Computação Evolucionária

2013/2014

Notes for the PL 7

Ernesto J. F. Costa

25 de Março de 2014

Class PL 7

Genetic Programming: a few experiments

7.1 Introduction

In the lectures we introduced the basics of genetic programming (GP). The most relevant aspect of this approach is the fact the the algorithm evolves programs which, when executed, solve a particular problem. In this lab we are going to do some experiments using a simple implementation of GP, in order to grasp the way it works. We will also try to do some modifications to the provided algorithm, and test the best configuration for the several parameters. We will use classical benchmark problems.

You must download from InforEstudante the following files:

1. gp_2014.py
2. gen_data.py

In order to succeed you must study previously the given code to understand how it works. The data to be used are generated by a program and stored into a file. The file has an header that defines the number of input variables of the problem, followed by pairs indicating a function's name and arity. After the header we have the examples, one by line, of pairs inputs-output. The figure 7.2 show the example of symbolic regression.

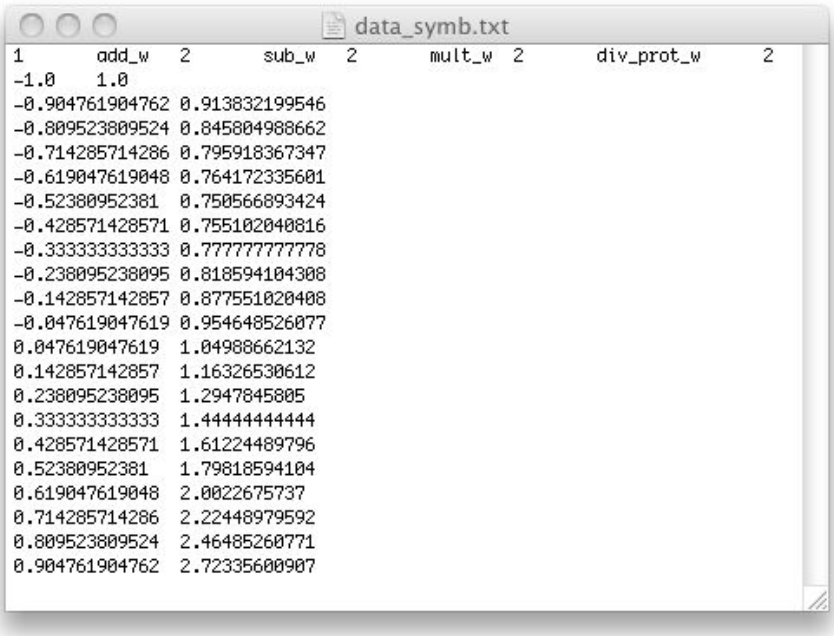

Figura 7.1: Symbolic Regression

7.2 The problems

7.2.1 Symbolic Regression

The problem of symbolic regression consists in finding the best polynomial that approximates a given set of points. In our case we will deal with the case of the polynomial

$$x^2 + x + 1$$

with $x \in [-1, 1]$. You should use the programs provided in file **gen_data.py** to generate the training set of data. The figure 7.2 show the polynome when you use a set of 21 points.

7.2.2 The `sin` function

The `sin` function is just the ... the `sin` function. We will work with a training set of 62 points in the interval $x \in [-3.14, 3.14]$. The classical figure in shown in 7.3.

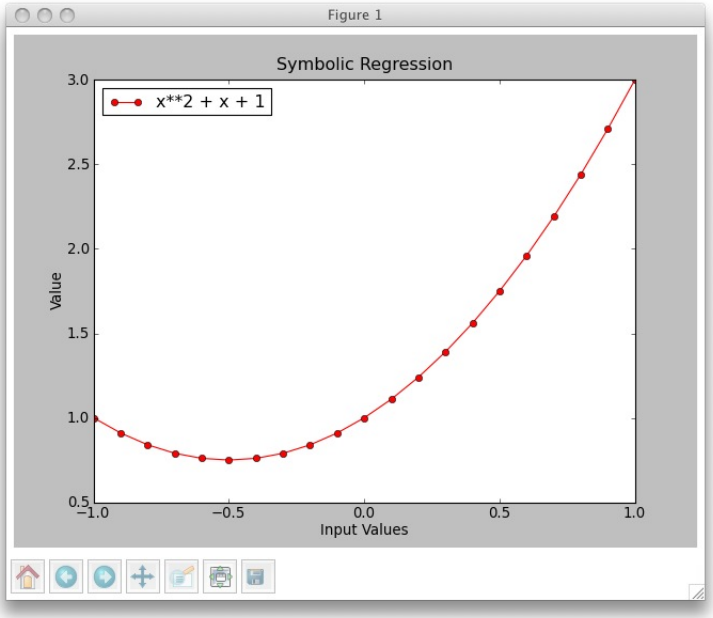

Figura 7.2: Symbolic Regression

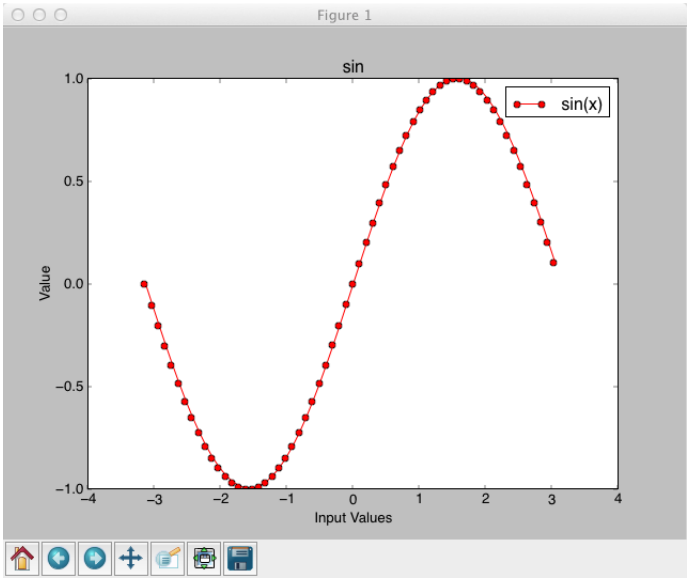

Figura 7.3: The `sin` function

7.3 Exercices

Problema 7.1 MF

Use the code given to generate the data for the two problems that we are going to use: symbolic regression and the sin function.

Problema 7.2 F

The given code of GP uses the following:

- ramped-half-and-half to initialize the population
- tournament selection of the parents
- sub-tree crossover
- point-mutation
- selection of survivors to type generational

Test the code for different values of the parameters, such as the population and tournament sizes and the probabilities of mutation and crossover. What conclusions can you draw?

Problema 7.3 M

Now is time to implement some variants of the provided basic algorithm. You should proceed incrementally making one change at a time. These are the options.

- Survivors' selection of type elitist
- Survivors' selection of type steady state
- Using the reproduction operator
- Using the headless chicken crossover

For the selection of survivors by steady state you should do it in two steps:

1. Randomly select some individuals of the initial population. They enter a tournament and the **worst** is withdraw. Repeat this process a certain number of times.
2. Replace these individuals by the same number of best individuals of the offspring population.

Run the code for the different variants with the parameters that give the best results. What conclusions can you draw?