Evolutionary Algorithms 101

António Leitão

CISUC

Outline

- Evolutionary algorithms
- Biological background
- An application example on a design task
- GAs for hyper-parameterization
- Training a classifier
- Challenges for GA

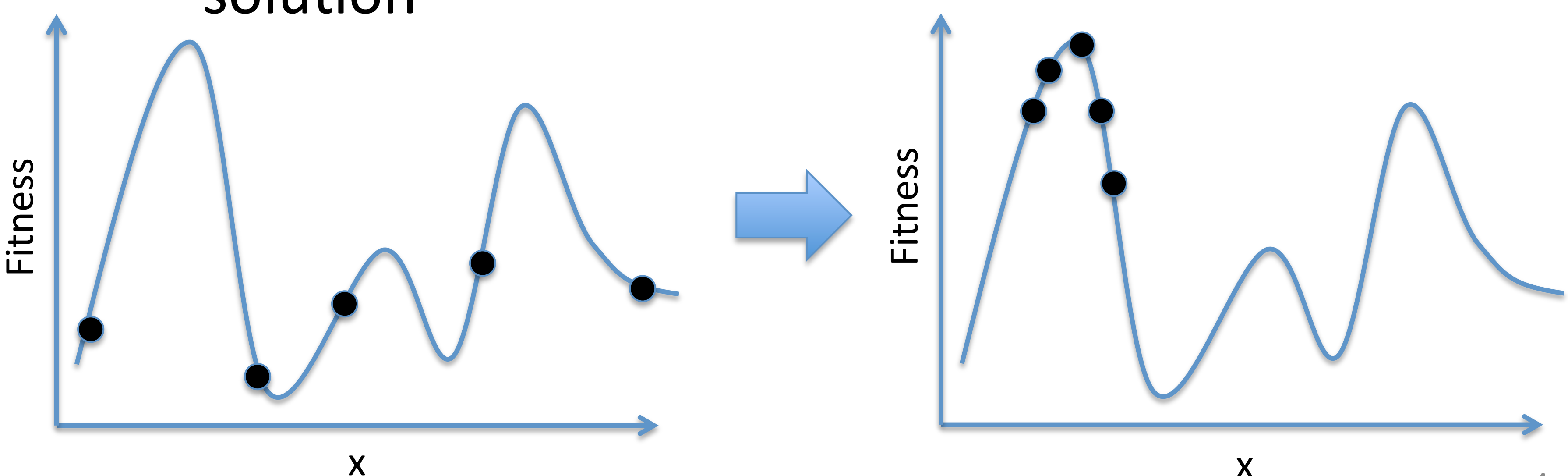
Evolutionary Algorithms

- Nature inspired approach for optimization
  - Darwin’s model on natural selection
  - Mendel’s genetics ideas of inheritance

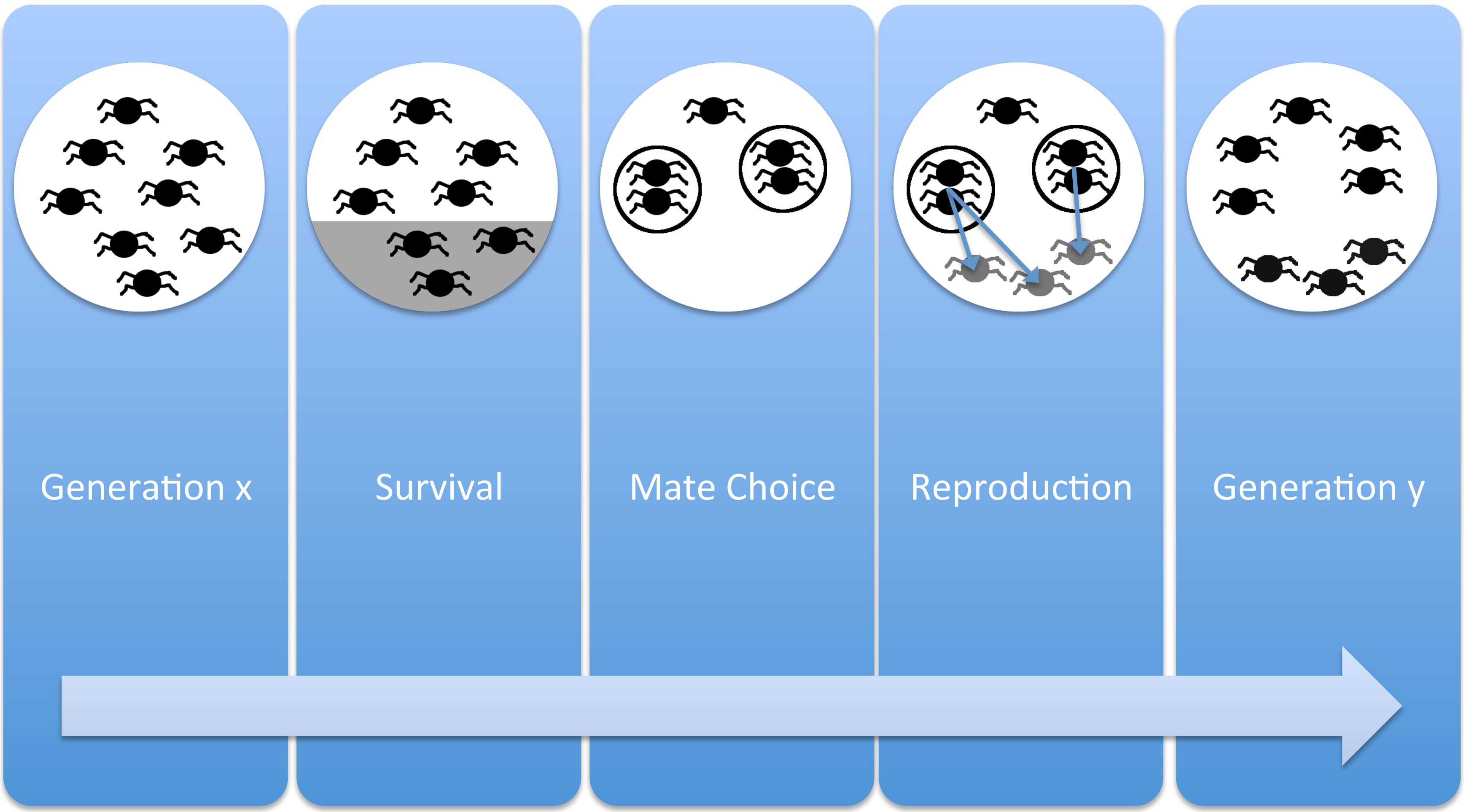
- Different flavors:
  - Most popular ones: Genetic Algorithms and Genetic Programming

