

# MẠNG NƠON NHÂN TẠO và GIẢI THUẬT DI TRUYỀN

*Neural Network & Genetic Algorithm*



Biên soạn: ThS. Phạm Đình Tài  
pdtai@ntt.edu.vn  
0985.73.39.39

# **CHƯƠNG 7**

## **Giải thuật di truyền (GAs)**



- ✓ **Giới thiệu về giải thuật di truyền**
- ✓ **Giả lập thuật giải di truyền**
- ✓ **Các phép lựa chọn trong giải thuật di truyền**



# TỔNG QUAN THUẬT GIẢI DI TRUYỀN

- Thuật giải di truyền là một kỹ thuật của khoa học máy tính nhằm tìm kiếm giải pháp thích hợp cho các bài toán tối ưu tổ hợp.
- Thuật giải di truyền là một phân ngành của giải thuật tiến hóa vận dụng các nguyên lý của tiến hóa như di truyền, đột biến, chọn lọc tự nhiên, và trao đổi chéo.
- Ngày nay, thuật giải di truyền được dùng phổ biến trong một số ngành như tin sinh học, khoa học máy tính, trí tuệ nhân tạo, tài chính và một số ngành khác.
- Thuật giải di truyền sử dụng một số thuật ngữ của ngành di truyền học, như: Gen, nhiễm sắc thể, quần thể, thế hệ.

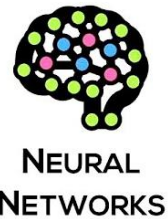
# TỔNG QUAN THUẬT GIẢI DI TRUYỀN

- Một nhiễm sắc thể (NST) được tạo thành từ nhiều Gen.
- Mỗi mỗi gen là một phần tử của NST, có vị trí và giá trị khác nhau để quy định một trạng thái.
- Quần thể: Một tập hợp các cá thể (NST) có cùng một số đặc điểm.
- Trong thuật giải di truyền, quần thể là một tập hợp các lời giải bài toán.
- Thế hệ: là tập các cá thể ứng với một giai đoạn phát triển nào đó.

