# An Introduction to GA and EDA

GA – Genetic Algorithm
EDA – Estimation of Distribution Algorithm

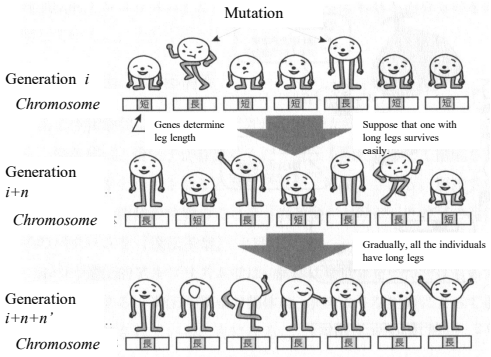
1

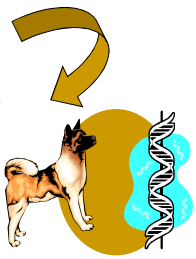
## Bioinspiration

- An example of evolution

Characteristics of living things are determined by genes

Evolution gives inherent characteristics and Functions



Evolution is realized the following components

- selection
- crossover
- mutation

2

# Concept of evolution

- Three evolution operators

Selection

Those who fit into an environment survive, otherwise die out.

Crossover

Genes are exchanged between two individuals

⇒ New individuals are produced

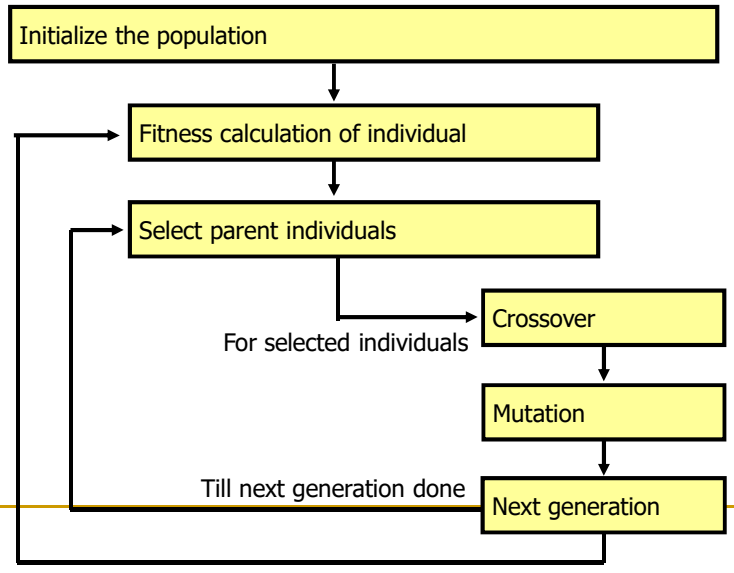
Mutation

Some of the genes of the selected individuals are changed to other ones

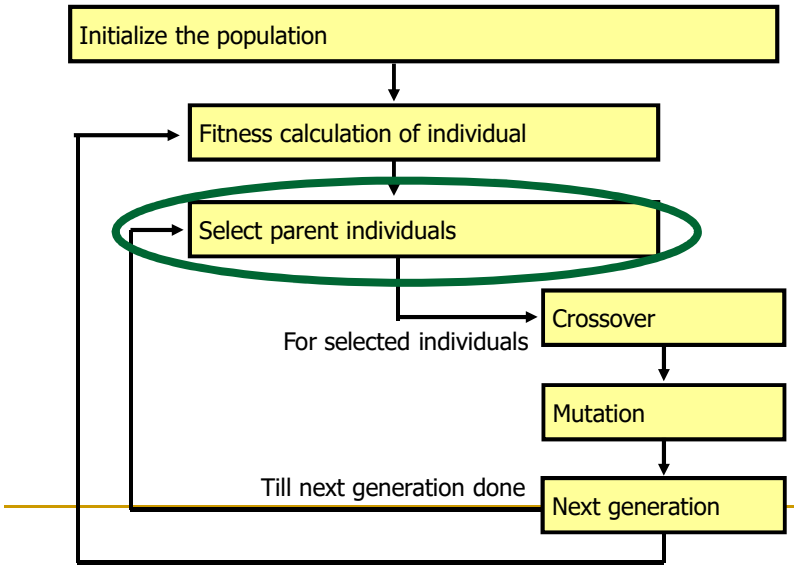
⇒ New individuals are produced

3

# Flowchart of Genetic Algorithm



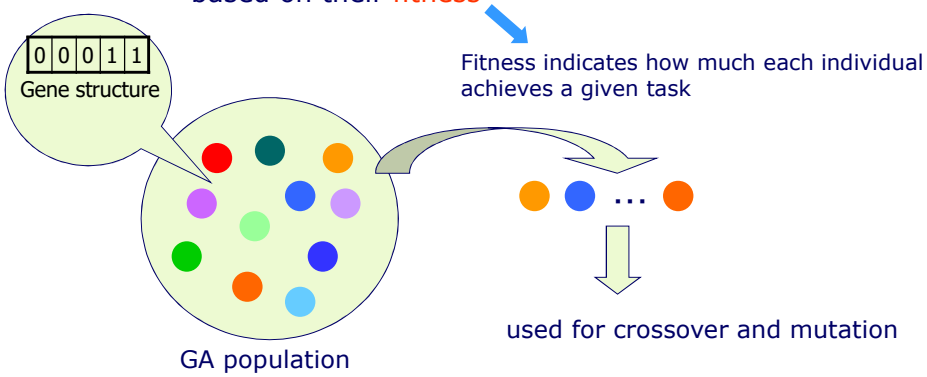
# Flowchart of Genetic Algorithm




# Concept of selection in Genetic Algorithm

Select good individuals (genes) from the population based on their fitness

Fitness indicates how much each individual achieves a given task



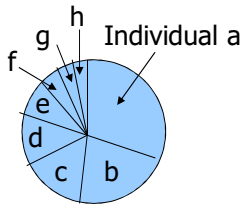

6

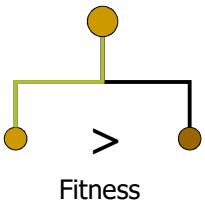
# Selection

How to select individuals based on the fitness

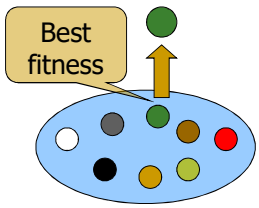
There are many kinds of selection methods, and each designer choose appropriate methods for each problem.

- Roulette Selection
- Tournament Selection
- Elite Selection




Fitness




7

# Selection (cont'd)

- **『Fitness Proportional Strategy』**

