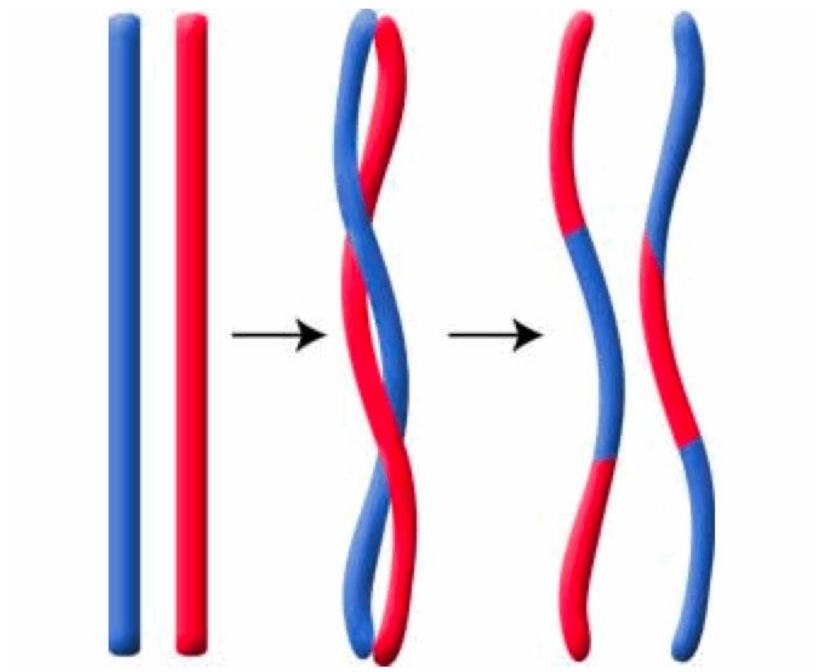
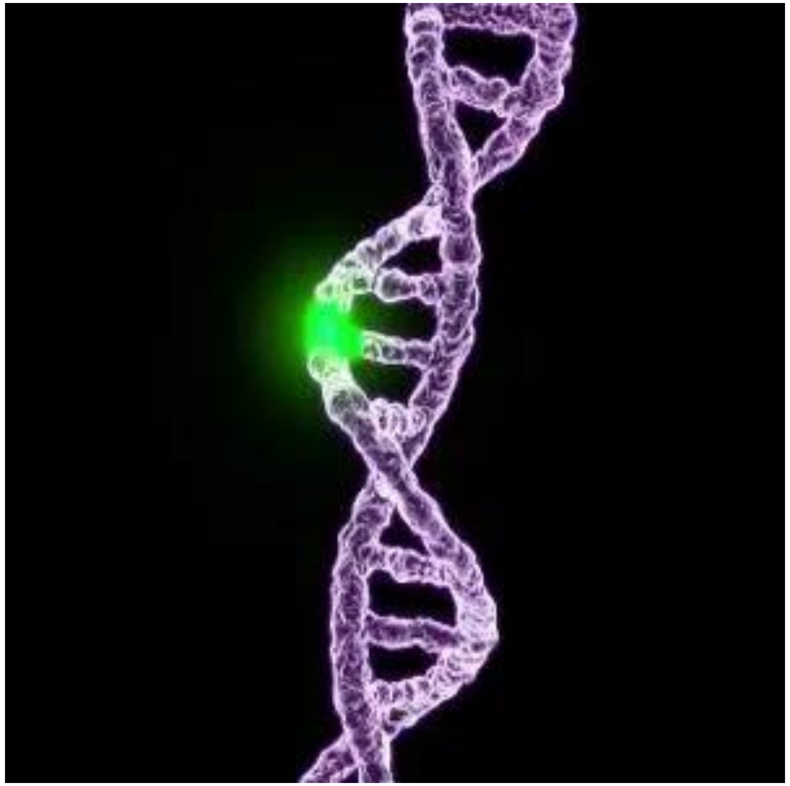
Genetic/Evolutionary Algorithms

