

Software Development

Project

Sergio Peignier

sergio.peignier@insa-lyon.fr

Associate Professor

INSA Lyon

Biosciences department

Table of contents

1. Organization
2. Problem Statement
3. Product Owner Requirements

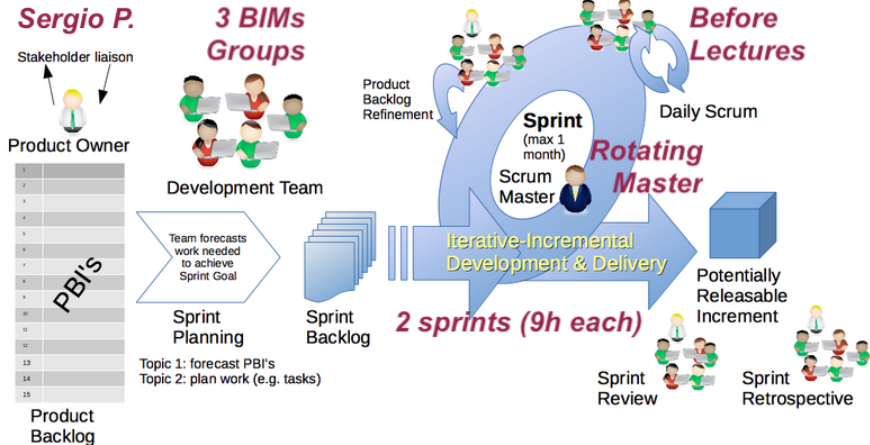
Organization

Randomized Groups

Why? Closer to real life conditions...

Check your groups directly on discord

Scrum-Flavored Organization

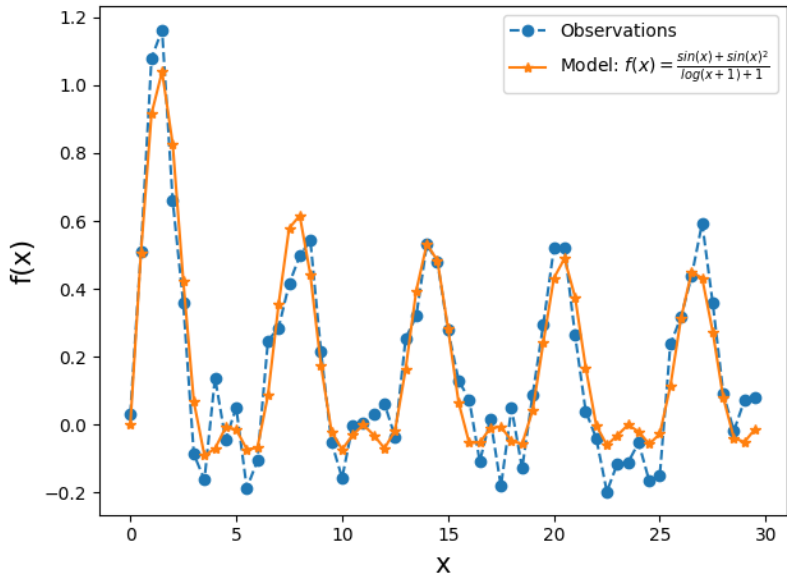


Evaluation

- Product and sprint backlogs (1/6)
- Scrum dynamic (1/6)
- Product presentation (1/6)
- Product (code, documentation) (1/2)

