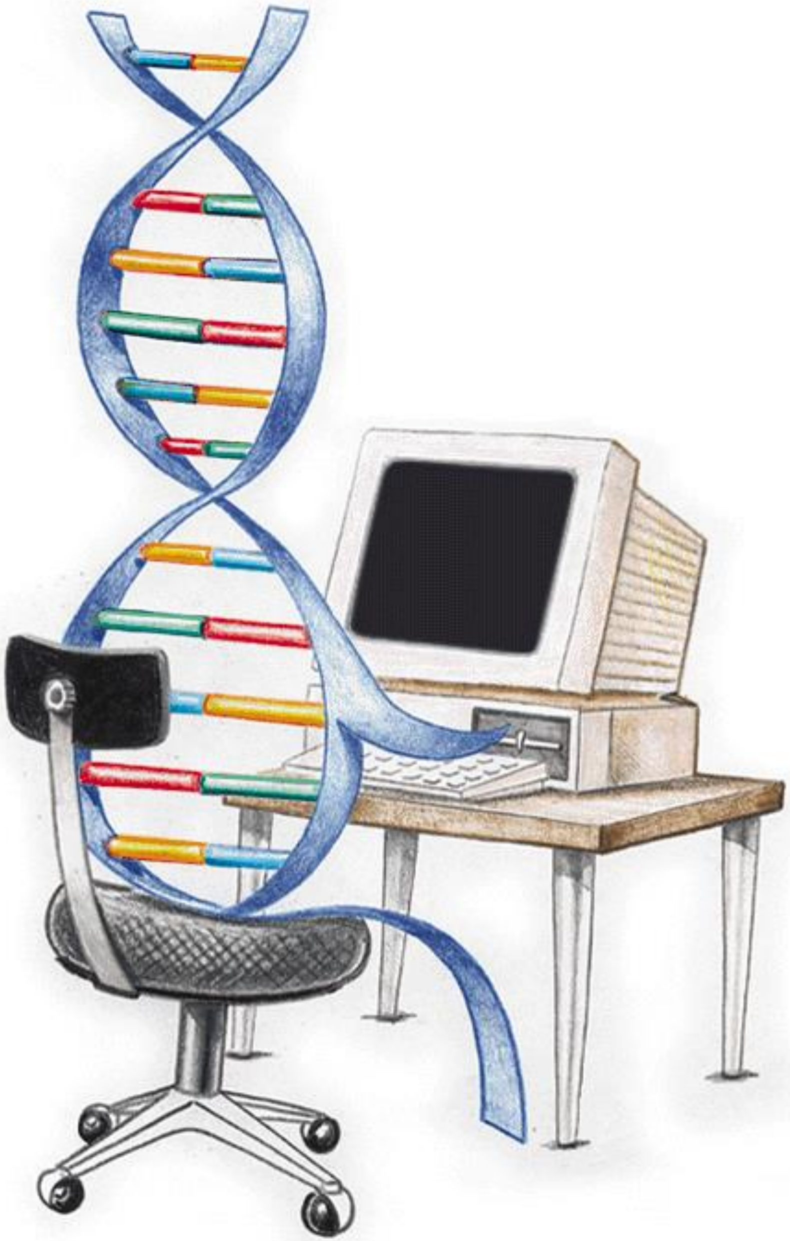
Nature inspired computing



Prof Dr Marko Robnik Šikonja
October 2018

Evolutionary and natural computation

- Many engineering and computational ideas from nature work fantastically!
- Evolution as an algorithm
- Abstraction of the idea:
  - progress, adaptation - learning, optimization
- Survival of the fittest - competition of agents, programs, solutions
- Populations – parallelization
- (Over)specialization – local extremes

Template of evolutionary program

generate a population of agents (objects, data structures)

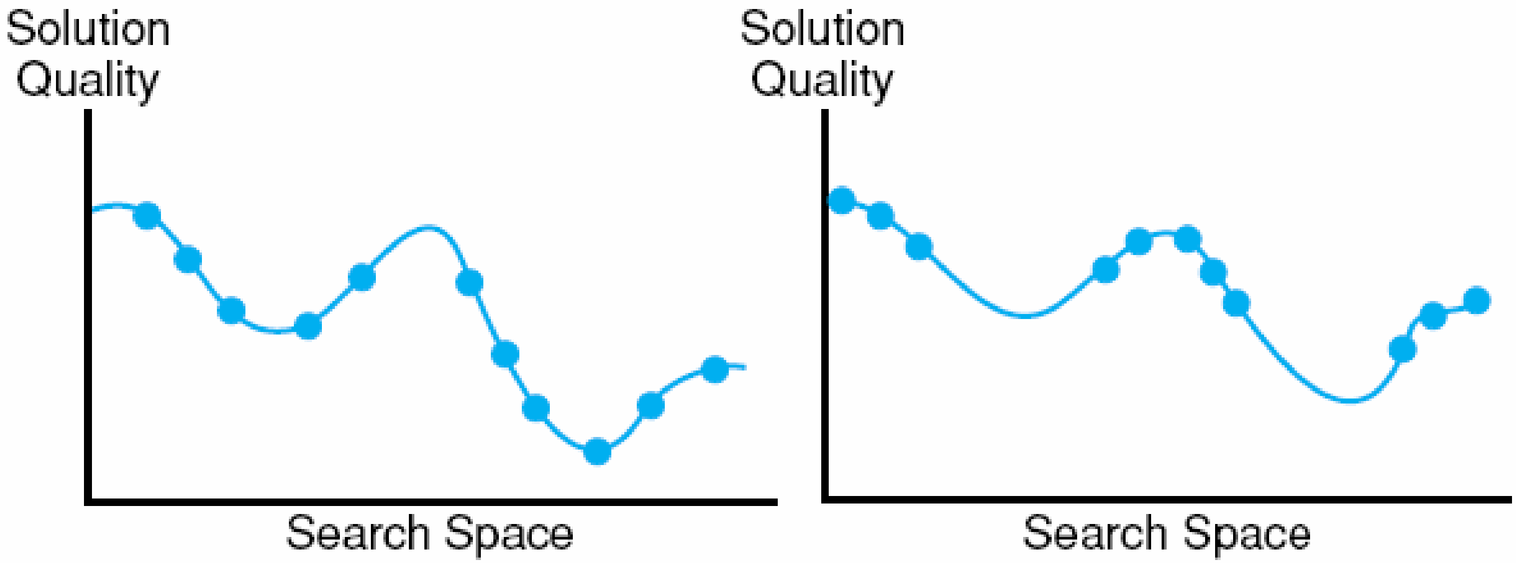
do {

compute fitness (quality) of the agents
select candidates for the reproduction using fitness
create new agents by combining the candidates
replace old agents with new ones

} while (not satisfied)

- immensely general -> many variants

A result of successful evolutionary program



a. The beginning search space

b. The search space after n generations

Strengths and weaknesses

- robust, adaptable, general
- requires only weak knowledge of the problem (fitness function and representation of genes)
- several alternative solutions
- hybridization and parallelization

- suboptimal solutions
- possibly many parameters
- computationally expensive

- no-free-lunch theorem

Main approaches

- Genetic algorithms
- Genetic programming
- Swarm methods (particles, ants, bees, …)
- Self organized fields
- Differential evolution
- ….

Genetic Algorithms - History

- Pioneered by John Holland in the 1970's
- Got popular in the late 1980's
- Based on ideas from Darwinian Evolution
- Can be used to solve a variety of problems that are not easy to solve using other techniques

