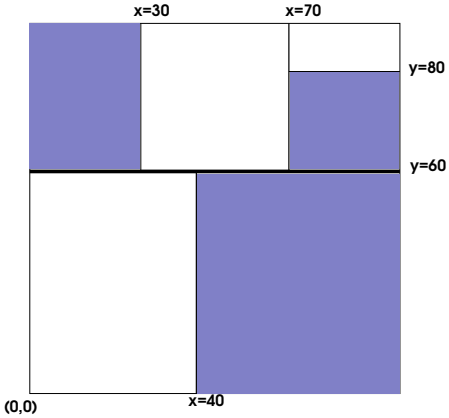
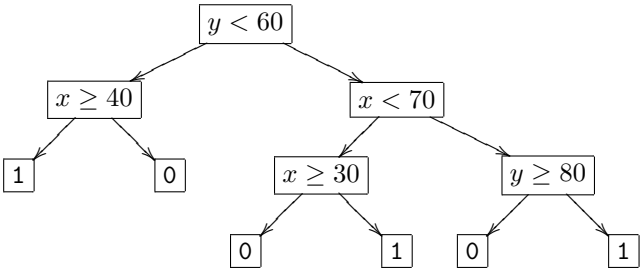
CEng 713 Evolutionary Computation, Homework 2

Due: 16th December 2007

In this homework, you will have a genetic programming classifier in a very simple setup. At each level the classified space is divided into 2 in one of the dimensions only. Sublevels continue dividing. Following are the example of such classifier in 2D and the corresspond-ing shape:





A terminal node contains the identifier of the class. The example is a 2 class problem, 0 and 1. Nonterminal nodes contain the following information:

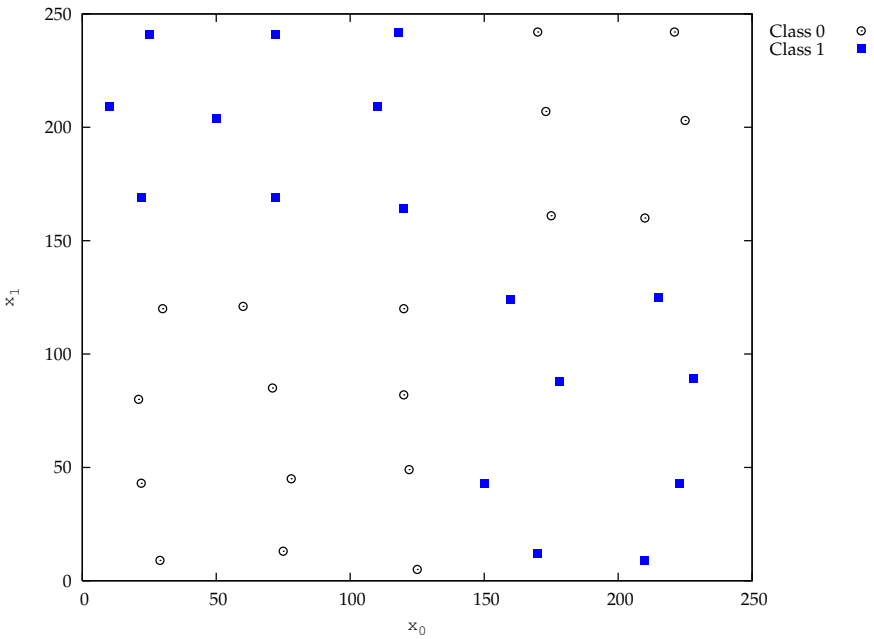
- The dimension of division (x, y or $x_1, x_2, x_3, ...$)
- The comparison operator ($<$ or $\geq$)
- The position of division (any integer in 0 to 255)

You are given a set of of sample points and their classes as input and your task is to find a classifier tree classifying as much points as possible correctly.
Your input format is:

dim	(number of dimensions in your search space)
$x_{1\,1}\ x_{1\,2}\ x_{1\,3}\ ...\ x_{1\,dim}\ \ class_1$	(coordinates of point 1 and the class of it)
$x_{2\,1}\ x_{2\,2}\ x_{2\,3}\ ...\ x_{2\,dim}\ \ class_2$	(coordinates of point 1 and the class of
....	

A sample input and its plot:

```
2
10 209 1
21 80 0
22 169 1
22 43 0
25 241 1
29 9 0
30 120 0
50 204 1
60 121 0
71 85 0
72 169 1
72 241 1
75 13 0
78 45 0
110 209 1
118 242 1
120 120 0
120 164 1
120 82 0
122 49 0
125 5 0
150 43 1
160 124 1
170 12 1
170 242 0
173 207 0
175 161 0
178 88 1
210 160 0
210 9 1
215 125 1
221 242 0
223 43 1
225 203 0
228 89 1
```



You are first asked to implement a simple genetic programming application (default of the library you used) solving this problem for two dimensions . Then suggest a simple improvement like a special mutator, a local search on one of the classifier boundary or leaf nodes, special crossover, a representation enhancement (add slope to your classifier line at each non-terminal), greedy initializer, etc.. You can use any evolutionary parameter, operator etc. as long as you explicitly state them. You can use galib tree encoding or other system like Open Beagle (`http://beagle.gel.ulaval.ca/index.html`), Beagle Puppy (a limited version of Open Beagle).

You will be given 3 test data. One is a simple set of two class points that can be classified with 2 lines (like in the example). The other two will be harder (with multiple classes and non-linear boundaries)

Please repeat your two experiments for at least 5 times for each of 3 test data. Report your results with the following format:

Topic: Simple Classifier with GP
Name surname: **Id:**

GP Parameters:

Population size:
Crossover/mutation rates:
Type of crossover/mutation:
Selection mechanism:
% of Elitism:
Maximum depth allowed:
Fitness function:

Enhanced version:

Description of the enhancement
.....

Results

....

Figure 1: Generation vs best and average fitness (simple and enhanced versions, test data 1)

Figure 2: Generation vs tree depth of the best individual (simple and enhanced versions, test data 1)

....