Genetic Algorithms

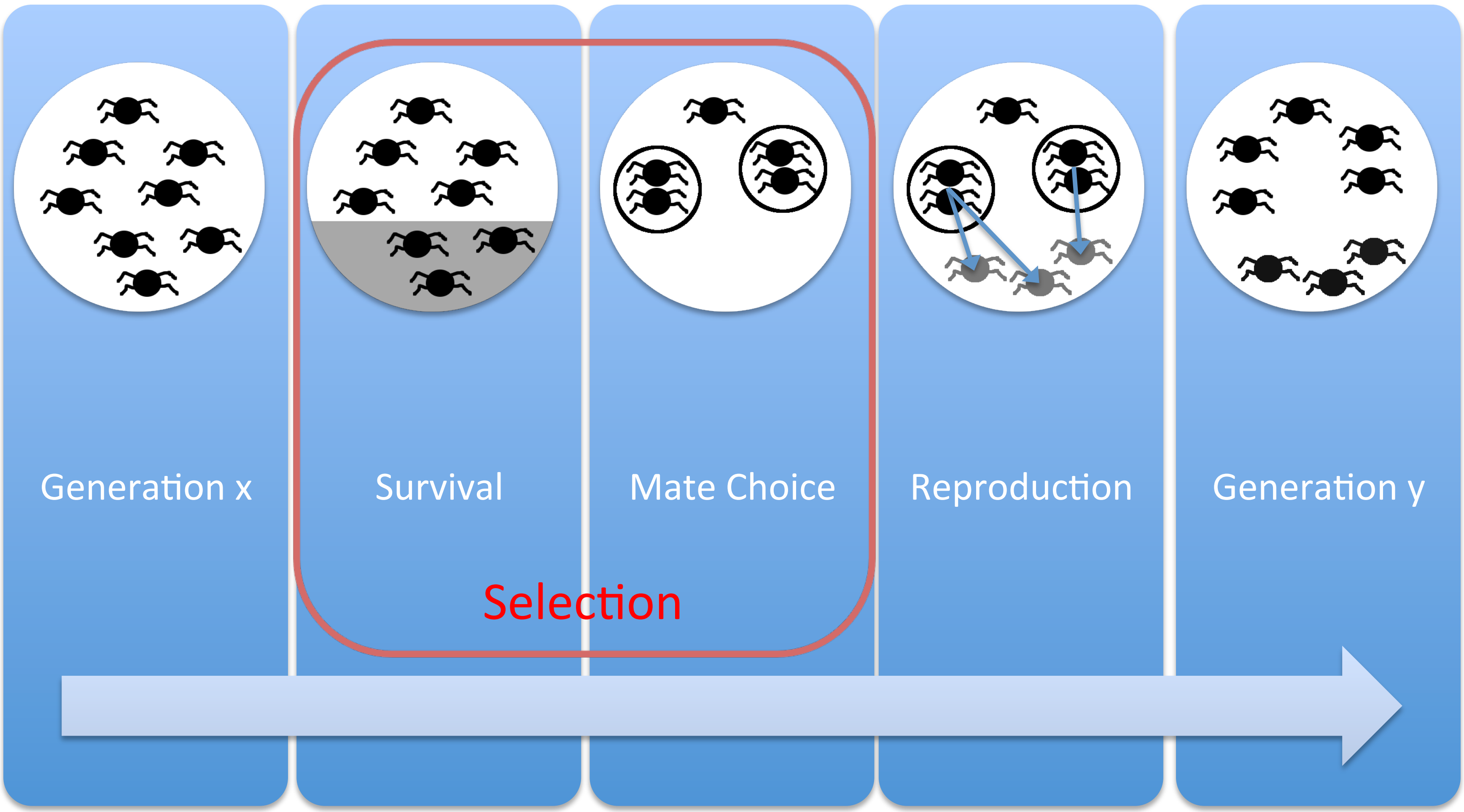
- Guided Search
  - Profit from population dynamics
  - Drift its candidate solutions toward relevant parts of the search space
  - Hopefully find a global optima or approximate solution