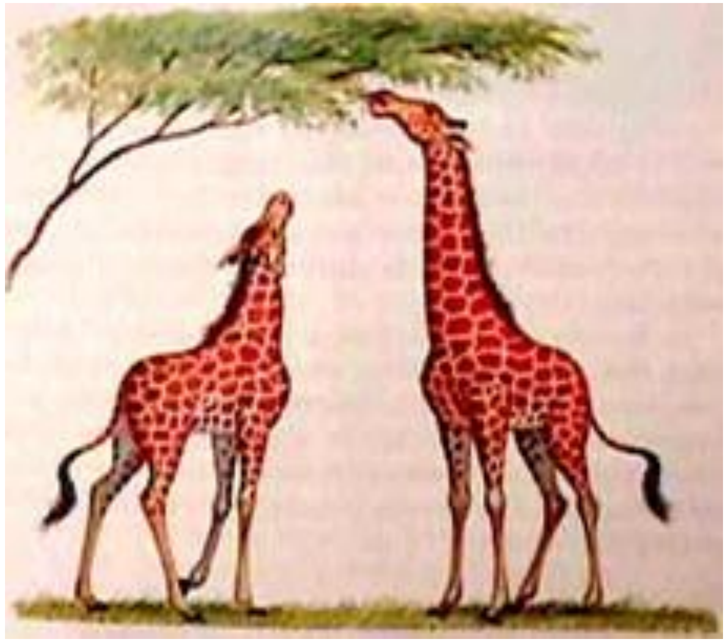
Paul Bodily

Evolution and Natural Selection



http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp3_2.shtml



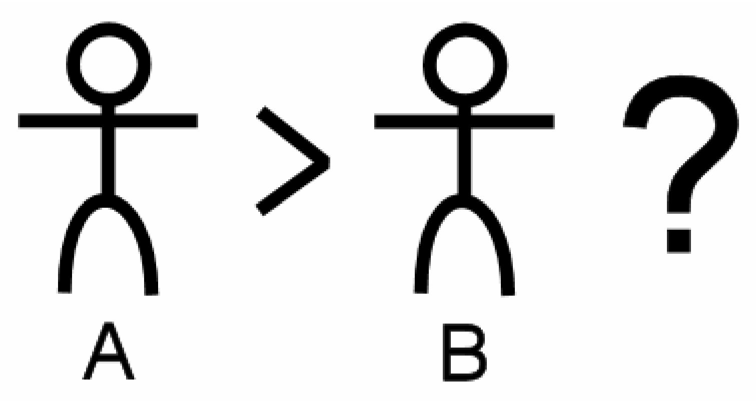


Genetic Algorithms – Guided Trial and Error

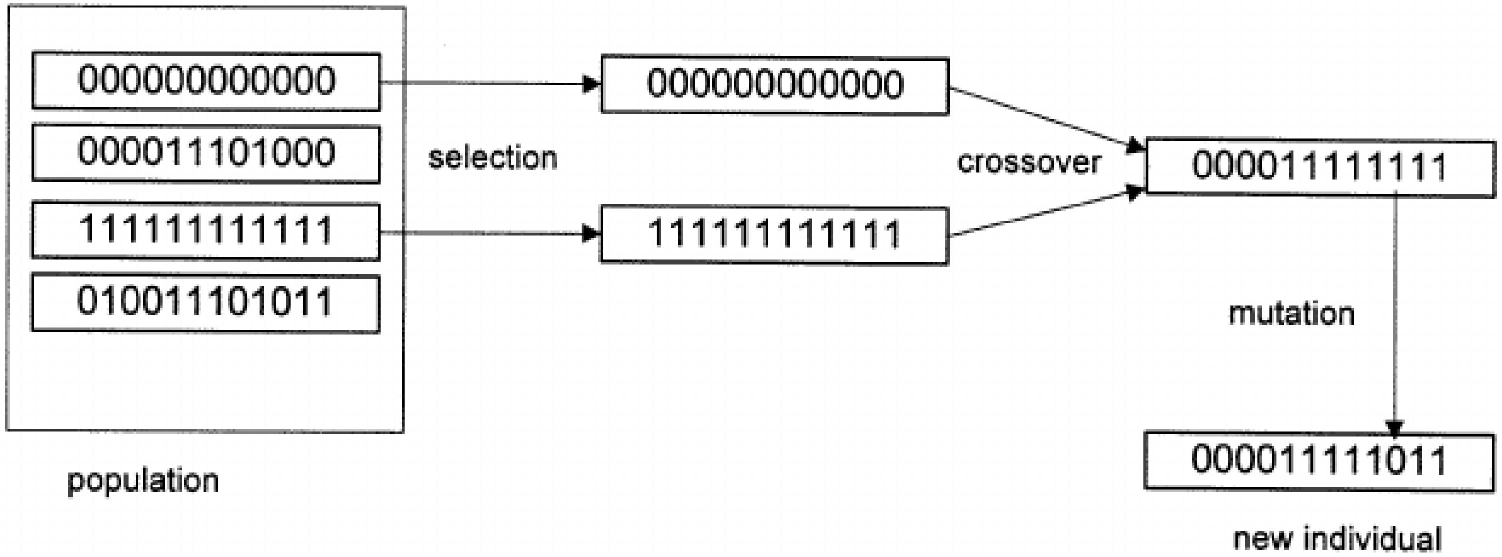
1. Genetic Representation

	$\ell_{i,1}$	$\ell_{i,2}$	$\ell_{i,3}$	$\ell_{i,j}$	...	$\ell_{i,N}$
chromosome 1	1	0	0	1	...	1
chromosome 2	0	1	0	0	...	0
. . .						
chromosome i	1	1	1	0	...	1
:						
chromosome Λ	0	1	1	1	...	0

3. Fitness Function



2. Selection, Crossover & Mutation



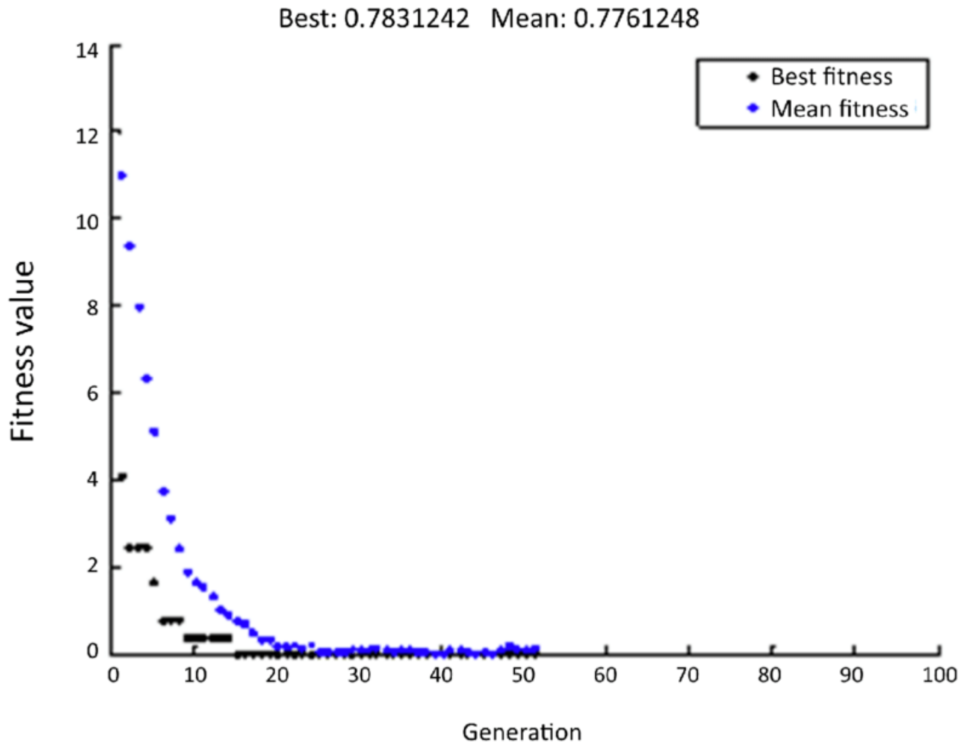
ROAR

Example: String discovery problem

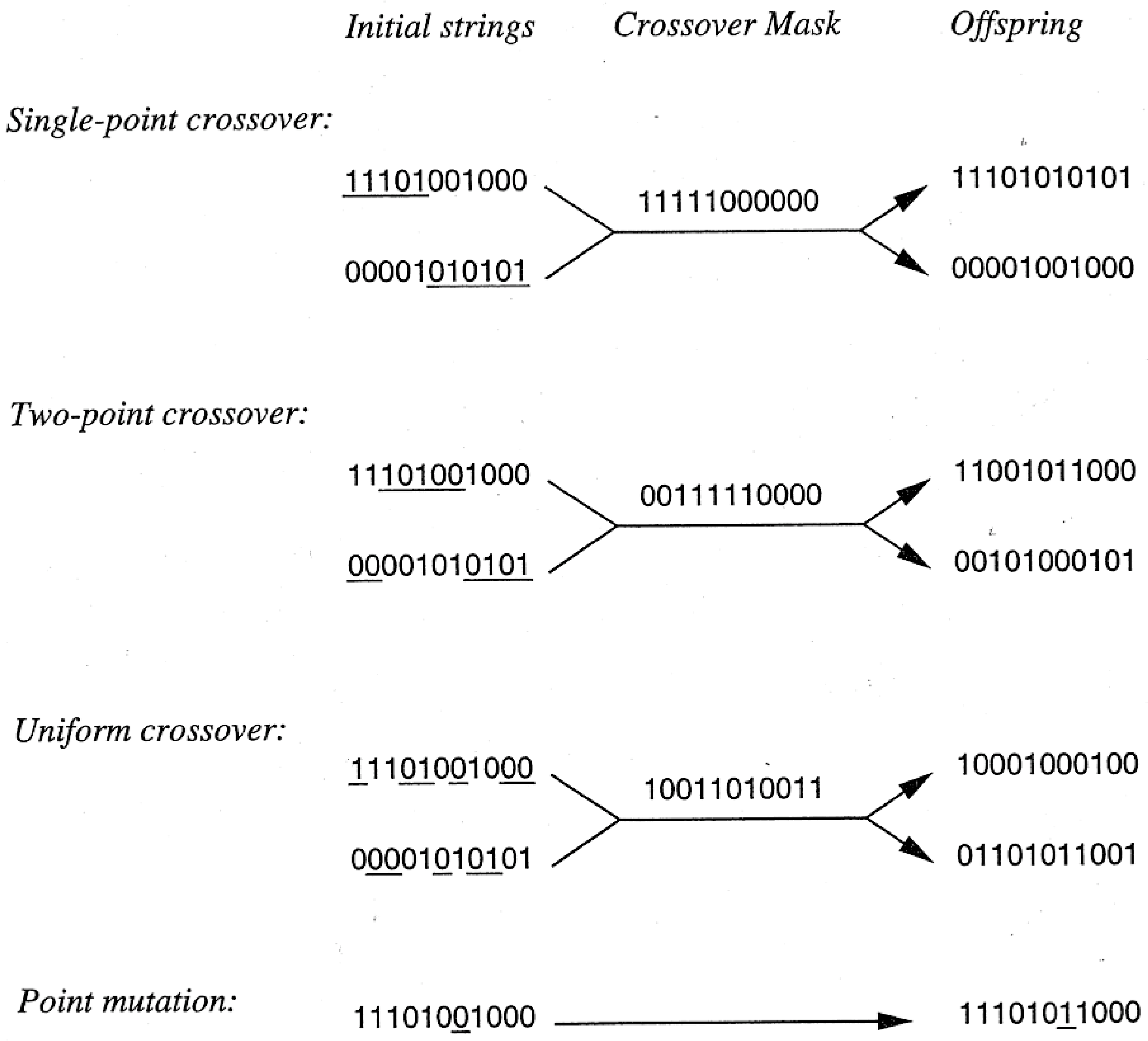
1. Randomly pick a 5-letter “word” from **letters a-f** (e.g., “dbdae”). **Write it down.**
2. Get a *fitness* score from tinyurl.com/gademopb for your word (e.g., f(“dbdae”) = 0). Hold up fingers indicating your *fitness*.
3. While TRUE:
   1. Find another person
   2. Create a **new word** by *crossing over* the first two of yours and last three of theirs (e.g., “dbdae” + “eacdf” -> “dbcdf”).
   3. Mutate one letter (e.g., “dbcdf” -> “dbadf”)
   4. Get a score for the new word (e.g., f(“dbadf”) = 1)
   5. Choose whether to keep the new or the old word

