COMSM0302 - GP COURSEWORK

JAMES MARSHALL
OCTOBER 23, 2008

Submission deadline: Midnight Friday 12th December 2008
Late submission deadline: Midnight Monday 15th December 2008
Weighting: 20% of total course marks

1. Problem Description

You will develop a GP algorithm to perform symbolic regression. The function $f(x)$ is defined for $x > 0$, and is graphed below for $x \in (0, 100]$. Data are provided for you as 5000 fitness cases, where each fitness case is a pair of numbers x and $f(x)$. Your GP algorithm must try to induce the function $f(x)$. It is up to you to choose the function and terminal sets F and T.

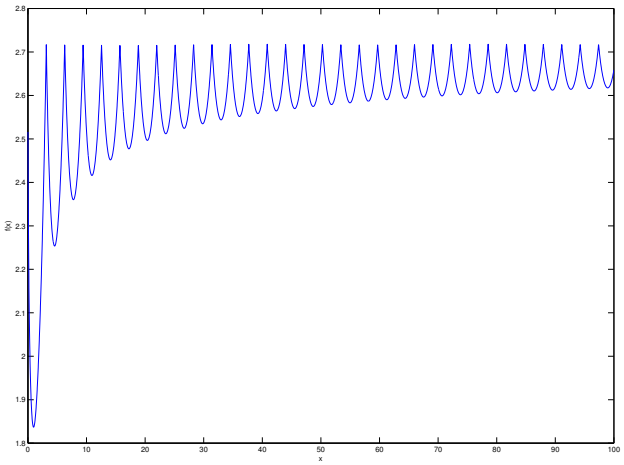

Figure 1. $f(x)$

2. Submission

You should submit a report, your code, and the output of your best-of-run individual for each of the fitness cases, via the online submission system.

The report should be a PDF of no more than 2 A4 sides, with a minimum font size of 10 point. Additional pages over the page limit will not be read, and submissions with a font size smaller than 10 point will not be read at all. The report should include at the top the best-of-run function induced by your algorithm (in standard mathematical notation), the number of individuals evaluated to find it, a graph of your function over $x \in (0, 100]$, and the sum of squares error when evaluating it against the 5000 fitness cases. The remainder of the report should briefly

explain the function and terminal set used, your choices on population initialisation, selection and genetic operators, termination condition, and anything else you consider relevant.

The output of your best-of-run individual should be submitted as a text file of 5000 lines, with each line containing the output of that individual on each of the fitness cases in order, starting from $x = 0.02$ and ending with $x = 100$.

3. Marking

Marks will be allocated as follows

Overall report quality	5%
Choice and justification of selection and genetic operators	20%
Empirical performance	75%