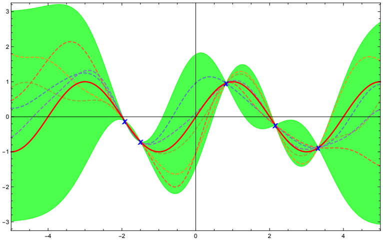
Baryons from Mesons: A Machine Learning Perspective



Challenger Mishra

Accelerate Science Research Fellow
University of Cambridge

Accelerate Science School
Cambridge, UK

Feb 4, 2021

Baryons from Mesons: A Machine Learning Perspective, arXiv:2003.10445

– Yarin Gal (Oxford), Vishnu Jejjala, and Damián Kaloni Mayorga Peña (Wits).

Baryons from Mesons: A Machine Learning Perspective, arXiv:2003.10445

– Yarin Gal (Oxford), Vishnu Jejjala, and Damián Kaloni Mayorga Peña (Wits).

particlezoo.net, Particle Zoo App

Outline

- The Standard Model of particle physics and Quantum Chromodynamics.
- A machine learning model of particle masses.

What ties all of known Physics together?

Invariance of natural laws under symmetries of Nature.

- Newtonian Mechanics: Galilean invariance.
- Electromagnetism and Relativistic Mechanics: Poincaré invariance.
- General Relativity: Invariance under diffeomorphisms of spacetime.
- Standard Model of Particle Physics: Invariance under gauge symmetry.

The Standard Model of Particle Physics is a theory that unifies Electromagnetism, Weak, and Strong nuclear forces.

What ties all of known Physics together?

Invariance of natural laws under symmetries of Nature:

- Standard Model of Particle Physics (SM): Invariance under the gauge symmetry $SU(3) \times SU(2) \times U(1)$.
- Grand Unified Theory (GUT): Invariance under gauge symmetry $SU(5)/SO(10)/E_6/(?)$.
- Quantum Gravity: Invariance under gauge symmetry $E_8 \times E_8/SO(32)/(?)$.

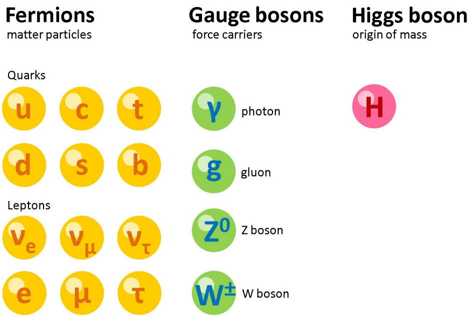
Unification: What ties all of known Physics together?

Invariance of natural laws under symmetries of Nature:

- Standard Model of Particle Physics (SM): Invariance under the gauge symmetry $SU(3) \times SU(2) \times U(1)$.
- Grand Unified Theory (GUT): Invariance under gauge symmetry $SU(5)/SO(10)/E_6/(?)$.
- Quantum Gravity: Invariance under gauge symmetry $E_8 \times E_8/SO(32)/(?)$.

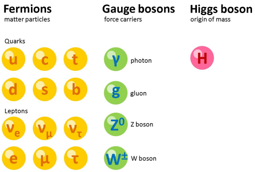
The mathematician in the picture?

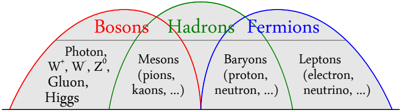
The Standard Model of particle physics



Quantum Chromodynamics $\subset$ Standard Model

- Standard Model of Particle Physics (SM): Invariance under the gauge symmetry $SU(3) \times SU(2) \times U(1)$.
- The SU(3) corresponds to the Quantum Field Theory of strong interactions, called Quantum Chromodynamics, or, QCD.
- QCD is the theory of strong interactions between quarks and gluons (carriers of strong force).
- Quarks and gluons combine to form composite particles called hadrons.