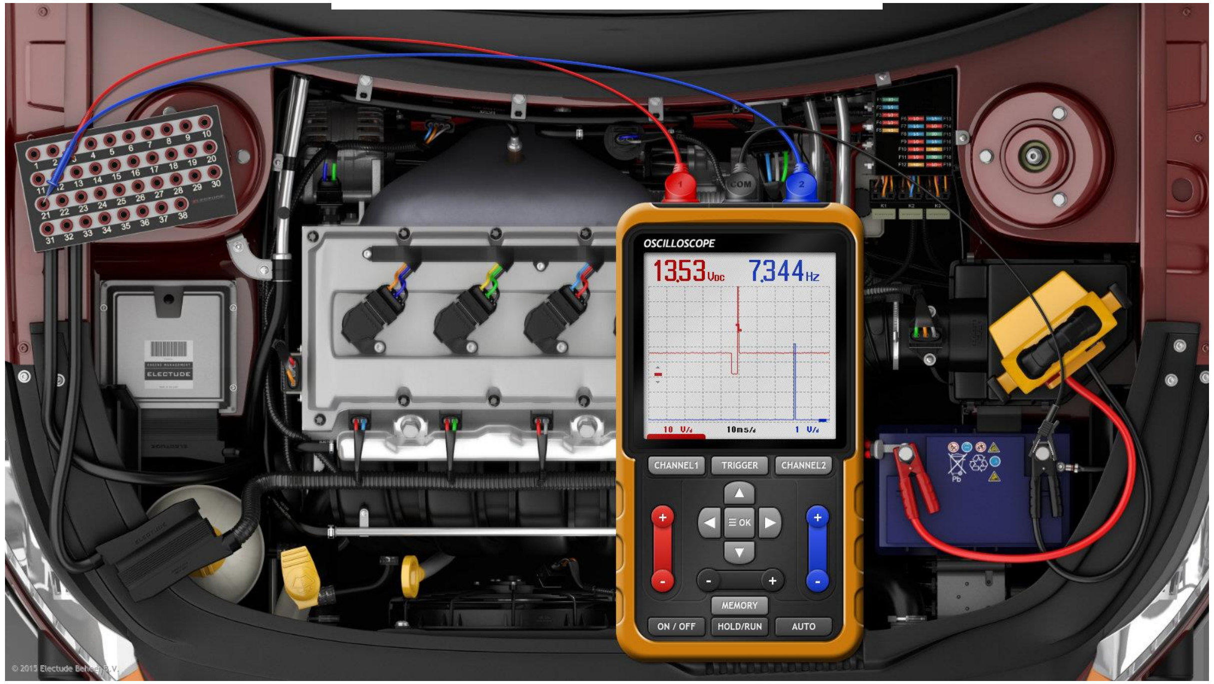
Pseudocode

```
START
Generate the initial population
Compute fitness
REPEAT
    Selection
    Crossover
    Mutation
    Compute fitness
    Prune population
UNTIL population has converged
STOP
```



For detailed code example check out *Intro to Genetic Algorithms*, towardsdatascience.com


Initial strings
Crossover Mask
Offspring
Single-point crossover:
11101001000
00001010101
11111000000
11101010101
00001001000
Two-point crossover:
11101001000
00001010101
00111110000
11001011000
00101000101
Uniform crossover:
11101001000
00001010101
10011010011
10001000100
01101011001
Point mutation:
11101001000
11101011000



Individual	CID	Fuel System	Turbo	Valves	Cylinders
1	350	4 Barrels	Yes	16	8
2	250	Mech. Inject.	No	12	6
3	150	Elect. Inject.	Yes	12	4
4	200	2 Barrels	No	8	4

We now evaluate each individual with the engine simulator. Each individual receives a fitness score (the higher the better):

Individual	CID	Fuel System	Turbo	Valves	Cylinders	Score
1	350	4 Barrels	Yes	16	8	50
2	250	Mech. Inject.	No	12	6	100
3	150	Elect. Inject.	Yes	12	4	300
4	200	2 Barrels	No	8	4	150

Parent selection decides who has children and how many to have. For example, we could decide that individual 3 deserves two children, because it is so much better than the other individuals. Children are created through recombination and mutation. As mentioned above, recombination exchanges information between individuals, while mutation perturbs individuals, thereby increasing diversity. For example, recombination of individuals 3 and 4 could produce the two children:

Individual	CID	Fuel System	Turbo	Valves	Cylinders
3’	200	Elect. Inject.	Yes	8	4
4’	150	2 Barrels	No	12	4

Note that the children are composed of elements of the two parents. Further note that the number of cylinders must be four, because individuals 3 and 4 both had four cylinders. Mutation might further perturb these children, yielding:

Individual	CID	Fuel System	Turbo	Valves	Cylinders
3’	250	Elect. Inject.	Yes	8	4
4’	150	2 Barrels	No	12	6

We now evaluate the children, giving perhaps:

Individual	CID	Fuel System	Turbo	Valves	Cylinders	Score
3’	250	Elect. Inject.	Yes	8	4	250
4’	150	2 Barrels	No	12	6	350

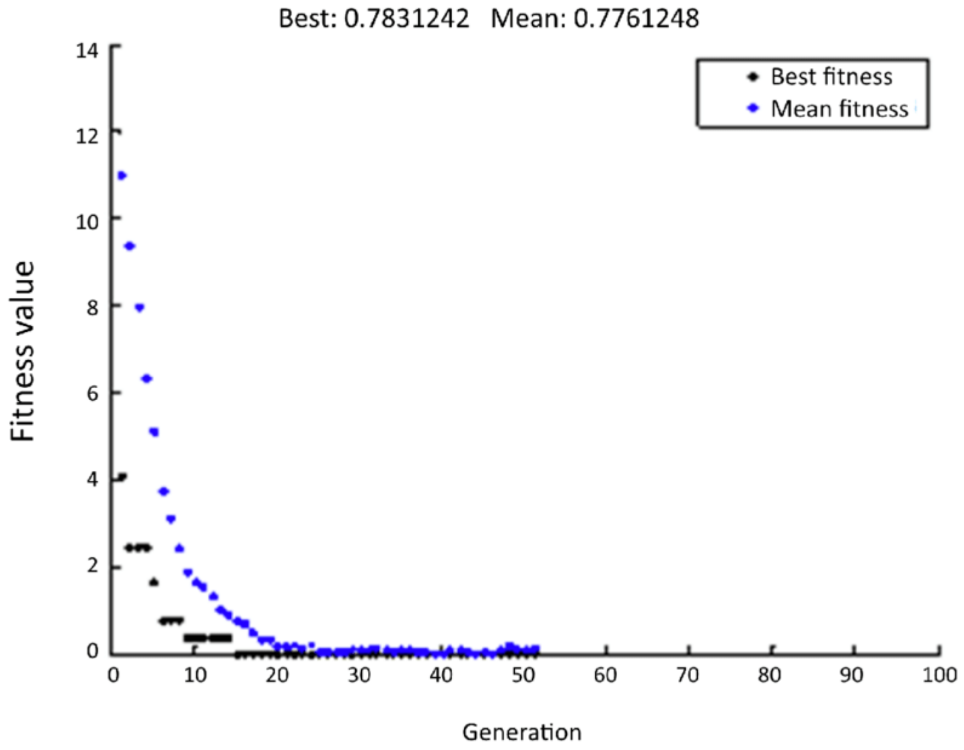
Finally we decide who will survive. In our constant population size example, which is typical of most EAs, we need to select four individuals to survive. How this is accomplished varies considerably in different EAs. If, for example, only the best individuals survive, our population would become:

Individual	CID	Fuel System	Turbo	Valves	Cylinders	Score
3	150	Elect. Inject.	Yes	12	4	300
4	200	2 Barrels	No	8	4	150
3’	250	Elect. Inject.	Yes	8	4	250
4’	150	2 Barrels	No	12	6	350

This cycle of evaluation, selection, recombination, mutation, and survival continues until some termination criterion is met.