  It first calculates the probability of being selected as a parent in proportion to its fitness, and then randomly selects individuals according to that probability.

  Let's denote the number of individuals by *n*, the fitness value of individual *i* by $f_i$. The probability of individual *i* to be selected, $p_i$ is defined by

$$p_i = \frac{f_i}{\sum_{j=1}^{n} f_j}$$

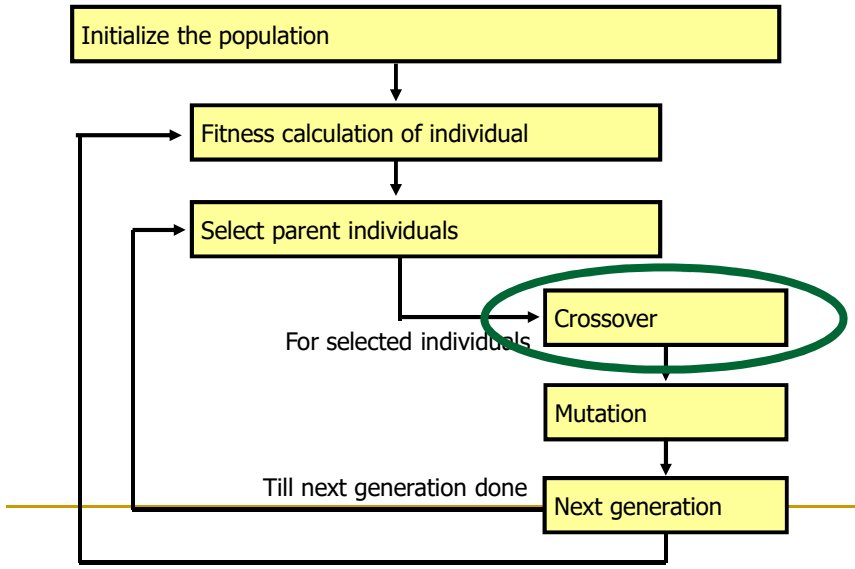
- **『Elite preservation strategy』**

  It tries to replicate the individuals with the highest fitness values to the next generation. Doing this way has the advantage that the best solutions at that time are selected and not destroyed by mutation and crossover. The elite strategy is usually used with other selection strategies.

- **『Tournament strategy』**

  Randomly select multiple individuals and leave the one with the highest fitness as the parent. The number of individuals to play a tournament is generally 2, but may be set higher.

# Flowchart of Genetic Algorithm




# Crossover

Crossover point is selected based on the probability (crossover rate $P_c$) and the selected genes are exchanged.

individual 1 (parent 1)

| 1 | 1 | 1 | 1 | 0 |
|---|---|---|---|---|

individual 1 (parent 2)

| 1 | 0 | 1 | 0 | 0 |
|---|---|---|---|---|

10

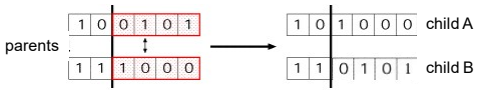
# Crossover (cont'd)

- Crossover

Once the individual pair to be selectively mated is determined, the chromosomes are crossed. Crossover basically takes a part of both chromosomes at a certain crossing position for an individual selected and makes a offspring. However, at certain locus, the gene of either parent is replicated from the same locus.

