING $\in \mathbb{Z}^N$ VIA DOUBLE FOR-LOOP: $O(N^2)$
SHORTEST PATH ON (G, V, E) VIA DIJKSTRA: $O(|V|^2)$

PHYSICS EX: PROT-FOLD :



t_{fold}(RANDOM PROT.) $\sim e^N$ t_{fold}(OUR PROTEINS) $\sim N^k$ (at least in some cases)

Q: WHY DO OURS FOLD QUICKLY? (Levinthal's Paradox)

Complexity Basics

Defn A problem $F: I \to B$, I set of instances

Defn A decision problem has $B = \{yes, no\}$
(eg, given (N_1, N_2, N_3), is $N_3 = N_2 \cdot N_1$?)

EX: clique: mutually conn. nodes of graph G
DEFN: CLIQUE $(G, k) \to$ { yes, G has k-clique; no, doesn't

Remarks
- "alg" computes F. ($\exists$ deterministic TM / classical computer that halts)
- $\exists$ diff algs (some better than others)
- polytime alg = produces $F(x \in I)$ in $\leq p(n)$ time
- exp time alg, not polynomial!

Defn P = class of problems w/ polytime sol. alg

Defn NP = " " " " " verifier.

Q: IS P = NP? (is solving a problem as hard as checking a soln? prob not, but open)

(Clay Millenium problem, expect $P \neq NP$)

NP-hardness (idea: at least as hard as any problem in NP)

KEY: POLYTIME ALG FOR NP-HARD $\Rightarrow P = NP$. $\therefore$ $P \neq NP \Rightarrow \nexists$ POLYTIME ALG FOR NP-HARD

Defn: $F: I \to \{yes, no\}$ can be polynomially transformed or "reduced" to $G: I' \to \{yes, no\}$ if $\exists\ f: I' \to I$ | f has polytime alg & $F(x) = yes \iff G(f(x)) = yes$

WRITE: $F \xrightarrow{red} G$

Defn $F: I \to B$ is NP-hard if $\exists\ G \xrightarrow{red} F\ \forall G \in NP$

"NP hard probs at least as hard as any prob $\in NP$"

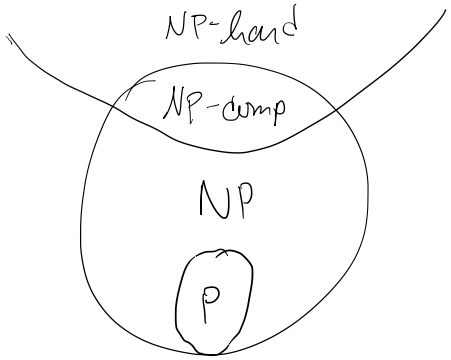
NB: 1) $(F \xrightarrow{red} G)$ & polytime G-alg $\Rightarrow$ polytime F-alg

2) $\therefore$ G NP-hard & polytime G-alg $\Rightarrow$ polytime alg $\forall F \in NP \Rightarrow P = NP$

3) expect $P \neq NP \Rightarrow \nexists$ polytime G-alg $\forall G \in$ NP-hard

4) $F \in$ NP-hard & $F \xrightarrow{red} G \Rightarrow G \in$ NP-hard

Defn: NP-complete := NP ∩ NP-hard



EX: $x_i \in \{0,1\}$, **literal** x_i, $\sim x_i$

Clause: "or" of literals eg. $x_1 \vee x_3 \vee \sim x_{10}$

CNF-formula: ("conjunctive normal form") "and" of clauses e.g. $x_1 \wedge (x_2 \vee \sim x_1) \wedge (\sim x_1 \vee \sim x_3)$

satisfiable: a CNF-formula φ | $\exists$ assign. A | $\varphi(A) = 1$

Defn: SAT given CNF-formula, det whether satisfiable

Theorem: (Cook) $\exists\ F \xrightarrow{red} SAT \quad \forall F \in NP$.