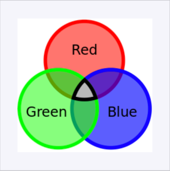


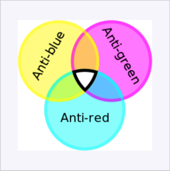
Color charge in Quantum Chromodynamics

- Each quark can have $N = 3$ color charges. Their anti-particles (anti-quarks) have opposite color charges. Gluons carry a color-anticolor charge. In comparison, there is only one type of electric charge.

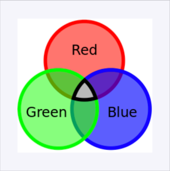
- Quarks and gluons combine to form composite particles called hadrons at energies below $\Lambda_{QCD} \sim 330MeV$. Compare this to the energy scale of operation of the Large Hadron Collider ($\sim 14TeV$).

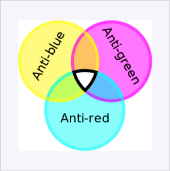
- Loosely speaking, hadrons can be composed of either equal number of quarks and anti-quarks (mesons), or an odd number of quarks (baryons), or anti-quarks (anti-baryons). These are colorless bound states at low energies.





Color charge in Quantum Chromodynamics





The idiot physicists, unable to come up with any wonderful Greek words anymore, call this type of polarization by the unfortunate name of 'color', which has nothing to do with color in the normal sense.

Feynman, Richard (1985), QED: The Strange Theory of Light and Matter.

A zoo of particles

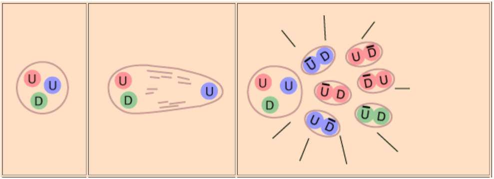
[particlezoo.net]

Summary from the zoo of particles

- Mesons are made of equal number of quarks and anti-quarks (usually one of each).
- Baryons are made up of an odd number of quarks (usually three).

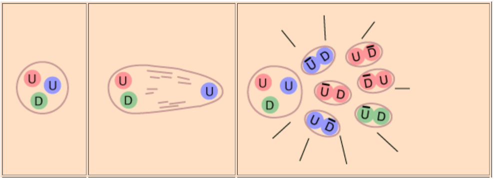
Quantum Chromodynamics: Color confinement

- Quarks have not been observed in isolation.
- Quarks cannot be separated from their parent hadron; instead new hadrons are produced in the process.



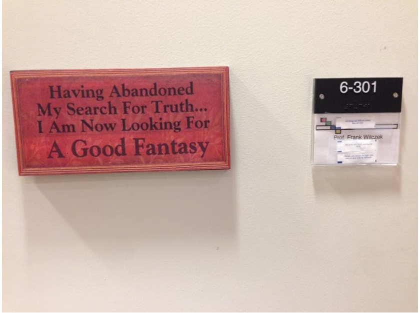
Quantum Chromodynamics: Asymptotic Freedom

- The strength of the strong interaction becomes asymptotically weaker as the energy scale increases, or correspondingly, the length scale decreases.
- Within the confines of the parent hadron, quarks are essentially free to move around, or, asymptotically free.
- This was discovered in 1973 by David Gross and Frank Wilczek, and independently by David Politzer.



Quantum Chromodynamics: Asymptotic Freedom

- This was discovered in 1973 by David Gross and Frank Wilczek, and independently by David Politzer.



(from Nov 2018 @ MIT)

Baryons as solitons in the mesonic spectrum

- Witten (*Baryons in the 1/N Expansion, 1979*):

  In the limit $N \to \infty$ (N being the number of color charges), baryons appear as solitons in the mesonic spectrum.

- The mesons are effectively free particles with couplings of $\mathcal{O}(1/N)$.
- The baryons take masses $\mathcal{O}(N)$.

- Can we then recover masses of baryons from properties of mesons alone?

- We will devise a ML model trained using mesonic properties, and use the model to predict masses of all baryons, and exotic quark states (hypothesised or otherwise!).

