

GAGP Tutorial 2

1. (Goldberg) You are asked to minimize a function $f(x,y,z)$ where
 $-20 < x < 125$, $0 < y < 1200000$, $-0.1 < z < 1.0$
and the desired precisions for x , y and z are 0.5, 10000 and 0.001
respectively. Using the 'customary' grid-based binary encoding idea of dividing
ranges into some power-of-2 number of points, how many bits are needed?
2. You want to try to represent the value of some integer quantity n in some
binary-encoded way, but n only ranges over the integers 1..17. How might you
do it? Discuss the advantages and disadvantages.
3. Proteins are made up of sequences of amino acids connected by chemical
bonds. The protein sequence "folds up" into a three-dimensional structure of
low energy by rotation of the chemical bonds connecting the amino acid
groups. The three-dimensional structure will then determine the biological
function of the protein, so it is important to be able to predict this structure from
the sequence of amino acids in the protein. Consider how you could apply a
genetic algorithm to find the three-dimensional structure of lowest energy for a
given protein sequence. Pay particular attention to how you would represent
the candidate structures, the fitness function you would use, and the types of
crossover and mutation.
4. (Mitchell) When is the union of two schemas also a schema? When is the
intersection of two schemas also a schema? E.g. the union of 1^* and 0^* is ** ;
the union of 10 and 01 is not a schema.
5. (Computer exercise from Mitchell) Implement a simple GA with fitness-
proportional selection, roulette-wheel sampling, population size 100, single-
point crossover rate $p_c=0.7$, and bitwise mutation rate $p_m=0.001$. Try it on the
following fitness function: $f(x)=$ number of ones in x , where x is a binary
chromosome of length 20.
Perform 20 runs, and measure the average generation at which the string of
all ones is discovered. Perform the same experiment with crossover turned off
(i.e. $p_c=0.0$). Do similar experiments, varying the mutation and crossover rates,
to see how the variations affect the average time required for the GA to find
the optimal string. If it turns out that mutation with crossover is better than
mutation alone, why is this the case?