Discovering Cosmological Relations With Symbolic Machine Learning

Kerry Zhao, School of Engineering and Applied Science, COL 2026
Mentor: Mathew Madhavacheril, School of Arts and Sciences, Department of Physics and Astronomy
Funded by the Penn Undergraduate Research Mentoring Program

PURM Penn Undergraduate Research Mentoring Program

Penn CURF CENTER for UNDERGRADUATE RESEARCH & FELLOWSHIPS

Abstract

Understanding the intricate connection between cosmological parameters is crucial for gaining insights into the evolution and properties of the universe. This research project employs the Python Symbolic Regression (PySR) package within a Jupyter Notebook environment to explore the relationship between two fundamental cosmological parameters: σ8 (the amplitude of the matter power spectrum) and ΩM (the fraction of the universe's total mass-energy content due to matter).

PySR is a powerful tool that utilizes symbolic regression to automatically discover mathematical expressions that best fit a given dataset. This work leverages PySR to construct a symbolic model that accurately describes the interdependence between σ8 and ΩM based on simulated cosmological data.

Results

This project is still ongoing, and a conclusive predictive model is still in the process of being found for the target set of cosmological data. However, predictive models for other sample sets of data have yielded promising results, showing that PySR is able to successfully predict complex expressions with little to no manual guidance. Once a result is obtained, the symbolic model can be integrated into cosmological analyses, aiding parameter estimation, model comparison, and predictions of cosmic phenomena.

Introduction

Cosmology, the scientific study of the universe's origin, evolution, and large-scale structure, has made remarkable strides in unraveling the mysteries of our cosmos. Key to this pursuit are cosmological parameters, fundamental values that characterize the universe's properties. Among these, σ8 and ΩM stand out as cornerstones in our understanding of cosmic structure formation and expansion. σ8 represents the amplitude of the matter power spectrum, quantifying the level of fluctuations in the distribution of matter. ΩM, on the other hand, denotes the fraction of the universe's total mass-energy content attributed to matter.

The relationship between σ8 and ΩM is of paramount importance, influencing a wide array of cosmological phenomena, from the formation of galaxies and galaxy clusters to the overall expansion history of the universe. A comprehensive grasp of this relationship can provide insights into the physical processes underlying cosmic evolution and aid in refining cosmological models. As observational and computational techniques continue to advance, the ability to accurately model and interpret the interplay between these parameters becomes increasingly vital.

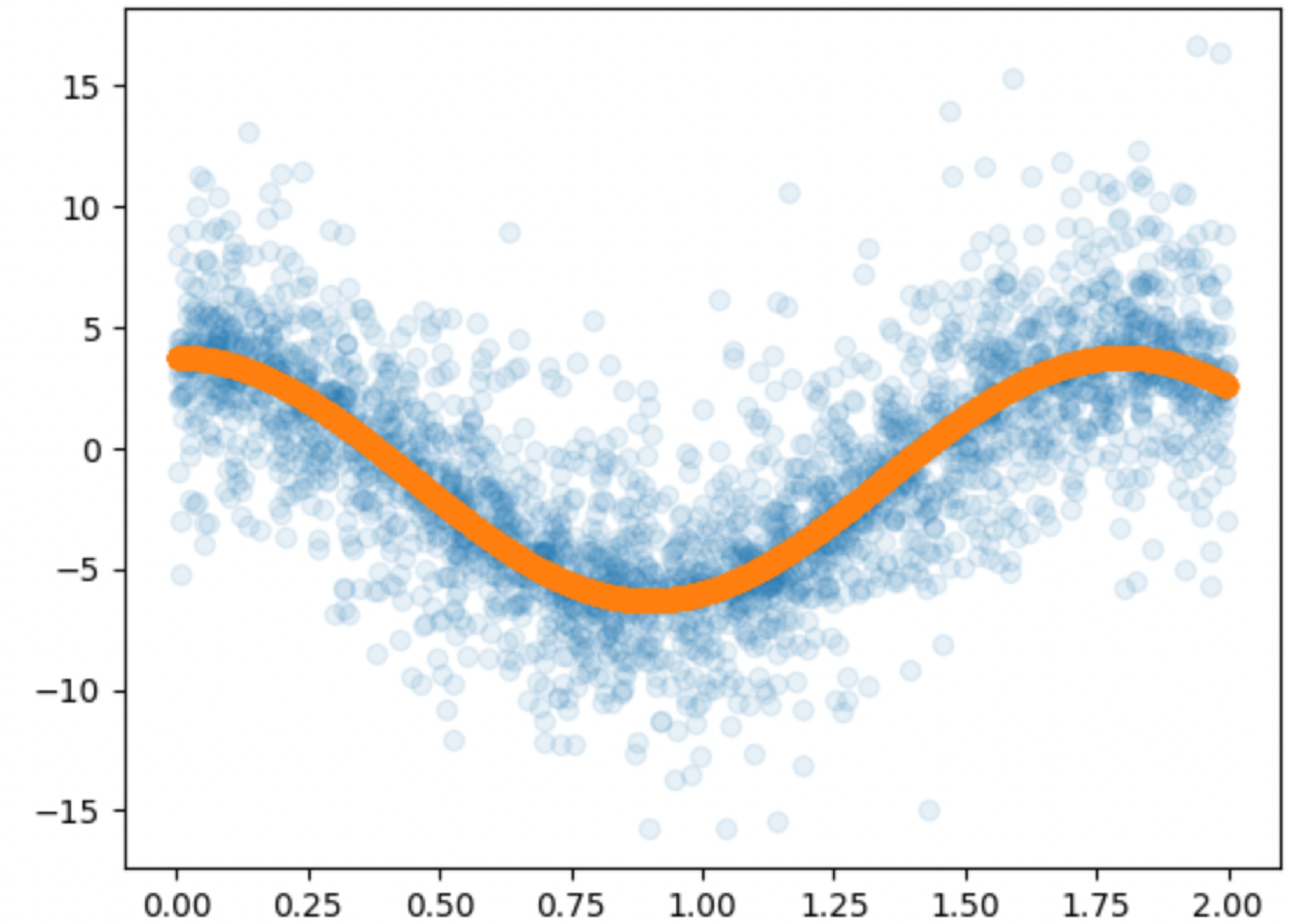
The significance of this project lies in its potential to provide a compact and interpretable mathematical representation of the σ8-ΩM relationship, shedding light on the complex interactions between these parameters in cosmological models. The symbolic model may offer insights into the underlying physical mechanisms governing cosmic structure formation and evolution. Moreover, it could contribute to enhancing the accuracy and efficiency of parameter estimation and model comparison in cosmological studies.

Methods and Materials

Symbolic regression is a technique used to find a mathematical expression as a model for a given dataset. In contrast to ordinary regression, which requires a model structure to fit to the dataset, symbolic regression does not require an existing model, as it uses machine learning to generate both the model and its parameters.

First, a high-quality dataset of cosmological simulations varying σ8 and ΩM is generated. Next, PySR is employed to perform symbolic regression on this dataset, searching for mathematical expressions that best capture the relationship between the two parameters. The resulting symbolic model is analyzed and validated using statistical measures and visualizations.

<matplotlib.collections.PathCollection at 0x7fee8ba66320>



A sample set of data points along with the predictive expression generated by PySR