ie **SAT ∈ NP-complete**

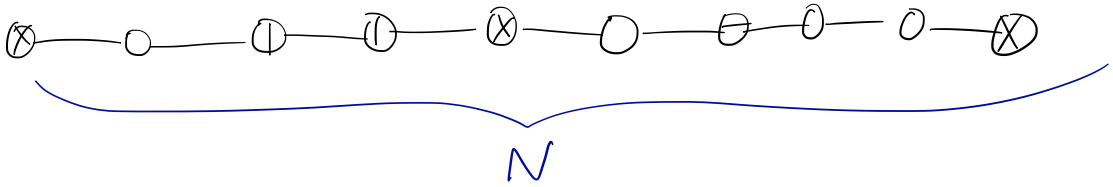
EX: clique: mutually conn. nodes of graph G

DEFN: CLIQUE $(G,k) \longrightarrow$ { yes, G has k-clique; no, doesn't

THM: (KARP) $SAT \xrightarrow{red} CLIQUE$, $\therefore$ NP-hard (act, complete)

Physical Examples

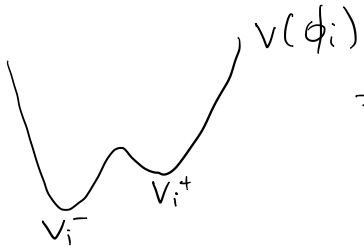
PROTEIN



- GLOB-MIN OF PROT-FOLD IS NP-HARD
- EVOLUTION CHOSE OURS

ADK CCs (Arkani-Hamed, Dimop., Kachru)

$$V(\phi) = \sum_{i=1}^{N} V(\phi_i)$$



$\Longrightarrow 2^N$ Vacua,

CC $\quad \Lambda_m = \sum m_i (V_i^+ - V_i^-) + V_i^- =: \sum m_i \Delta V_i + V_{min} \qquad m_i \in \{0,1\}$

DEF'N (ADKCC) $(\Delta V_i, V_{min}, \Lambda_0, \varepsilon) \longrightarrow$ does $\exists\ m \mid \Lambda_m \in (\Lambda_0, \Lambda_0 + \varepsilon)$?

(Brute force $\sim 2^N$, @ 10^9/sec, $N = 50 \Rightarrow 2$ weeks
$N = 400 \Rightarrow 10^{104}$ years)

DEF'N (KNAPSACK)

(cost c_i, values v_i, C_{max}, V_{min}) $\longrightarrow$ does $\exists\ m \in \{0,1\}^n \mid \begin{array}{l} \sum m_i c_i < C_{max} \\ \sum m_i v_i > V_{min} \end{array}$?