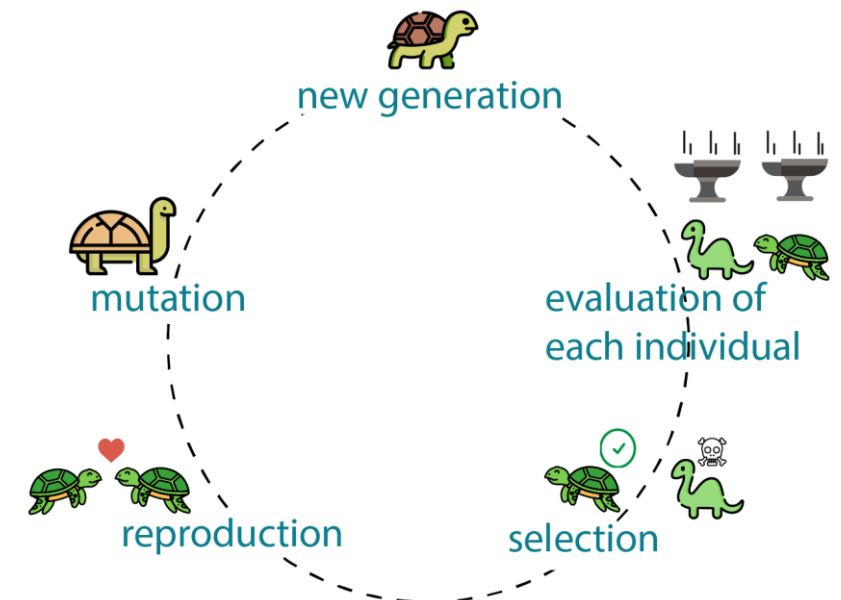


- Optimization
- Classification (GBML)
- Human Comparable Design

VS

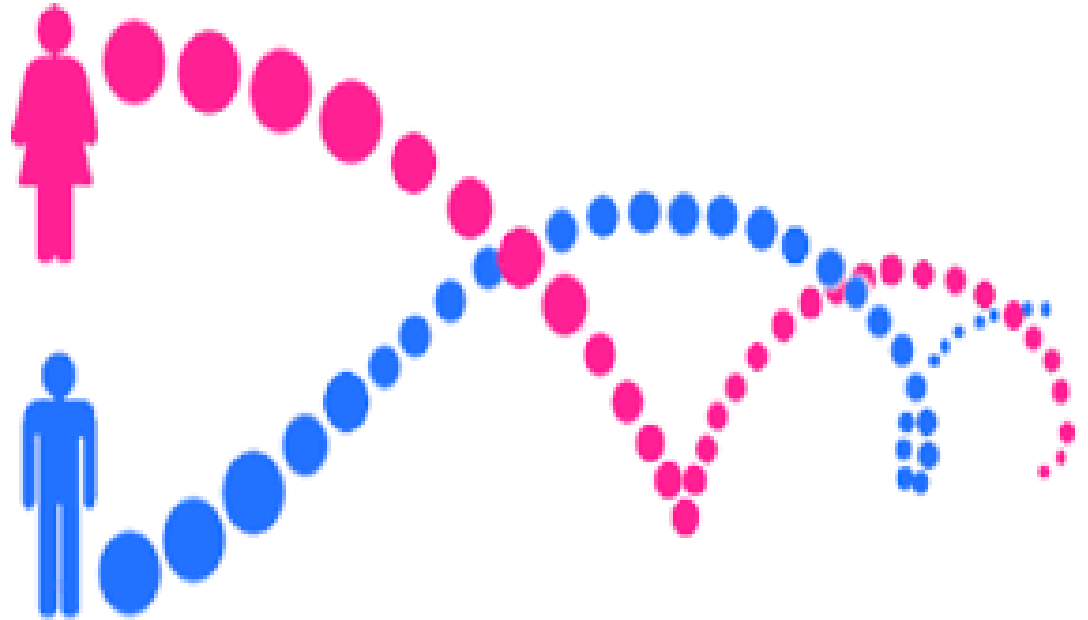


- Classification<sup>†</sup>
- Pattern recognition

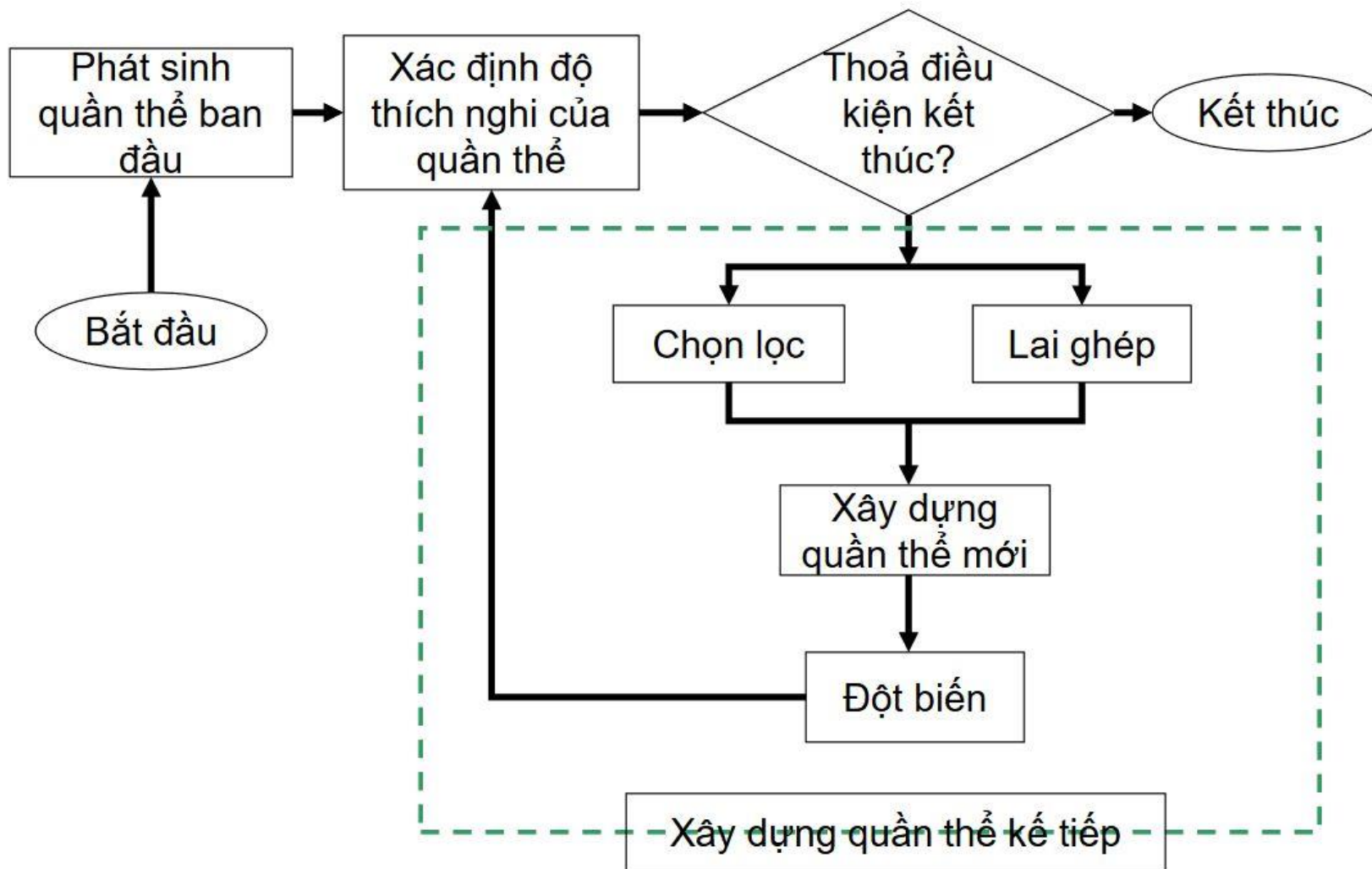


# TỔNG QUAN THUẬT GIẢI DI TRUYỀN

- **Mã hóa:** là quá trình biểu diễn gen của cá thể hay nhiễm sắc thể.
- **Phép chọn:** là quá trình chọn cá thể thích nghi tốt và loại bỏ cá thể kém thích nghi khỏi quần thể.
- **Lai ghép:** Hai cá thể cha, mẹ trao đổi các gen để tạo ra hai cá thể con.
- **Đột biến:** Một cá thể thay đổi một số gen để tạo thành cá thể mới.
- **Giải mã:** là quá trình biến đổi cá thể thành lời giải bài toán.



# MÔ HÌNH THUẬT GIẢI DI TRUYỀN





# GIẢ LẬP THUẬT GIẢI DI TRUYỀN

```
procedure Genetic_Algorithm;  
begin  
     $t \leftarrow 0$ ;  
    Khởi tạo thể hệ ban đầu  $P(t)$ ;  
    Đánh giá  $P(t)$  (theo hàm thích nghi);  
repeat  
     $t \leftarrow t + 1$ ;  
    Sinh ra thể hệ mới  $P(t)$  từ  $P(t-1)$  bởi:  
        Chọn lọc  
        Lai ghép  
        Đột biến;  
    Đánh giá  $P(t)$ ;  
until đạt điều kiện kết thúc;  
end;
```