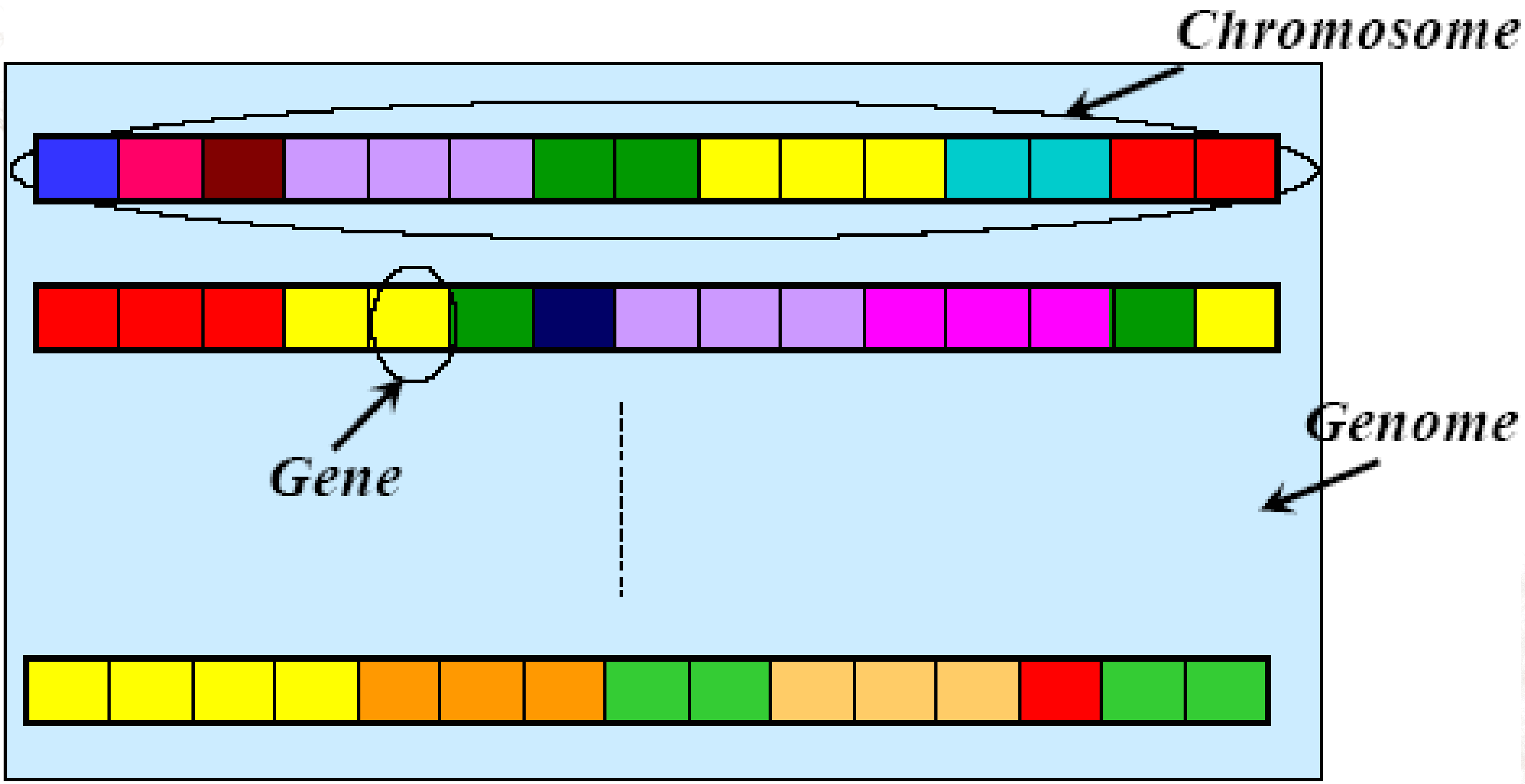
Evolution in the real world

- Each cell of a living thing contains ***chromosomes*** - strings of *DNA*
- Each chromosome contains a set of ***genes*** - blocks of DNA
- Each gene determines some aspect of the organism (like eye colour)
- A collection of genes is sometimes called a ***genotype***
- A collection of aspects (like eye colour) is sometimes called a ***phenotype***
- Reproduction involves recombination of genes from parents and then small amounts of ***mutation*** (errors) in copying
- The ***fitness*** of an organism is how much it can reproduce before it dies
- Evolution based on "survival of the fittest"

Key terms

- **Individual** - Any possible solution
- **Population** - Group of all *individuals*
- **Search Space** - All possible solutions to the problem
- **Chromosome** - Blueprint for an *individual*
- **Trait** - Possible aspect (*features)* of an *individual*
- **Allele** - Possible settings of trait (black, blond, etc.)
- **Locus** - The position of a *gene* on the *chromosome*
- **Genome** - Collection of all *chromosomes* for an *individual*

Chromosome, Genes and Genomes



Biological equivalents

- Evolution is a variation of alleles frequencies through time.
- Reproduction, variation (mutation, crossover), selection

Evolutionary computation keywords

- Representation: data structures, operations
- Fitness, heuristics
- Population variability
- Local and global extremes
- Coevolution
- Variability of fitness function

Gen representation

- Bit vector
- Numeric vectors
- Strings
- Permutations
- Trees: functions, expressions, programs
- ...

Crossover

- Single point/multipoint
- Shall preserve individual objects

Crossover: bit representation

Parents: 1101011100 0111000101

Children: 1101010101 0111001100

A demo: smart rockets

- a evolution of navigational skills
- each spaceship has 5 motors (rockets)
- motors are placed anywhere on the spaceship, at any angle, variable in strength
- the direction of the movement depends:
  - on the placement of the rockets
  - on the firing pattern
- initially the rockets are randomly placed, firing pattern is randomly assigned
- data structure: firing signature, motor angle, motor strength – bit arrays
- fitness function: the minimal distance to the target, minimal fuel

Crossover: vector representation

Simplest form

Parents: (6.13, 4.89, 17.6, 8.2) (5.3, 22.9, 28.0, 3.9)

Children: (6.13 , 22.9, 28.0, 3.9) (5.3, 4.89, 17.6, 8.2)

In reality: linear combination of parents

Linear crossover

- The linear crossover simply takes a linear combination of the two individuals.
- Let $x = (x_1, \ldots x_N)$ and $y = (y_1, \ldots y_N)$
- Select α in $(0, 1)$
- The results of the crossover is $\alpha x + (1 - \alpha) y$.
- Possible variation: choose a different α for each position.

Linear crossover example

- Let $\alpha = 0.75$ and we have this two individuals:

  $A = (5, 1, 2, 10)$ and $B = (2, 8, 4, 5)$

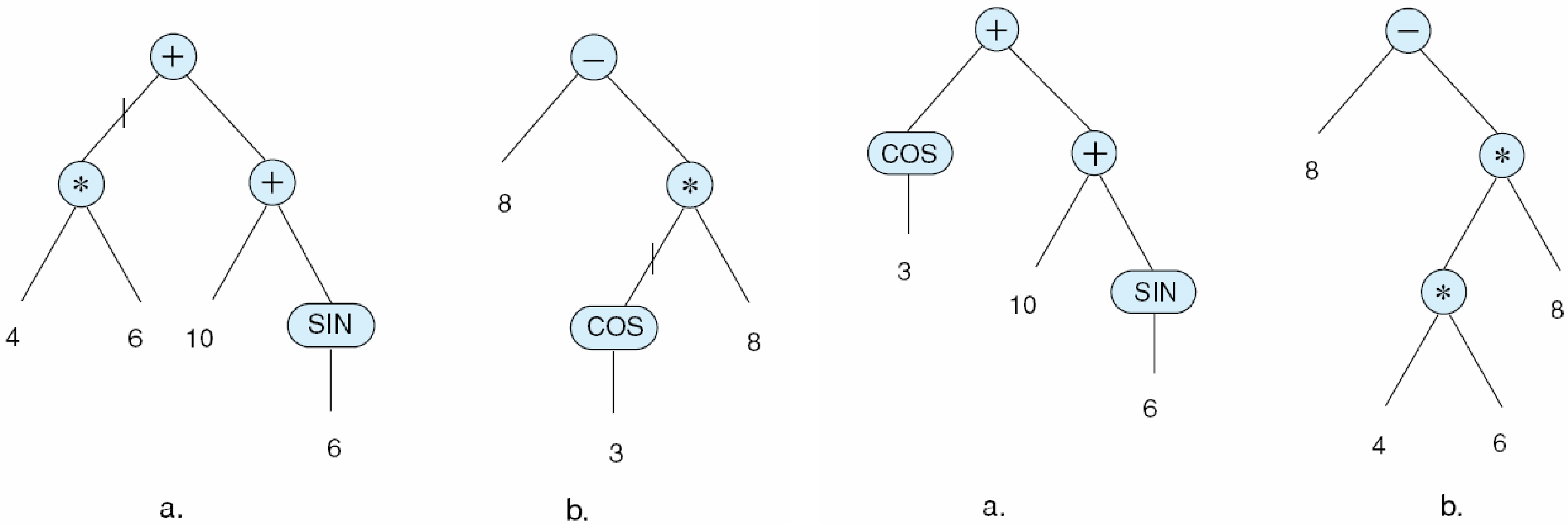
- then the result of the crossover is:

$(3.75 + 0.5, 0.75 + 2, 1.5 + 1, 7.5 + 1.25) = (4.25, 2.75, 2.5, 8.75)$

- If we use the variation and we have $\alpha = (0.5, 0.25, 0.75, 0.5)$, the result is:

$(2.5 + 1, 0.25 + 6, 1.5 + 1, 5 + 2.5) = (3.5, 6.25, 2.5, 7.5)$

Crossover: trees



Permutations: travelling salesman problem

- 9 cities: 1,2 ..9
- bit representation using 4 bits?
  - 0001 0010 0011 0100 0101 0110 0111 1000 1001
  - crossover would give invalid genes
- permutation and ordered crossover
  - keep (part of) sequences
  - use the sequence from second cut, keep already existing

1 9 2 | 4 6 5 7 | 8 3 → x x x | 4 6 5 7 | x x ↘ <u>2 3 9</u> | 4 6 5 7 | <u>1 8</u>

4 5 9 | 1 8 7 6 | 2 3 → x x x | 1 8 7 6 | x x ↗ <u>3 9 2</u> | 1 8 7 6 | <u>4 5</u>

Gray coding of binary numbers

- Keeping similarity

Binary	Gray
0000	0000
0001	0001
0010	0011
0011	0010
0100	0110
0101	0111
0110	0101
0111	0100
1000	1100
1001	1101
1010	1111
1011	1110
1100	1010
1101	1011
1110	1001
1111	1000

Adaptive crossover

- Different evolution phases
- Crossover templates
- 0 – first parent, 1 second parent
- Different dynamics of template crossover

	Gene	Template
Parent 1	1.2 3.4 5.6 4.5 7.9 6.8	010101
Parent 2	4.7 2.3 1.6 3.2 6.4 7.7	011100
Child 1	1.2 2.3 5.6 3.2 7.9 7.7	010100
Child 2	4.7 3.4 1.6 4.5 6.4 6.8	011101