- **『1 Position Crossover』**

The simplest method is to select one intersection and swap the genes before and after. This is called one-point crossing.

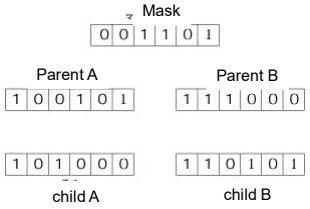



- **『Multiple Position Crossover』**

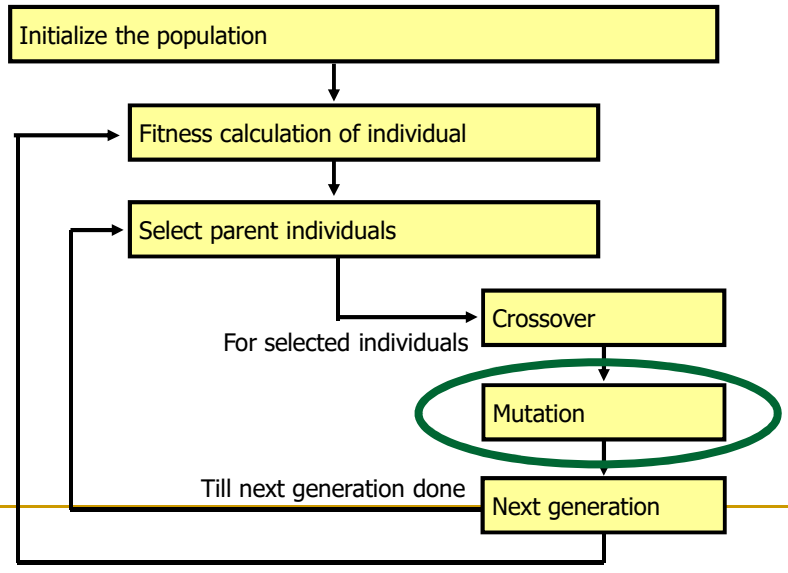
The one-position intersection had one intersection, but the multi-position crossover is made up of multiple positions.

- **『Uniform Crossover』**

Use a mask prepared in advance (generally a random bit string). At the position where the mask pattern is 0, the gene of parent A is given to child A; At position 1, it copies the gene for parent B. For child B, do the opposite of this. Uniform crossover can be thought of as a type of multi-position crossover.




# Flowchart of Genetic Algorithm



# Mutation

Mutation genes are selected based on the probability (mutation rate $P_m$) and the selected genes are randomly changed.

individual (parent)

| 1 | 0 | 1 | 1 | 0 |
|---|---|---|---|---|

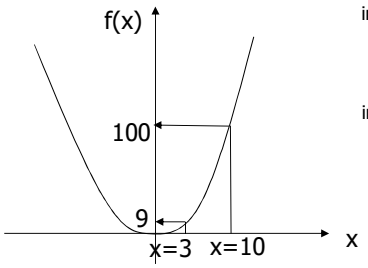
13

# Genetic Algorithm (GA): an simple example

- gene expression
  - binary code (0 and 1)

| 0 | 0 | 0 | 1 | 1 |
|---|---|---|---|---|

An example: the aim of GA is to find a maximum value of $f(x)=x^2$ ( $0 \le x \le 31$ )




individual 1

| 0 | 0 | 0 | 1 | 1 |
|---|---|---|---|---|

x=3

fitness=$3^2$=9

individual 2

| 0 | 1 | 0 | 1 | 0 |
|---|---|---|---|---|

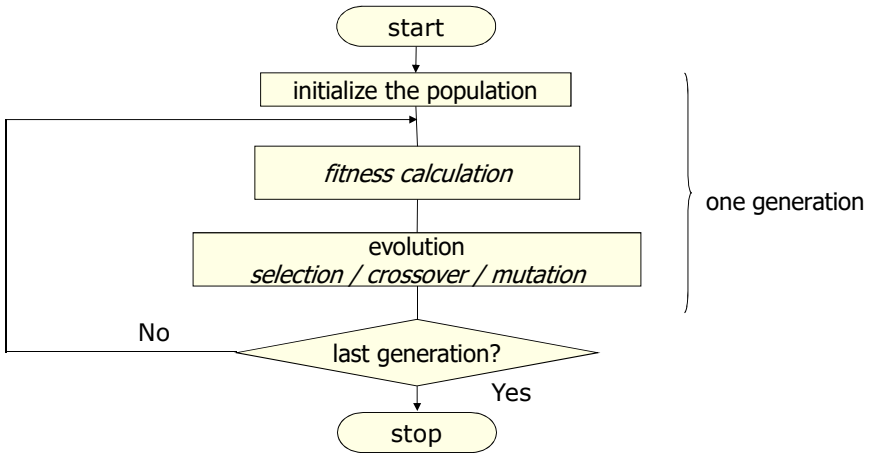
x=10

fitness=$10^2$=100

better individual

14

# Flowchart




15

# The flow of GA

- initialize the population
  - determine the number of individuals
    - Here, five individuals for simplicity

| individual number | gene (genetic code) |
|---|---|
| 0 | 01100 |
| 1 | 10110 |
| 2 | 00011 |
| 3 | 10001 |
| 4 | 00010 |

