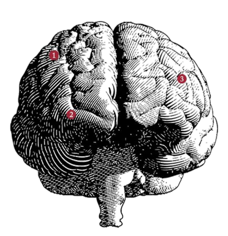
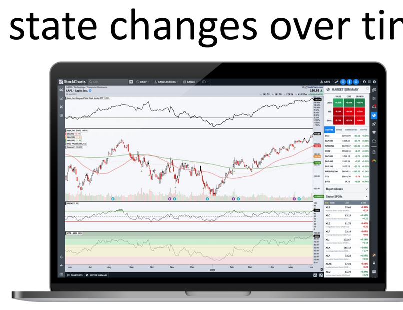
Improving Dynamical Systems Benchmarking: Extending Libraries and Enhancing OOD Capabilities.

Hadil Owda

Brains in Silicon, Boahen Lab, Advised by Max Kanwal

Introduction

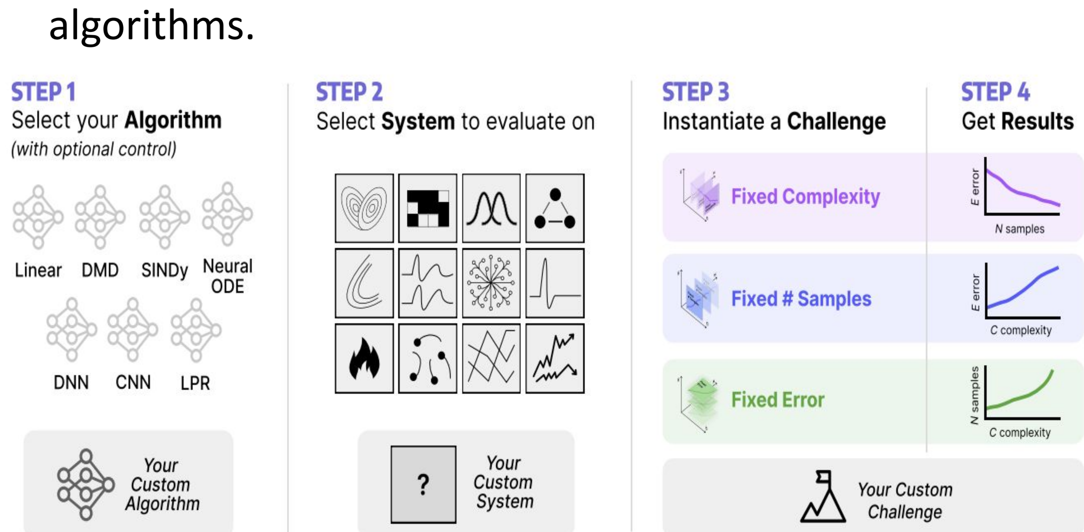
Dynamical systems are mathematical models that describe how a state changes over time.


They exhibit complex behaviors that are difficult to predict and control due to the lack of predefined equations. Solution is ***data-driven ML***, but its effectiveness relies on selecting the appropriate algorithm, so we need ***benchmarking***!

What is DynaDojo?

An ***open and extensible*** benchmarking platform for ML algorithms.

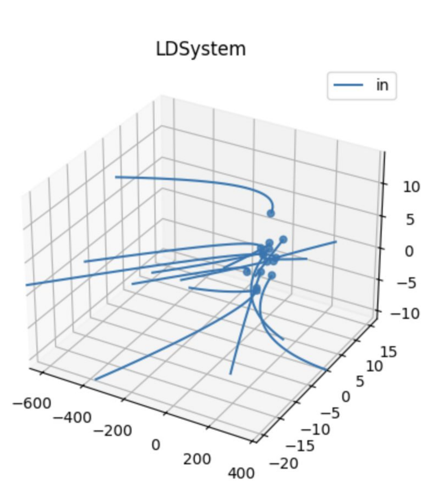


Extending Systems Library

Previously, Dynadojo had only 20 systems and relied on known equations to compute trajectories.

However, research is also interested in complex systems that lack known equations but have precise data.