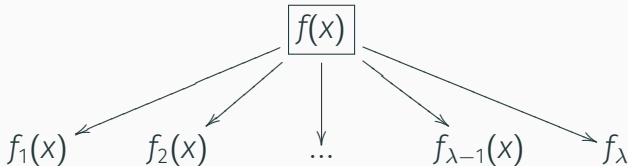
Problem Statement

Symbolic Regression



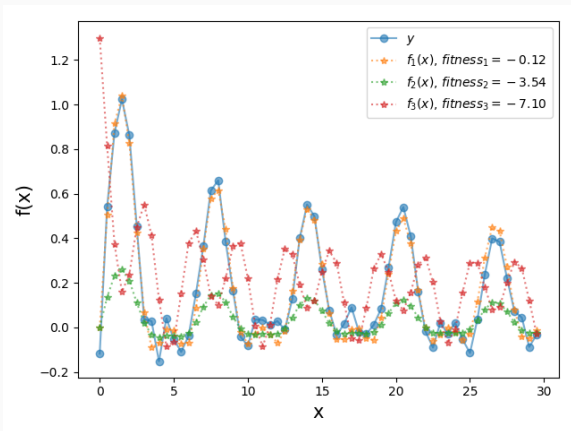
Evolution Strategy

- Formula \leftrightarrow Organism
- 1 parent: $f(x)$
- λ children $f_1(x), \dots, f_\lambda(x)$ (replication and mutation)
- Select best child (highest fitness) \rightarrow next parent.



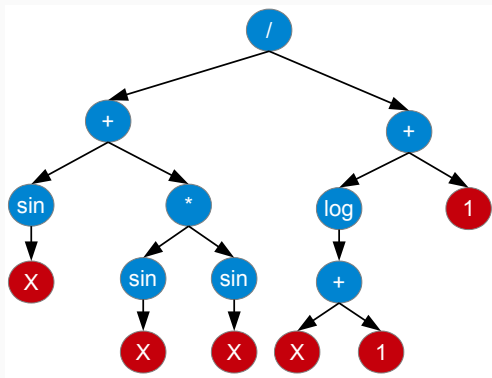
Fitness

- Variables: (x_1, x_2, \dots, x_n)
- With observations (y_1, y_2, \dots, y_n) .
- $fitness_k = -\sum_i (y_i - f_k(x_i))^2$



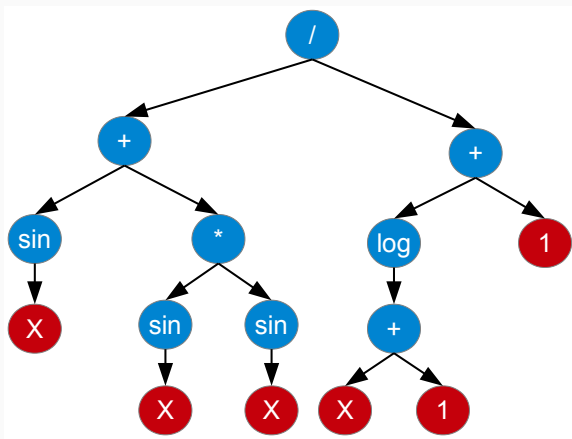
Tree Representation

- Formula \leftrightarrow Tree
- **Operations** internal nodes (blue)
- **Variables/constants** leaf nodes (red)
- **Bottom-up recursive** evaluation of the tree.



Tree Representation | Example

$$f(x) = \frac{\sin(x) + \sin(x)^2}{\log(x+1) + 1}$$

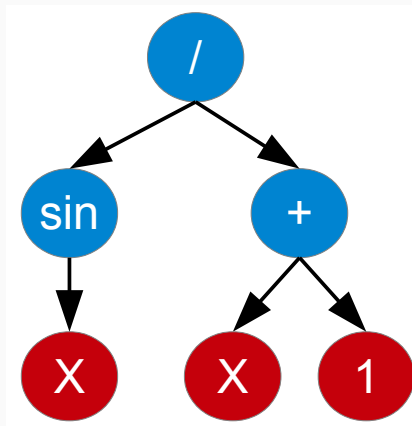


Mutations

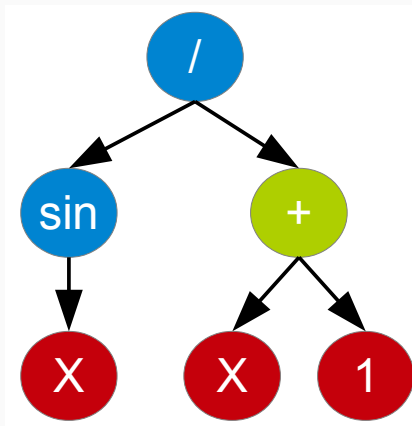
Some examples:

- Insertions
- Deletions
- Replacements
- ...