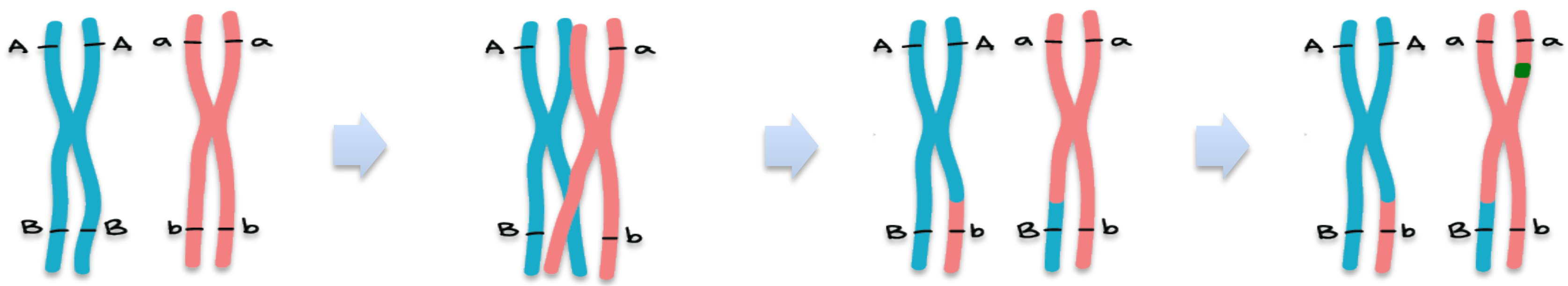
Naive model of evolution



Driving forces of evolution

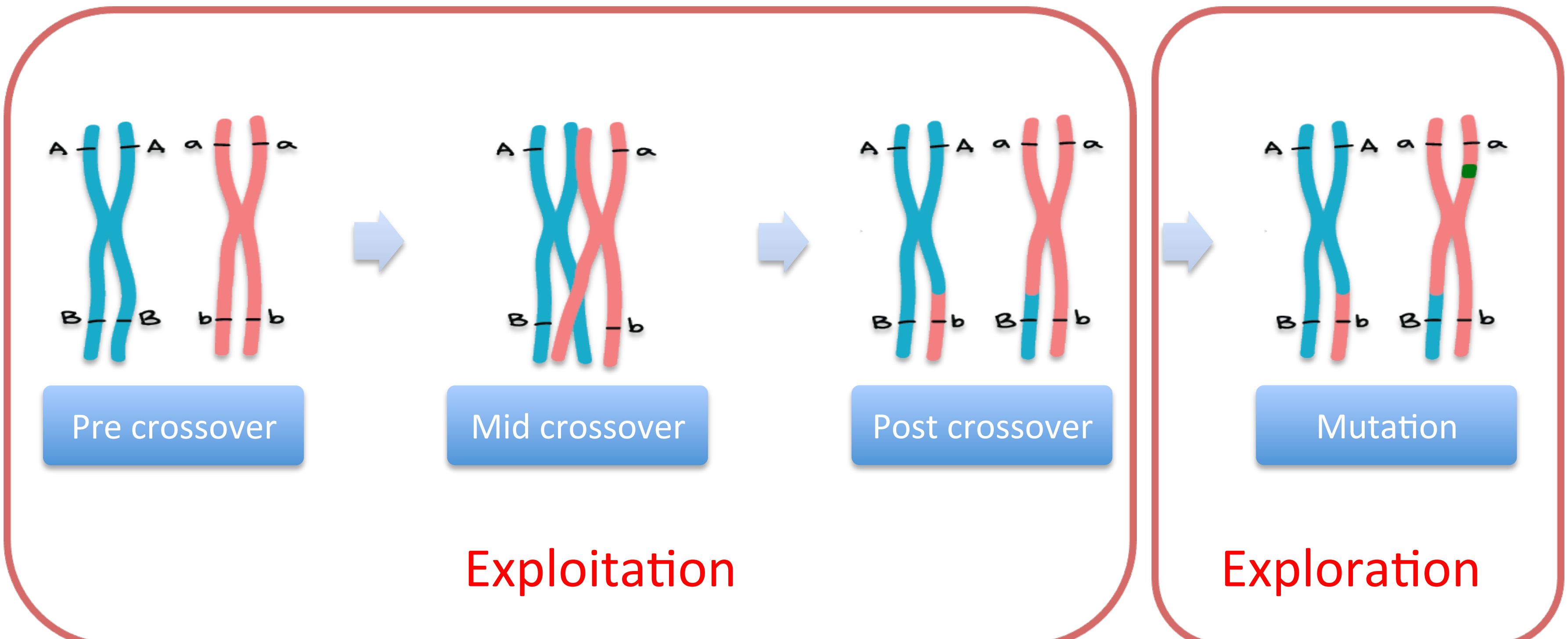


Genetic level

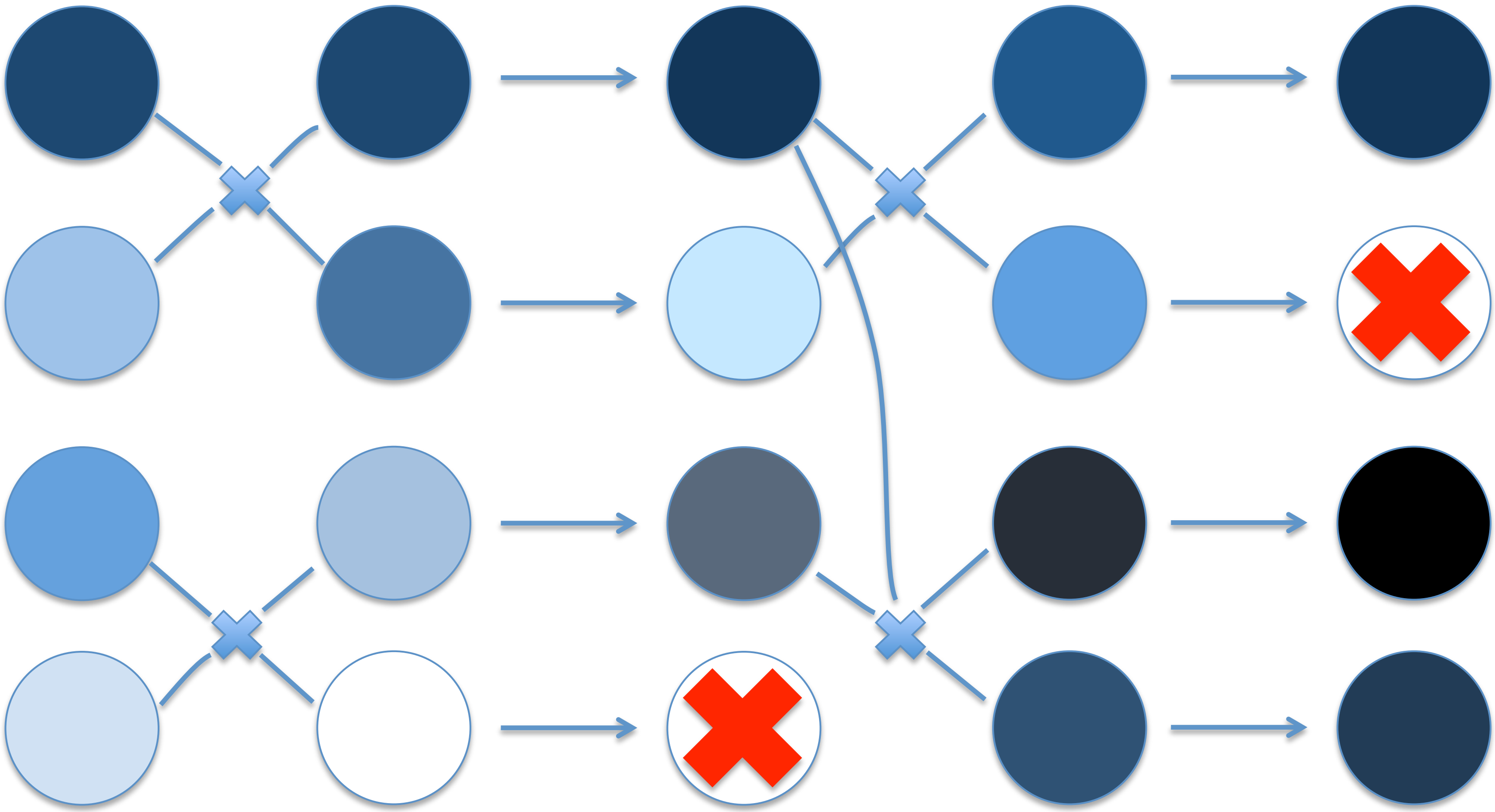


Pre crossover | Mid crossover | Post crossover | Mutation

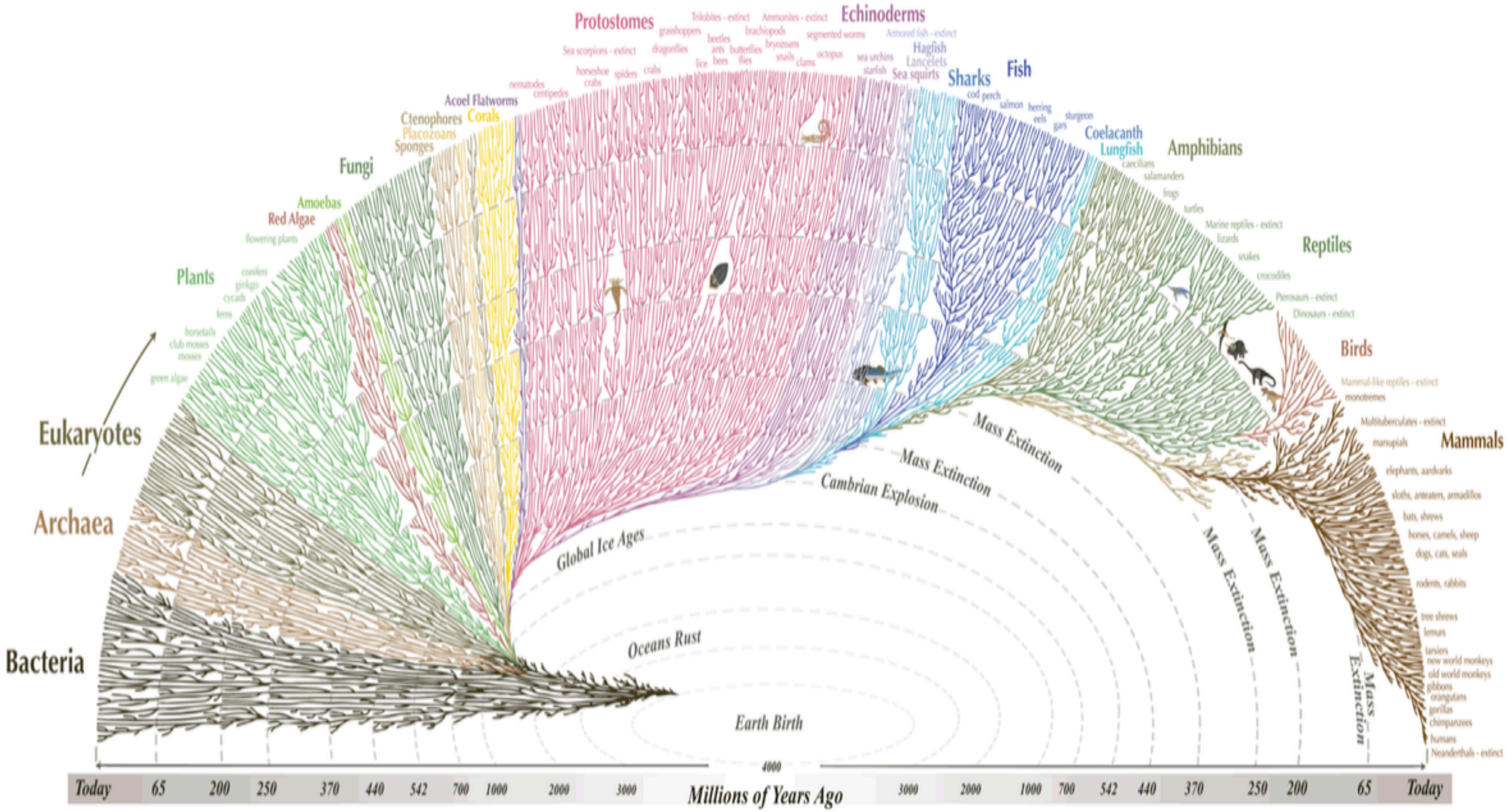
Two relevant behaviors



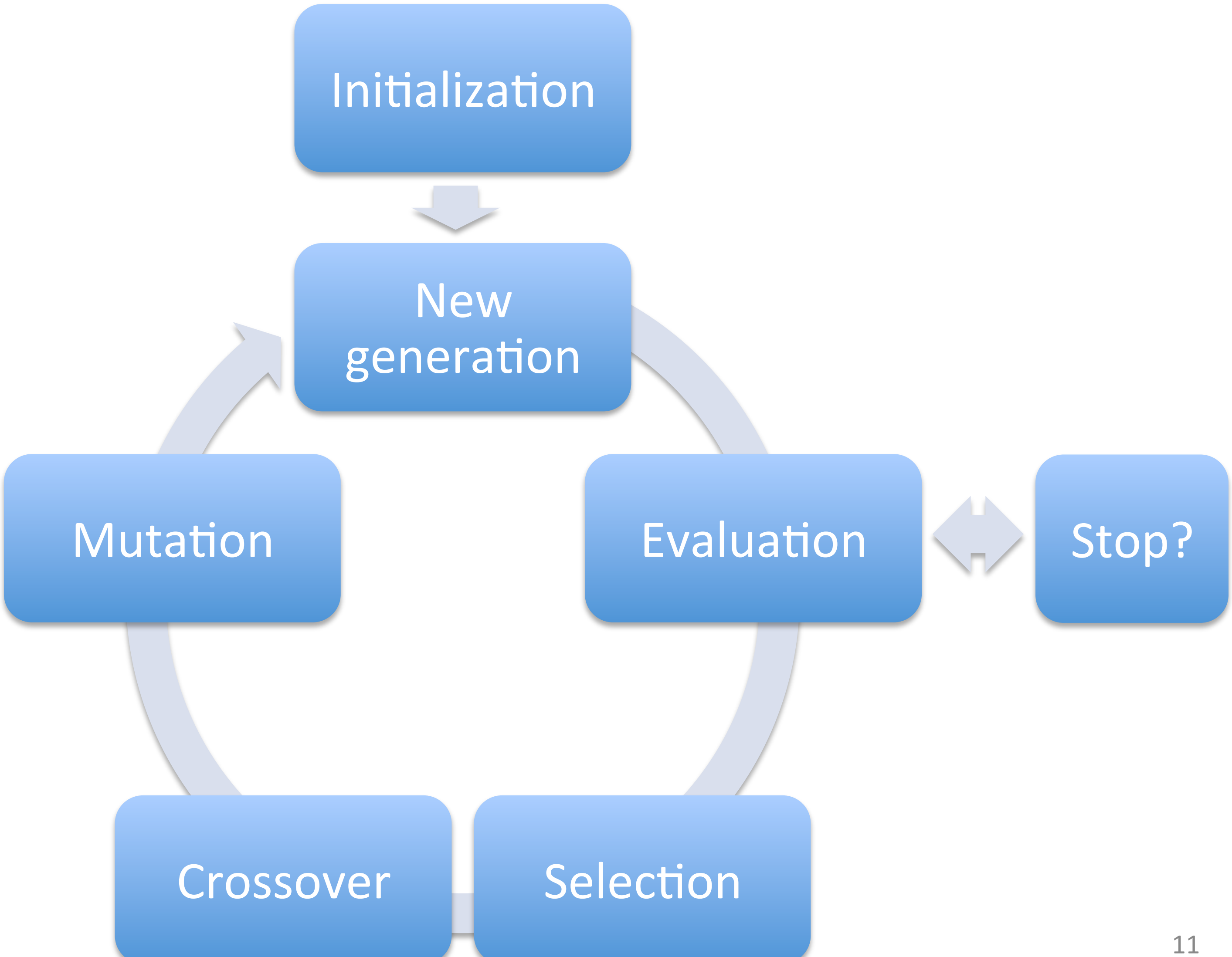
Individual / Population level



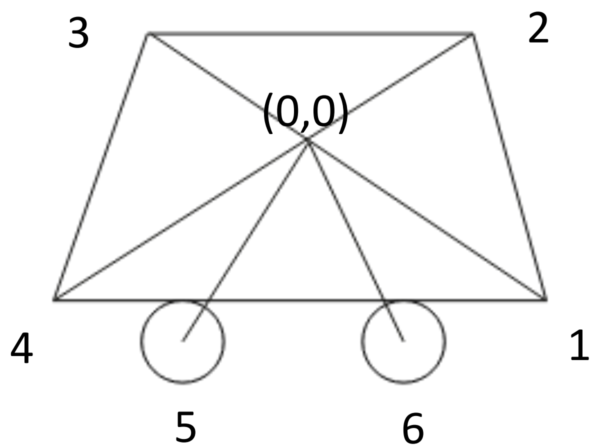
Macro evolution



Genetic Algorithm



Carbox2D example



X1	Y1	X2	Y2	X3	Y3	X4	Y4	X5	Y5	X6	Y6

Ang1	Dist1	Ang2	Dist2	Ang3	Dist3	Ang4	Dist4	Ang5	Dist5	Ang6	Dist6

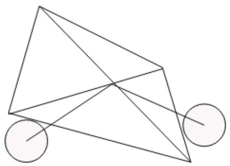
Initialization

- Population
  - 5 individuals

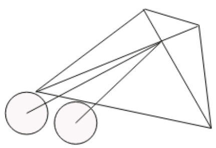
- Each individual