Mutation

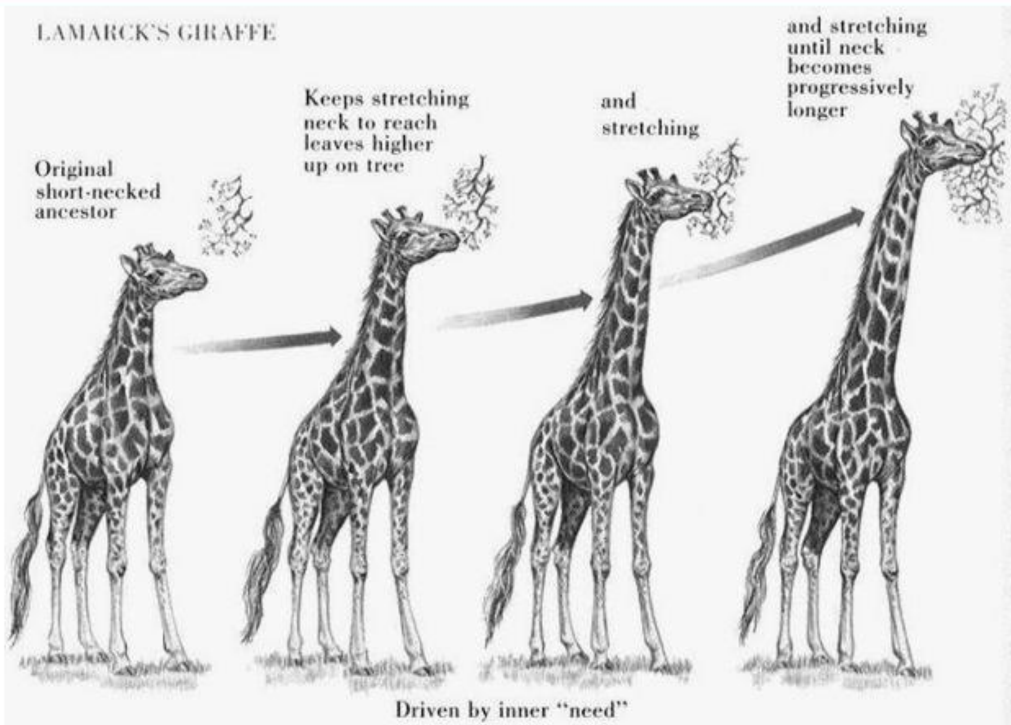
- Adding new information
- Random search?
- Binary representation:

  0111001100 --> 0011001100
- Single point/multipoint
- Lamarckian (searching for locally best mutation)

Lamarckianism

Lamarckism is the hypothesis that an organism can pass on characteristics that it has acquired through use or disuse during its lifetime to its offspring.

An Early Proposal of Evolution: Theory of Acquired Characteristics



Jean Baptiste Lamarck (~ 1800) : Theory of Acquired Characteristics

- Use and disuse alter shape and form in an animal
- Changes wrought by use and disuse are heritable
- Explained how a horse-like animal evolved into a giraffe

Gaussian mutation

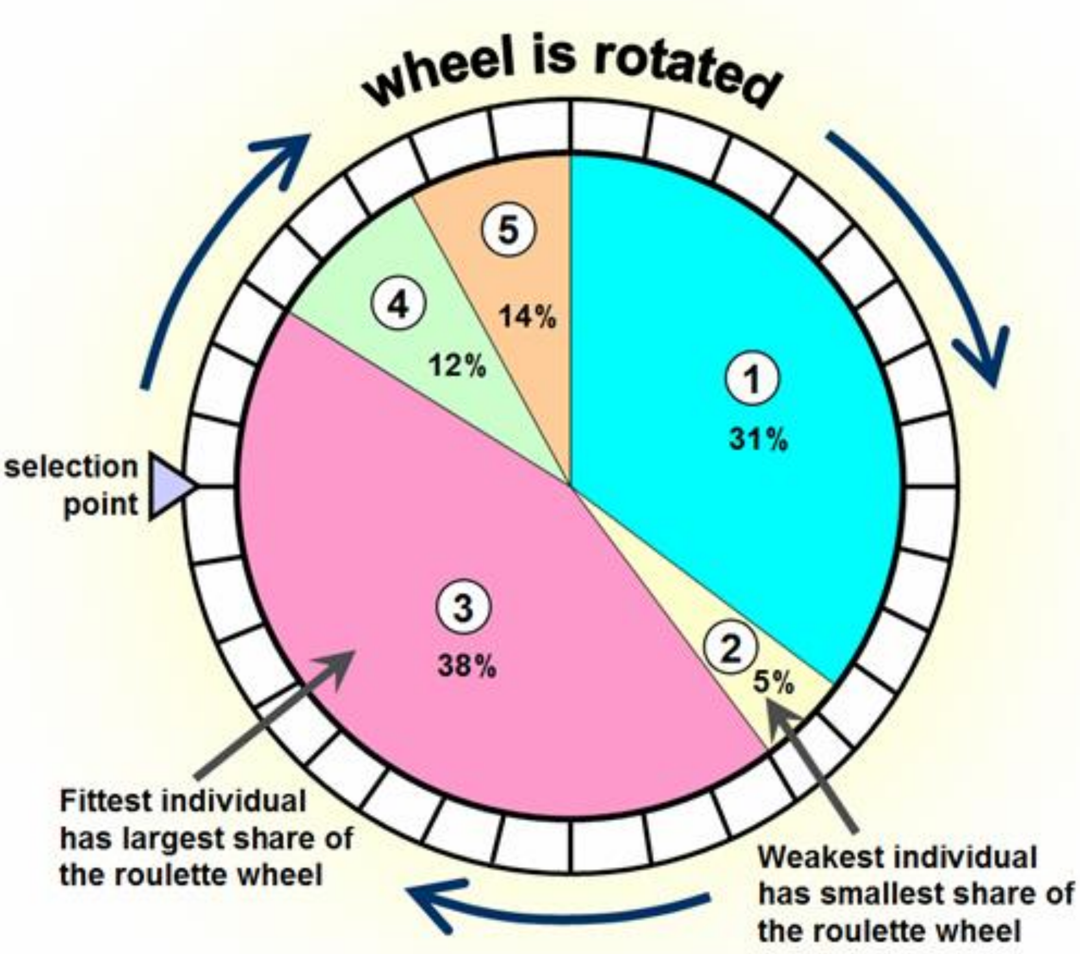
- When mutating one gene, selecting the new value by choosing uniformly among all the possible values is not the best choice (empirically).
- The mutation selects a position i in the vector of floats and mutates it by adding a Gaussian error: a value extracted according to a normal distribution with mean 0 and variance depending on the problem.

Evolutional model

- Keeping the good
- Prevent premature convergence
- heterogeneity of population

Selection

- Proportional
- Rank proportional
- Tournament
- Single tournament
- Stochastic universal sampling

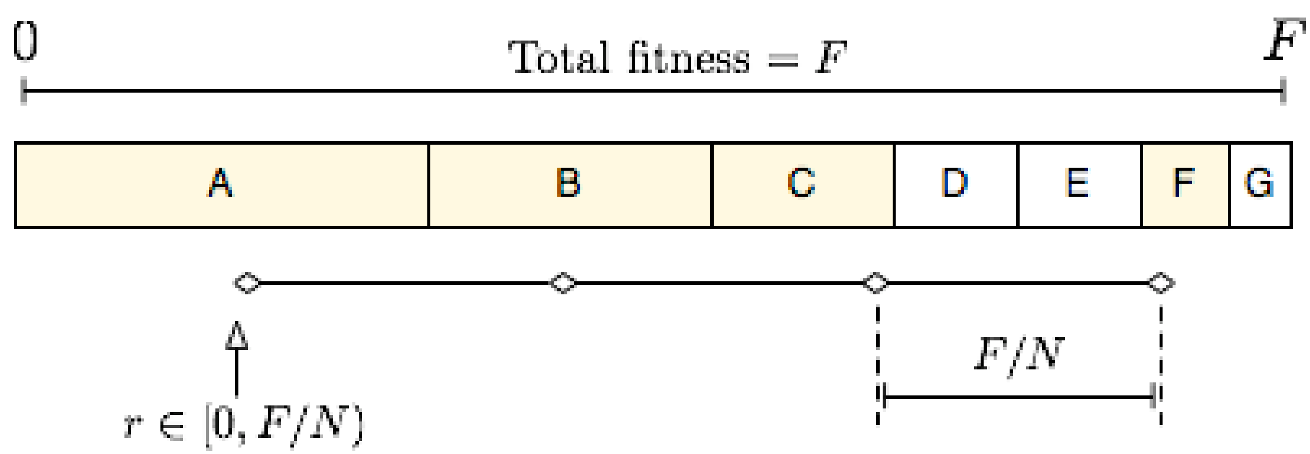


Tournament selection

1. set t=size of the tournament,
   p=probability of a choice
2. randomly sample t agents from population forming a tournament
3. select the best with probability p
4. select second best with probability $p(1-p)$
5. select third best with probability $p(1-p)^2$
6. …

Stochastic universal sampling (SUS)

- unbiased
- r ∈ [o, F/N]
- r + i*N, i ∈ o, 1,…, N-1]



Replacement

- All
- According to fitness (roulette, rang, tournament, randomly)
- Elitism (keep a portion of the best)
- Local elitism (children replace parents if they are better)

Single tournament selection

1. randomly split the population into small groups
2. apply crossover to two best agents from each group; their offspring replace two worst agents from the group

- advantage: in groups of size *g* the best g-2 progress to next generation (we do not use good agents, maximal quality does not decrease)
- no matter the quality even the best agents have no more than two offspring (we do not loose population diversity)

Population size

- small, large?

Niche specialization

- evolutionary niches are generally undesired
- punish too similar agents

$$f'_i = f_i / q(r,i)$$

$$q(r,i) = \begin{cases} 1 & ; sim(i) <= 4, \\ sim(i)/4 & ; otherwise \end{cases}$$

Stopping criteria

- number of generations, track progress, availability of computational resources, etc.

Why genetic algorithms work?