Goal was to build the functionality to incorporate non-procedurally generated dynamical systems.



Acknowledgements

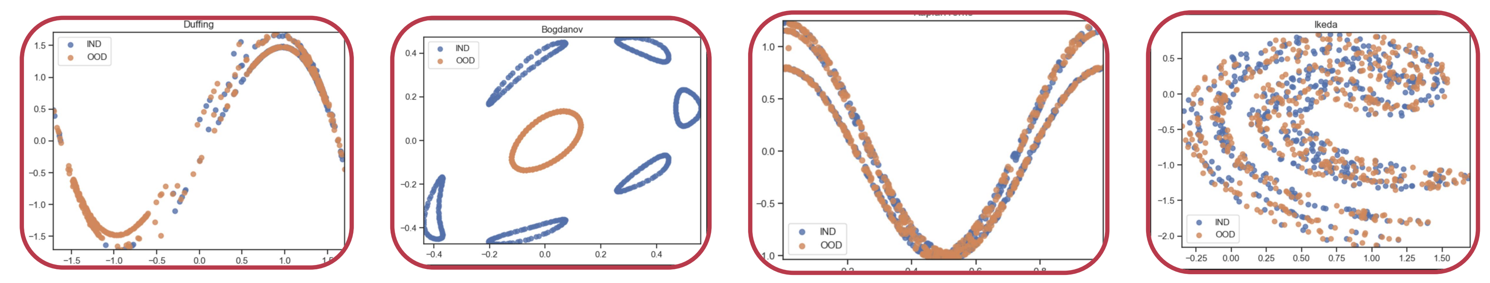
I thank my mentor, Max Kanwal, for the great mentorship and help during my first research experience, and Professor Kwabena Boahen for having me in the Brains in Silicon lab. I also thank the Electrical Engineering department and the REU program for funding my research. Lastly, thanks to my teammates!

References:

- William Gilpin. "Chaos as an interpretable benchmark for forecasting and data-driven modelling" Advances in Neural Information Processing Systems (NeurIPS) 2021 https://arxiv.org/abs/2110.05266
- Bhamidipaty, L. M., Bruzzese, T., Tran, C., Mrad, R. R., & Kanwal, M. (2023, November 2). Dynadojo: An extensible benchmarking platform for scalable... OpenReview. https://openreview.net/forum?id=pTSNoBTk8E

Extending Systems Library: Process & Results

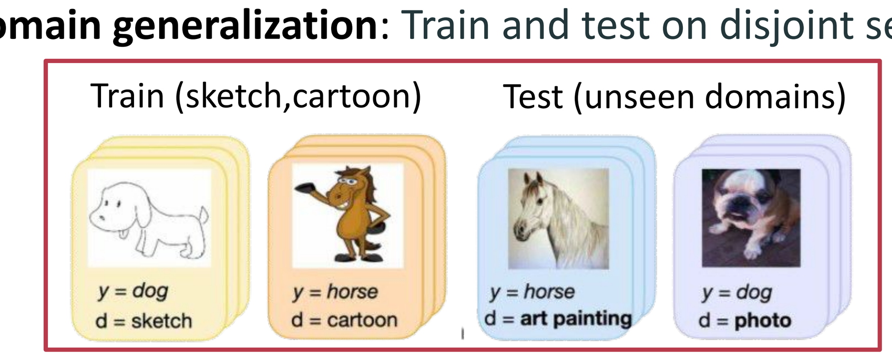
Adapted Gilpin's chaotic dynamical systems database into DynaDojo, vastly expanding the benchmarking and analysis options by ***100+ systems.***



Compatible with DynaDojo algorithms and Fixed Complexity challenge.