Pseudocode

```
START
Generate the initial population
Compute fitness
REPEAT
    Selection
    Crossover
    Mutation
    Compute fitness
    Prune population
UNTIL population has converged
STOP
```



For detailed code example check out *Intro to Genetic Algorithms*, towardsdatascience.com

Generic GA implementation

```
START
Generate the initial population
Compute fitness
REPEAT
    Selection
    Crossover
    Mutation
    Compute fitness
    Prune population
UNTIL population has converged
STOP
```

- Break

TAPPS – Recreate the string discovery problem

- Get in pairs, choose a driver, and code up a genetic algorithm

```
START
Generate the initial population
Compute fitness
REPEAT
    Selection
    Crossover
    Mutation
    Compute fitness
    Prune population
UNTIL population has converged
STOP
```

```
Make len and char set variables
10 random “words” made of up a-e
Char by char comparison with solution

Randomly select from population
Randomly select an index for crossover
Foreach index, mutate with low prob
Char by char comparison with solution
Sort by fitness and keep top 10
Until fitness = word len
```

Idaho State University

GA Project

- Due 1 week from today before class
- Creating Crosswords Minis with Genetic Algorithms

Idaho State University

GA Project Ideation (possibilities)

- What is a genome?

O	W	L	T
U	J	P	I
S	P	S	F
T	Z	S	F

- Mutation?

O	W	X	T
I	J	P	I
O	O	S	F
P	E	A	F

- Crossover?

O	W	L	T
U	J	P	I
S	P	S	F
T	Z	S	F

+

U	C	R	L
I	M	J	U
N	O	A	W
P	E	C	S

=

O	W	L	T
I	J	P	I
N	O	S	F
P	E	C	F

- Fitness?

ROAR

GA Project: Fitness

P	A	G	E
N	B	S	S
W	Y	A	T
I	E	S	S

Fitness: 7

- Clues:
  - 1 Across: one side of one leaf (of a book or magazine or newspaper or letter etc.) or the written or pictorial matter it contains (PAGE)
  - 2 Across: a soft grey ductile metallic element used in alloys; occurs in niobite; formerly called columbium (NBSS)
  - 3 Across: English poet who introduced the sonnet form to English literature (1503-1542) (WYAT)
  - 4 Across: a commercial browser (IESS)

1 Down: PNWI is invalid

2 Down: make amends for (ABYE)

3 Down: a central management agency that sets Federal policy for Federal procurement and real property management and information resources management (GSAS)

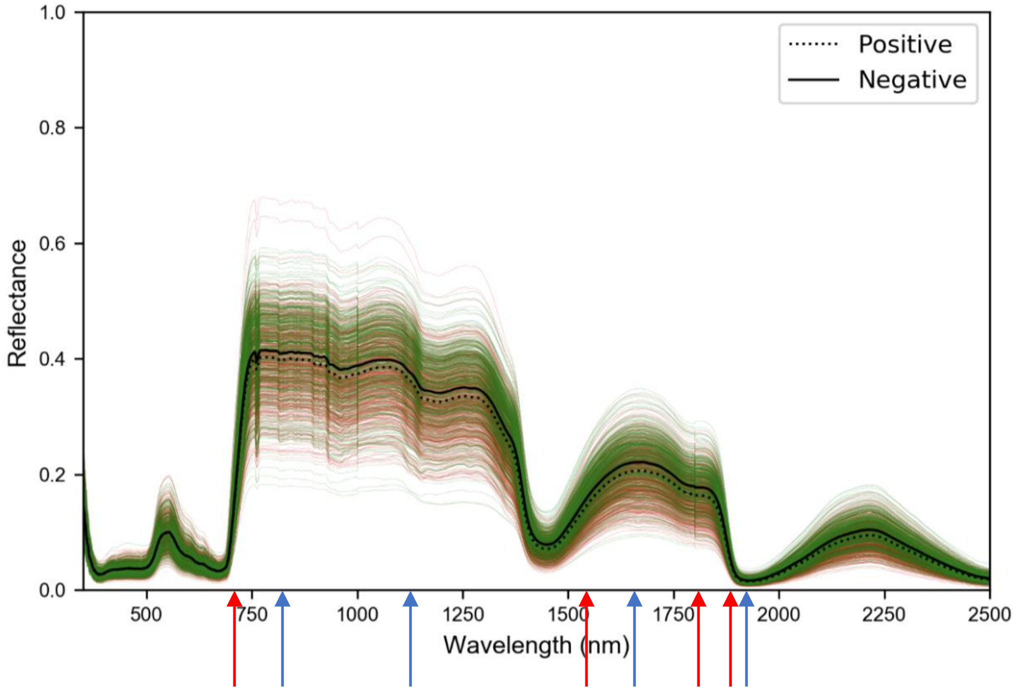
4 Down: standard time in the 5th time zone west of Greenwich, reckoned at the 75th meridian; used in the eastern United States (ESTS)