  - Distances: random sample [0,max]
  - Angles: random sample [0,2] pi rad

Ang1	Dist1	Ang2	Dist2	Ang3	Dist3	Ang4	Dist4	Ang5	Dist5	Ang6	Dist6

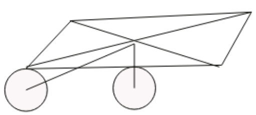
Initialization



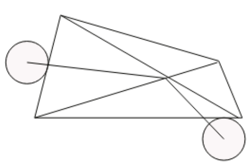
1.0	0.7	1.5	1.3	1.1	1.4	0.5	1.9	0.7	1.0	1.3	0.9



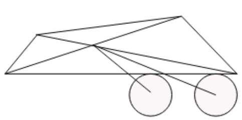
1.3	0.9	1.7	1.1	1.3	0.5	1.7	0.3	0.8	0.8	1.6	0.3



1.4	1.7	1.2	1.4	0.1	1.0	0.9	2.0	1.7	0.6	1.5	1.9



1.3	0.9	1.1	1.1	2.0	1.4	0.2	0.3	0.7	1.1	1.2	0.6

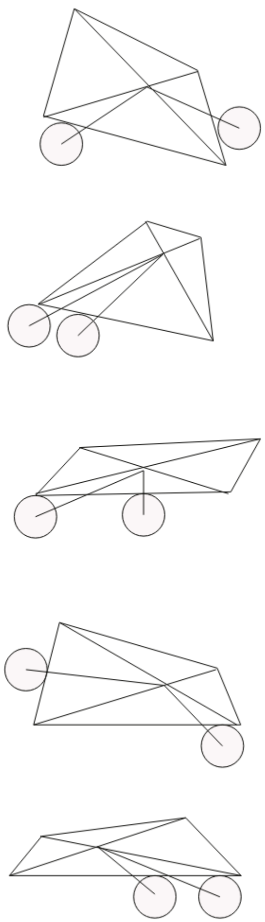


0.3	1.2	0.5	1.3	0.1	1.7	0.8	1.0	1.2	1.1	1.5	0.9

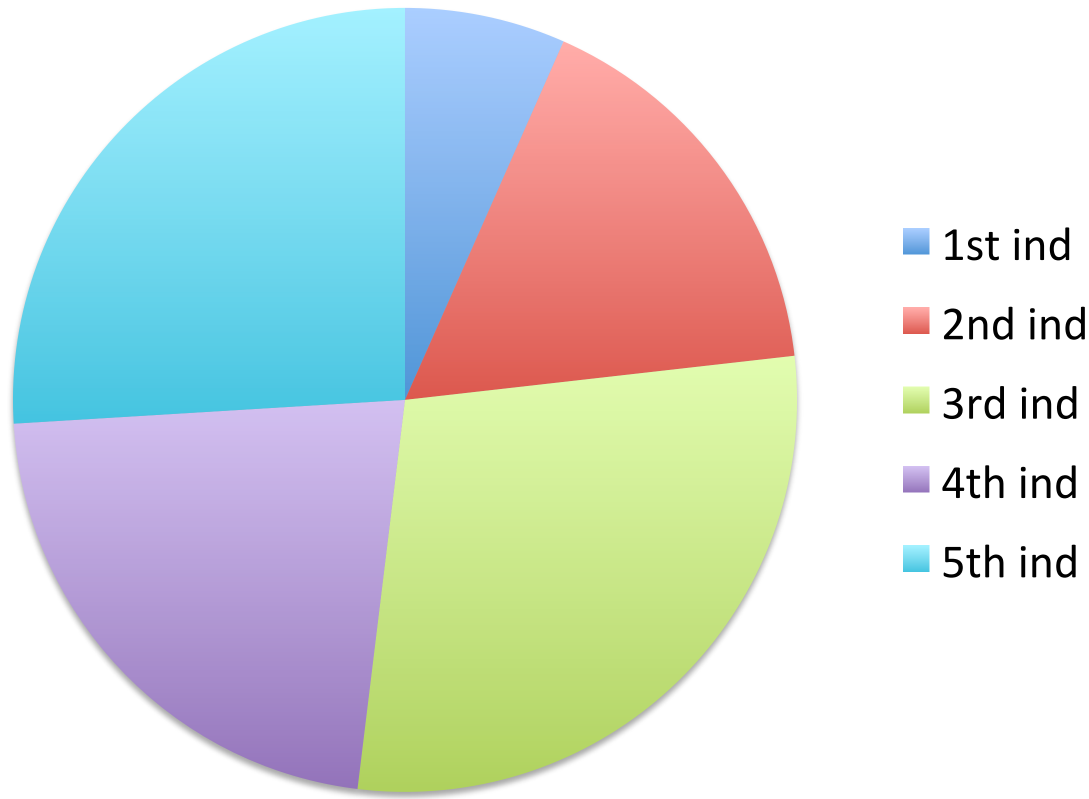
Evaluation

- Multiple possibilities
  - Distance traveled in x seconds
  - Time before stopping
  - Combination of both