- Lattice QCD, a discrete formulation of QCD, can compute masses of such hadronic states, but such computations involving Monte Carlo methods are notoriously expensive.

- Recently, ML approaches to Lattice Field theories in general has received a lot of attention. See e.g., the collaboration between MIT, DeepMind, and NYU:

  (1) *Sampling using SU(N) gauge equivariant flows*, arxiv:2008.05456,
  (2) *Equivariant flow-based sampling for lattice gauge theory*, arxiv:2003.06413.

Baryons from Mesons: the dataset

Dataset of 196 mesons (training set) and 43 baryons (test set).

Inputs (quantum numbers): $\vec{v} = (d, \bar{d}, u, \bar{u}, s, \bar{s}, c, \bar{c}, b, \bar{b}, I, J, P)^1$.

Outputs (log mass/1 MeV): $log\ m_s$.

Representation of a meson (training examples):

charged pion: $\pi^+ = (0, 1, 1, \vec{0}_7, 1, 0, -1)$, charged pion: $\pi^- = (1, \vec{0}_2, 1, \vec{0}_6, 1, 0, -1)$,

Representation of a baryon (test examples):

Proton: $p = (1, 0, 2, \vec{0}_7, \frac{1}{2}, \frac{1}{2}, 1)$, Neutron: $n = (2, 0, 1, \vec{0}_7, \frac{1}{2}, \frac{1}{2}, 1)$.

---

[1] For $1 \leq i \leq 10$, $\vec{v}_i \in \{0, 1, 2, 3, 4\}$, $I \in \{0, \frac{1}{2}, 1, \frac{3}{2}\}$, $J \in \{0, \frac{1}{2}, 1, \frac{3}{2}, \ldots, 4\}$, $P \in \{-1, 1\}$

Baryons from Mesons: selecting features

Input 13-vectors do not uniquely identify a particle:

$$\begin{aligned} \text{Train input features (GP):} \quad \vec{v}_{\text{tr}} &= (d, \bar{d}, u, \bar{u}, s, \bar{s}, c, \bar{c}, b, \bar{b}, I, J, P, rk) \ , \\ \text{Test input features (GP):} \quad \vec{v}_{\text{test}} &= (d, \bar{d}, u, \bar{u}, s, \bar{s}, c, \bar{c}, b, \bar{b}, I, J, P, \ 0) \ . \end{aligned}$$

$$\begin{aligned} \text{Train input features (NN):} \quad \vec{v}_{\text{tr}} &= (d, \bar{d}, u, \bar{u}, s, \bar{s}, c, \bar{c}, b, \bar{b}, I, J, P) \ , \\ \text{Test input features (NN):} \quad \vec{v}_{\text{test}} &= (d, \bar{d}, u, \bar{u}, s, \bar{s}, c, \bar{c}, b, \bar{b}, I, J, P) \ . \end{aligned}$$

Baryons from Mesons: baseline model

The constituent quark model assigns the following masses to the *constituent* quarks:

$m_u = 336$ MeV, $m_d = 340$ MeV, $m_s = 486$ MeV,

$m_c = 1550$ MeV, $m_b = 4730$ MeV

Baryons from Mesons: Experiments using a Neural Network

Neural network: 5 fold cross-validation yielded a single-layer neural network with 50 neurons, and logistic-sigmoid activation. We used ADAM optimiser and the *mean squared loss*:

$$g(\theta) = \sum_s ||f_\theta(\vec{v}_s) - \log m_s||^2 \ .$$

Experiment : Train on meson data and then test on the baryons and exotics hadrons. We predict on the test sets by repeating our experiment 10^3 times, with the neural network parameters randomly initialised in each run.

Baryons from Mesons: Experiments using a Gaussian Process

- The Gaussian process is equivalent to a fully connected single layer neural network with an independent and identically distributed (i.i.d.) prior over its parameters in the limit of infinite width.
- Thus, Gaussian processes could be used to make exact Bayesian inferences for infinite width neural networks.
- Making inference using a Gaussian process involves inverting matrices that are the size of the dataset. A Gaussian process is therefore particularly well-suited to relatively small datasets, such as ours.
- Further, unlike neural networks, Gaussian processes do not require a validation set to tune hyperparameters. Instead, Gaussian processes use their own marginal likelihood – a quantity that captures how well each subset of the training points can predict the rest of the training points.