THM: (Karp?) KNAPSACK $\in$ NP-comp

THM: (PD) ADKCC $\in$ NP-complete

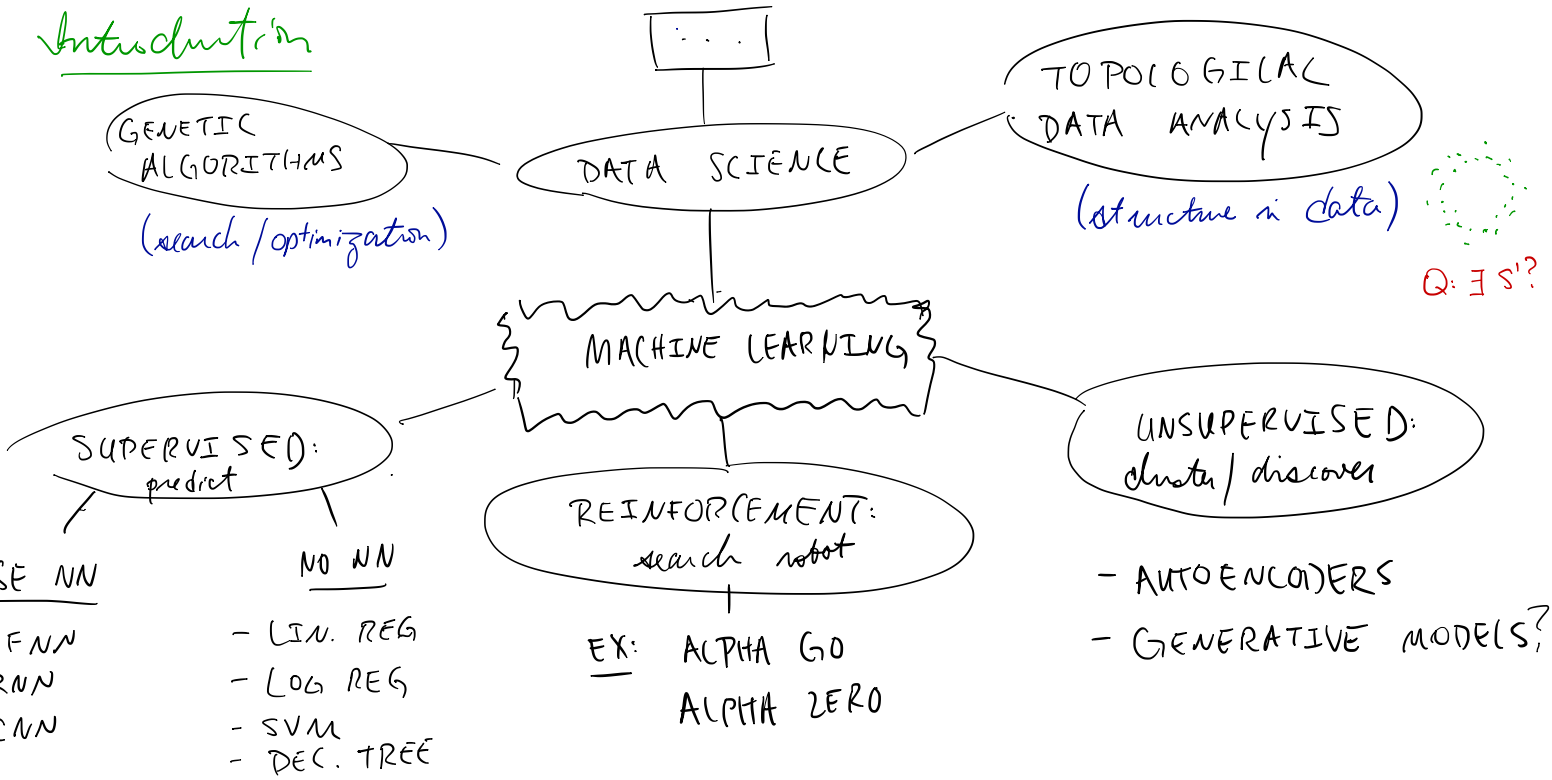
Pf: $\quad v_i = c_i = \Delta V_i \qquad V_{min} = \Lambda_0 - V_{min} \qquad C_{max} = \Lambda_0 - V_{min} + \varepsilon$

$$\Longrightarrow \quad \text{KNAP} \xrightarrow{\text{red}} \text{ADKCC}$$

Lecture 2: Introduction to Machine Learning

i) intro ii) supervised & NN's iii) reinforcement learning iv) generative models

Introduction



Supervised Learning & Neural Networks

Given S_{data}: test/train, generalization, pictorial rep

deep nn: Hughes, width, depth, univ approx, act fn's, RNN, CNN, training & eval, generalization error, optimizers

GOAL: LEARN $F: D \to R$ S.T. $(i, o) \in S_{data}$ HAS $F(i) \simeq o$

Q: DID IT WORK!? TEST "UNSEEN" DATA

EX 1:

i	o
0	32
100	212
30	86
10	50

i) $F \to F_\theta$, FAMILY OF F's

TAKE $F_\theta(i) = \theta_1 i + \theta_2$

ii) START w/ SOME θ_1, θ_2

iii) LOOP:
- CHECK IF $F_\theta(i) \simeq o$
- DET $\Delta\theta$ THAT IMPROVES F_θ
- $\theta \to \theta + \Delta\theta$

HERE: LOOP = LINEAR REGRESSION

& F_θ w/ $\theta_1 = \frac{9}{5}$ & $\theta_2 = 32$ PERFECT

(silly example makes key point, function is learned from data, given S_{data}.)

CAN FAIL TO GENERALIZE!

eg. $S_1 =$

i	.001	.002	.008
0	1.0005...	1.001...	1.004...

$\xrightarrow{\text{LIN REG}}$ $F_\theta = 1 + \frac{1}{2} i$ HAS $F_\theta(i) \simeq 0$

BUT FRIEND GIVES MORE DATA $S_2 =$

i	3	8	15	24
0	2	3	4	5

$F_\theta = 1 + \frac{1}{2} i$ TERRIBLE FIT!