GA Project

- https://www2.cose.isu.edu/~bodipaul/courses/sp21/4473/projects/ga.php

- Break

PVY Example

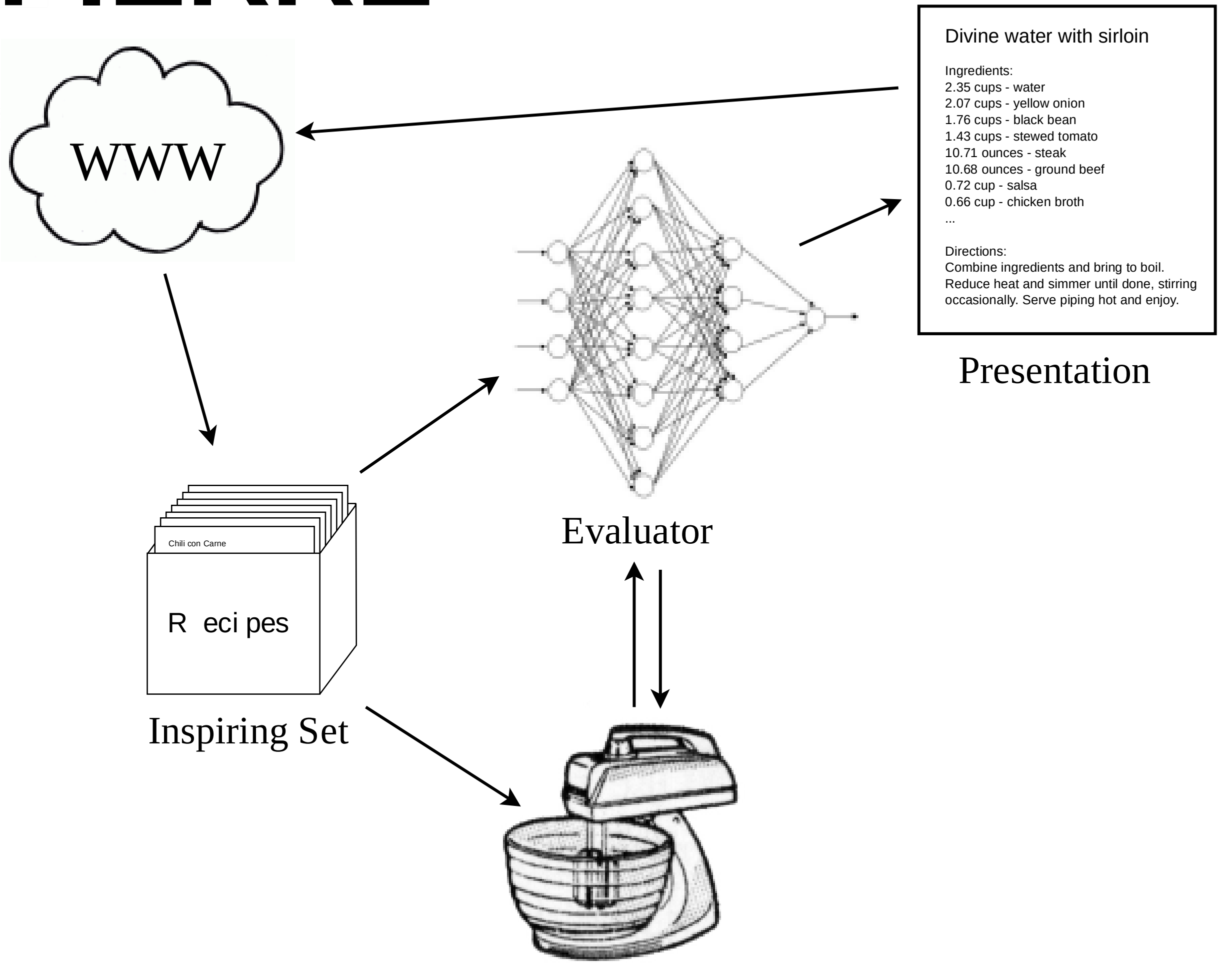


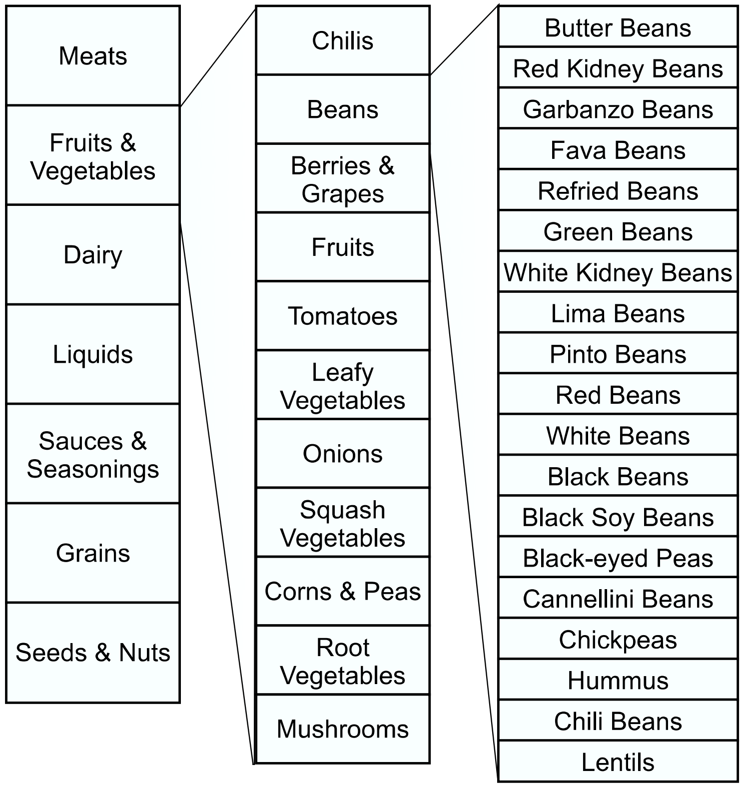
E.g., (750,1246,1876,1976)

- Genome: A subset of wavelengths
- Fitness: How well they distinguish +/-
- Cross-over: Recombine subsets
- Mutation: Randomly resample a wavelength
- Population size: Boh? 100?
- Initialization: Random subsets of wavelengths

Idaho State University

PIERRE





Abstraction 2

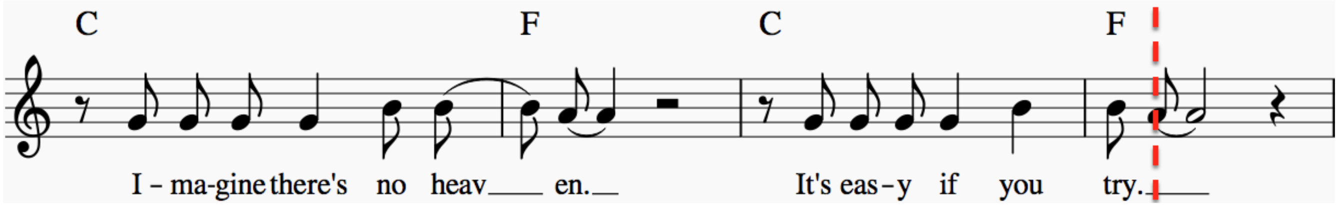
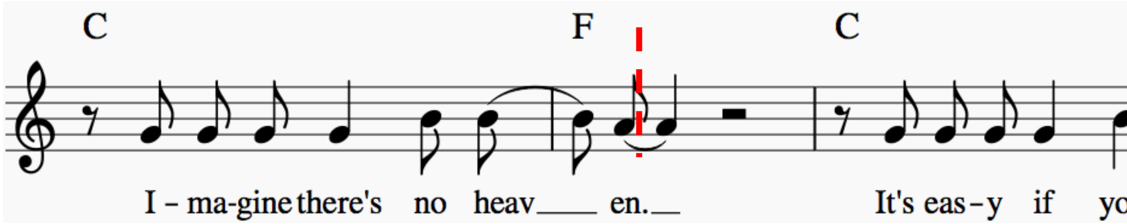
22.55 oz Meats
44.5 oz Fruits & Veggies
8.99 oz Sauces & Seasonings
23.96 oz Liquids

Abstraction 1

22.36 oz Beef
0.19 oz Pork
13.35 oz Beans
12.04 oz Tomatoes
19.11 oz Onions
8.99 oz Spices
23.96 oz Broths

Recipe

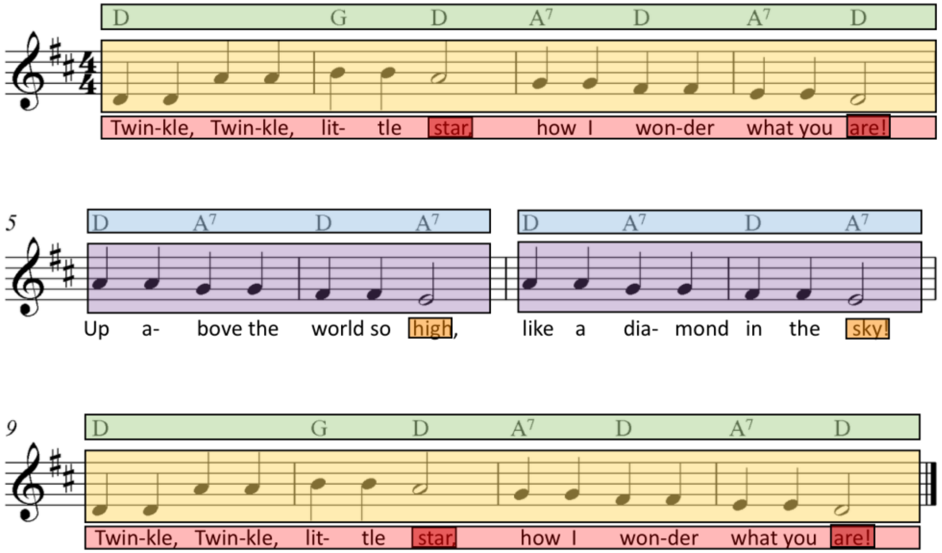
17.63 oz ground beef
4.73 oz steak
0.19 oz pork sparerib
6.24 oz red kidney bean
0.25 oz garbanzo bean
0.28 oz lima bean
6.58 oz chickpea
0.33 oz crushed tomato
0.59 oz chopped tomato
1.98 oz tomato puree
1.31 oz diced tomato
0.39 oz roma tomato
7.44 oz spaghetti sauce
12.13 oz yellow onion
6.75 oz white onion
0.23 oz chive baton
8.86 oz garlic
0.13 oz fresh parsley
23.96 oz chicken broth

