

## Final Project

---

### Genetic Programming

Use genetic programming to find symbolic formula  $f(x, y)$  that most accurately describe the given data. The data (datapoint3d.txt) consists of 300 x,y,z coordinates, where

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

Implement the genetic programming algorithm with some restriction below:

- Maximum Population Size: 500
- Maximum Generation: 500

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

The usage of GP Library is **prohibited**

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

1. Symbolic equation (equation.txt): mathematics model that most accurately describe data
2. Python function of the mathematics model (model.py):
  - Create one function f with parameter x and y, that output z. The pseudocode given below and example "model.py" given along with this pdf

```
Function f(x,y)
```

```
    z = compute symbolic function given x and y
    return z
```

3. Report (report.pdf), containing:
  - Explain method used for solving the problem
  - Compute the **root mean square error (RMSE)** of the symbolic formulas
  - The number of **nodes and in the generated trees**
4. Source Code (Python, C, C++)

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

**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 name is also cheating. Using GP library also get **0 points**.

The problem must be solved by Genetic Programming. If you solve without GP (optimizing function coefficient with GA), you will also get 0.

Due: 12/21 23:59 (No late submission)

Email: [anandaphan@gm.gist.ac.kr](mailto:anandaphan@gm.gist.ac.kr)