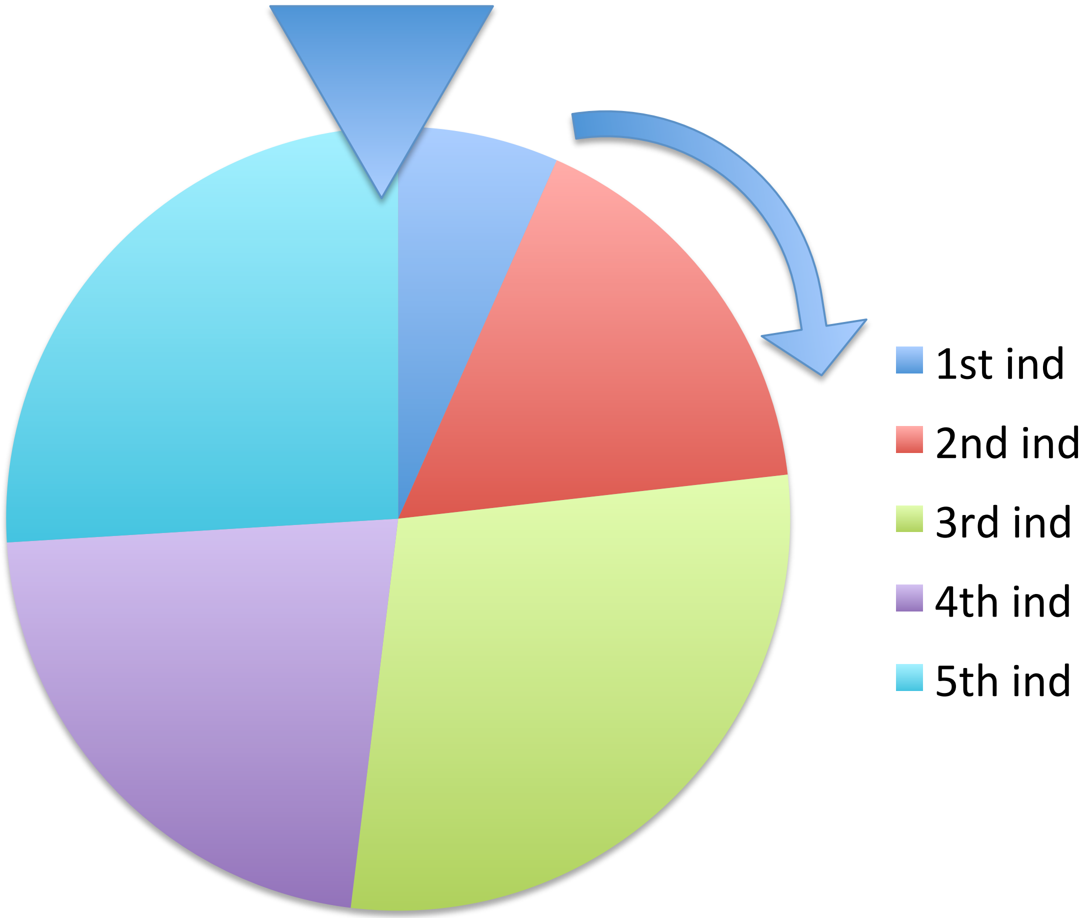
Fitness

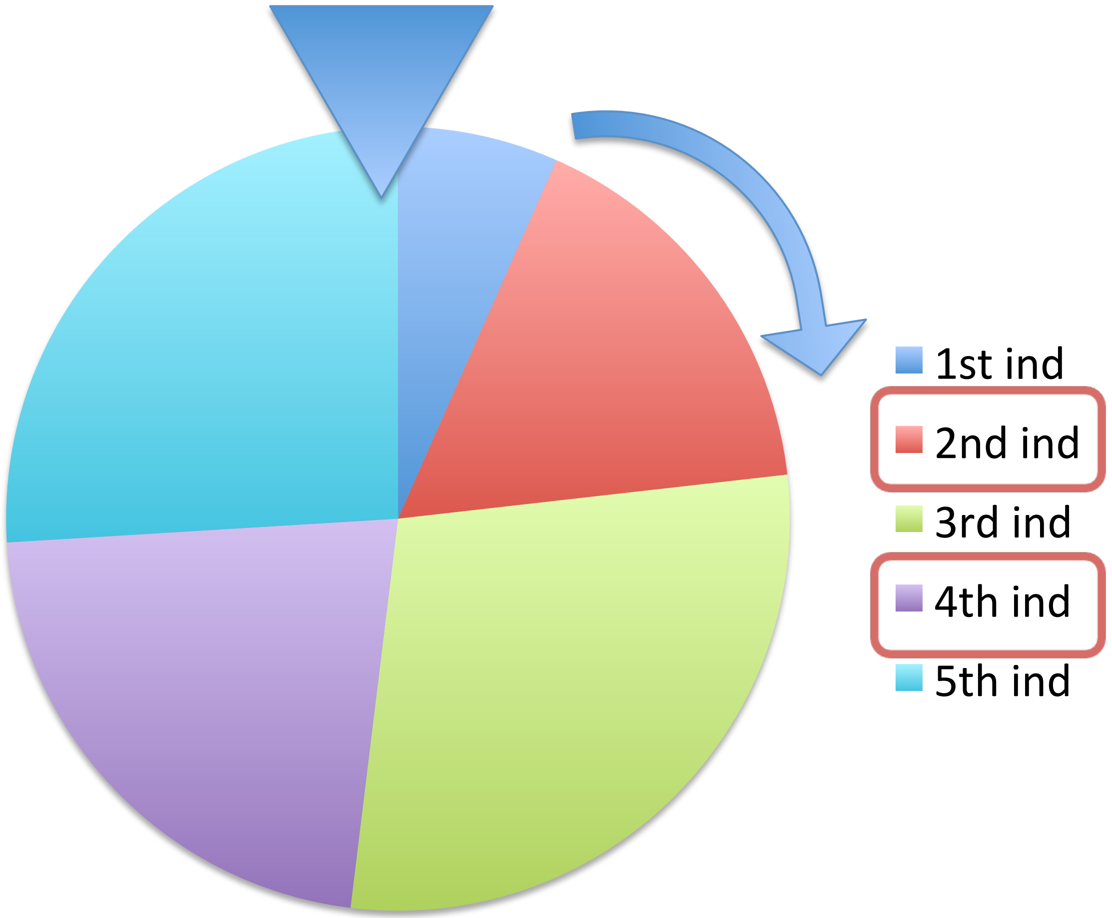


120
300
520
400
470



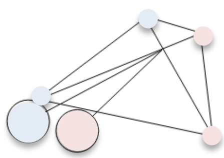
Selection



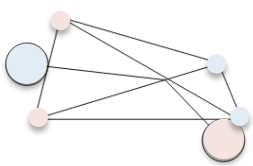

1st ind
2nd ind
3rd ind
4th ind
5th ind

Crossover

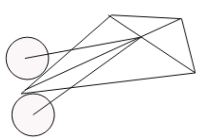
Crossover points



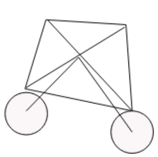
1.3	0.9	1.7	1.1	1.3	0.5	1.7	0.3	0.8	0.8	1.6	0.3



1.3	0.9	1.1	1.1	2.0	1.4	0.2	0.3	0.7	1.1	1.2	0.6



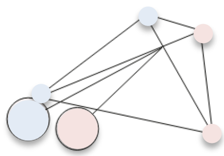
1.3	0.9	1.1	1.1	1.3	0.5	1.7	0.3	0.8	0.8	1.2	0.6