- building blocks hypothesis
- … is controversial (mutations)
- sampling based hypothesis

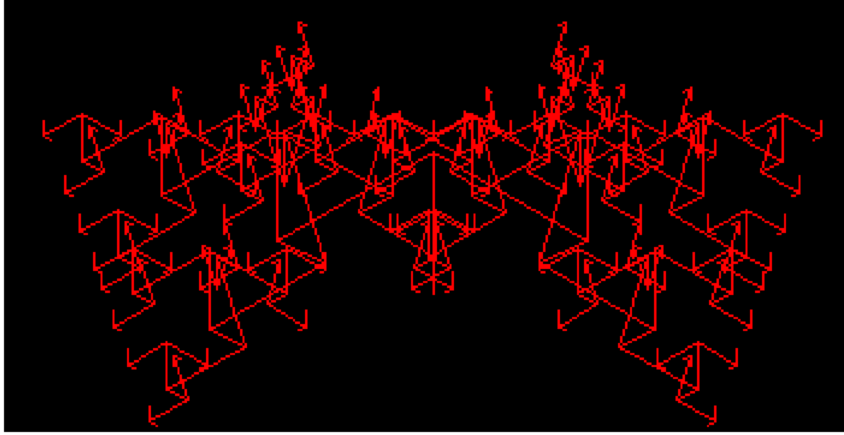
Parameters of GA

- encoding into fixed length strings
- Length of the strings;
- Size of the population;
- Selection method;
- Probability of performing crossover (p_c);
- Probability of performing mutation (p_m);
- Termination criteria (usually a number of generations and/or a target fitness).

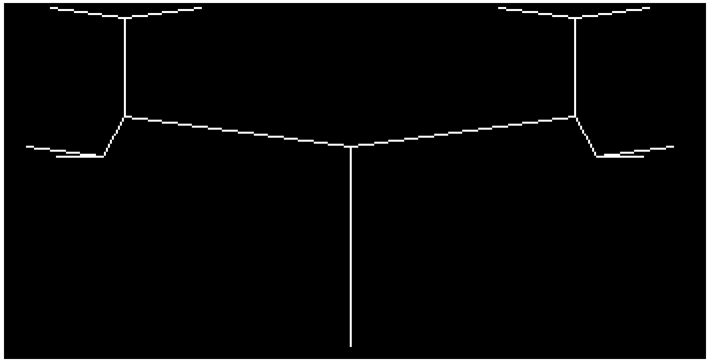
Usual settings of GA parameters

- Population size: from 20–50 to a few thousands individuals;
- Crossover probability: high (around 0.9);
- Mutation probability: low (below 0.1).

Demo: find genome of a biomorph



- A biomorph is a graphic configuration generated from nine genes.
- The first eight genes each encode a length and a direction.
- The ninth gene encodes the depth of branching.
- Each gene is encoded with five bits.
  - The four first bits represent the value, the fifth its sign.
  - Each gene can get a value from -15 to +15.
  - value of gen nine is limited to 2-9.



- There are : 8 (number of possible depths) x 2^{40} (the 8 * 5 =40 bits encoding basic genes) = 8.8 x10^{12} possible biomorphs. If we were able to test 1000 genomes every second, we would need about 280 years to complete the whole search.
- At the beginning, the drawing algorithm being known, we get the image of a biomorph. The only informations directly measurable are the positions of branching points and their number. The basic algorithm simulates the collecting of these informations.
- fitness function: the distance of the generated biomorph from the target one.

Applications

- optimization
- scheduling
- bioinformatics, machine learning
- planning
- multicriteria optimization

Where to use evolutionary algorithms?

- Many local extremes
- Just fitness, without derivations
- No specialized methods
- Multiobjective optimization
- Robustness
- Combined approaches

Multiobjective optimization

- Fitness function with several objectives
- cost, energy, environmental impact, social acceptability, human friendliness
- min $F(x)=\min (f_1(x), f_2(x), \ldots, f_n(x))$
- Pareto optimal solution: we cannot improve one criteria without getting worse on others
- GA: in reproduction use all criteria

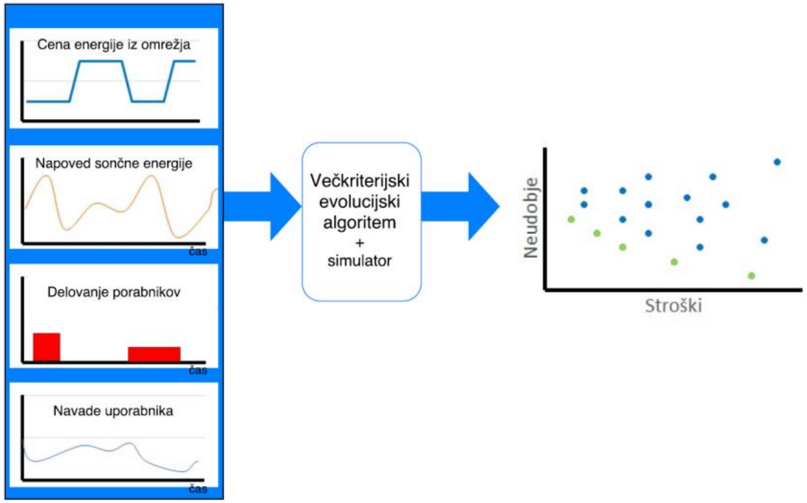
An example: smart buildings

- simple scenario: heater, accumulator, solar panels, electricity from grid
- criteria: price, comfort of users (as the difference in temperature to the desired one)
- chromosome: shall encode schedule of charging and discharging the battery, heating on/off
- operational time is discretized to 15min intervals

Control problem for smart buildings

Parameters:

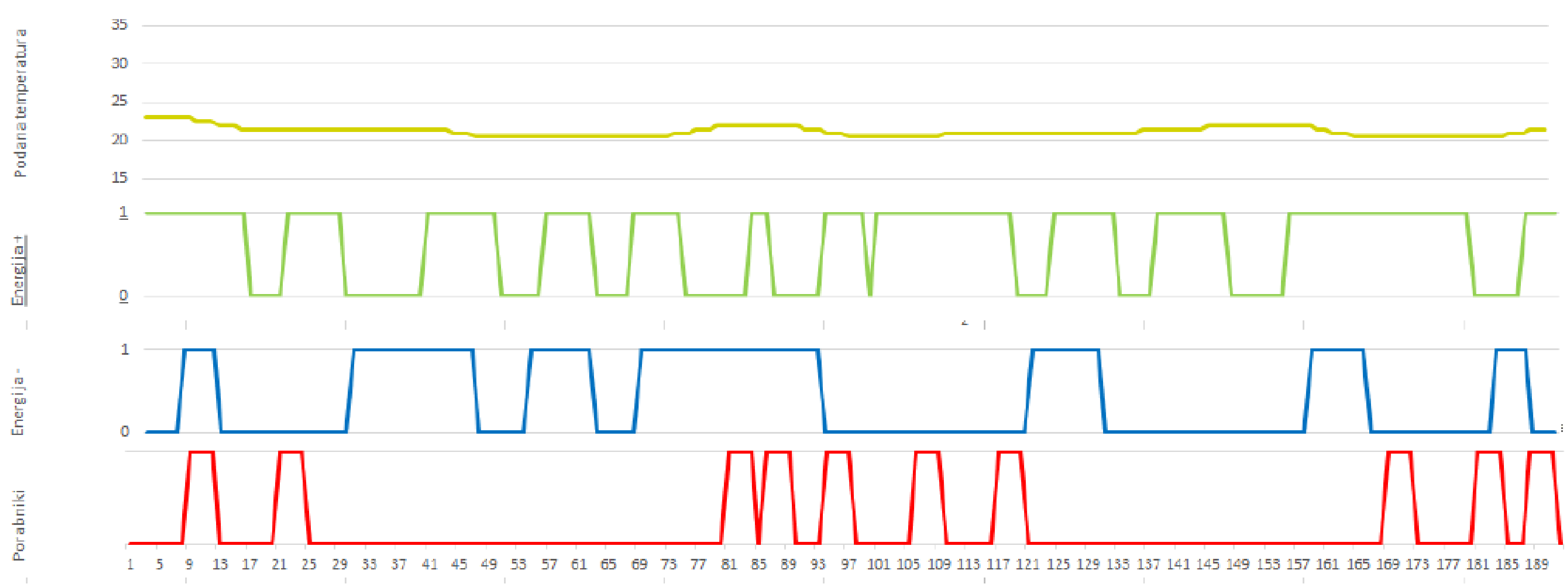
- the price of energy from the grid varies during the day
- the prediction of solar activity
- schedule of heater and battey
- usual activities of a user