REALITY: $0 = \sqrt{1+i} \simeq 1 + \frac{1}{2} i$ FOR $i \ll 1$ (only looked good b/c of biased sample)

LESSONS: i) F_θ GOOD ON S_1 WAS FAKE! SAMPLE BIAS

ii) $F_\theta = \theta_1 i + \theta_2$ CAN'T EXPRESS $\sqrt{1+i}$ $\forall i$

pro F_θ simple con F_θ not expressive

(will deal with "expressiveness" soon, but for now assume F_θ expressive)

LEARNING PROCESS & ML LANGUAGE

"train set" "test set"

1) OBTAIN S_{data} ("dataset") & SPLIT $S_{data} = S_{train} \cup S_{test}$

2) FAMILY F_θ, PARAM'D BY θ ("architecture" of F

3) OPTIMIZE θ TO MIN $L(F_\theta(i), o)$ ("loss function")

"train loop" $\longrightarrow$ FOR EPOCH $\in \{1, \ldots, N_{epoch}\}$

"epoch" = pass through S_{train}

FOR BATCH $\in$ PARTITION OF S_{train}

"batch" or "mini-batch"

$I_s, O_s = B$

$PREDS = F_\theta(I_s)$

$ERR = L(PREDS, O_s)$

$\theta \xrightarrow{OPT} \theta + \alpha \Delta\theta$

$\alpha \in \mathbb{R}$: "learning rate"

"optimizer" determines θ update

↳ eg. grad. desc., "Adam", "RMSProp"

4) TEST TRAINED F_θ ON S_{test}

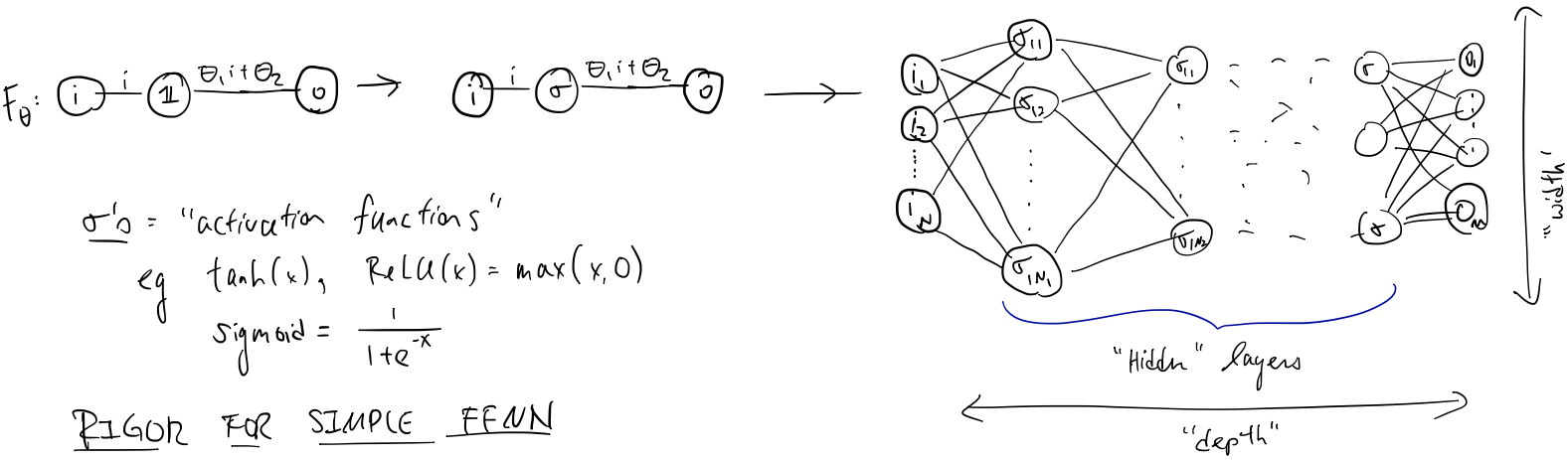
I.E. $I_s, O_s = S_{test}$

$L(F_\theta(I_s), O_s)$ small $\Longrightarrow$ "generalization" outside train set

THIS = "SUPERVISED LEARNING"

NEURAL NETWORKS: expressive F_θ.