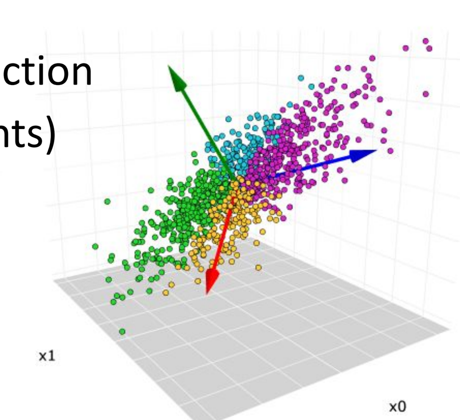
What is Out-of-distribution (OOD) Data for Time Series?

Covariate shifts/Domain generalization: Train and test on disjoint sets of domains.

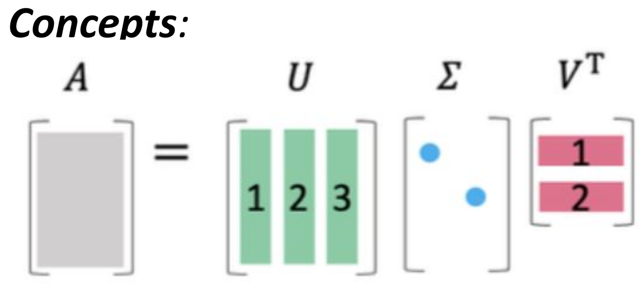


OOD Generation

Principal Component Analysis (PCA): A dimensionality reduction technique that identifies the directions (principal components) capturing the most variance in the data.



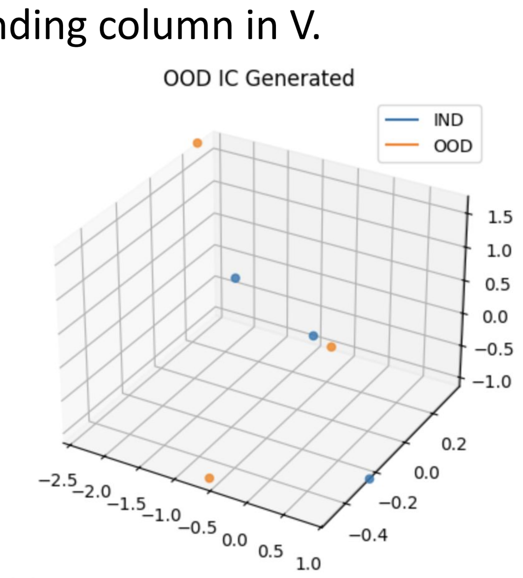
Key Mathematical Concepts:



- **Maximum Variance**: corresponds to the largest singular value in Σ.
- **Direction of Maximum Variance**: given by the corresponding column in V.

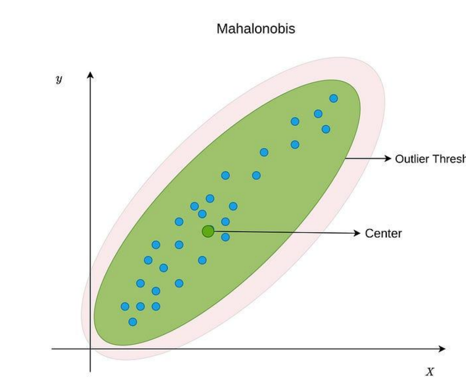
Procedure for OOD IC Generation:

- **Select Principal Components**: choose components that explain at least 90% of the variance.
- **Identify Remaining Components**: remove the selected components from the total set.
- **Generate OOD Points**: perturb data along these directions of minimal variance.



OOD Detection

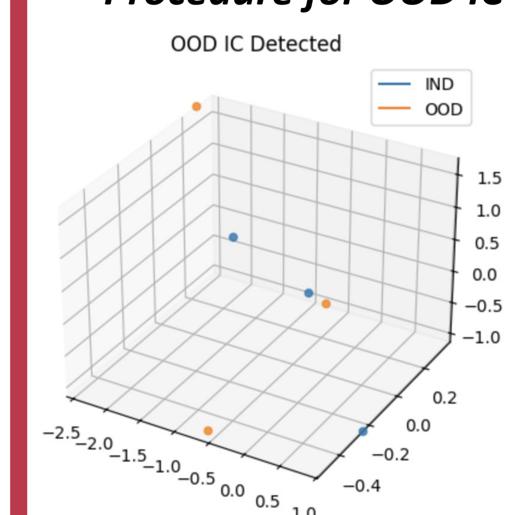
Mahalanobis Distance: measures how far a point is from a distribution, considering correlations between variables. Unlike Euclidean distance, it accounts for the data's structure.



