Data Science Problems and Hidden Dynamical Systems

DESCI LONDON HACKATHON

Matteo Manzi, Enzo Caceres

2023/01/02

Data Science Problems and Hidden Dynamical Systems

abstract

DeSci London Hackathon is a hackathon event taking place on the 12th and 13th of January 2023. This hack is aimed at anyone interested in DeSci. In this small work we investigate the use of IPFS to foster the reproducibility of machine learning papers.

1. Machine Learning and Dynamical Systems: Supervised Learning

Building on the problem statement presented (Vahid Nateghi 2022) and (Champion et al. 2019), we assume a causal, deterministic system underlying a set of noisy, high-dimensional observations.

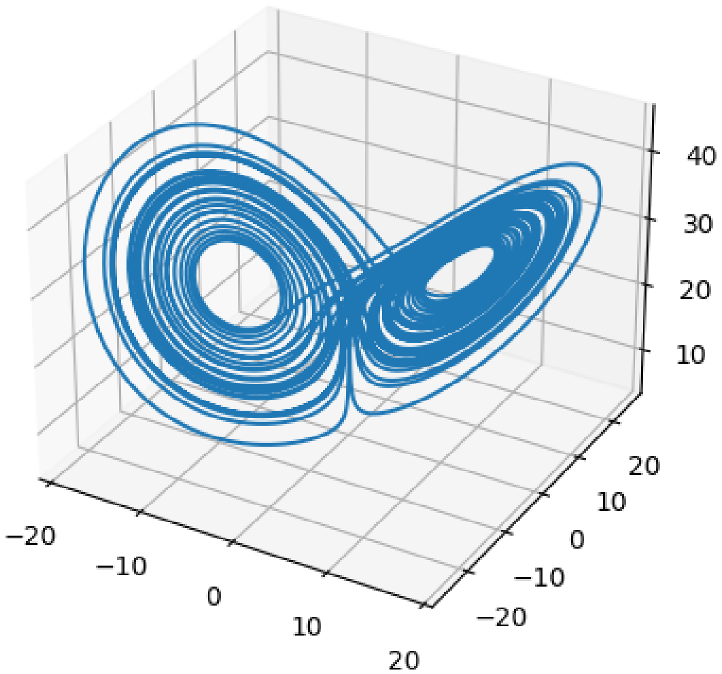

Figure 1: Sample trajectory for the Lorenz system

We are interested in solving a supervised learning problem, given a scalar field to be best fitted, in a mean squared error sense, using the observed features in time. This is a common problem statement in the machine learning literature. In order for claims associated with such works to be scientific, i.e., falsifiable (Prado 2022), (Popper 1963), it is necessary to attach to publications not only the source code written in the context of the work, but also data and outputs.

We enable this using an IPFS node, and including all the necessary CIDs into the source code. In this way it is also possible to save trained machine learning models.

2. Convex Optimization

This enables incremental works: assume one is interested in making use of the trained machine learning models in order to integrate it into a more complex pipeline. As an example, we use the trained model to estimate targets and use them as an input to a convex optimizer (Diamond and Boyd 2016).

3. Chaos

Following the methodology discussed in (Matteo Manzi 2022), it is possible to use the tools of chaos theory in order to investigate the existance of coherent structures in the underlying dynamical system (Froyland and Padberg 2009). As we are interested in exploring bifucating phenomena of such structures in a two-dimensional space, an animation can be used to do this. The use of IPFS enables the integration of multimedia contents like this in this work, accessible both from the paper and from the source code.

References

Champion, Kathleen, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. 2019. "Data-Driven Discovery of Coordinates and Governing Equations." *Proceedings of the National Academy of Sciences* 116 (45): 22445–51. https://doi.org/10.1073/pnas.1906995116.

Diamond, Steven, and Stephen Boyd. 2016. "CVXPY: A Python-Embedded Modeling Language for Convex Optimization." arXiv. https://doi.org/10.48550/ARXIV.1603.00943.

Froyland, Gary, and Kathrin Padberg. 2009. "Almost-Invariant Sets and Invariant Manifolds — Connecting Probabilistic and Geometric Descriptions of Coherent Structures in Flows." *Physica D: Nonlinear Phenomena* 238 (16): 1507–23. https://doi.org/https://doi.org/10.1016/j.physd.2009.03.002.

Matteo Manzi, Mattia Petrini, Stefano Guidolotti. 2022. "Interplay Between Chaos and Stochasticity in Celestial Mechanics." *Julia Conference 2022 Proceedings.*

Popper, Karl. 1963. "Conjectures and Refutations." *Routledge and Kegan Paul, 1st Edition, London, UK.*

Prado, Marcos M. Lopez De. 2022. "Causal Factor Investing: Can Factor Investing Become Scientific?" *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4205613.

Vahid Nateghi, Matteo Manzi. 2022. "Machine Learning Methods for Nonlinear Reduced-Order Modeling of the Thermospheric Density Field." *Preprint*. https://doi.org/10.13140/RG.2.2.12431.05280.

crunch
dao