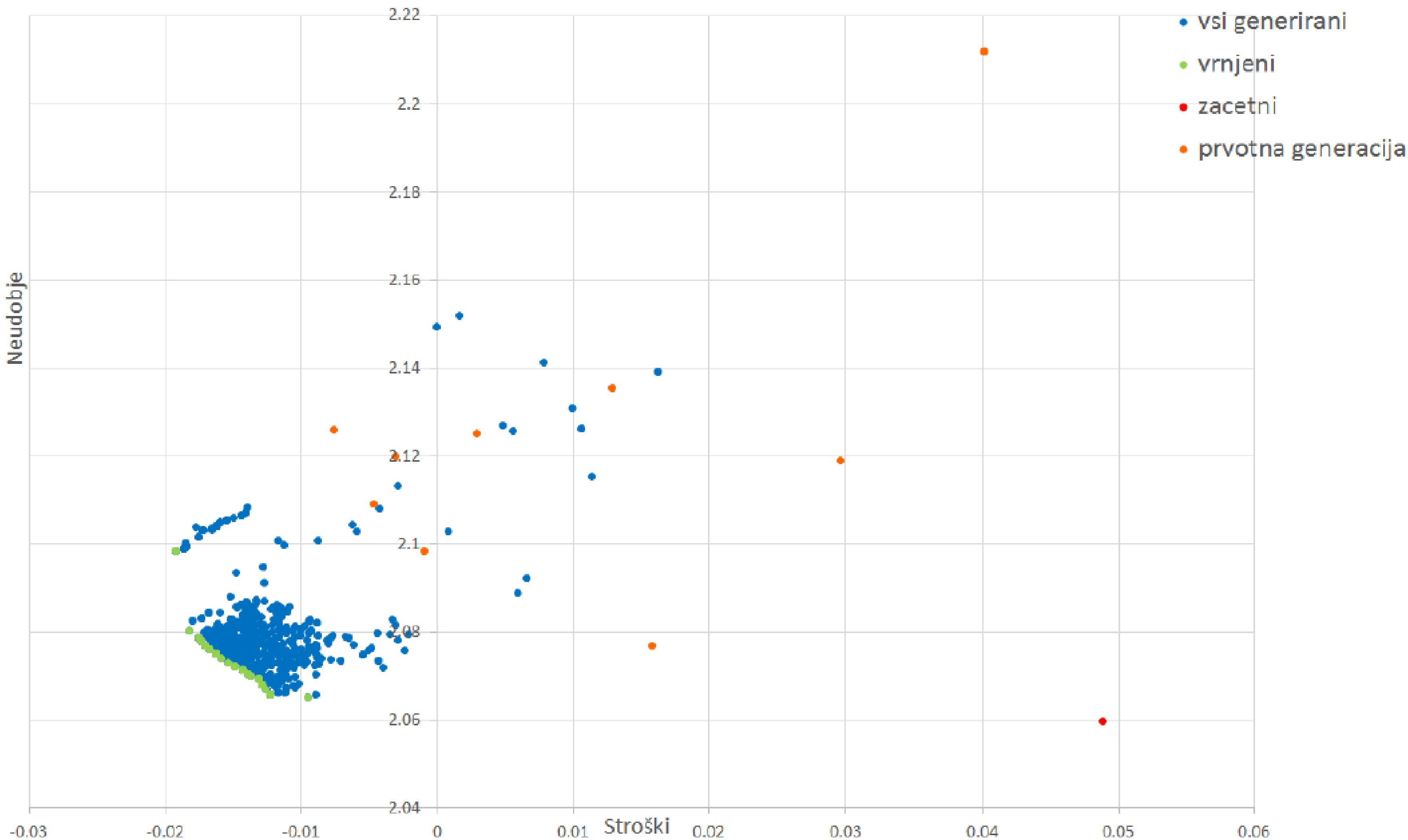
Smart building: structure of the chromosome

- temperature: for each interval we set the desired temperature between Tmin and Tmax interval
- battery+: if photovoltaic panels produce enough energy we set: 1 charging, 0 no charging
- battery-: if photovoltaic panels do not produce enough energy, we set: 1 battery shall discharge, 0 battery is not used
- appliances: each has its schedule when it is used (1) and when it is off (0)

Example of schedule



Example of solutions and optimal front



Toolboxes and libraries

- CIlib – computational intelligence library
- EO (C++) - evolutionary computation library
- ECJ (Java)
- EvA2 (Java),
- JAGA (Java)
- ECF- Evolutionary Computation Framework (C++)
- Matlab, ...
- R: Rfreak, ppso, numDeriv,...

Pros and Cons

- Pros
  - Faster (and lower memory requirements) than searching a very large search space.
  - Easy, in that if your candidate representation and fitness function are correct, a solution can be found without any explicit analytical work.
- Cons
  - Randomized – not optimal or even complete.
  - Can get stuck on local maxima, though crossover can help mitigate this.
  - It can be hard to work out how best to represent a candidate as a bit string (or otherwise).

Genetic programming

- Functions, programs, expression trees
- Keep the structures valid
- Tree crossover, type closure
- applications

GP quick overview

- Developed: USA in the 1990's
- Early names: J. Koza
- Typically applied to:
  - machine learning tasks (prediction, classification…)
- Attributed features:
  - competes with neural nets and alike
  - needs huge populations (thousands)
  - slow
- Special:
  - non-linear chromosomes: trees, graphs
  - mutation possible but not necessary (disputed!)

GP technical summary tableau

Representation	Tree structures
Recombination	Exchange of subtrees
Mutation	Random change in trees
Parent selection	Fitness proportional
Survivor selection	Generational replacement

Introductory example: credit scoring

- Bank wants to distinguish good from bad loan applicants
- Model needed that matches historical data

ID	No of children	Salary	Marital status	OK?
ID-1	2	45000	Married	0
ID-2	0	30000	Single	1
ID-3	1	40000	Divorced	1
…				

Introductory example: credit scoring

- A possible model:

  IF (NOC = 2) AND (S > 80000) THEN good ELSE bad

- In general:

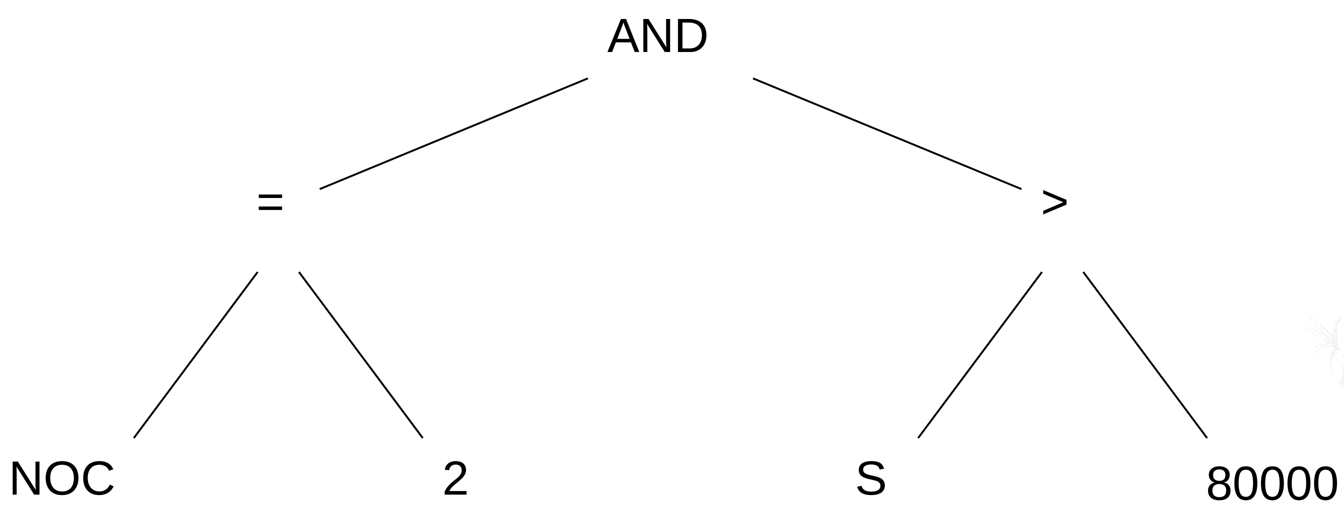
  IF formula THEN good ELSE bad

- Only unknown is the right formula, hence
- Our search space (phenotypes) is the set of formulas
- Natural fitness of a formula: percentage of well classified cases of the model it stands for
- Natural representation of formulas (genotypes) is: parse trees

Introductory example: credit scoring

IF (NOC = 2) AND (S > 80000) THEN good ELSE bad

can be represented by the following tree



Tree based representation

- Trees are a universal form, e.g. consider
- Arithmetic formula

$$2 \cdot \pi + \left((x+3) - \frac{y}{5+1} \right)$$

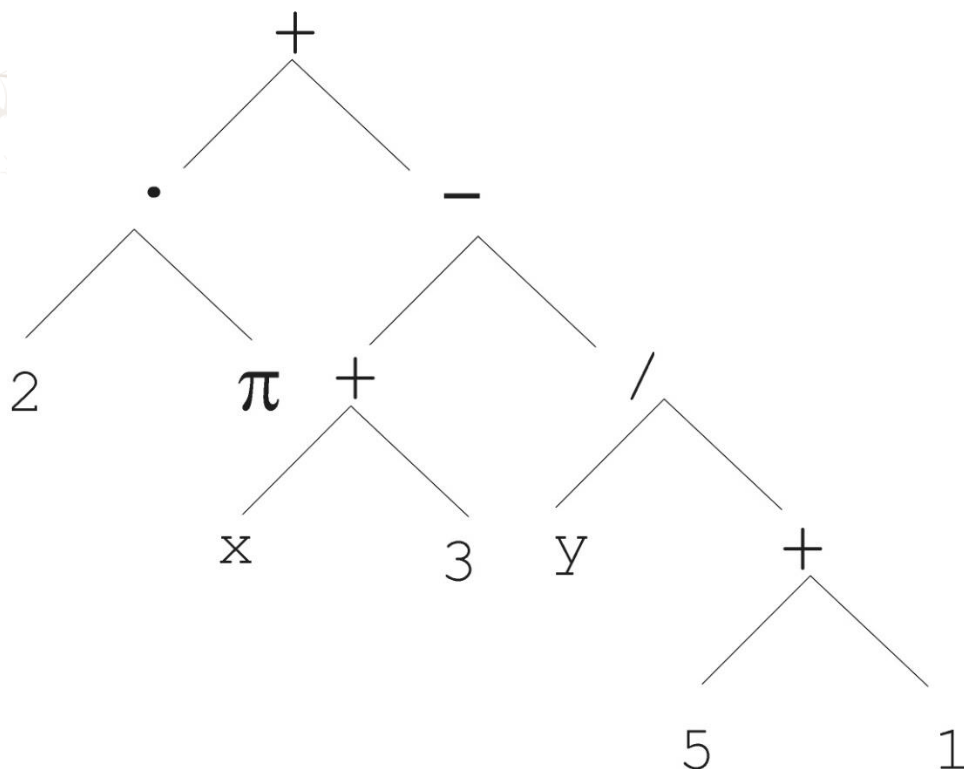
- Logical formula

$$(x \wedge \text{true}) \rightarrow ((x \vee y) \vee (z \leftrightarrow (x \wedge y)))$$