Baryons from Mesons: Experiments using a Gaussian Process

Constructing a Gaussian process proceeds as follows:

- We first specify a positive definite *kernel*

$$K_{ij} := k(x_i, x_j) \ ,$$

where $i = 1, \ldots, M$ runs over the set of training data (mesons), and each x_i is a D-dimensional input vector. The *covariance* function defines the prior on noisy observations:

$$\mathrm{cov}(x_i, x_j) = k(x_i, x_j) + N_{i,j} \ , \quad \text{where} \quad N_{i,j} = \sigma_n^2 \delta_{i,j} \ .$$

Each positive definite covariance function lends itself to an expansion in terms of basis functions.

Baryons from Mesons: Experiments using a Gaussian Process

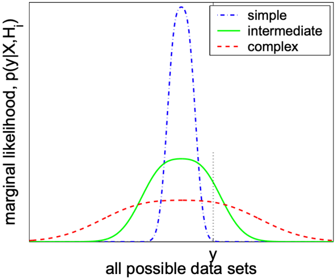
- Two of the simplest *kernels* are the *squared exponential (SE)* and the *rational quadratic (RQ) kernels* defined below,

$$
\begin{aligned}
k_{\mathrm{SE}}(x_i, x_j) &= \sigma_f^2 \exp\left(-\frac{1}{2}(x_i - x_j)^{\mathsf{T}}\Lambda^{-1}\,(x_i - x_j)\right) \quad ,\\
k_{\mathrm{RQ}}(x_i, x_j) &= \sigma_f^2 \left(1 + \frac{1}{2\alpha}(x_i - x_j)^{\mathsf{T}}\Lambda^{-1}\,(x_i - x_j)\right)^{-\alpha},
\end{aligned}
$$

  where Λ is a diagonal matrix with entries $\{\lambda_i^2\}_{i=1}^{D}$, where λ_i is the characteristic length scale for the i^{th} feature, and σ_f is an overall scale. We also have $\alpha > 0$.

- Since the D distinct characteristic length scales determine the relevance of each input feature, the kernels above implement automatic relevance determination (ARD) in our experiments.

Experiments using a Gaussian Process: Model selection



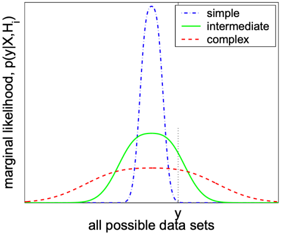
- Model selection in Gaussian processes is done by maximising the *log marginal likelihood* with respect to a training set $\{x_i \to y_i\}_{i=1}^{M}$ of size M. This is defined as

$$\log p(y|X) := -\frac{1}{2}y^T(K+N)^{-1}y - \frac{1}{2}\log|K+N| - \frac{M}{2}\log(2\pi) ,$$

  where y is the M-vector of training set output values and X is the collection of input vectors of size $D \times M$.

- Maximising the *log marginal likelihood* sets the hyperparameters of the covariance function.

Experiments using a Gaussian Process: Model selection

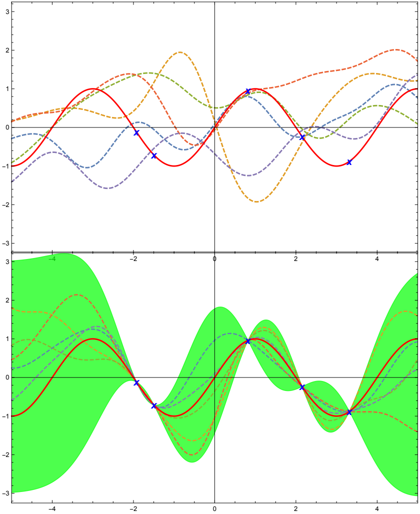


- Maximising the *log marginal likelihood* sets the hyperparameters of the covariance function.
- At the optimal value of the hyperparameters, a trade off is achieved between the complexity of the model and the model fit.
- Once the hyperparameters are set, one can predict the mean ($\mu_\star$) and the variance ($\mathrm{var}_\star$) of the distribution for an unseen input vector $x_\star$, using the formulae