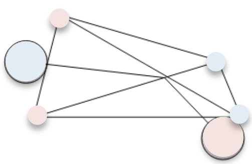
1.3	0.9	1.7	1.1	2.0	1.4	0.2	0.3	0.7	1.1	1.6	0.3

Mutation

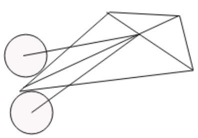
Crossover points



1.3	0.9	1.7	1.1	1.3	0.5	1.7	0.3	0.8	0.8	1.6	0.3

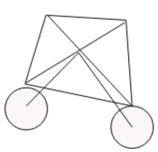


1.3	0.9	1.1	1.1	2.0	1.4	0.2	0.3	0.7	1.1	1.2	0.6



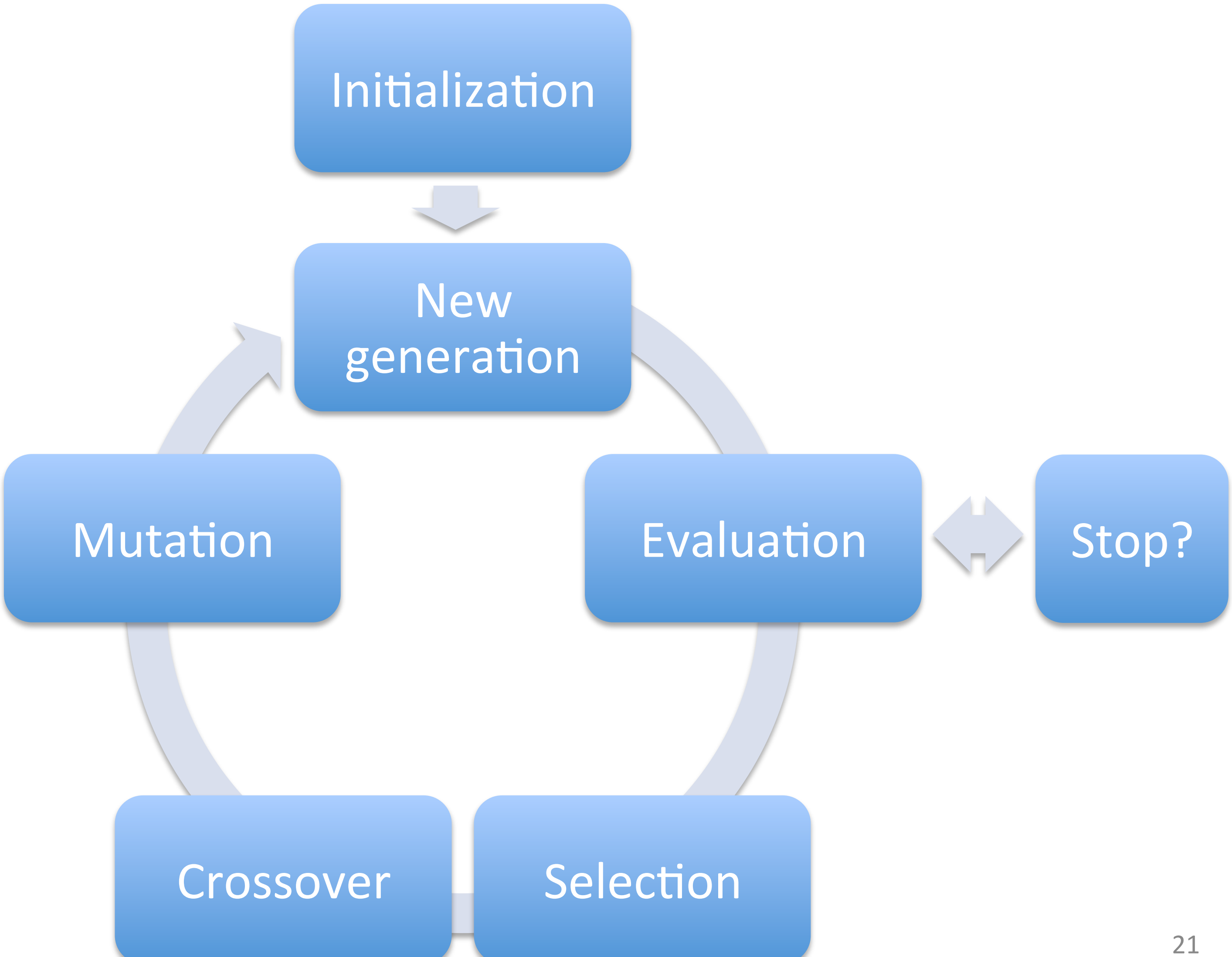
1.3	0.9	1.1	1.1	1.3	0.6	1.7	0.3	0.8	0.8	1.2	0.6

Gaussian mutation



1.3	0.9	1.7	1.1	2.0	1.4	0.2	0.3	0.9	1.1	1.6	0.3

Stop?


Initialization
New generation
Evaluation
Stop?
Selection
Crossover
Mutation

Configuring a NN

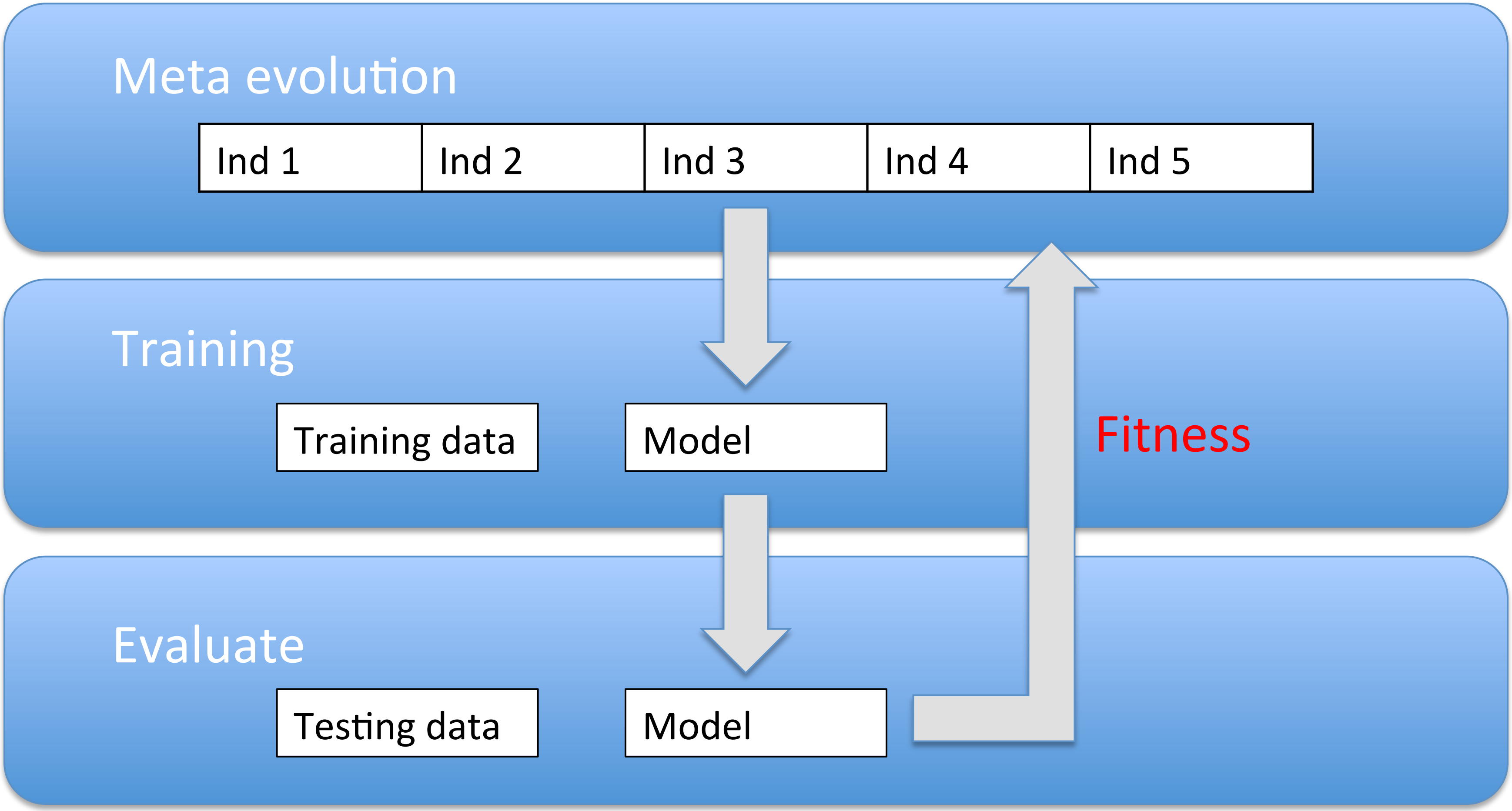
- Various parameters set by
  - Rules of thumb
  - Guidelines
  - Knowledge on the target problem
  - Experience
  - Random initialization

Evolving NN parameters

N neurons in hidden layer
Learning rate
Momentum
Training type
Epoch
Minimum error

Activation function 1
Bias 1
Activation function 2
Bias 2
…
Initial weight 1
Initial weight 2
Initial weight 3
….