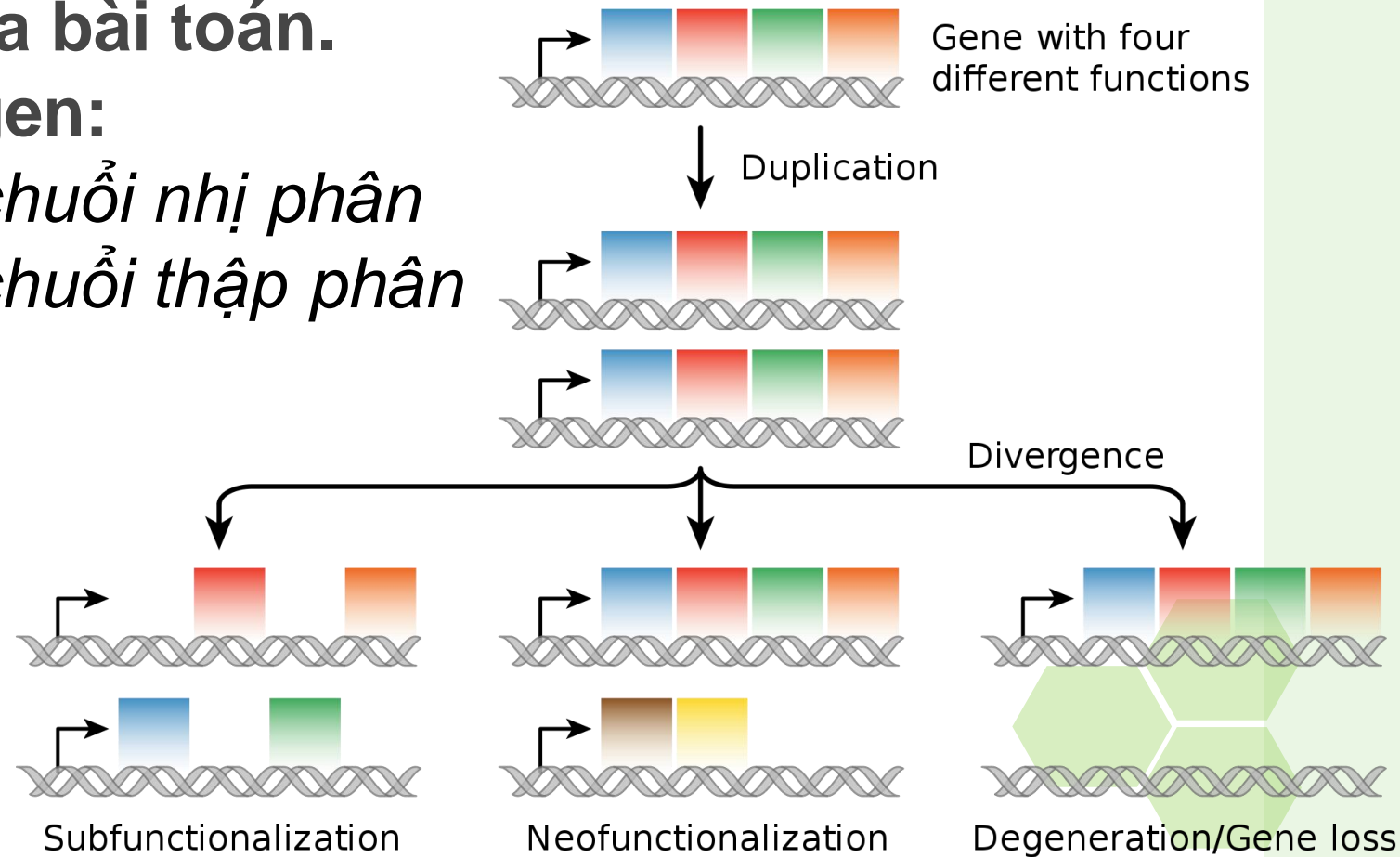
**Cá thể tốt nhất ở bước cuối cùng  
sẽ được chọn**



- Khi sử dụng thuật giải di truyền để giải bài toán, phải làm rõ các vấn đề:
  1. Chọn cách biểu diễn di truyền nào đối với những lời giải có thể của bài toán?
  2. Tạo tập lời giải ban đầu (thế hệ) như thế nào?
  3. Xác định hàm đánh giá để đánh giá độ thích nghi của các cá thể và quần thể ở thế hệ?
  4. Xác định các toán tử di truyền để sản sinh các thế hệ?
  5. Xác định giá trị các tham số sử dụng như: kích thước tập lời giải, xác suất các toán tử di truyền,...

# CÁCH BIỂU DIỄN GEN

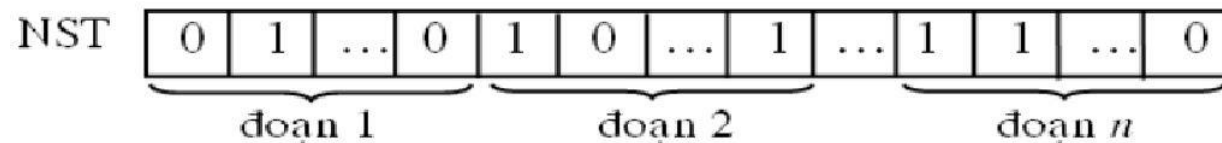
- Để có thể giải bài toán bằng thuật giải di truyền ta phải gen hóa cấu trúc dữ liệu của bài toán.
- Có hai cách biểu diễn gen:
  1. *Biểu diễn gen bằng chuỗi nhị phân*
  2. *Biểu diễn gen bằng chuỗi thập phân*



# CÁCH BIỂU DIỄN GEN

## a) Biểu diễn gen bằng chuỗi nhị phân

❖ Bộ gen gồm 2 ký hiệu 0, 1. Mỗi biến được mã hóa thành một đoạn gen, chuỗi NST gồm nhiều đoạn gen



- ❖ Giả sử biến  $\theta_i$  cần tìm trong đoạn  $\theta_{i \min} < \theta_i < \theta_{i \max}$  và được mã hóa thành chuỗi nhị phân có độ dài  $L_i$
- ❖ Chiều dài của đoạn gen được xác định dựa trên độ chính xác mong muốn  $\varepsilon_i$  tương ứng với mỗi biến

$$L_i = \log_2 \left( \frac{\theta_{i \max} - \theta_{i \min}}{\varepsilon_i} \right)$$