$$\mu_\star = k_\star^T (K + N)^{-1} y \ , \qquad \mathrm{var}_\star = k(x_\star, x_\star) - k_\star^T (K + N)^{-1} k_\star \ ,$$

where the components of the M-vector $k_\star$ are given by $k_{\star,i} = k(x_\star, x_i)$.

Experiments using a Gaussian Process: Model selection



A toy example: A squared exponential kernel is used to fit the function $\sin\left(\frac{\pi}{2}x\right)$ shown in red.

Baryons from Mesons: Experiments using a Gaussian Process

Gaussian Process: We use a rational-quadratic kernel function with ARD distance measures. At the cost of using a more complex model, we also tried a kernel search using GPML, but that did not yield significantly better marginal likelihoods.

$$k_{\mathrm{RQ}}(x_i, x_j) \;=\; \sigma_f^2 \left(1 + \frac{1}{2\alpha}(x_i - x_j)^{\mathsf{T}} \Lambda^{-1} (x_i - x_j)\right)^{-\alpha}.$$

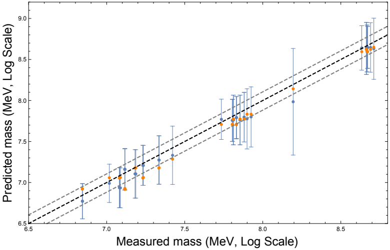
where $\Lambda = diag(\{\lambda_i^2\}_{i=1}^{D})$. λ_i is the *characteristic length scale* for the i^{th} feature. σ_f is an overall scale. $\alpha > 0$.

Baryons from Mesons: NN and GP predictions on lightest baryons

$$
\begin{array}{lll}
\text{NN:} & m_p = 1068 \pm 183 \text{ MeV}, & m_n = 1205 \pm 206 \text{ MeV}. \\
\text{GP:} & m_p = 893.8 \pm 194.4 \text{ MeV}, & m_n = 892.8 \pm 193.9 \text{ MeV}. \\
\text{actual:} & m_p = 938.28 \text{ MeV}\ , & m_n = 939.57 \text{ MeV}.
\end{array}
$$

- The GP predicts masses of proton and neutron to within 5%!.
- The neural network accurately predicts the proton to be the lightest in the spectrum.
- Gaussian process predicts the proton and neutron to be almost degenerate.

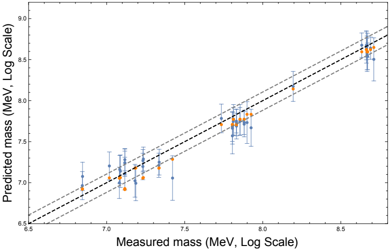
Baryons from Mesons: GP predictions on baryons



- We notice the clustering of masses into three groups.
- The first corresponds to baryons composed of light quarks, u, d, or s.
- The second group corresponds to baryons containing a c quark, while the third group corresponds to baryons containing b quarks.

Absolute errors: baseline model (*shown in orange*) 0.087, GP **0.034** $\pm$ **0.032**.

Baryons from Mesons: NN predictions on baryons



- We notice the clustering of masses into three groups.
- The neural network successfully predicts that the proton is the lightest baryon 79% of the time and in a plurality of cases concludes that the neutron is the next lightest particle.
- The first corresponds to baryons composed of light quarks, u, d, or s.
- The second group corresponds to baryons containing a c quark, while the third group corresponds to baryons containing b quarks.

Absolute errors: baseline model (*shown in orange*) 0.087, neural network 0.097 ± 0.074.

Baryons from Mesons: light–light and heavy–light systems

Hypotheses testing: conventional meson, or tetraquark?