How do humans recognize musical ***motifs?***



Event Feature	Description	Range	Feature Value for E in Figure 1
$is\_rest(E)$	$True$ if E occurs during a rest	$[True, False]$	$False$
$pitch(E)$	the MIDI note value being voiced at E	[0,127] (? if $is\_rest(E)$)	69
$measure(E)$	the measure in which E occurs (0-based)	$Z_{>0}$	3
$beat(E)$	the offset in beats within measure $measure(E)$ (0-based)	$R_{>0}$	0.5
$duration(E)$	the duration in beats of the note or rest being voiced at E	$R_{>0}$	2.5
$is\_note\_onset(E)$	$True$ if the measure and beat of the onset of the note or rest being voiced at E equals $measure(E)$ and $beat(E)$	$[True, False]$	$True$
$lyric(E)$	the lyric being sung at E	Set of all valid syllables [?	"try"
$is\_lyric\_onset(E)$	$True$ if the measure and beat of the onset of the lyric being voiced at E equals $measure(E)$ and $beat(E)$	$[True, False]$ (? if $lyric(E)$ = ?)	$False$
$harmony(E)$	the harmony (represented using chord symbols) being voiced at E	Set of all valid chord symbols [?	F
$is\_harmony\_onset(E)$	$True$ if the measure and beat of onset of the harmony being voiced at E equals $measure(E)$ and $beat(E)$	$[True, False]$ (? if $harmony(E)$ = ?)	$False$

Table 1: Features for a music sequence event

Fitness function: How accurately did it predict motifs?



$$Genetic\ Representation: \Gamma_P = (G_o, G_e, \tau, \iota_B, \Delta_B, \delta_B, R, \rho, \gamma_R, \iota_P, \Delta_P, \delta_P, \Omega_P, \omega_P, \gamma_P)$$

ROAR

Idaho State

Genetic Representation: $\Gamma_P = (G_o, G_e, \tau, \iota_B, \Delta_B, \delta_B, R, \rho, \gamma_R, \iota_P, \Delta_P, \delta_P, \Omega_P, \omega_P, \gamma_P)$