- Program

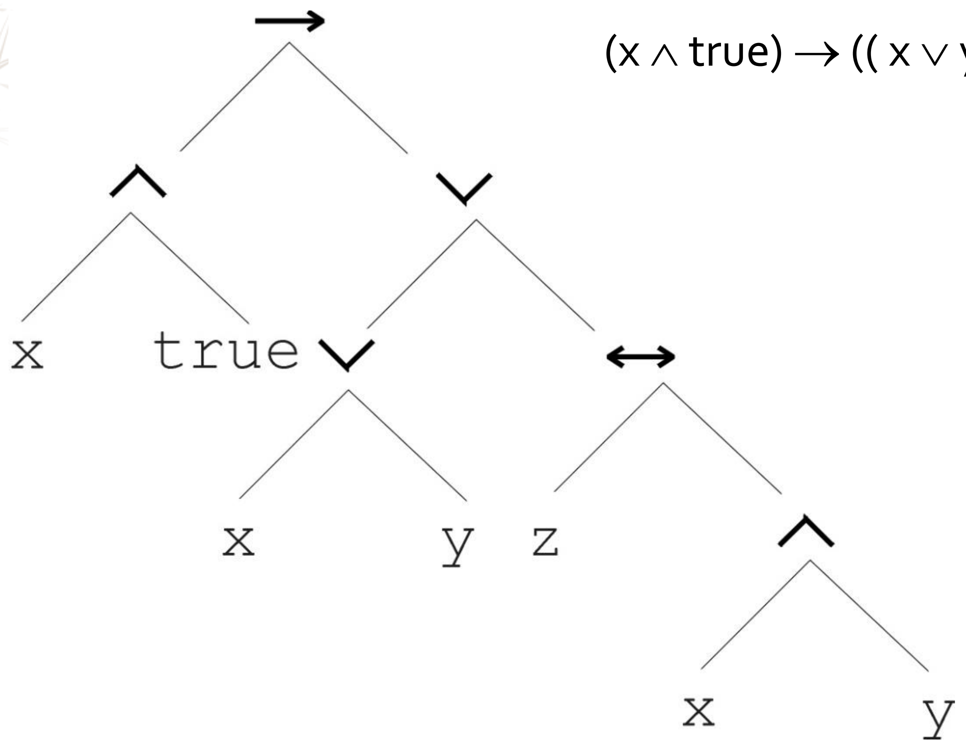
```
i =1;
while (i < 20)
{
        i = i +1
}
```

Tree based representation



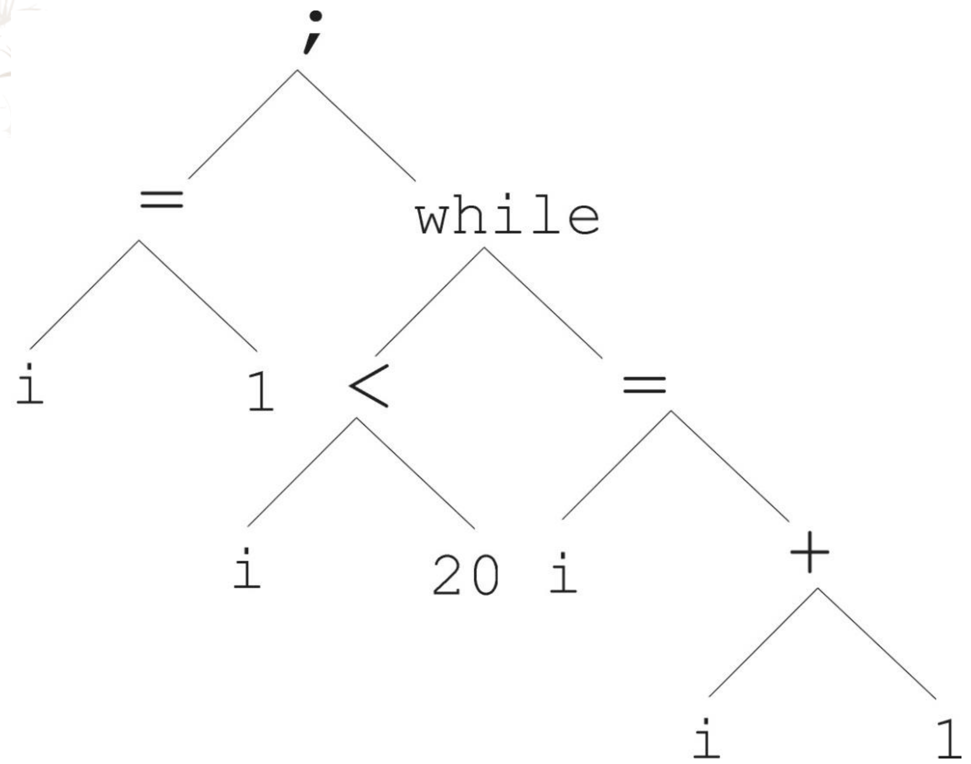
$$2 \cdot \pi + \left((x+3) - \frac{y}{5+1} \right)$$

Tree based representation



$(x \wedge true) \rightarrow ((x \vee y) \vee (z \leftrightarrow (x \wedge y)))$

Tree based representation




```
i =1;
while (i < 20)
{
        i = i +1
}
```

Tree based representation

- In GA chromosomes are linear structures (bit strings, integer string, real-valued vectors, permutations)
- Tree shaped chromosomes are non-linear structures
- In GA the size of the chromosomes is fixed
- Trees in GP may vary in depth and width

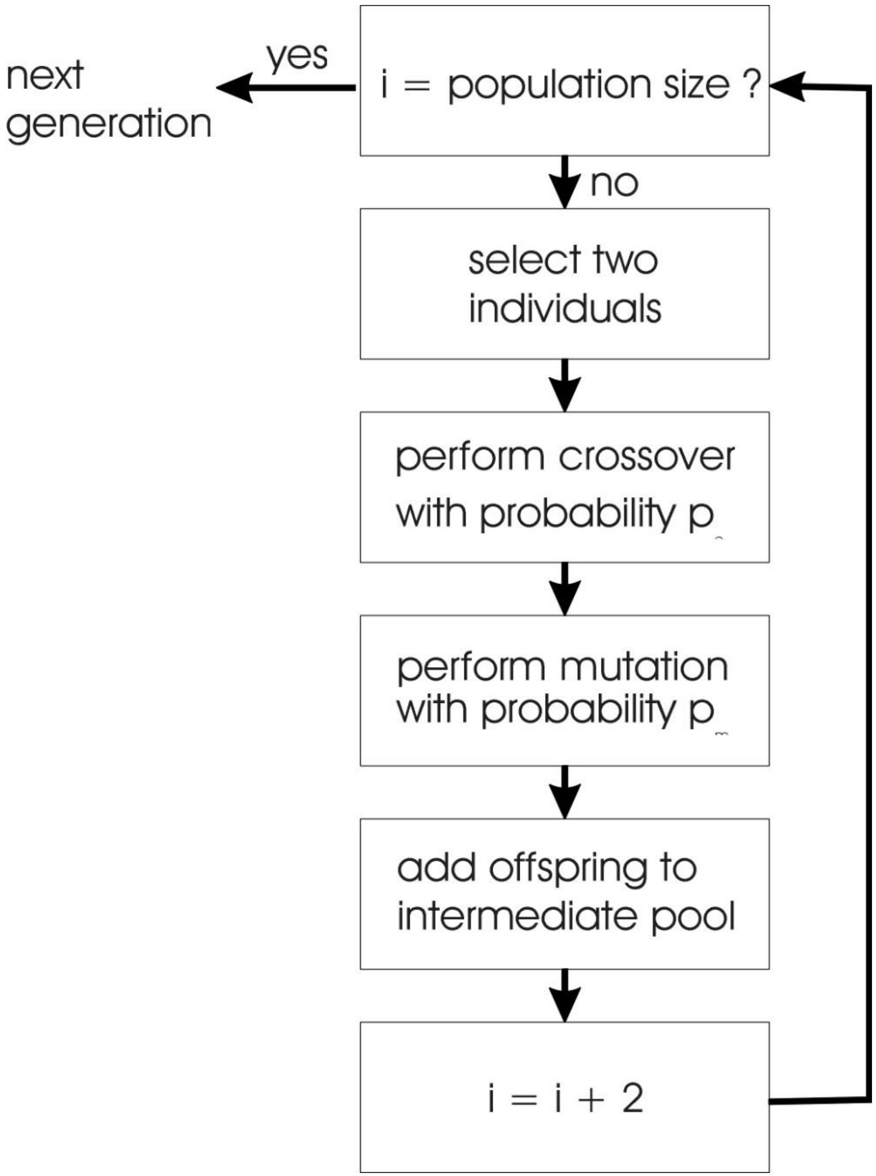
Tree based representation

- Symbolic expressions can be defined by
  - Terminal set T
  - Function set F (with the arities of function symbols)
- Adopting the following general recursive definition:
  1. Every $t \in T$ is a correct expression
  2. $f(e_1, \ldots, e_n)$ is a correct expression if $f \in F$, arity(f)=n and $e_1, \ldots, e_n$ are correct expressions
  3. There are no other forms of correct expressions
- In general, expressions in GP are not typed (closure property: any $f \in F$ can take any $g \in F$ as argument)