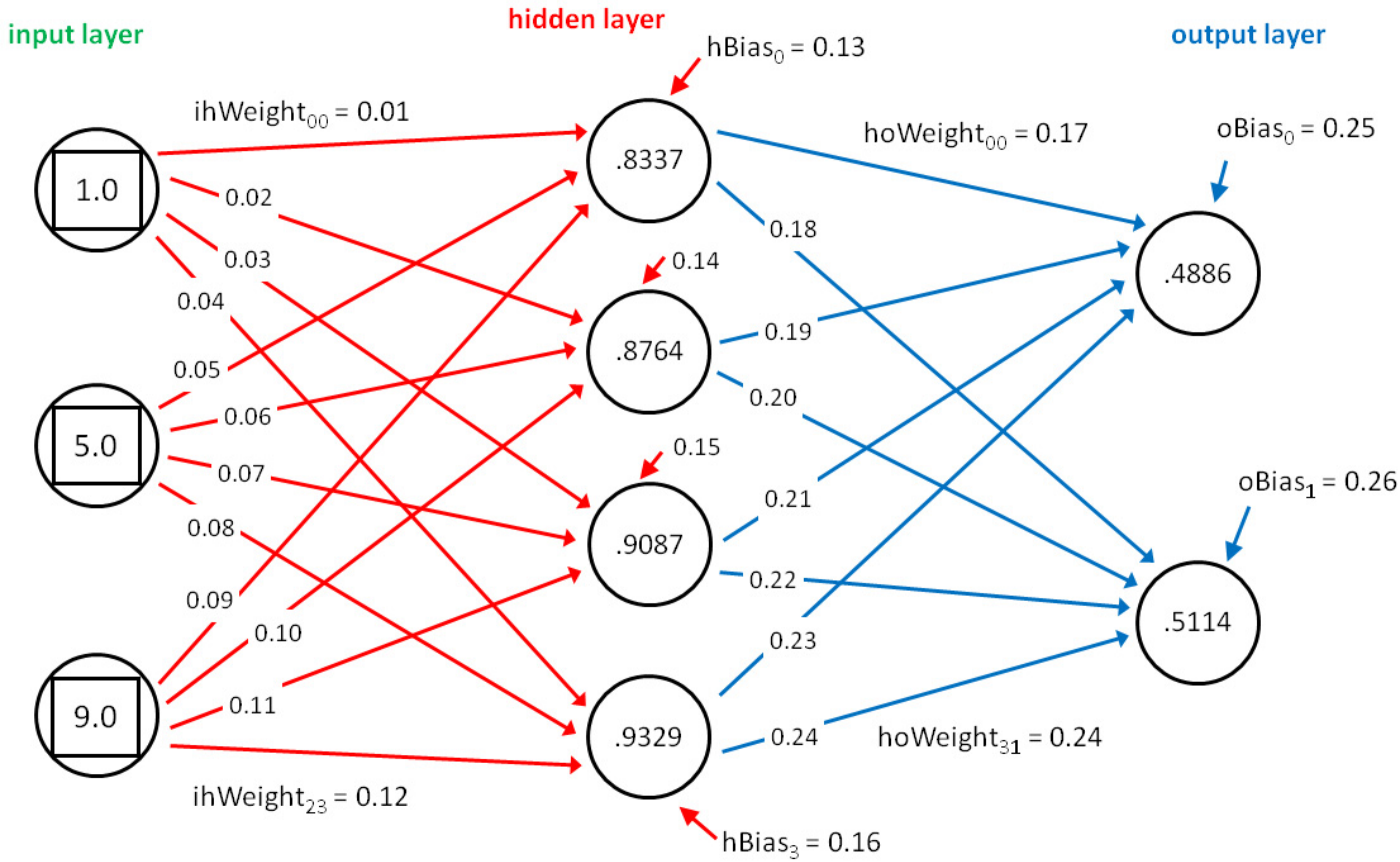
Hyper parameterization



Should we use this?

- When to use
  - Little knowledge of the problem / data
  - As a starting point to find good parameters
- Downsides
  - Computationally expensive
  - Search could be wider than needed
- Tip
  - Use reason to restrict genes to appropriate intervals
    - Customize initialization and variation operators

Training NN with GAs



GAs for small NNs

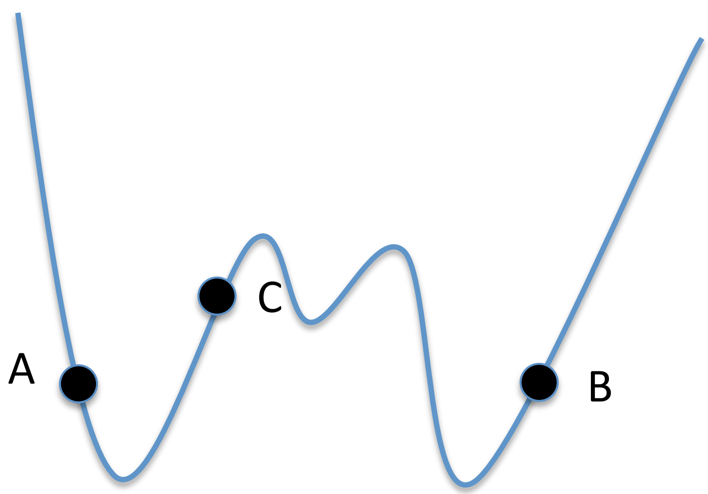
- Reported results are competitive
- Gas are capable of
  - Maintaining sturdiness through population size
  - Exploit genetic material through crossover
  - Explore the search space through mutation
  - Depend little on initial conditions to approximate global optima

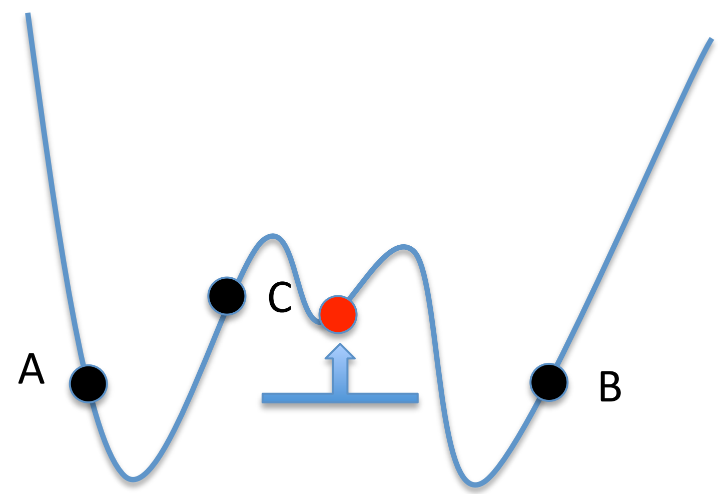
What about large NNs?