Mathematically, between a point **x** and a distribution with mean vector **μ** and covariance matrix ***S*** it is defined as:

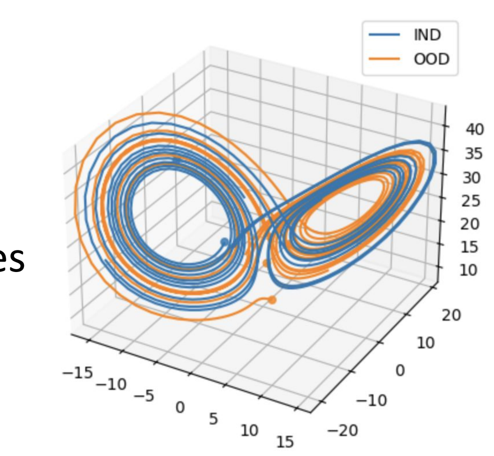
$$d_M(\vec{x}, Q) = \sqrt{(\vec{x} - \vec{\mu})^{\mathsf{T}} S^{-1} (\vec{x} - \vec{\mu})}.$$

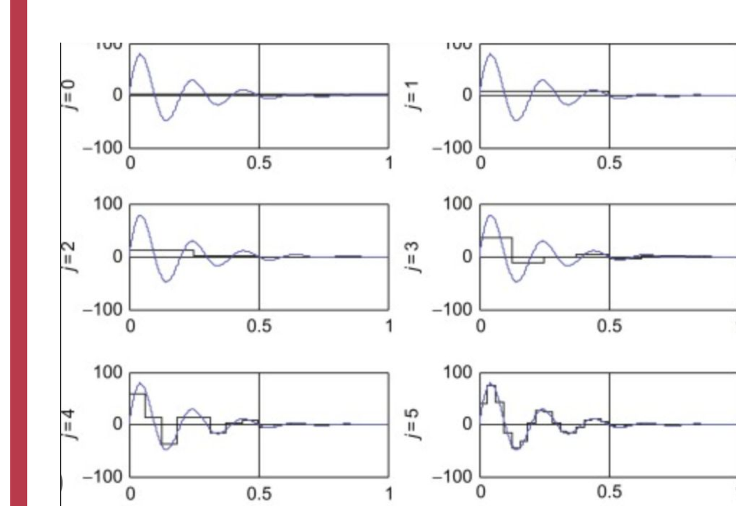
Procedure for OOD IC Detection:



- Calculate the mean and inverse covariance matrix from the in-distribution (IND) initial conditions.
- Compute Mahalanobis distances for each IND trajectory.
- Set a threshold for OOD detection based on the Mahalanobis distances of IND data.
- Classify initial conditions based on whether their Mahalanobis distance is below or above the threshold.

True OOD detection should be performed on the ***full trajectories***. To fit mahalanobis distance on trajectories accurately we need to extract ***specific features***.





Haar Wavelet Transform: extract features from time series trajectories by decomposing the signals into approximation and detail coefficients at multiple levels. Chosen for its ability to capture both ***spectral and temporal*** details, it provides ***approximation coefficients*** for broad trends (averages of pairs of data points) and ***detail coefficients*** for localized variations (differences between consecutive data points). These features were then used for out-of-distribution detection.

Next Steps

1. **Adversarial Training:** Implement an adversarial learning framework where PCA-based OOD generation and Mahalanobis distance detection with Haar wavelet transform are iteratively trained against each other to enhance both generation and detection accuracy.
2. **Static Dataset OOD Handling:** Extend the framework to handle cases where OOD data availability is unknown, ensuring robust OOD detection even in datasets without prior information about OOD samples.