0 or 1 are determined randomly

16

# The flow of GA

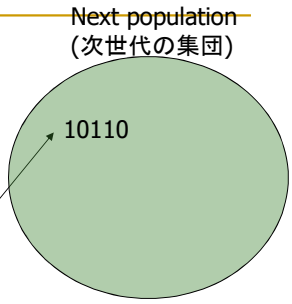
- fitness calculation
  - fitness is "$x^2$" in this example

| individual number | gene | fitness |
|---|---|---|
| 0 | 01100 | ($12^2$=) 144 |
| 1 | 10110 | ($22^2$=) 484 |
| 2 | 00011 | ($3^2$=) 9 |
| 3 | 10001 | ($17^2$=) 289 |
| 4 | 00010 | ($2^2$=) 4 |

17

# The flow of GA

- genetic operation
  - selection
    - elite selection
    - tournament selection




copy

elite individual
(the best one)

| individual number | gene | fitness |
|---|---|---|
| 0 | 01100 | ($12^2$=) 144 |
| 1 | 10110 | ($22^2$=) 484 |
| 2 | 00011 | ($3^2$=) 9 |
| 3 | 10001 | ($17^2$=) 289 |
| 4 | 00010 | ($2^2$=) 4 |

18

# The flow of GA

- genetic operation
  - selection
    - elite selection
    - tournament selection

select better one

| individual number | gene | fitness |
|---|---|---|
| 0 | 01100 | ($12^2$=) 144 |
| 1 | 10110 | ($22^2$=) 484 |
| 2 | 00011 | ($3^2$=) 9 |
| 3 | 10001 | ($17^2$=) 289 |
| 4 | 00010 | ($2^2$=) 4 |

484
gene (10110)

randomly pick up two* individuals

*sometimes more than two.
Here two individuals are selected (i.e., tournament size 2)

19

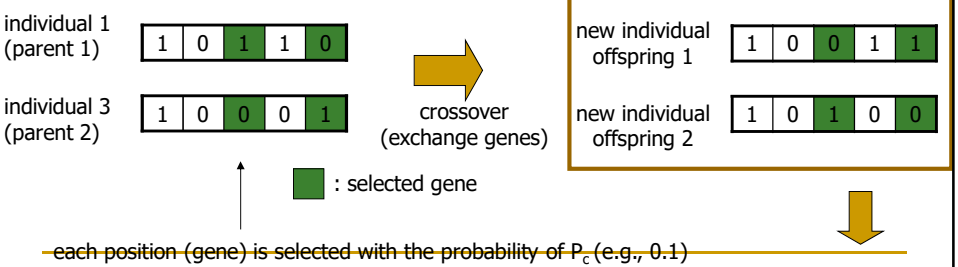
# The flow of GA

| individual number | gene | fitness |
|---|---|---|
| 0 | 01100 | ($12^2$=) 144 |
| 1 | 10110 | ($22^2$=) 484 |
| 2 | 00011 | ($3^2$=) 9 |
| 3 | 10001 | ($17^2$=) 289 |
| 4 | 00010 | ($2^2$=) 4 |