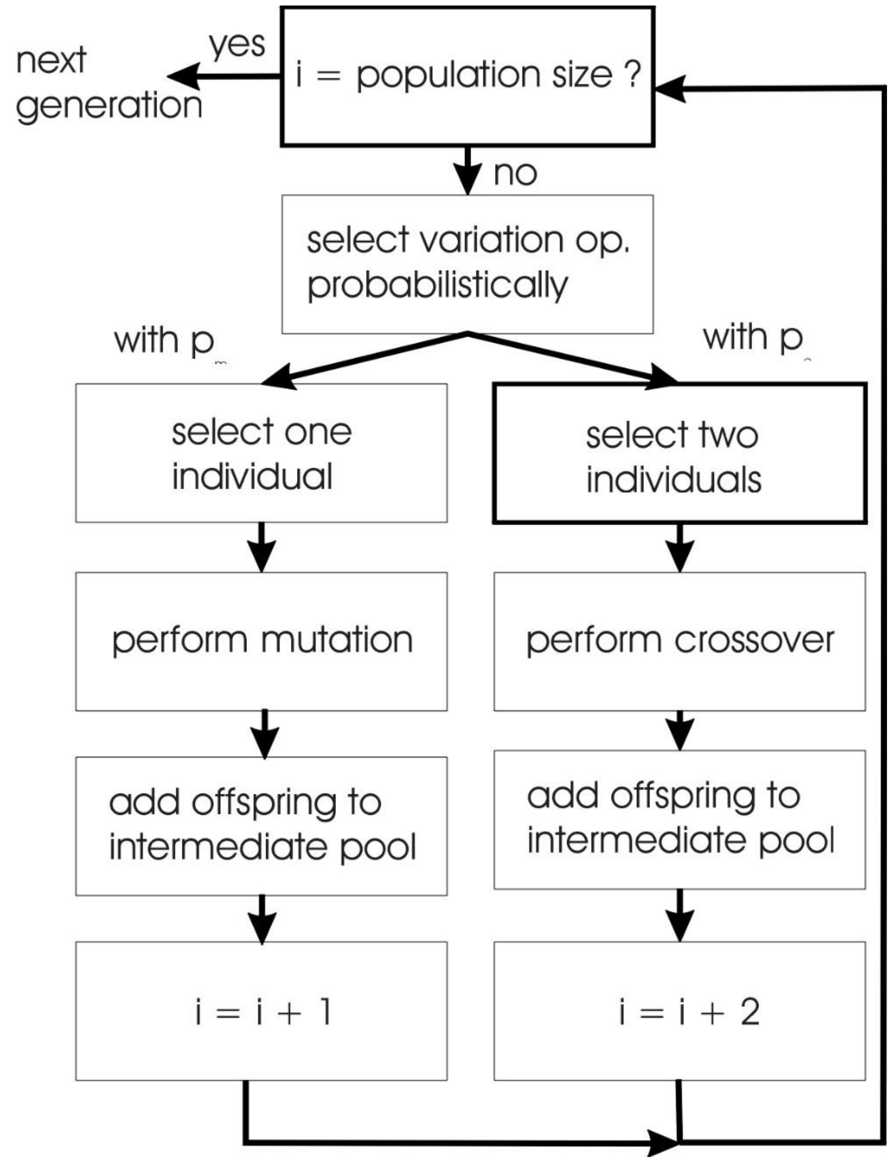
Offspring creation scheme

Compare

- GA scheme using crossover AND mutation sequentially (be it probabilistically)
- GP scheme using crossover OR mutation (chosen probabilistically)



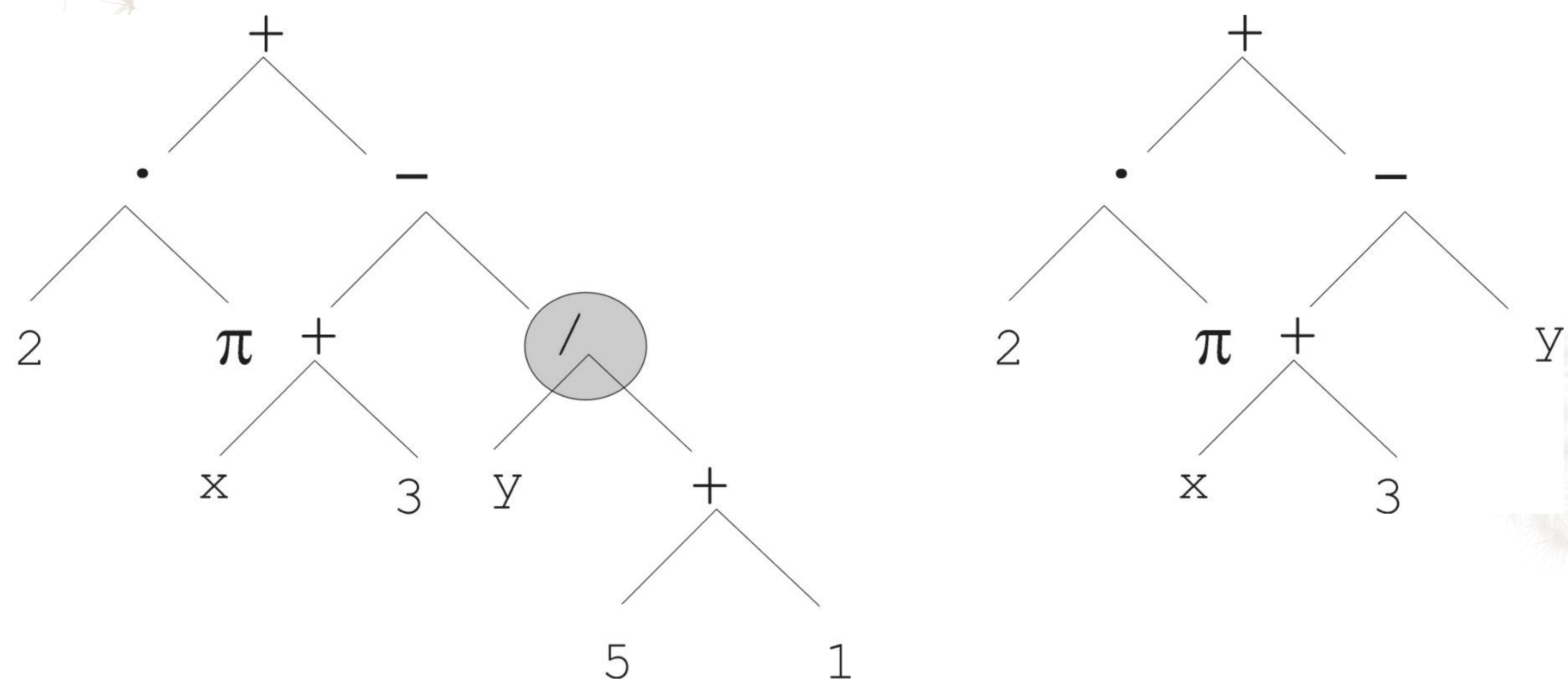
GA flowchart



GP flowchart

Mutation

- Most common mutation: replace randomly chosen subtree by randomly generated tree

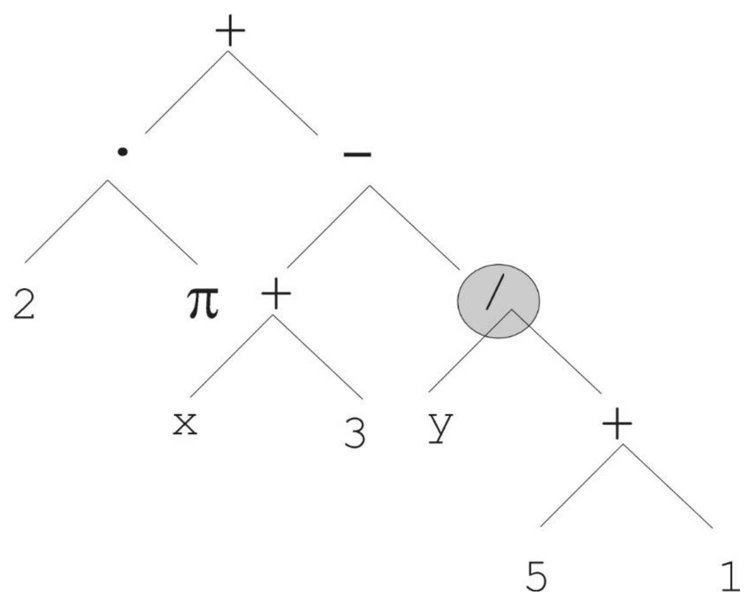


Mutation cont'd

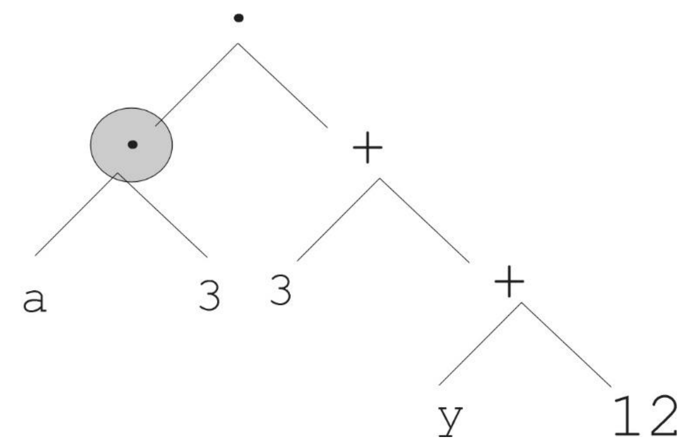
- Mutation has two parameters:
  - Probability p_m to choose mutation vs. recombination
  - Probability to chose an internal point as the root of the subtree to be replaced
- Remarkably p_m is advised to be 0 (Koza'92) or very small, like 0.05 (Banzhaf et al. '98)
- The size of the child can exceed the size of the parent