LIN REG → NON-LINEARITY → FEED FORWARD NEURAL NETWORK



σ's = "activation functions"

eg $\tanh(x)$, $\mathrm{ReLU}(x) = \max(x, 0)$

$\mathrm{sigmoid} = \frac{1}{1+e^{-x}}$

RIGOR FOR SIMPLE FFNN

$n_i \in \mathbb{N}$ $i = 0, \dots, g$ $\quad h_j : \mathbb{R} \to \mathbb{R}$ $\forall\ 1 \le j \le g$

$A_j \in \mathbb{R}^{n_j \times n_{j-1}}$ $\quad b_j \in \mathbb{R}^{n_j}$ $\quad T_j : \mathbb{R}^{n_{j-1}} \to \mathbb{R}^{n_j}$ $\quad T_j(x) = A_j x + b_j$

then $\psi : \mathbb{R}^{n_0} \to \mathbb{R}^{n_g}$ is a feed forward neural network, def by:

$$\mathbb{R}^{n_0} \xrightarrow{T_1} \mathbb{R}^{n_1} \xrightarrow{h_1} \mathbb{R}^{n_1} \xrightarrow{T_2} \mathbb{R}^{n_2} \xrightarrow{h_2} \mathbb{R}^{n_2} \to \dots \xrightarrow{T_g} \mathbb{R}^{n_g} \xrightarrow{h_g} \mathbb{R}^{n_g}$$

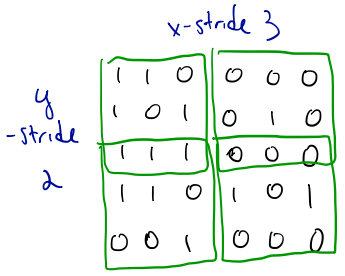
UNIVERSAL APPROX THM

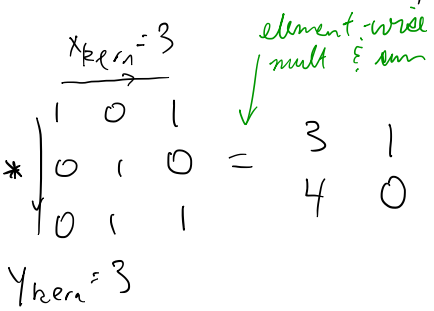
Q: HOW EXPRESSIVE?

A: ANY "SUFFICIENTLY NICE" $f(x)$ CAN BE APPROX'D TO ARB. PRECISION w/ $N_{hid} = 1$ (orig sigmoids)

CAVEAT: WIDTH $\sim e^{\text{precision}}$ (should combine w/ depth)

OTHER ARCHITECTURES IF VERY DIFF F_θ





i) CNN

LOCAL INF

ii) RNN

TIME SERIES

Reinforcement Learning

"search" (examples: AlphaGo & AlphaZero, Atari DQN)

IDEA: ROBOT STARTS EXPLORING, UPDATES BEHAVIOR BASED ON REWARDS

LANGUAGE: S = state space A = action space environment = E

policy $\pi: S \to A$ (dictates how to act) reward $r: S \to \mathbb{R}$ (good or bad state) discount $\gamma \in [0,1]$

$$S_{t=0} \to \pi \to a_0 \to S_1 \xrightarrow{r(s_1)} \to \pi \to a_1 \to S_2 \xrightarrow{r(s_2)} \to \cdots \to S_{end}$$

"episode"

end = natural endpoint or trunc. @ t_{max}

episodes: SAMPLE TRAJECTORIES THROUGH S.

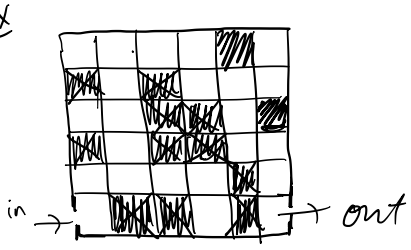
return $G_t = \sum_{k=0} \gamma^k r_{t+k}$ value function $V_\pi(s) = \mathbb{E}[G_t | s = s_t]$

action value function $q_\pi(s,a) = \mathbb{E}[G_t | s = s_t, a = a_t]$

optimality: $V_*(s) := \max_\pi V_\pi(s)$ $q_*(s,a) := \max_\pi q_\pi(s,a)$

$\pi_* \geq \pi \ \forall \pi$ via $V_\pi(s) \geq V_{\pi'}(s) \ \forall s \implies \pi > \pi'$

EX



E = MAZE

S = POSITION $\in \mathbb{Z}^2$

$A = \{U, D, L, R\}$

reward = what?

