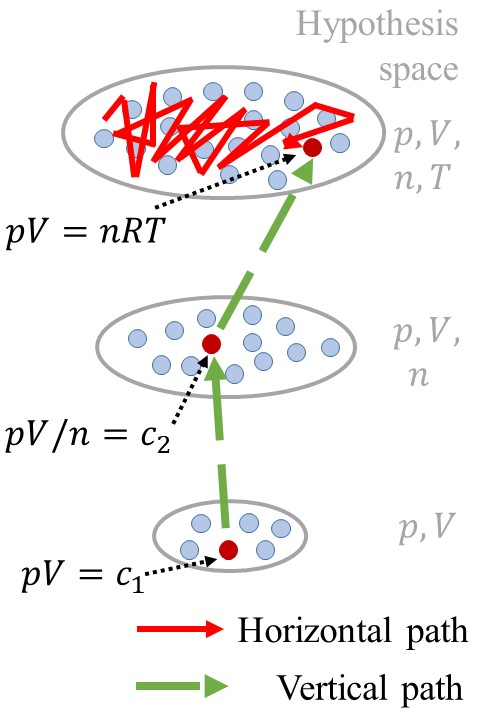
**Symbolic Regression Using Scientific Approaches (SRSci)**

****Automating scientific discovery has been a grand goal dating back to the founders of AI (Herbert Simon et. al. [14, 12, 30]) but remains a holy grail today. The underlying societal impact is immense because of its multiplier effect on applicational domains.

Symbolic regression discovers an underlying equation which maps the input to the output of a given dataset. It is one of the best benchmarking tasks for scientific discovery, as it mimics the process of discovering physics laws from experiment data. Over the years, much effort has been made toward symbolic regression, including search-based methods [13, 15], genetic programming [27, 29, 25, 5], reinforcement learning [21, 26, 20, 21], deep function approximation [19, 2, 23, 22, 17, 31, 3, 7, 1], integrated systems [28, 10, 9, 16]. Most endeavor focuses on horizontal discovery paths, i.e., they directly search for the best equation in the full hypothesis space involving all independent variables (red path in Figure 1). This path can be challenging because of the exponentially large full hypothesis space.

Interestingly, the widely applied scientific approaches imply a vertical discovery path, which may be more efficient. To discover the ideal gas law pV = nRT, scientists first held n (gas amount) and T (temperature) as constants and find p (pressure) is inversely proportional to V (volume). They then studied the relationship between pV and n, T. This led to a vertical discovery path (green long dashed path in Figure 1). The first few steps of a vertical path can be significantly cheaper than the horizontal path, because the searches are in reduced spaces involving a small number of independent variables. As a result, vertical discovery has the potential to supercharge state-of-the-art approaches in modeling complex scientific phenomena with more interlocking contributing factors or processes than what current approaches can handle.

This competition calls for **efficient symbolic regressors, which are free to follow vertical (scientific approach motivated), horizontal (machine learning driven) paths, or any new scientific discovery paths.** The goal of this challenge is to explore various approaches to querying data and fitting the model to the dataset. The learning algorithms can determine what training data is needed to uncover the underlying symbolic expression.

Design Your Experiments!

In order to discover a symbolic equation , we give you an oracle which will return the output given which you query. You are allowed to come up with multiple minibatches of X. The oracles return Y, which are noisy observations of .

Competition Details.

We evaluate your program on a Python platform. Within a given time limit, you are allowed to query the oracle as many times as possible. Your program needs to write the symbolic equation found to a pre-specified output file. The program is allowed to overwrite the output file for as many times as you’d like within the time limit. Only the equation in the last output file will be evaluated.

Competition Tracks

Provide a table of time limit, noise type, dataset.

How to Participate

[Go to the page of how to participate]

How we will evaluate

**Organizer**[**Permalink**](https://jiangnanhugo.github.io/scibench/#organizer)

* Nan Jiang (Purdue University), jiang631 at purdue dot edu
* Yexiang Xue (Purdue University), yexiang at purdue dot edu

**Discord Contacts**[**Permalink**](https://jiangnanhugo.github.io/scibench/#discord-contacts)

If you want to ask any question, provide some feedback or simply chit-chat, join us at the [Discord server](https://discord.gg/MeGnkHr4).