Recombination

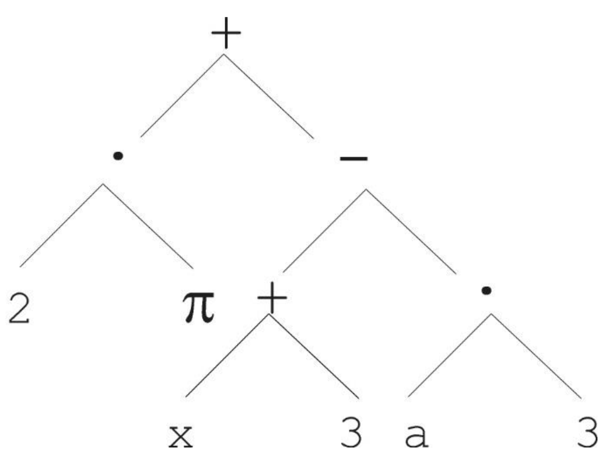
- Most common recombination: exchange two randomly chosen subtrees among the parents
- Recombination has two parameters:
  - Probability p_c to choose recombination vs. mutation
  - Probability to chose an internal point within each parent as crossover point
- The size of offspring can exceed that of the parents



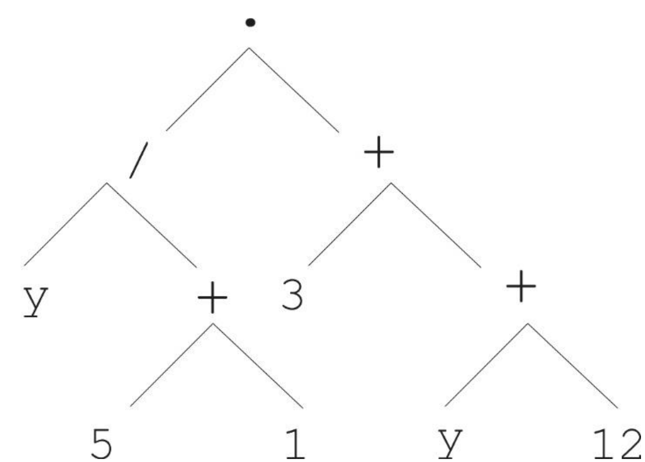
Parent 1



Parent 2



Child 1



Child 2

Selection

- Parent selection typically fitness proportionate
- Over-selection in very large populations
  - rank population by fitness and divide it into two groups:
  - group 1: best x% of population, group 2 other (100-x)%
  - 80% of selection operations chooses from group 1, 20% from group 2
  - for pop. size = 1000, 2000, 4000, 8000 x = 32%, 16%, 8%, 4%
  - motivation: to increase efficiency, %'s come from rule of thumb
- Survivor selection:
  - Typical: generational scheme (thus none)
  - Recently steady-state is becoming popular for its elitism

Initialisation

- Maximum initial depth of trees D_{max} is set
- Full method (each branch has depth = D_{max}):
  - nodes at depth $d < D_{max}$ randomly chosen from function set F
  - nodes at depth $d = D_{max}$ randomly chosen from terminal set T
- Grow method (each branch has depth $\leq D_{max}$):
  - nodes at depth $d < D_{max}$ randomly chosen from $F \cup T$
  - nodes at depth $d = D_{max}$ randomly chosen from T
- Common GP initialisation: ramped half-and-half, where grow & full method each deliver half of initial population

Bloat

- Bloat = “survival of the fattest”, i.e., the tree sizes in the population are increasing over time
- Ongoing research and debate about the reasons
- Needs countermeasures, e.g.
  - Prohibiting variation operators that would deliver “too big” children
  - Parsimony pressure: penalty for being oversized

Problems involving "physical" environments

- Trees for data fitting vs. trees (programs) that are "really" executable
- Execution can change the environment → the calculation of fitness
- Example: robot controller
- Fitness calculations mostly by simulation, ranging from expensive to extremely expensive (in time)
- But evolved controllers are often very good

Example application: symbolic regression

- Given some points in $\mathbf{R}^2$, $(x_1, y_1), \ldots, (x_n, y_n)$
- Find function $f(x)$ s.t. $\forall i = 1, \ldots, n : f(x_i) = y_i$
- Possible GP solution:
  - Representation by $F = \{+, -, /, \sin, \cos\}$, $T = \mathbf{R} \cup \{x\}$
  - Fitness is the error $$err(f) = \sum_{i=1}^{n} (f(x_i) - y_i)^2$$
  - All operators standard
  - pop.size = 1000, ramped half-half initialisation
  - Termination: n "hits" or 50000 fitness evaluations reached (where "hit" is if $| f(x_i) - y_i | < 0.0001$)

Discussion

Is GP:

The art of evolving computer programs ?

Means to automated programming of computers?

GA with another representation?

Evolving Neural Networks

- Evolving the architecture of neural network is slightly more complicated, and there have been several ways of doing it. For small nets, a simple matrix represents which neuron connects which, and then this matrix is, in turn, converted into the necessary 'genes', and various combinations of these are evolved.

Evolving Neural Networks

- Many would think that a learning function could be evolved via genetic programming. Unfortunately, genetic programming combined with neural networks could be *incredibly* slow, thus impractical.
- As with many problems, you have to constrain what you are attempting to create.
- For example, in 1990, David Chalmers attempted to evolve a function as good as the delta rule.
- He did this by creating a general equation based upon the delta rule with 8 unknowns, which the genetic algorithm then evolved.

Other Areas

- Genetic Algorithms can be applied to virtually any problem that has a large search space.
- Al Biles uses genetic algorithms to filter out 'good' and 'bad' riffs for jazz improvisation.
- The military uses GAs to evolve equations to differentiate between different radar returns.
- Stock companies use GA-powered programs to predict the stock market.

Example

- $f(x) = \{MAX(x^2): 0 <= x <= 32 \}$
- Encode Solution: Just use 5 bits (1 or 0).
- Generate initial population.

A	0	1	1	0	1
B	1	1	0	0	0
C	0	1	0	0	0
D	1	0	0	1	1

- Evaluate each solution against objective.

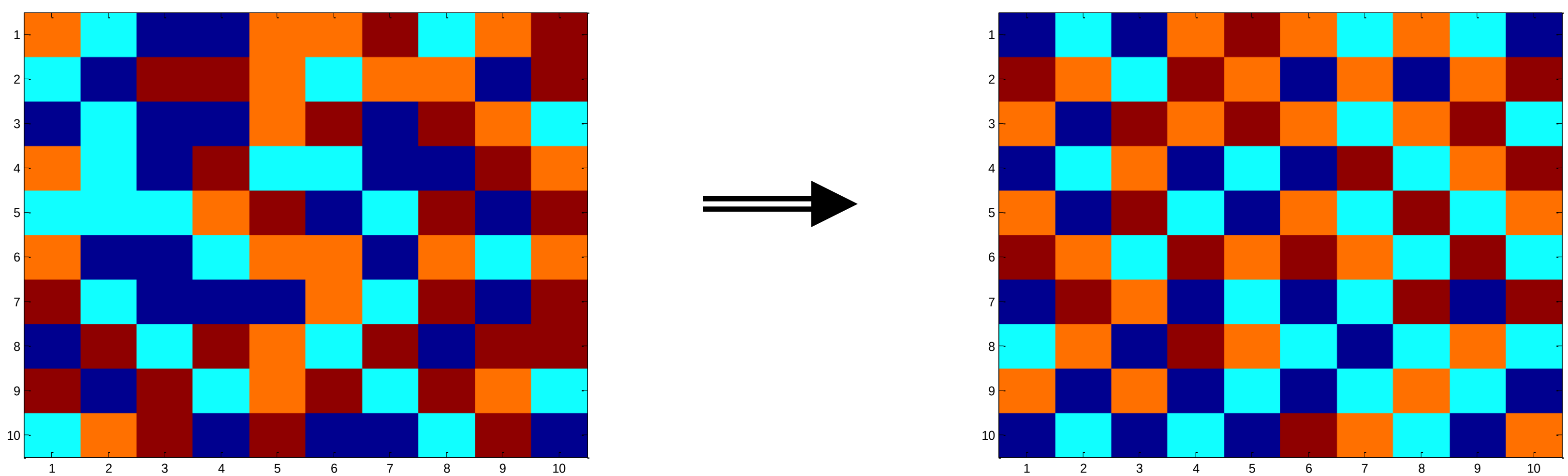
Sol.	**String**	**Fitness**	**% of Total**
A	01101	169	14.4
B	11000	576	49.2
C	01000	64	5.5
D	10011	361	30.9

Example Cont'd

- Create next generation of solutions
  - Probability of "being a parent" depends on the fitness.
- Ways for parents to create next generation
  - Reproduction
    - Use a string again unmodified.
  - Crossover
    - Cut and paste portions of one string to another.
  - Mutation
    - Randomly flip a bit.
  - COMBINATION of all of the above.

Checkboard example

- We are given an ***n*** by ***n*** checkboard in which every field can have a different colour from a set of four colors.
- Goal is to achieve a checkboard in a way that there are no neighbours with the same color (not diagonal)



Checkboard example Cont'd

- Chromosomes represent the way the checkboard is colored.
- Chromosomes are not represented by bitstrings but by **bitmatrices**
- The bits in the bitmatrix can have one of the four values 0, 1, 2 or 3, depending on the color.
- Crossing-over involves matrix manipulation instead of point wise operating.
- Crossing-over can be combining the parential matrices in a horizontal, vertical, triangular or square way.
- Mutation remains bitwise changing bits in either one of the other numbers.

Checkboard example Cont'd

- This problem can be seen as a graph with ***n*** nodes and ***(n-1)*** edges, so the fitness **f(x)** is defined as:

$$\mathbf{f(x) = 2 \cdot (n-1) \cdot n}$$

Checkboard example Cont'd

- Fitnesscurves for different cross-over rules: