Objectives

- Given a regression dataset, explore Genetic Programming (GP) strategies that generate individuals with competitive generalization

- Find ways to properly balance between learning the training data and generalizing to unseen data

Task

- You are provided with a dataset and are allowed to use it as you wish

- You can use any form of tree-based GP (e.g., standard, geometric semantic, etc)

- You can use methods from outside the GP area as long as the final individual is a GP individual

- In the end of the project, you are submitting a final individual to be evaluated on data that was withheld

- The final model must be a tree-based GP individual, defined by the same structure as provided in the baseline code

- You cannot change the targets (e.g., you cannot normalize the values)

Dataset

- Real-world regression dataset

- 8 attributes

- 300 instances available to build your individuals

- 200 instances left out to evaluate the final individual

Delivery

- Send me an email until 2017-06-07 23:59, with the following:

  - the proposed individual/model
  - the source code
  - a small report describing the experiments conducted

Evaluation

- 50%: generalization of the proposed individual

- 40%: ideas tested

- 10%: report