

# Application of Computer in Chemistry

## SSC 3533

## GENETIC ALGORITHMS

Prof. Mohamed Noor Hasan

Dr. Hasmerya Maarof  
Department of Chemistry



# Introduction

- Searching algorithm based on the process of biological evolution
- Developed by John Holland, University of Michigan (1970' s)
  - To understand the adaptive process in the natural system
  - To design an artificial software system that retain the robustness of natural system
- Effective method for optimization and machine learning
- Widely used in business, sciences and engineering

# GA Procedure

## 1. Population



## 2. Evaluation



## 3. Reproduction



## 4. Mutation

Generate the first generation using random parameters

Evaluate all individuals in the generation using the fitness criteria

Combine the best individuals to form the next generation

Perform mutation in the new generation

Repeat steps 2,3 and 4

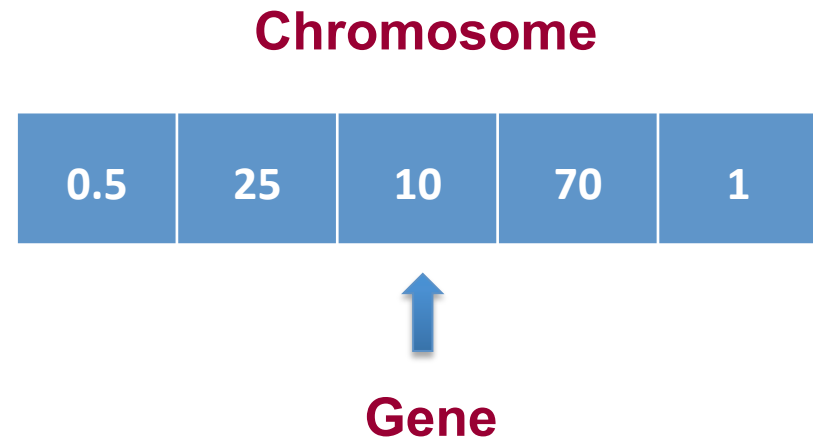
# Evaluation

- Giving fitness value which is a measure of success of a process, experiment, analysis, etc.
- Examples:
  - Analysis time, e.g. in chromatographic separation
  - Yield, e.g. in synthesis
  - SSE,  $R^2$ , e.g. in model development
  - % correct classification, e.g. in pattern recognition

# Coding

The chromosome might consist of

- Bit strings  
(0101 ... 1100)
- Real numbers (43.2  
-33.1 ... 0.0 89.2)
- Lists of rules  
(R1 R2 R3 ... R22 R23)
- Any data structure ...



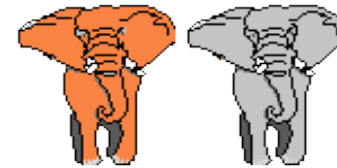
# Reproduction

- Parents are selected randomly based on the fitness values they get
- Combine the genes – crossover
- This step should produce better offspring (based on fitness value)

# Crossover

A 101001**01**

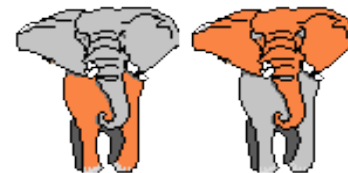
B 100110**00**



Crossover

A\* 101001**00**

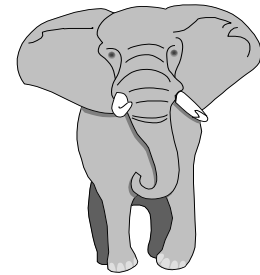
B\* 100110**01**



# Mutation

Before

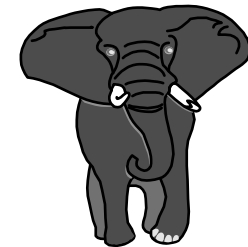
**10100100**



Mutation

After

**11100100**





# Example

The Traveling Salesman Problem:

Find visiting schedule to a number of cities so that:

- Each city is visited only once
- Number of travels made is minimum

# Representation

Representation is an ordered list of city numbers known as an *order-based GA*.

- |           |              |            |             |
|-----------|--------------|------------|-------------|
| 1) London | 3) Dunedin   | 5) Beijing | 7) Tokyo    |
| 2) Venice | 4) Singapore | 6) Phoenix | 8) Victoria |

CityList1 (3 5 7 2 1 6 4 8)

CityList2 (2 5 7 6 8 1 3 4)

# Crossover

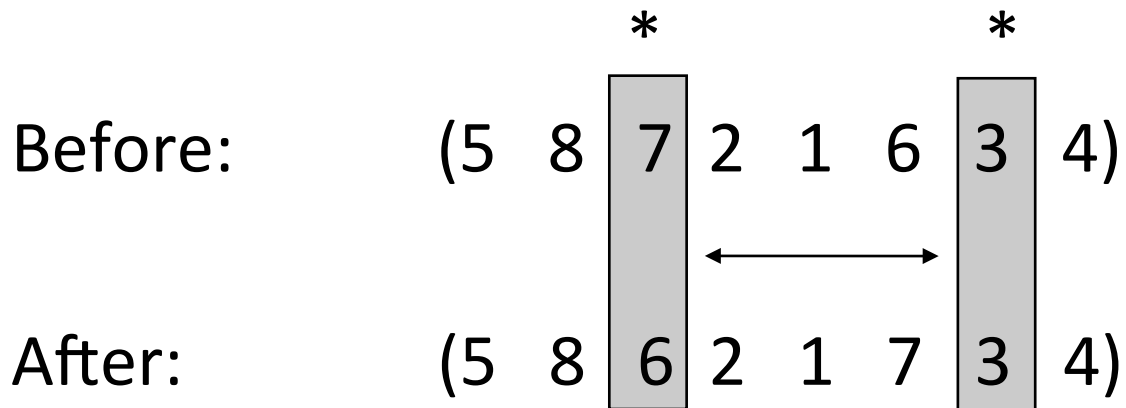
Crossover combines inversion and recombination:

	*	*							
Parent1	(	3	5	7	2	1	6	4	8)
Parent2	(	2	5	7	6	8	1	3	4)
Daughter	(	5	8	7	2	1	6	3	4)

This operator is called the *Order1* crossover

# Mutation

Mutation involves reordering of the list:



# Variable selection in regression

## Problem

Many variables can be used to develop a model.  
Choose the best 5 out of 10 variables

## Step 1

Code the variables into chromosome

1010010101

Value of 1 means selected

Value of 0 means not selected

# Variable selection (cont.)

## Step 2

Build regression model and use SSE or std. error as fitness criterion

## Step 3

Select a number of good models (lowest SSE)  
perform crossover to form a new generation

1010010101

1000011101

## Step 4

Perform mutation randomly to form a new generation

1000011100

# Other Examples

- Finding optimum conditions of an experiment
  - For example, finding the optimum conditions in an HPLC separation of a complex mixtures.
  - Parameters: pH, temperature, polarity of mobile phase, etc.
  - Fitness criteria: time of analysis, resolution
- Finding molecular models with lowest energy
  - With GA the process of finding lower energy conformation can be accelerated.
- Finding variables that form classes in pattern recognition
  - Similar to variable selection in regression analysis

# Advantages of GA

- Concept is easy to understand
- Modular, separate from application
- Support multi objects optimization
- Good for noisy data
- Always has an answer; the answer becomes better
- Many ways to accelerate solution when more information learned about the problem



# When to use GA

- Other methods are too slow or too complicated
- Need to explore new approaches to solve problems
- Similar problems can be solved using GA
- To be combined with existing methods