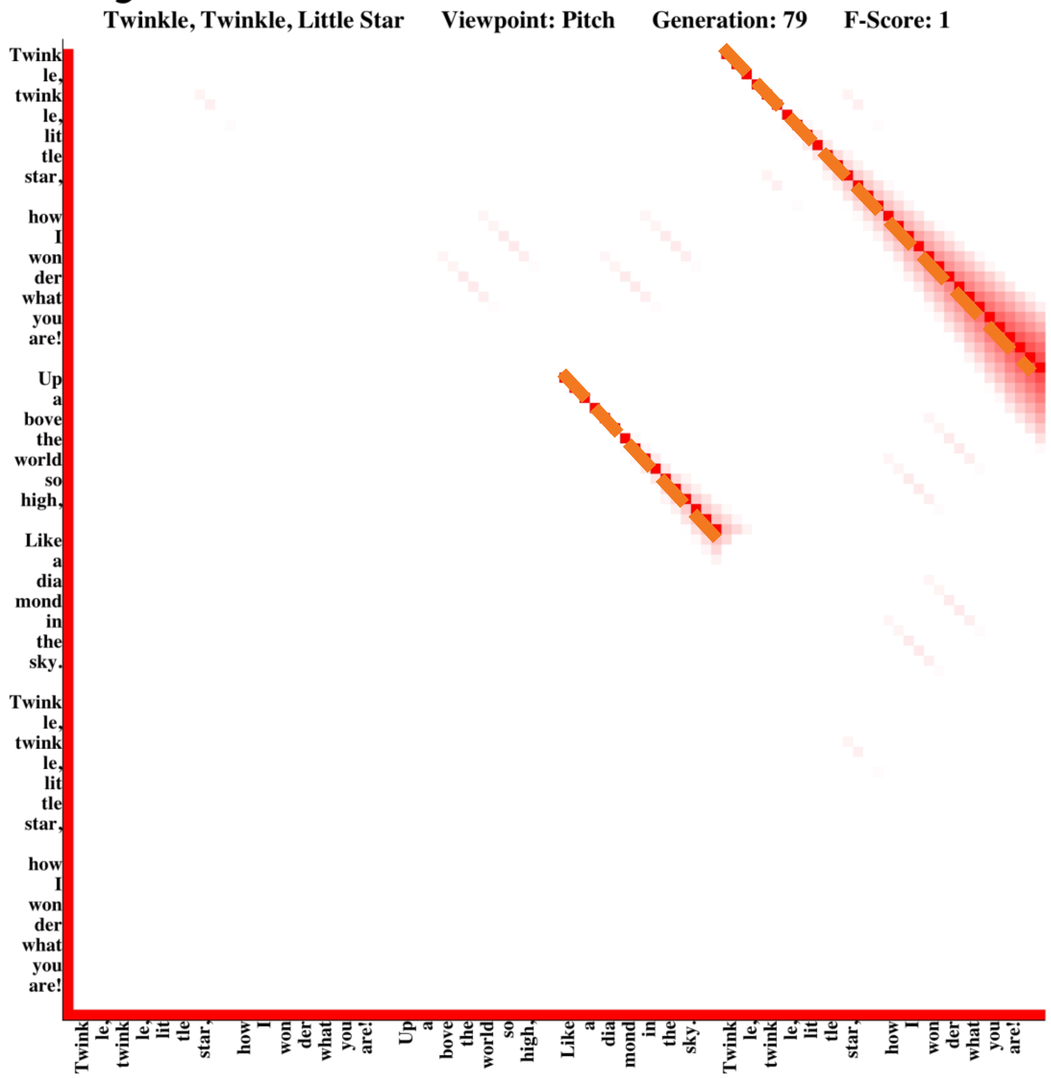
Generation 1

0	0.2027512171314219	0, -3, 18.82, 2, 2, -2, 0, 0, 1, -1, -1, 3, 2, -1, 3, -3, 0
1	0.15202181545737853	-1, -3, 12.5, 2, 2, 0, -3, 0, -2, -3, 2, 0, -3, 3, -2, 2, -3
2	0.1500895146270687	-1, 3, 0.54, 2, 2, -3, -1, 3, -2, -3, -1, -3, 1, -2, -3, 0, 3
3	0.13685356843018534	0, -1, 14.02, 2, 2, -3, -2, -1, -1, 1, 0, -2, 3, 3, 2, -2, -2
4	0.1302921377615676	3, 0, 14.25, 2, 2, -2, -1, 3, 0, 3, 0, 3, 1, -1, 0, 1, 1
5	0.13022695096121684	3, 3, 14.71, 2, 2, 0, 1, 1, 3, 1, 3, 0, -1, -1, -3, 2, -1
6	0.1206836584151453	-1, 2, 13.68, 2, 2, 2, -2, 1, -3, -2, 2, 2, -3, 1, 0, 0, 2
7	0.1147013498354312	1, -1, 0.58, 2, 2, 0, 2, 3, 1, 3, 3, 2, -3, 2, -2, -3, -2
8	0.10426969293165815	1, -2, 10.63, 2, 2, -1, -1, 0, 1, -3, 0, 2, 0, 1, 1, 2, 2
9	0.0838432148613218	3, 1, 9.24, 2, 2, 2, 0, 1, 0, 3, 0, -2, 3, 3, 2, 1, -3
10	0.08316844426032159	1, 3, 16.97, 2, 2, 1, -1, 3, 3, 3, 3, -1, 0, 2, -2, 1, 1
11	0.08132901298140642	2, 3, 17.31, 2, 2, 2, 1, -1, -3, 3, 1, 0, 3, 3, -1, -3, -3
12	0.08000864600198239	2, -1, 14.81, 2, 2, 3, -1, -1, 3, 3, -2, 2, -2, -2, -1, 2, 1
13	0.07504809366902623	2, 3, 2.86, 2, 2, 0, 0, 0, -3, 3, 0, -1, 3, 1, 0, -3, 0
14	0.07261799533358979	1, 1, 3.8, 2, 2, -2, 1, -3, -3, 1, -1, -3, 2, 3, -1, 2, 2
15	0.047138951981281305	3, -3, 16.78, 2, 2, 3, 2, 1, 0, 0, 1, -1, -1, -3, 2, 3, 3
16	0.027127919115237414	3, 2, 19.87, 2, 2, -2, -2, 3, -3, -3, -1, 2, 0, -3, -2, 1, 1
17	0.023305008373622853	2, -2, 9.94, 2, 2, 2, -2, -1, -2, 3, 1, -3, -2, 0, 3, -1, 0
18	0.023303477520424786	2, -1, 4.81, 2, 2, 1, -3, -1, -2, 3, -1, -2, -1, -3, -1, 1, -3
19	0.023303477520424786	1, 0, 5.3, 2, 2, -3, 1, -3, 0, 1, 0, -3, -3, -1, -1, 1, -1

ROAR

Idaho State University

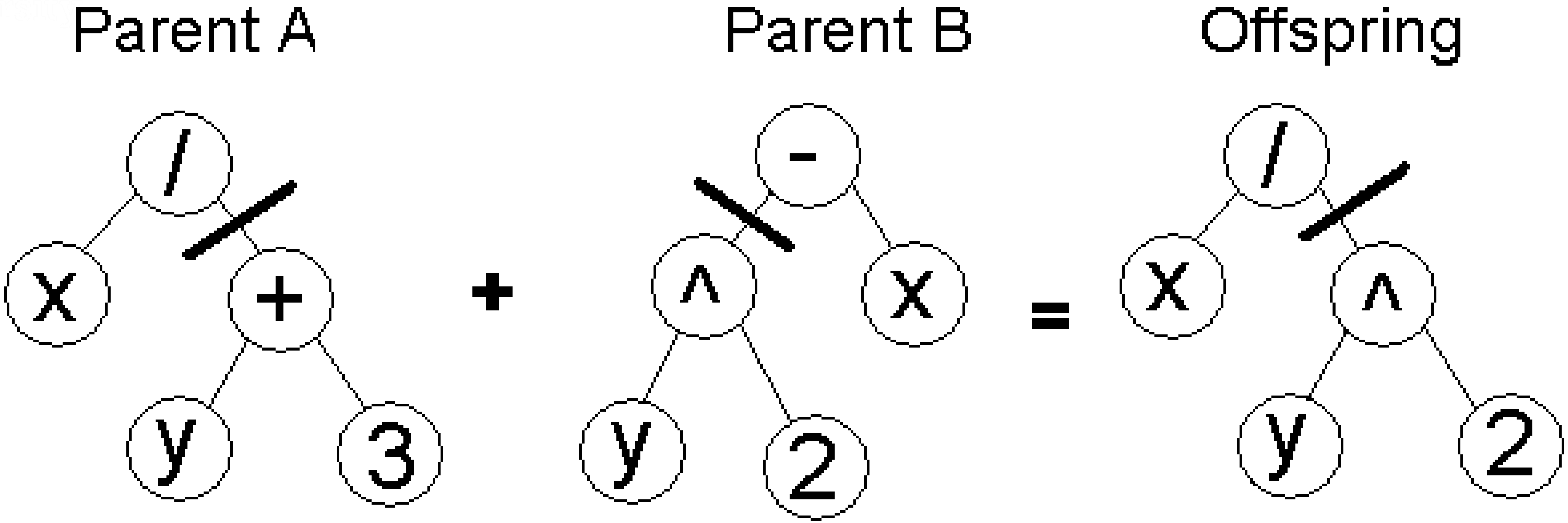
$$\Gamma_P = (G_o, G_e, \tau, \iota_B, \Delta_B, \delta_B, R, \rho, \gamma_R, \iota_P, \Delta_P, \delta_P, \Omega_P, \omega_P, \gamma_P)$$



Gen	Avg F-Score
1	0.202751217
2	0.283322436
3	0.348936672
5	0.419817123
6	0.583818628
8	0.626922611
18	0.630017273
22	0.822838916
44	0.877148402
59	0.899415259
67	0.903958945
79	0.91960754