Mutations | Insertions

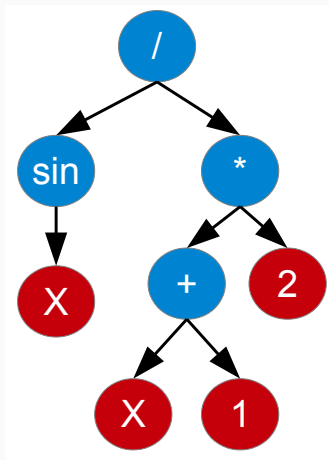


Mutations | Insertions



Pick a node n randomly

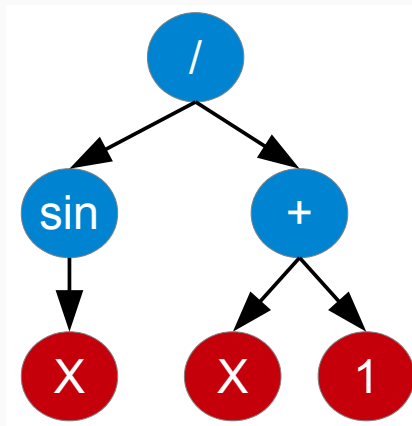
Mutations | Insertions



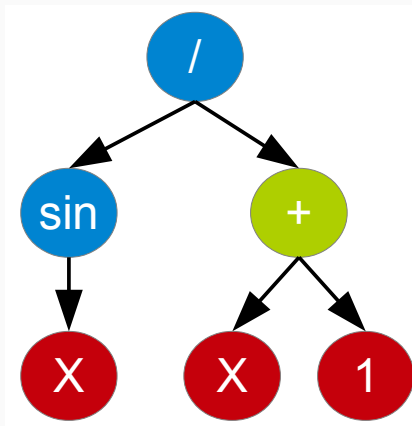
Insert a new function node before n

(Eventually insert a random constant or a variable node)

Mutations | Deletions

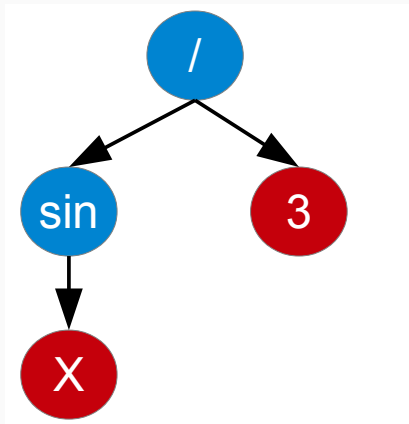


Mutations | Deletions



Pick a node n randomly

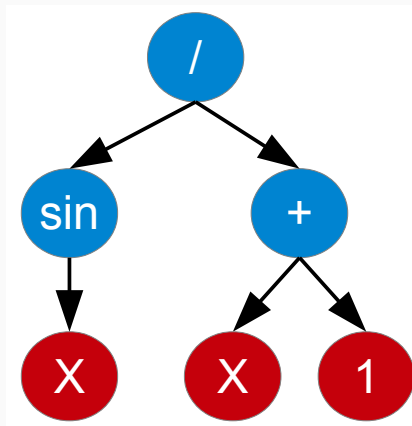
Mutations | Deletions



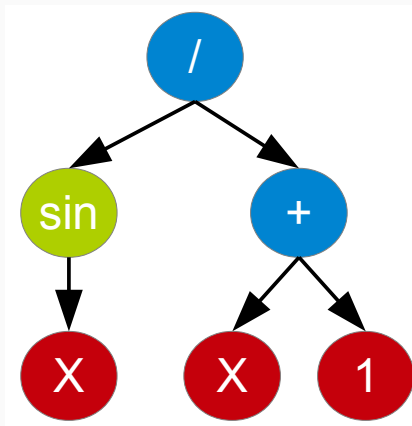
Delete the sub-tree at n

(Insert a random constant or a variable node instead)

