COSC 420/527: Biologically-Inspired Computation
Lab 2: Genetic Algorithms in LEAP

Due: February 23, 11:59 PM

Introduction

In this project, you will use the Library for Evolutionary Algorithms in Python (LEAP) to investigate the effect of different parameters of an evolutionary algorithm on the performance of that algorithm.

You will need to install LEAP before running the provided code. We recommend creating and activating a virtual environment and installing and running LEAP from there. Instructions for installation of LEAP are available at: `https://github.com/AureumChaos/LEAP`

Experimental Setup

The parameters that we will be varying and studying the effect of (and the values that will be evaluated for each) are:

- N: Population size: [25, 50, 75, 100]
- p_m: Mutation probability: [0, 0.01, 0.03, 0.05]
- p_c: Probability of uniform crossover: [0, 0.1, 0.3, 0.5]
- Tournament size: the size of the tournament in tournament selection: [2, 3, 4, 5]

You should evaluate all combinations of each of these. For each combination, you should run for 30 generations. You run should at least 20 different iterations with different random populations. **Note: This will take a considerable amount of compute time. Plan accordingly!**

We have provided the code using LEAP that you will use to capture the appropriate data. **You will update this code to use the following fitness function to maximize:**

$$F(S) = \left(\frac{x}{2^l - 1}\right)^{10} \tag{1}$$

Where S is the genome, x is the integer that results from interpreting S as an unsigned binary number, and l is the length of the genome. You should test your implementation of the fitness function to confirm it gives you the expected output.

Examine this code carefully. The code will create data files that store each of the individuals evaluated, the step they were evaluated in, and their fitness score.

Calculations

For each of these data files and each generation, you will write code to parse the file created by the LEAP code to calculate the following:

- Average fitness for that generation
- Best fitness for that generation
- Best genome for that generation
- Whether the solution (fitness = 1) was found in that generation
- How many solutions (fitness = 1) were found in that generation
- **For CS 527**: Measure of diversity of the population

You should create a CSV file that includes each of the above as a column that is calculated, as well as the parameter values that were used, the generation number, and the iteration number. You will submit this CSV file as part of your full submission.

Diversity Metric: To calculate the diversity metric for the a given population, you will first calculate the distance between every individual and every other individual in the population (treating them as a vector). You will calculate the average of these distances for each individual. Then the diversity of the population will be the average of these values across the whole population. The smaller this value is, the less diverse the population is.

Report Write-Up

Your report write-up should include the following information:

Graphs

Assume that the default parameters are $N = 50$, $p_m = 0.01$, $p_c = 0.3$, and tournament size at 2.

For each of the parameters (N, p_m, p_c, tournament size), create a plot where the x-axis is the generation and the y-axis is the fitness value. Show the different iterations as individual lines or as a mean line with a standard deviation shading and different colors for each of the different values (where the other parameter values are fixed at their defaults). Plot both the maximum fitness and average fitness. (You can do two separate plots for these if they get too crowded).

Add additional plots as necessary to visualize any of the results you would like to discuss. Additional plot types you might consider are heatmaps and boxplots. You should have plots

to back up any conclusions you draw from the data. Other types of plots you might include are:

- How the resulting fitness values change with different parameter combinations
- How the best performing individual in the population changes over time
- How quickly the GA converges with respect to the parameter values
- How homogeneous the population becomes when the best fitness is found

You may want to plot how the resulting fitness values change with different parameter combinations, or you might want to plot how the best performing individual in the population

For CS 527: Include plots to show how the diversity is affected by each of the different parameter values. Once again, look at each parameter in isolation and only look at the results where the other parameters are fixed at the default values.

Discussion

The questions you should address in your report are:

- Which parameters have the biggest impact on performance in terms of fitness achieved?
- How does selection pressure influence performance? Note that higher values of tournament size correspond to higher selection pressure.
- Are both crossover and mutation necessary?
- Which parameters interact with each other the most?
- How does the population size affect performance?
- **For CS 527:** How do the different parameter values affect the diversity of the population?

Submission

You will submit your report PDF, as well as the CSV file described above.

Grading

Undergraduate Grading Breakdown:

- **20 points** for the CSV file you will create.
- **40 points total**: 10 points per plot/discussion of each plot for each of the four parameters showing how the parameter value affects the training trajectory (if you do separate plots for maximum and average fitness, you will have 8 total plots).
- **40 points:** 8 points for each of the discussion questions listed above. You should include at least a paragraph (of at least 3-4 sentences) for each discussion point.

Total: 100 points

For undergraduates, if you complete the diversity of population analysis required for the graduate course, that will be an additional 15 points.

Graduate Grading Breakdown:

- **20 points** for the CSV file you will create.
- **40 points total**: 10 points per plot/discussion of each plot for each of the four parameters showing how the parameter value affects the training trajectory (if you do separate plots for maximum and average fitness, you will have 8 total plots).
- **25 points** 5 points for each of the discussion questions listed above. You should include at least a paragraph (of at least 3-4 sentences) for each discussion point.
- **15 points** Discussion of the impact of the different parameters with diversity, including plots to show impact.

Total: 100 points

Late Penalty

There will be a penalty of 10 points off per day late.