- genetic operation
  - crossover: exchange the genes

select two individuals using the tournament selection twice

suppose individual 1 and 3 are selected

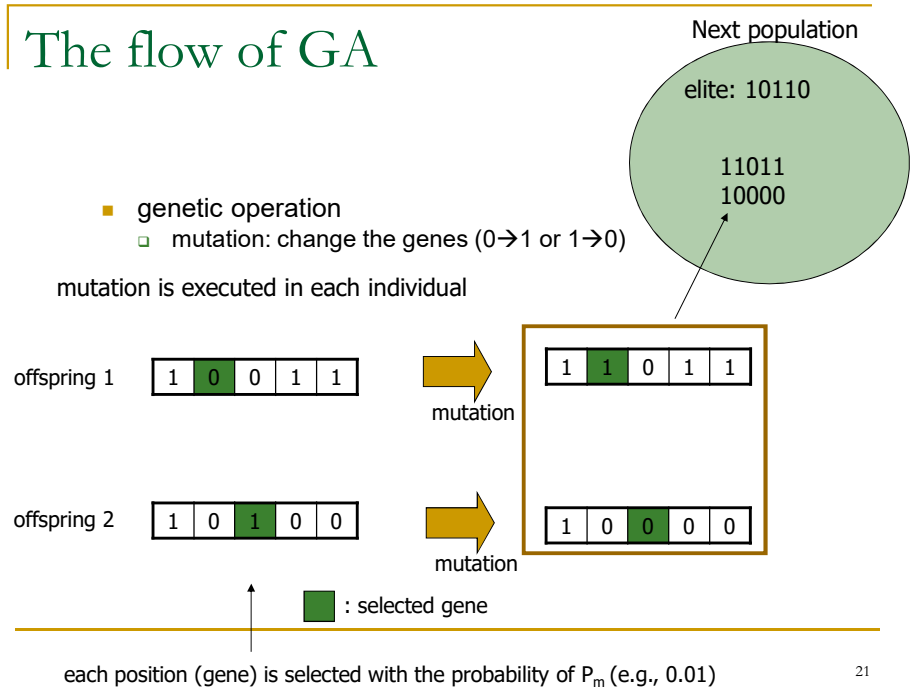



each position (gene) is selected with the probability of $P_c$ (e.g., 0.1)

Next, mutation

20

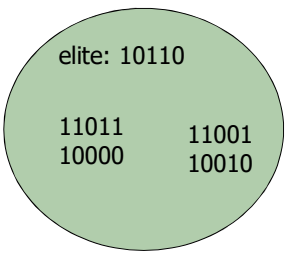
# The flow of GA




21

# The flow of GA

- genetic operation
  - repeat tournament selection, crossover and mutation until the next population becomes full with five individuals
  - In this example, execute selection, crossover and mutation once again

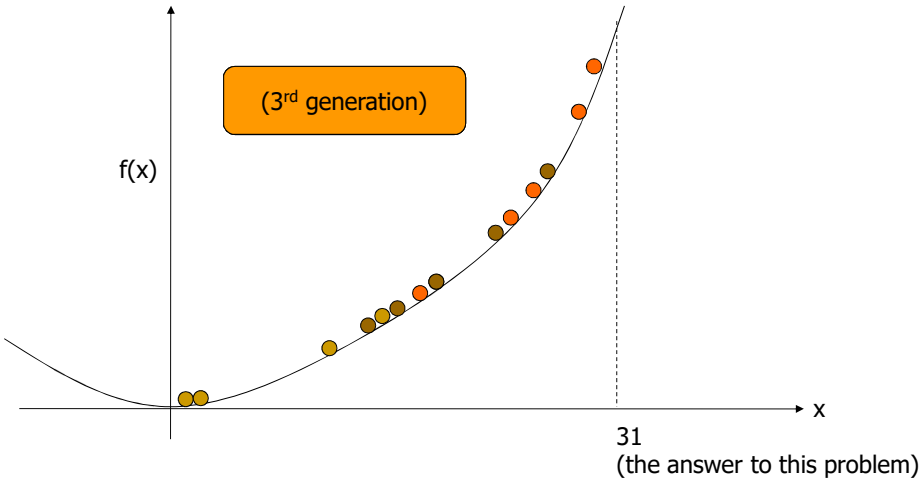



| individual number | gene | fitness |
|---|---|---|
| 0 | 10110 | ($22^2$=) 484 |
| 1 | 11011 | ($27^2$=) 729 |
| 2 | 10000 | ($16^2$=) 256 |
| 3 | 11001 | ($25^2$=) 625 |
| 4 | 10010 | ($18^2$=) 324 |

fitness increases

22

## The solutions after one generation




23

## Another Simple Example

The Traveling Salesman Problem:

Find a tour of a given set of cities so that

- each city is visited only once
- the total distance traveled is minimized

24

## 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) |

25

## Crossover

Crossover combines inversion and recombination:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | * | | | * | | |
| Parent1 | (3 | 5 | 7 | 2 | 1 | 6 | 4 | 8) |
| Parent2 | (2 | 5 | 7 | 6 | 8 | 1 | 3 | 4) |
| Child | (5 | 8 | 7 | 2 | 1 | 6 | 3 | 4) |

This operator is called the *Order1* crossover.

26

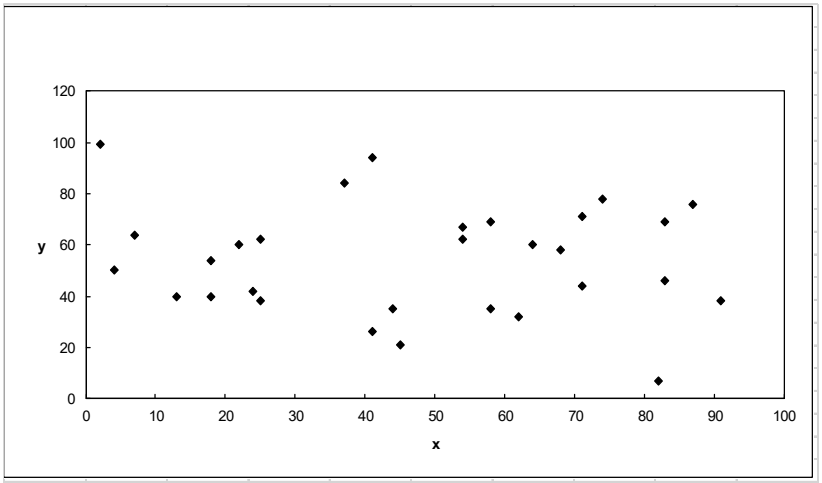
## Mutation

Mutation involves reordering of the list:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | * | | | * | | |
| Before: | (5 | 8 | 7 | 2 | 1 | 6 | 3 | 4) |
| After: | (5 | 8 | 6 | 2 | 1 | 7 | 3 | 4) |

