1. **Short Description of the Problem Explored (Symbolic Regression):**

Symbolic regression is an approach used in genetic programming (GP) to find a mathematical model that best fits a given dataset. Unlike traditional regression, symbolic regression does not require a predefined model structure but instead evolves the structure along with the parameters. The goal is to discover the underlying mathematical relationship between independent variables and dependent variables by evolving expressions in the form of computation trees. The process mimics natural selection to evolve models that minimize error in predicting the output.

1. **Table of GP Parameters, Including User and Defaults. Indicate the Parameter Variants:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parameter | Description | Default Value | Variant 1 | Variant 2 | Variant 3 | Variant 4 |
| Population Size | The size of the Population | 500 | 500 | 500 | 500 | 500 |
| Number of Generations | The number of generations | 30 | 30 | 30 | 30 | 30 |
| Tournament Size | The selection pool size for best fitness | 5 | 5 | 5 | 5 | 5 |
| Crossover Probability | The probability of crossover | N/A | 0.9 | 1.0 | 0.0 | 0.9 |
| Mutation Probability | The probability of mutation | N/A | 0.1 | 0.0 | 1.0 | 0.1 |
| Elitism | Number of top individuals to keep | N/A | 2 | 2 | 2 | 0 |

1. **Short Description of Fitness Evaluation: How Data Processed, Fitness Formula:**

Fitness evaluation in this GP task is done by assessing how well a candidate program predicts the output variable from the input variables. The data is processed by simulating each candidate program with input values and comparing the program's output against the true output values from a dataset. The fitness formula used is the mean squared error (MSE), calculated as the average of squared differences between the predicted and actual values. A size penalty is added to the fitness to control for bloat, ensuring that simpler, more generalizable solutions are favored.

1. **Short Description of the Independent Variable Changed:**

The independent variables experimented with in this GP task are the probabilities of crossover and mutation, along with the use of elitism. These parameters influence the genetic diversity of the population and the balance between exploration and exploitation in the evolutionary search process:

* Crossover Probability: Influences the rate at which new offspring are created by recombining parts of two-parent programs.
* Mutation Probability: Affects the rate at which new offspring are created by randomly altering parts of a single-parent program.
* Elitism: Determines the number of top-performing individuals preserved from one generation to the next.

The experiment that gave the best fitness performance was the one with a crossover probability of 0.9, a mutation probability of 0.1, and an elitism count of 2.

1. **Two Fitness Plots Averaged Over the 10 Runs:**

**For Crossover/Mutation Variants:**

A graph of a graph

Description automatically generated

A graph with a line

Description automatically generated

**For Elitism Variants:**

A graph with a line

Description automatically generatedA graph with a line

Description automatically generated

1. **A Short Conclusion:**

The overall results of the comparisons suggest that a balanced approach to crossover and mutation probabilities with the inclusion of elitism yields the best performance in symbolic regression tasks. Specifically, a higher crossover probability (0.9) combined with a lower mutation probability (0.1) and moderate elitism (2) led to the most successful outcomes. This parameter choice promotes both the exploration of new areas of the search space and the exploitation of existing good solutions. The use of elitism helps preserve high-quality solutions across generations, contributing to steady progress towards optimal fitness. However, the absence of elitism did not significantly degrade performance, which could imply that the GP was effective at exploring the search space even without directly preserving the best individuals.