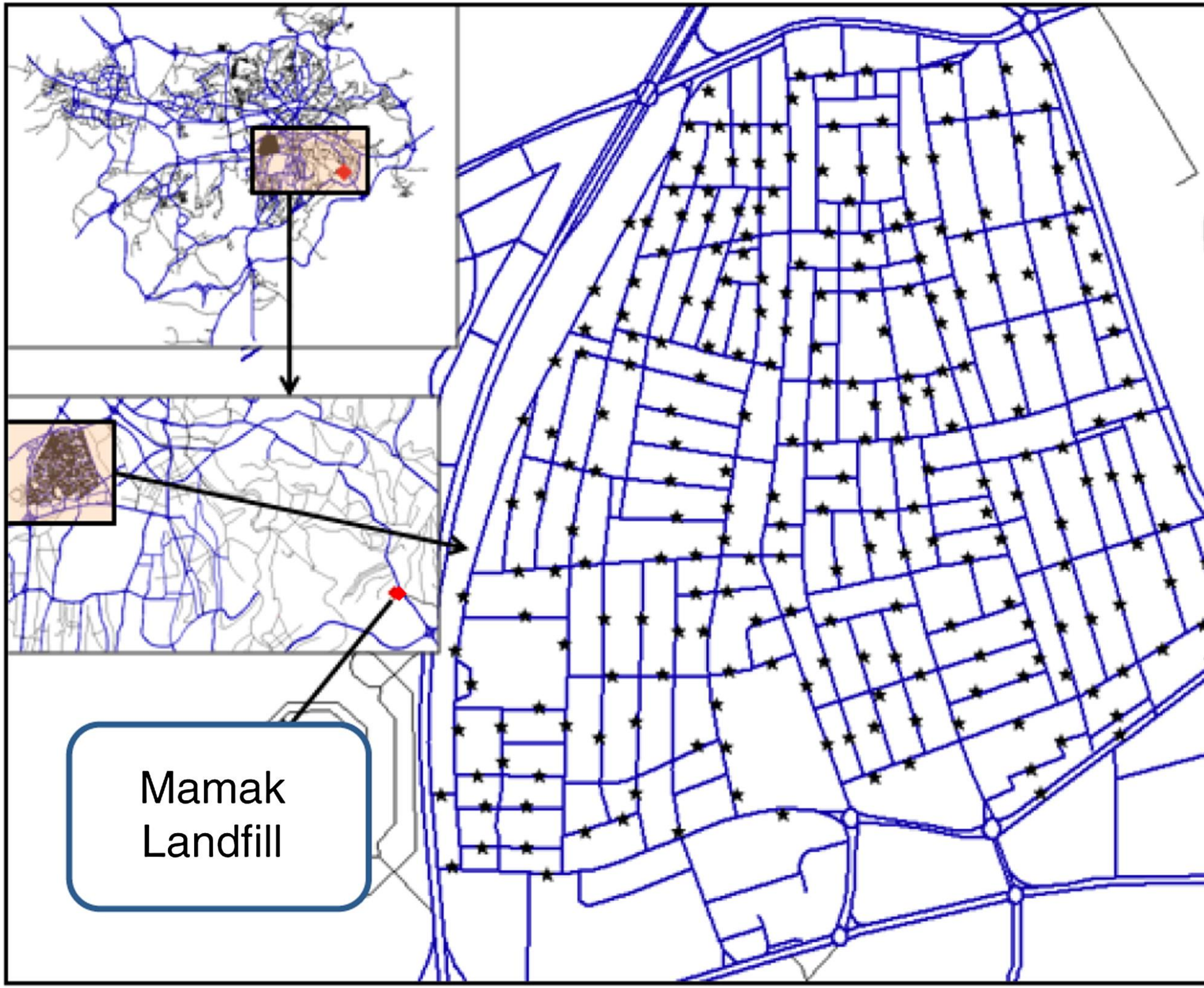
ROAR

Evolutionary Programming



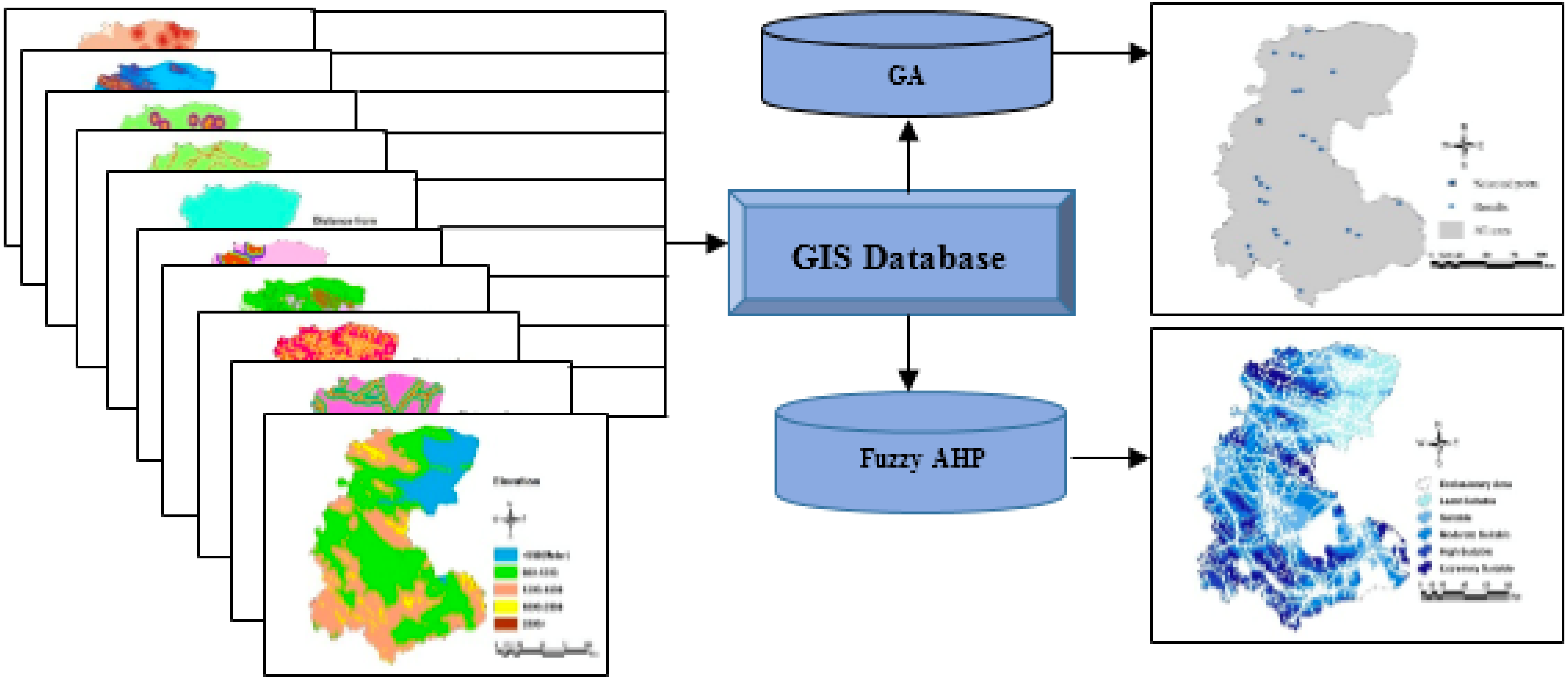
Idaho State University

Municipal solid waste collection routing



Idaho State University

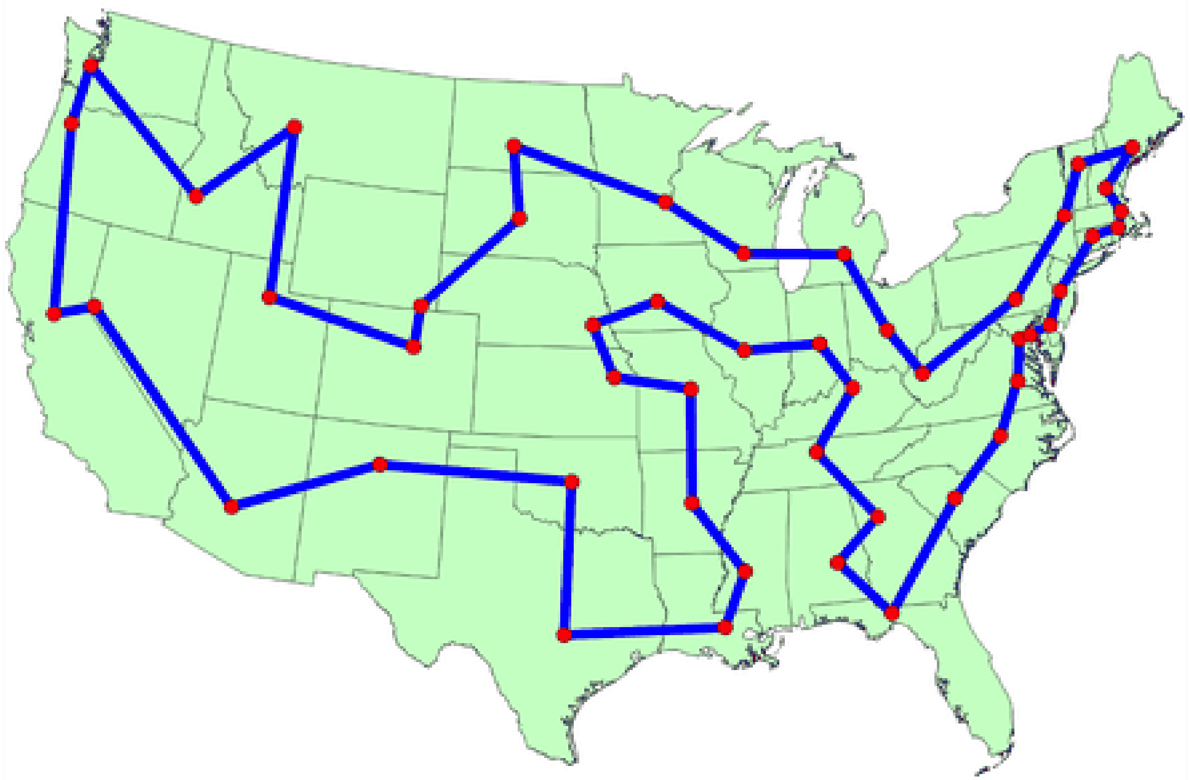
City Planning



Idaho State University

NP-Complete Problems

- Finding optimal solutions in reasonable time is not possible for large instances
- GAs can be applied to find good solutions



ROAR

Q&A

- What problems are you interested in