	$I\,(J^P)$	Measured mass (MeV)	Composition	NN Pred	GP Pred	CQM Pred
$a_0(980)$	$1\,(0^+)$	980 ± 20	$u\bar{u}$	1277 ± 246	511 ± 34	680
			$u\bar{u}s\bar{s}\ (K\overline{K})$	2172 ± 466	1713 ± 68	1652
$f_0(980)$	$0\,(0^+)$	990 ± 20	$d\bar{d}$	921 ± 117	977 ± 37	672
			$d\bar{d}s\bar{s}\ (K\overline{K})$	1592 ± 401	1132 ± 312	1644

	$I\,(J^P)$	Measured mass (MeV)	Composition	NN Pred	GP Pred	CQM Pred
$D_{s0}^*(2317)^\pm$	$0\,(0^+)$	2317.8 ± 0.5	$c\bar{s}$	2640 ± 433	2434 ± 700	2036
			$c\bar{u}u\bar{s}\ (DK)$	4326 ± 925	2474 ± 826	2858
$D_{s1}(2460)^\pm$	$0\,(1^+)$	2459.5 ± 0.6	$c\bar{s}$	2547 ± 39	2535 ± 0.003	2036
			$c\bar{u}u\bar{s}\ (D^*K)$	3431 ± 544	2560 ± 788	2858

Baryons from Mesons: heavy–heavy systems

	$I\,(J^P)$	Measured mass (MeV)	Composition	NN Pred (MeV)	GP Pred (MeV)	CQM Pred (MeV)
$X(3872)$	$0\,(1^+)$	3871.69 ± 0.17	$c\bar{u}\bar{c}u$ $(D^0\,\bar{D}^{*0})$	4815 ± 786	3514 ± 190	3772
$Y(4260)$	$0\,(1^-)$	4230 ± 8	$c\bar{s}s\bar{c}$ $(D_s\,\bar{D}_s)$	$(5.4 \pm 1.1) \times 10^3$	3543 ± 1167	4072
$Y(4360)$	$0\,(1^-)$	4368 ± 13	$c\bar{u}\bar{c}u$ $(D_1\,\bar{D}^*)$	4940 ± 903	3107 ± 168	3772
$Y(4660)$	$0\,(1^-)$	4643 ± 9	$u\bar{u}c\bar{c}$ $(f_0(980)\,\psi')$			
$Z_c(3900)^\pm$	$0\,(1^+)$	3886.6 ± 2.4	$c\bar{d}\bar{c}u$ $(D\,\bar{D}^*)$	4991 ± 815	3515 ± 199	3776
$Z_c(4020)^\pm$	$0\,(1^+)$	4024.1 ± 1.9	$c\bar{d}\bar{c}u$ $(D^*\,\bar{D}^*)$			
$Z_c(4200)\pm$	$0\,(1^+)$	4196^{+35}_{-32}	$cu\bar{c}\bar{d}$			
$Z_c(4430)^\pm$	$0\,(1^+)$	4478^{+15}_{-18}	$c\bar{d}\bar{c}u$ $(D_1 D^*\,,\ D_1' D^*)$			
$Z_b(100610)^\pm$	$0\,(1^+)$	10607.2 ± 2.0	$b\bar{d}\bar{b}u$ $(B\bar{B}^*)$	$(1.47 \pm 0.17) \times 10^4$	9907 ± 560	10136
$Z_b(100650)^\pm$	$0\,(1^+)$	10652.2 ± 1.5	$b\bar{d}\bar{b}u$ $(B^*\bar{B}^*)$			