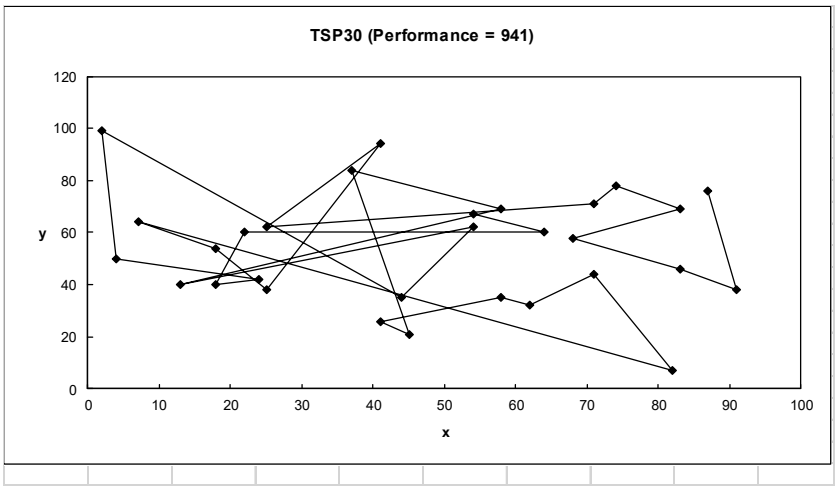
27

## TSP Example: 30 Cities



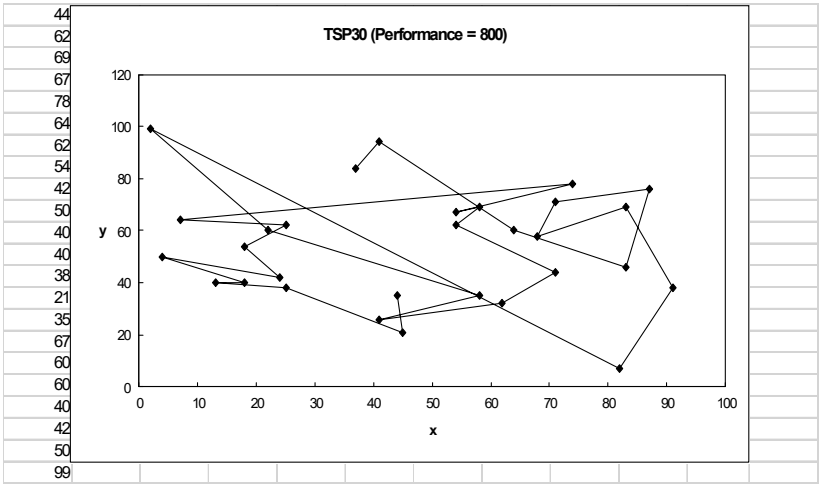

28

# Solution $_i$ (Distance = 941)



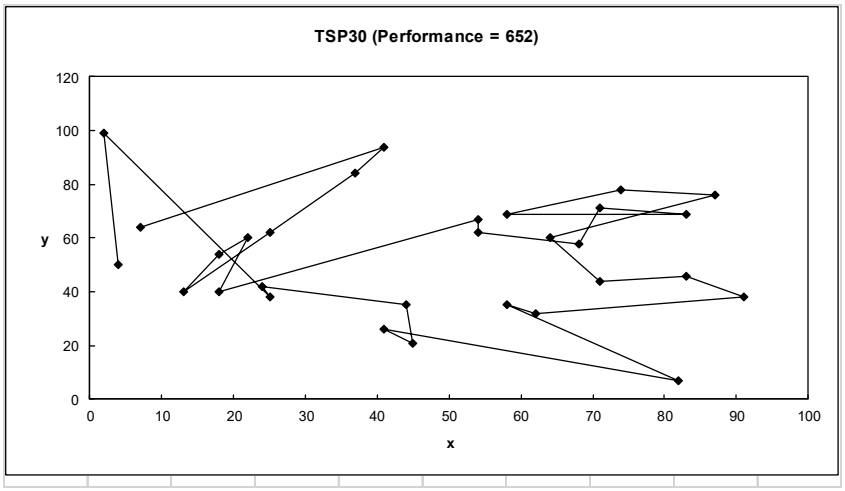

29

# Solution $_j$ (Distance = 800)



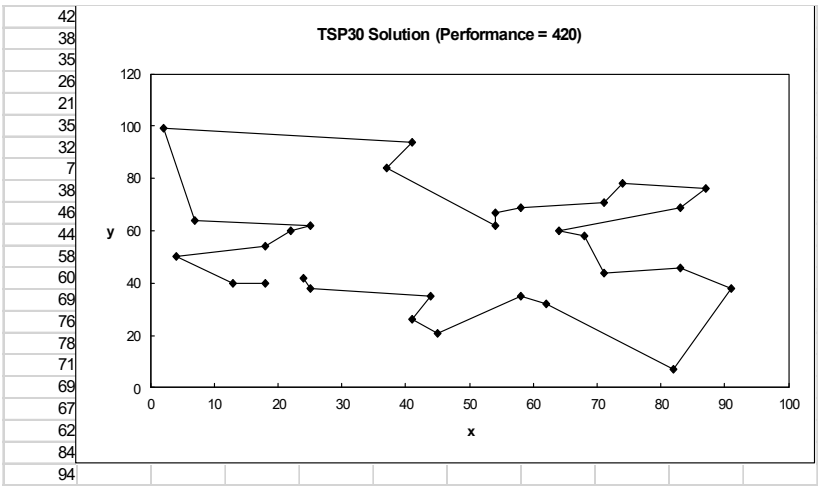

30

# Solution $_k$(Distance = 652)



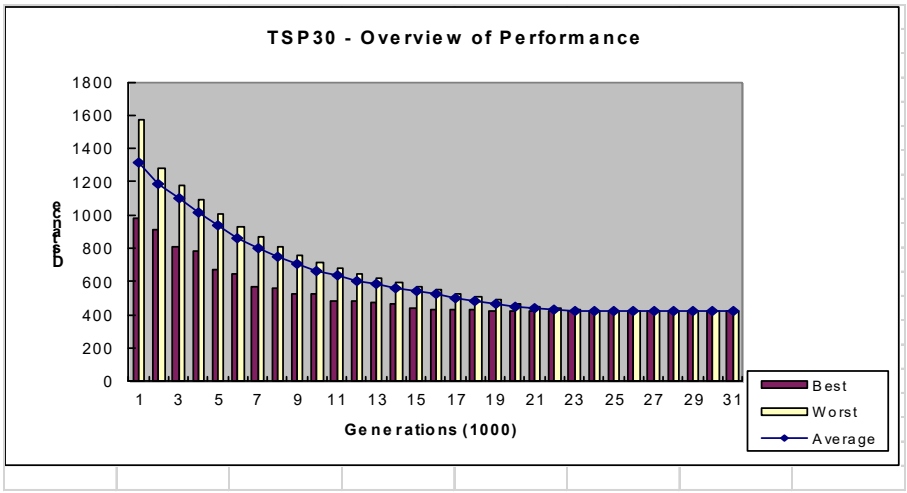

31

# Best Solution (Distance = 420)




32

# Overview of Performance




33

# Some Variants of GA

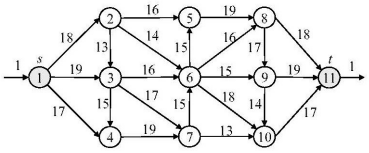
- Genetic Representation
  - Binary encoding
  - Real number encoding
  - Integer/literal permutation encoding
  - A general data structure encoding
- Selection Strategies
  - Roulette wheel selection
  - Tournament selection
  - (μ+λ)-selection
  - Ranking and scaling
  - Sharing
- Hybrid GA
  - Local search

34

# Some Variants of GA

## ➢ Priority-based GA

- Shortest path routing problem

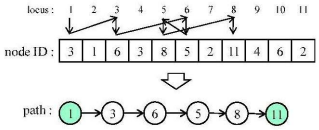



$$\min \quad z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} - \sum_{k=1}^{n} x_{ki} = \begin{cases} 1 & (i=1) \\ 0 & (i=2,3,\cdots,n-1) \\ -1 & (i=n) \end{cases}$$

$$x_{ij} = 0 \text{ or } 1 \quad \forall\, i,\ j$$

- Priority-based representation



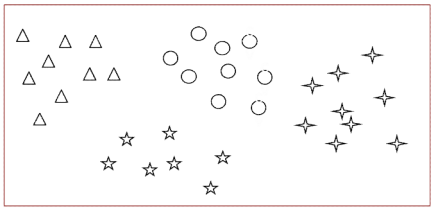

35

# Some Variants of GA

## ➢ Selection Strategy

- Niching technique

A niche →a subspace in the environment that can support different types of life.