# CÁCH BIỂU DIỄN GEN

## a) Biểu diễn gen bằng chuỗi nhị phân

- ❖ Mỗi đoạn mã nhị phân  $S_i$  được giải mã thành giá trị của biến

$\theta_i$  như sau: 
$$\theta_i = \theta_{i \min} + \frac{\theta_{i \max} - \theta_{i \min}}{2^{L_i - 1}} DV(s_i)$$

- ❖ Trong đó:  $DV(s_i)$  là giá trị thập phân của chuỗi  $s_i$

$$DV(s_i) = \sum_{j=0}^{L_i - 1} 2^j . s_{ij}$$

- ❖ **Thí dụ:** Giả sử miền tìm kiếm giá trị tối ưu của biến  $\theta_i$  là  $\theta_{i \min} = -1.5$  và  $\theta_{i \max} = 4.6$ , độ chính xác mong muốn là  $\varepsilon_i = 0.1$  thì độ dài của chuỗi nhị phân biểu diễn biến  $\theta_i$  là:

$$L_i = \log_2 \left( \frac{4.6 - (-1.5)}{0.1} \right) = 5.93$$

## a) Biểu diễn gen bằng chuỗi nhị phân

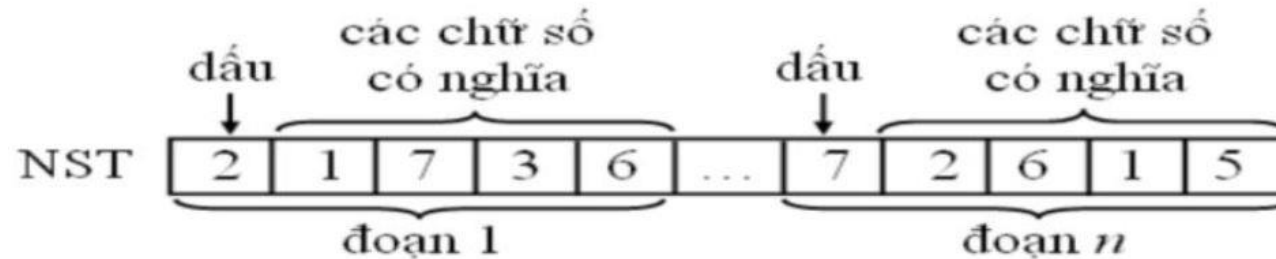
❖ Do số gen là số nguyên nên chọn  $L_i = 6$ . Chuỗi (000000) biểu diễn -1.5, chuỗi (111111) biểu diễn 4.6. Chuỗi bất kỳ, chẳng hạn  $S_i = (101110)$  biểu diễn giá trị:

$$\theta_i = -1.5 + \frac{4.5 - (-1.5)}{2^6 - 1} \cdot (2^5 + 2^3 + 2^2 + 2^1) = 2.8571$$

# CÁCH BIỂU DIỄN GEN

## a) Biểu diễn gen bằng chuỗi thập phân

- ❖ Bộ gen gồm **10** ký hiệu **0,1, ... 9**. Mỗi biến được mã hóa thành một đoạn gen, mỗi đoạn gen gồm có: 1 gen biểu diễn dấu, các gen còn lại biểu diễn các chữ số có nghĩa.



- ❖ Quy ước gen mã hóa dấu: Dấu **-** là **0-4** và dấu **+** là **5-9**.
- ❖ Vị trí dấu chấm thập phân của mỗi gen được lưu trữ để sử dụng khi giải mã.
- ❖ **Ví dụ:** Mã hóa thập phân 5 chữ số có nghĩa thì đoạn gen biểu diễn  $\theta_i = -56.4172$  là  $S_i = (056417)$ , đồng thời lưu vị trí dấu chấm thập phân là 2