Baryons from Mesons: pentaquarks

$uudc\bar{c}$	$I\,(J^P)$	Measured Mass (MeV)	NN Pred	GP Pred	CQM Pred
$\mathrm{P}_c(4312)^+$	$\frac{1}{2}\,(\frac{1}{2}^+)$	$4311.9 \pm 0.7^{+6.8}_{-0.6}$	$(4.2 \pm 1.2) \times 10^3$	3544 ± 923	4112
$\mathrm{P}_c(4440)^+$	$\frac{1}{2}\,(\frac{1}{2}^-)$	$4440.3 \pm 1.3^{+4.1}_{-4.7}$	$(4.1 \pm 1.1) \times 10^3$	3253 ± 846	
$\mathrm{P}_c(4457)^+$	$\frac{1}{2}\,(\frac{3}{2}^-)$	$4457.3 \pm 0.6^{+4.1}_{-1.7}$	$(4.5 \pm 1.1) \times 10^3$	3581 ± 932	

Baryons from Mesons: baryon–antibaryon and dibaryon states

	$I\,(J^P)$	Measured mass (MeV)	Composition	NN Pred	GP Pred	CQM Pred
$f_2(1565)$	$0\,(2^+)$	1562 ± 13	$u\bar{u}$	1879 ± 70	1275 ± 0.002	672
			$uu\bar{u}\bar{u}d\bar{d}$ $(p\bar{p})$	2648 ± 613	1284 ± 85	2024
$^2\mathrm{H}^+$	$0\,(1^+)$	1875	$uuuddd$ (pn)	1585 ± 551	1167 ± 138	2028
$^2\mathrm{He}^{++}$	$0\,(1^+)$	1878	$uuuudd$ (pp)	1409 ± 484	1166 ± 163	2024

Summary and Outlook

- Knowledge of the meson spectrum alone is sufficient to approximate masses of both baryons and exotic colour singlet bound states of more than three quarks.
- Our ML outcome is consistent with the theoretical idea that baryons appear as solitons in the mesonic spectrum.
- We have shown that simple ML tools for widely available particle physics data can perform hypothesis testing, and make predictions for masses of (unobserved) states of quarks at the LHC.
- Further understand different aspects of QCD using machine learning, e.g., by including gluons in the picture.
- Akin to the Gell-Mann–Okubo mass formula, attempt to find an analytic expression for particle masses using symbolic regression, other methods for interpretation of ML models.

Acknowledgments

Chris Bouchard, Glasgow

Amanda Cooper–Sarkar, Oxford

Robert de Mello Koch, Witwatersrand

Amihay Hanany, Imperial

Koji Hashimoto, Osaka

Djordje Minic, Virginia Tech

Brent Nelson, Northeastern

Tim Rudner, Oxford

Subir Sarkar, Oxford

Lewis Smith, Oxford

Wati Taylor, MIT

Mark van der Wilk, Imperial

A second case of ML driven scientific discovery

Neural Network Approximations for Calabi–Yau Metrics

Vishnu Jejjala[a] , **Damián Kaloni Mayorga Peña**[a,b] , **Challenger Mishra**[c]

[a] *Mandelstam Institute for Theoretical Physics, School of Physics, NITheP, and CoE-MaSS, University of the Witwatersrand, Johannesburg, WITS 2050, South Africa*

[b] *Data Laboratory, Universidad de Guanajuato, Loma del Bosque No. 103 Col. Lomas del Campestre C.P. 37150 Leon, Guanajuato, Mexico*

[c] *Department of Computer Science & Technology, University of Cambridge, 15 J.J. Thomson Ave., Cambridge CB3 0FD, United Kingdom*

E-mail: vishnu@neo.phys.wits.ac.za, damian.mayorgapena@wits.ac.za, cm2099@cam.ac.uk

arXiv:2012.15821v1 [hep-th] 31 Dec 2020

Abstract: Ricci flat metrics for Calabi–Yau threefolds are not known analytically. In this work, we employ techniques from machine learning to deduce numerical flat metrics for the Fermat quintic, for the Dwork quintic, and for the Tian–Yau manifold. This investigation employs a single neural network architecture that is capable of approximating Ricci flat Kähler metrics for several Calabi–Yau manifolds of dimensions two and three. We show that measures that assess the Ricci flatness of the geometry decrease after training by three orders of magnitude. This is corroborated on the validation set, where the improvement is more modest.