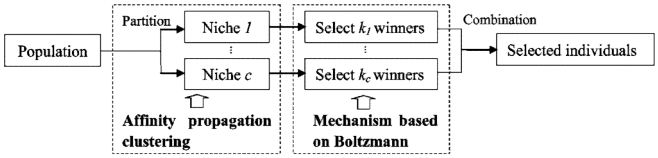
Niching method → Find target vectors → Keep diversity

36

# Some Variants of GA

## ➢ Selection Strategy

- Niching technique (clearing niche)




$$\kappa_t(n_i) = M \times [\omega_t \times Ei_t(n_i) + (1-\omega_t) \times Er_t(n_i)]$$

$$\omega_t = \left[\frac{e^{\left(\frac{t}{T_{\max}}\right)} - 1}{e-1}\right]^{\alpha}, \quad 1 \le i \le cl_t$$

$$Ei_t(n_i) = \frac{f_t(n_i)}{\sum_{j=1}^{cl_t} f_t(n_j)}$$

$$Er_t(n_i) = \frac{nov_t(n_i)}{\sum_{j=1}^{cl_t} nov_t(n_j)}$$

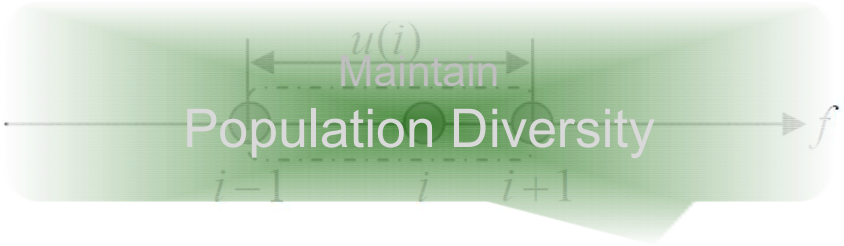
$$nov_t(n_i) = \beta^{Q_t(n_i)}, \quad 0 < \beta < 1$$

37

# Some Variants of GA

## ➢ Selection Strategy

- Niching technique

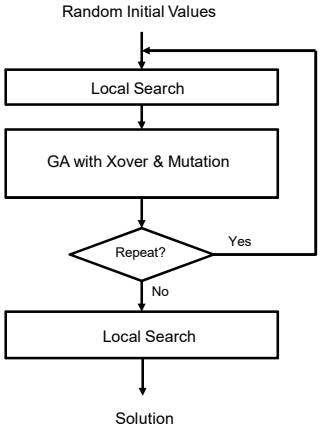



- Individuals with higher **uniqueness** are preferred in selection procedure.

38

# Some Variants of GA

- Hybrid GA with Local Search

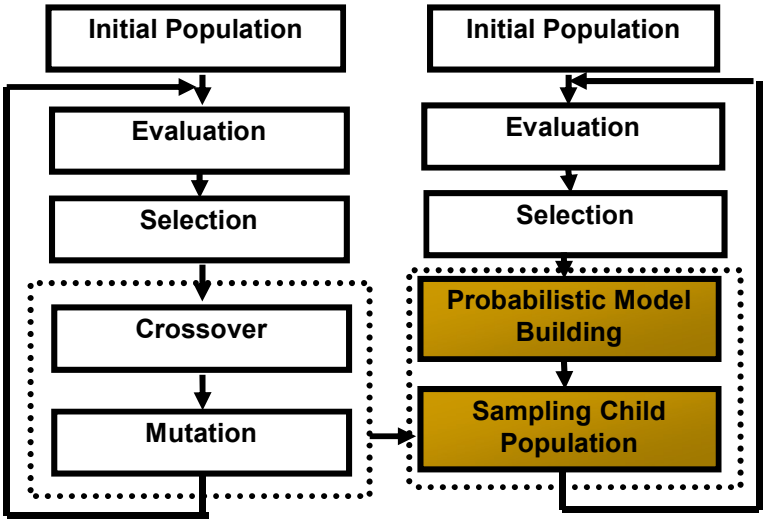



Typically, local search can be considered as another GA operator: Adaptation.

GA operators: Selection, Crossover Mutation, Adaptation.

39

# GA to EDA




*Simple GA framework* *EDA framework*

# Overview of EDAs (BOA)

(an example)

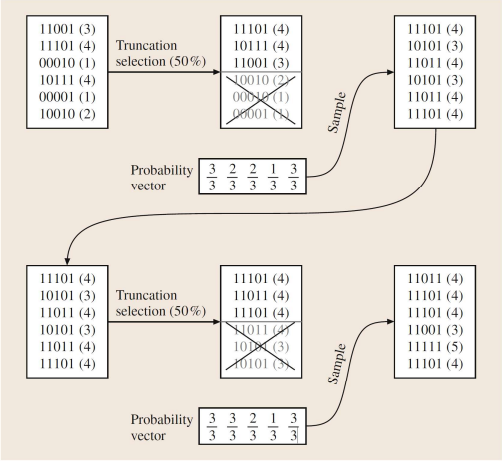



**Fig. 45.1** Simple simulation of an EDA based on the probability-vector model for onemax. The fitness values of candidate solutions are shown inside parentheses

41