BEST: -1 per step

$\exists$ Reward engineering

Deep RL: some comb of V, q, π are NNs

Algorithms: optimize V, q → "value based methods" eg DQN

optimize π → "policy gradient methods" eg PPO

NNs for q & π → "actor critic methods" eg A3C

Generative Models "generate" or "fake"

= "NOISE"

= "PERSON"

BOTH $\in \{0,\dots,255\}^{N\times N} = G$ (grayscale)
B W

$|G| = 256^{N^2}$

Q: HOW TO PRODUCE $I \in G$?

i) N^2 DRAWS $z \sim \mathbb{P}_{\text{flat}}(\{0,255\})$
z's = PIXELS

ii) $z \sim P_z \longrightarrow \boxed{F_\theta} \xrightarrow{\text{output}} G$

$\Rightarrow$ ALMOST ALWAYS NOISE

FACT ii) GIVES IMPLICIT DIST P_{F_θ} ON G, $F_\theta(z)$ A SAMPLE

Q: GIVEN S_{data} = PEOPLE $\subsetneq G$ SAMPLES OF P_{data}, TRAIN SO $P_{F_\theta} \simeq P_{data}$?

IE SO $F_\theta(z) \simeq$ PERSON $\neq$ NOISE?

$\Rightarrow$ GENERATIVE MODELS eg. GAN, VAE

GENERATIVE ADVERSARIAL NETWORKS

Lecture 3: Machine Learning & the String Landscape

tomorrow: generative models

4 papers summer 2017, see wiki in upcoming Ruehle review

i) Fundamental Physics & String Theory ii) Statistical Predictions & ML

iii) 10^{755} Geometries iv) BP CC's v) Consistency & Particles

Fundamental Physics & String Theory

"Fundamental" Physics:

- particles: spin 0: h; spin $\frac{1}{2}$: $u\ c\ t$, $d\ s\ b$, $e\ \mu\ \tau$, $\nu_e\ \nu_\mu\ \nu_\tau$; spin 1: $W_\pm, g, \gamma$
- gravity $S_{EH} = \frac{1}{2\kappa}\int d^4x \sqrt{-g} R$
- cosmology
  - inflation (?)
  - D.M. $\frac{E_{DM}}{E_{baryons}} \simeq 5$
  - D.E. $E_{DE}/E_{TOT} \simeq .7$

String Theory

open: ℓ_s → quarks & gluons

closed: ℓ_s → gravitons!

D-branes

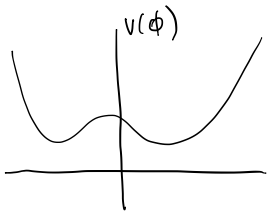
fluxes

9+1 D $\Rightarrow$ $\mathbb{R}^{3,1} \times X_6$

LOW E: gravity + gauge theories + ...

(somewhat realistic)

KEY PT: $N_{Sol.} >> 1 \Rightarrow$ "String Landscape"

NOT 

IS ϕ_i $i = 1 \ldots$ 100's OR 1000's

NOT ONE G.S. $|0\rangle$ IS MANY VACUA $|0\rangle_i$

$\Rightarrow$ S.T. PREDS STATISTICAL Q: WHICH ONE? $P(i) := P(|0\rangle_i)$?

Statistical Predictions & ML

- ENSEMBLE $S = \{|0_{Y_i}\rangle\}$
- COSMOLOGY $\Rightarrow$ $P(|0_{Y_i}\rangle) = p(i) \stackrel{?}{=} D(i)\, F(i)$ (Dynamic factor, Flat draw factor)
- WANT: EXPECTED OBSERVABLES $\langle O \rangle = \dfrac{\sum_{i \in S} p(i)\, O(i)}{\sum_{i \in S} p(i)}$
- REALITY S, D, F HARD (size & instances of NP-hard)
- IDEA: FAKE IT! LEARN GOOD GENERATIVE MODELS

$\Rightarrow \hat{S} \simeq S \quad \hat{D} \simeq D \quad \hat{F} \simeq F$

SIMPLE VERSION $S \simeq \{\mathcal{L}_i\}$ $\mathcal{L}$ LAGRANGIAN, DET BY m's T's

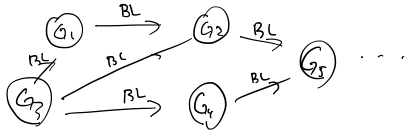
$\Rightarrow$ LEARN RANDOM APPROX'S OF m's, T's!

