

Final Project

Genetic Programming vs Multi-Layer Perceptron

1. (Same as HW4) Use genetic programming to find the symbolic formula (f) that most accurately describes the given data. Each data (data-gp1.txt, data-gp2.txt) contains 200 xy coordinates, where

$$y = f(x) + \text{noise}.$$

Restrictions on population size and generation include:

- Maximum Population Size = 500
- Maximum Generation = 500

Print the two symbolic formulas that most accurately describe each piece of data.

2. Submit a brief report, containing:

- (a) From the output of problem 1, calculate (1) **the root mean square error** of the symbolic formulas, (2) **the number of nodes** in the generated trees.
- (b) Solve problem 1 using Multi-Layer Perceptron (MLP). Fit the given data (data-gp1.txt, data-gp2.txt) using MLP of your own structure. Here you can use a deep learning library (e.g., PyTorch, TensorFlow, etc.). **Compare the performance of the GP and MLP and write a conclusion.**

Notice:

Implement it yourself. No cheating. No plagiarism. If you copy code from the others (including Google and GitHub), you will get **0 points**. Changing only variable names is also cheating. Use of the GP library will give you **0 points**.

Problem 1 must be solved using Genetic Programming. If you solve without GP (e.g., optimizing function coefficients using GA), you will get **0 points**.

Submit a zip file to the TA's email containing the following:

- Output: Two symbolic formulas that most accurately describe each piece of data.
- Source code (Python, C, or C++): **Source code of Problem 1** that reproduces the output.

- Report (PDF or Word)

The name of the zip file must include **your name** and **your student id**.

Due: 12/22 23:59 (**no late submission**)

Email: cheetos@gm.gist.ac.kr