- NNs with huge number of inputs, dealing with large data widen the search space
- Very large populations may be prohibitive
  - Due to computational effort
- There's a chance of losing appropriate coverage of the search space
  - Individuals become sparse
  - Risking beneficial exploitation

Destructive crossover

A	B	C
50	100	50





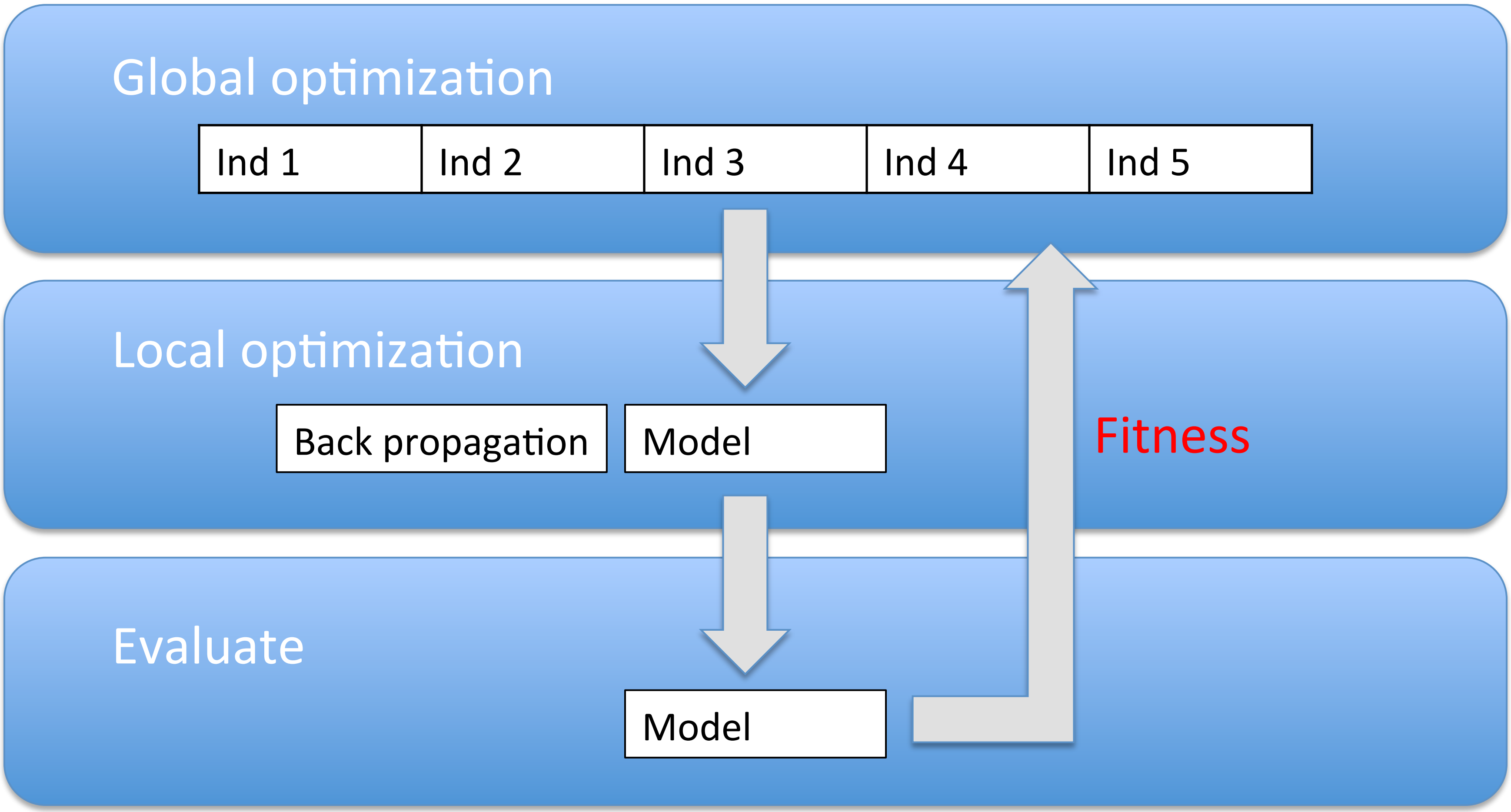
Challenges for GA

- Can we offset the need for computational effort?
  - Island Models (parallelization)
  - Fitness sharing
  - Restricted Mating
  - Fitness scaling or transformation
  - Local search methods

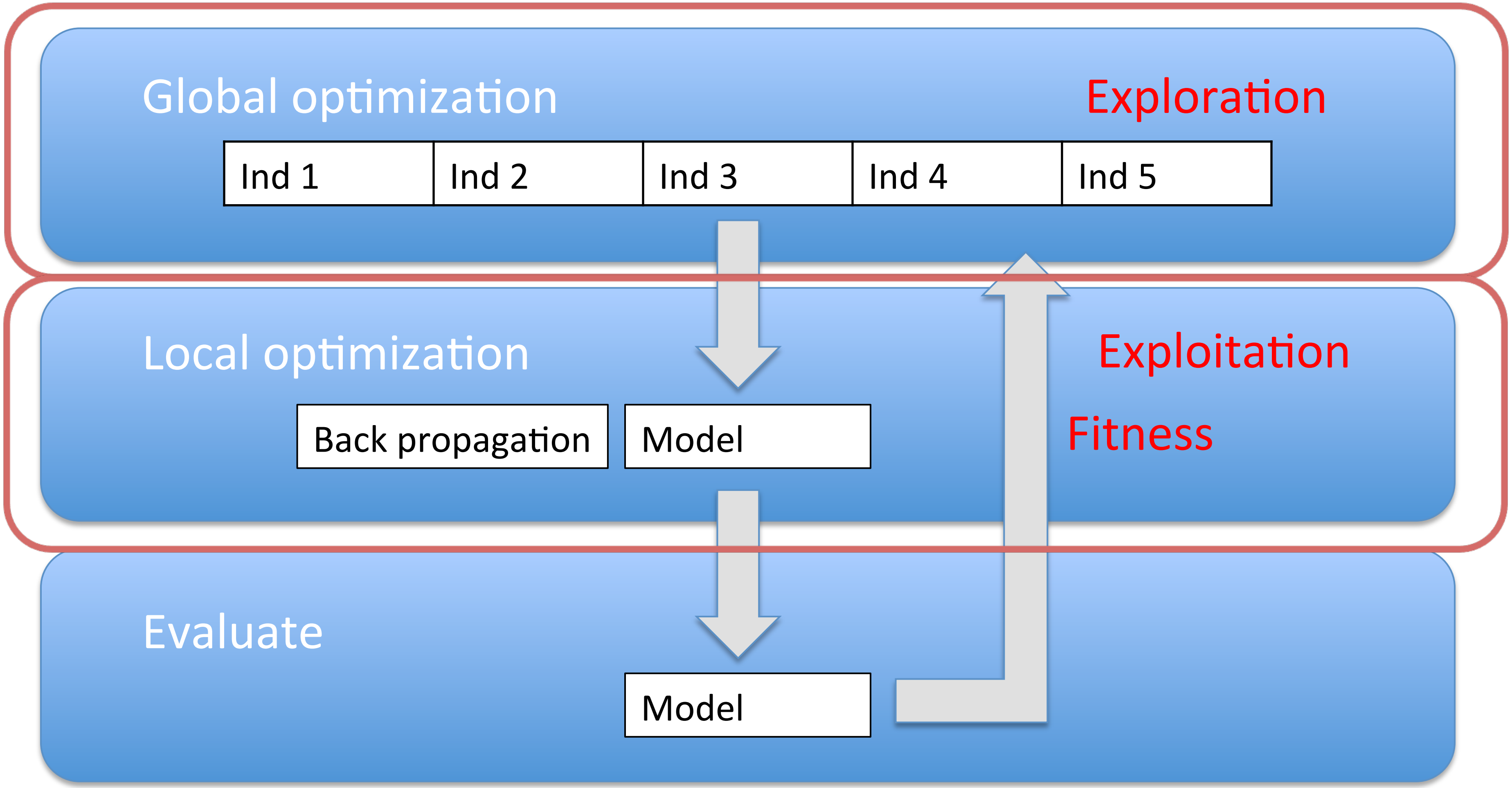
Gradient methods vs Population Dynamics

- Gradient based methods use information on expected output
  - Adjust the weights in the right direction
- GAs rely on population dynamics and fitness information
  - Drifting genes toward seemingly good areas of the search space

Combining GAs with local search



Combining GAs with local search



Conclusions

- GAs have proven valuable on various classification tasks (mostly academic)
- Large data problems present a number of new challenges
- Further studies are needed to assess its usefulness