Mutations | Replacements

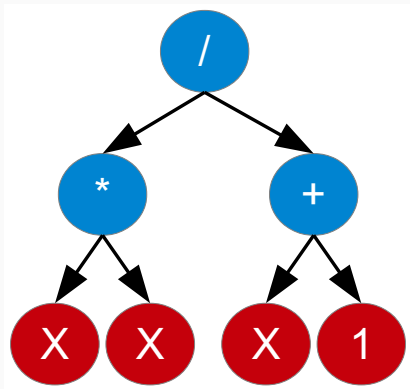


Mutations | Replacements



Pick a node n randomly

Mutations | Replacements



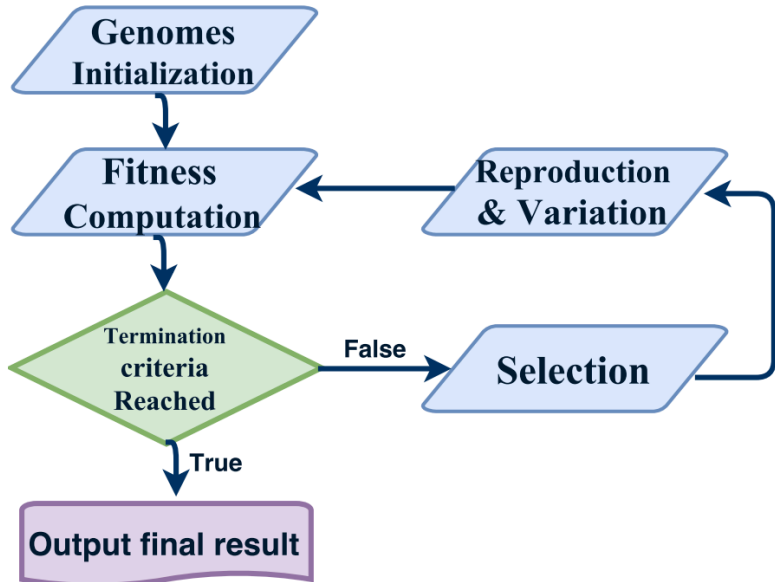
Replace node n by another node

(Eventually insert a random constant or a variable)

Unsuitable individuals

- Beware of **unsuitable individuals** e.g.:
 - Invalid input value: $\sqrt{-3}$
 - Invalid number of inputs: $\sin(x, 3)$
 - No input value: function node in leaf: $\sin()$
 - Const./var. as function internal node: $12(\cos(x))$
 - ...
- Possible **solutions**:
 - Assign a **strongly negative fitness** $-\infty$
 - **Guided mutations**
 - Post-mutation **corrections**

Overall Scheme



Product Owner Requirements

Requirements

- First use case - **Boolean regression**:
 - m Boolean input variables $x_i \in \{0, 1\}, \forall i \in \{1, \dots, m\}$
 - Boolean constants $\in \{0, 1\}$
 - Basic Boolean Operations: **AND, OR, NOT**
- **Modular/Extensible architecture (not this year)**:
 - New sets of functions (e.g., defined in \mathbb{R}).
 - New fitness functions.
 - New mutation operations.
 - New selection procedures.

C/C++ Extension Module for Python

Core computations in C/C++:

- Mutations
- Fitness computation
- Selection

Python Wrapper interface (not this year):

- Get individual formula/tree.
- Send variables/observations and get fitness.
- Send input variables and get result.

Simple Python auxiliary functions:

- Visualize best individual.
- Plot fitness evolution.
- Simplify formula using SymPy

2019-2020 Specific Steps

- Compile and execute one or many executable files, sending the parameters settings and input file names (matrix X and vector y) as command line arguments
- These programs can be called from python
- Results should be written in files
- Python functions for visualization or post-processing should access such files.

Resources

- Minimal C++/Python Wrapper example (not this year)
- Some auxiliary python functions

Some Links

- <http://groups.csail.mit.edu/EVO-DesignOpt/groupWebSite/uploads/Site/UM-GP-tutorialGECC02011.pdf>
- <https://jankrepl.github.io/symbolic-regression/>
- <https://www.youtube.com/watch?v=9TycLR0TqFA>