10^{755} geometries: 2 APPLICATIONS

DATA: - 10^{755} GEOMS, GEN'S OF $\mathbb{CP}^n := \dfrac{\mathbb{C}^{n+1} \setminus \{0\}}{\mathbb{C}^*}$

- TOPOLOGICAL NETWORK

BL: TOPOLOGY CHANGE



- UNIVERSAL PHYSICS $\simeq$ 750 GAUGE FACTORS
  $\simeq$ 2000 AXIONS

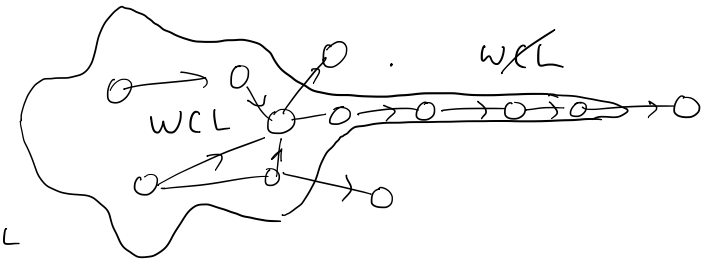
APP 1: "DEEP BOUNDARY"

RANDOM WALK: 3-4 STEPS $\Rightarrow$ ~~WCL~~

RL: 100 PTS IF s_{t+1} WCL

GAME OVER IF NOT

$\Rightarrow$ LEARNS 108 STEPS w/ WCL



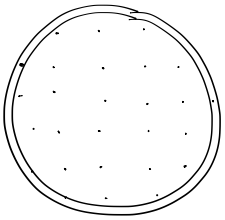
APP 2: "E_6 ON D" (relevant for grand unification)

I: geom data O: yes or no

LOG REG ON 50/50 Y/N SET $\Rightarrow$ 98% ACC $\Rightarrow$ CONJECTURE $\Rightarrow$ THM

(possible because interpretable supervised learning)

Bousso-Polchinski Cosmological Constant



$$\Lambda = -\Lambda_0 + \vec{N}^T \cdot q \cdot \vec{N}, \quad \vec{N} \in \mathbb{Z}^k$$

BPCC: given $(\Lambda_0, q, \Lambda_1, \Lambda_2)$ does $\exists\ \vec{N}$ s.t. $\Lambda_1 \leq \Lambda(\vec{N}) \leq \Lambda_2$?

$\Rightarrow$ DOES $\exists\ \vec{N}$ IN k-DIM THIN SHELL?

DENEF-DOUGLAS PROMISE SUBSET SUM $\xrightarrow{\text{red}}$ BPCC $\therefore$ BPCC NP-complete

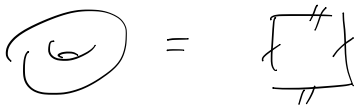
Q: HOW DID NATURE FIND SMALL CC?

RL: CAN IT FIND GOOD $\vec{N}$, SMALL CC?

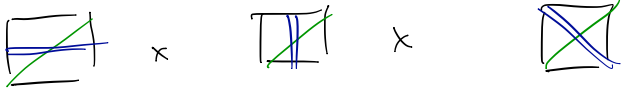
↳ A3C: FACTOR OF 1000 OR SO SMALLER

Consistency & Particles ("Branes w/ Brains")

(open AI gym)



ENVIRONMENT $\simeq$

STATE = D-BRANE PLACEMENT $\simeq$  $\xrightarrow{\text{RED}}$ $7 \times N_{stacks}$ MATRIX

ACTIONS: ADD OR REMOVE BRANE @ POS

2 TASKS:

CONSISTENCY:
- i) D-BR. GAUSS LAW
- ii) K-THEORY
- iii) SUSY

PARTICLES
- i) $G \simeq G_{SM}$?
- ii) PARTICLES $\simeq$ PARTS$_{SM}$?

A B

REWARD POSS: i) A & B ii) A THEN B iii) B THEN A

RESULTS
- LEARNS HUMAN STRATEGY
- FACTOR OF $O(700)$ IMPROVEMENT