# PHÉP CHỌN LỌC

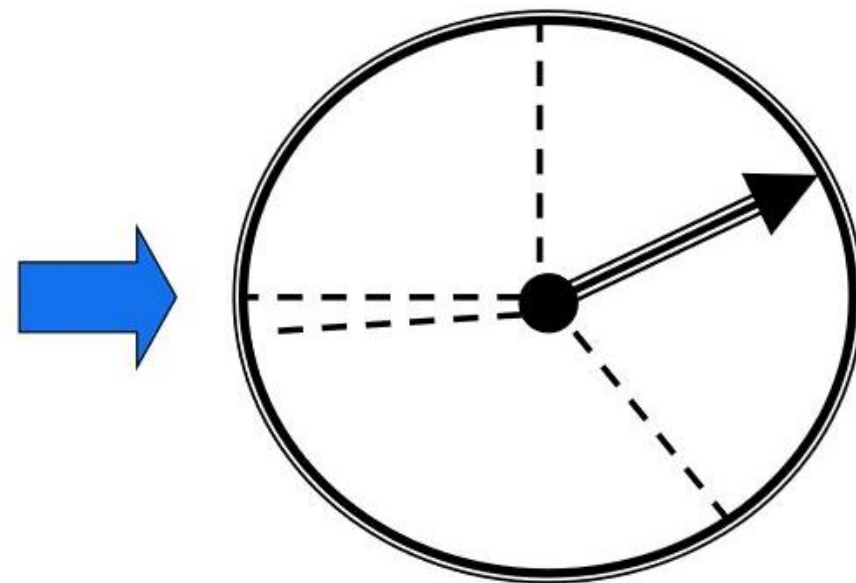
- Nguyên tắc cơ bản của phép chọn lọc là NST có độ thích nghi càng cao thì có xác suất chọn càng lớn.
- Có cách phương pháp chọn: chọn lọc tỉ lệ, chọn lọc vòng, chọn lọc cắt, chọn lọc xếp hạng tuyến tính và chọn lọc xếp hạng lũy thừa.
- Cường độ chọn lọc:  $I = \frac{M^* - M}{\sigma}$ . Trong đó M và  $M_*$  là độ thích

nhì trung bình của quần thể trước và sau chọn lọc,  $\sigma_2$  là phương sai của độ thích nghi trước chọn lọc.

# PHÉP CHỌN LỌC

- Một trong các cách đơn giản và hiệu quả để chọn cá thể thích nghi là dùng vòng tròn Rulet. Trong vòng tròn Rulet, mỗi cá thể chiếm 1 vùng có diện tích tỉ lệ với độ thích nghi của chúng.
- Tính tổng độ thích nghi của quần thể đồng thời mỗi cá thể.
- Sinh số ngẫu nhiên  $n$  trong khoảng từ 0 đến tổng thích nghi.

STT	Cá thể	ĐTN chuẩn
1	0010001	0,4
2	0010101	0,3
3	0101000	0.05
4	1100011	0.25



# PHÉP LAI GHÉP

- Lai ghép là kết hợp hai NST cha mẹ để tạo NST con với triển vọng cha mẹ tốt sẽ tạo con tốt hơn.
- Phép lai ghép xảy ra giữa hai NST cha mẹ được chọn ngẫu nhiên với xác suất  $P_c$  (gọi là xác suất lai ghép).
- Các NST con thừa hưởng một phần gen của cha và một phần gen của mẹ.
- Các phương pháp lai ghép: lai ghép một điểm, lai ghép nhiều điểm và lai ghép đều.

# PHÉP LAI GHÉP

## Lai ghép một điểm

### Điểm lai ghép

Cha mẹ 1 (A)

2	1	4	5	9	8	6	1	7	5	3
---	---	---	---	---	---	---	---	---	---	---

Cha mẹ 2 (B)

3	5	2	1	7	6	0	9	8	1	2
---	---	---	---	---	---	---	---	---	---	---

Con 1 (A')

2	1	4	5	7	6	0	9	8	1	2
---	---	---	---	---	---	---	---	---	---	---

Con 2 (B')

3	5	2	1	9	8	6	1	7	5	3
---	---	---	---	---	---	---	---	---	---	---

## Lai ghép nhiều điểm

### Điểm lai ghép

Cha mẹ 1 (A)

2	1	4	5	9	8	6	1	7	5	3
---	---	---	---	---	---	---	---	---	---	---

Cha mẹ 2 (B)

3	5	2	1	7	6	0	9	8	1	2
---	---	---	---	---	---	---	---	---	---	---

Con 1 (A')

2	1	2	1	7	6	6	1	8	1	3
---	---	---	---	---	---	---	---	---	---	---

Con 2 (B')

3	5	4	5	9	8	0	9	7	5	2
---	---	---	---	---	---	---	---	---	---	---

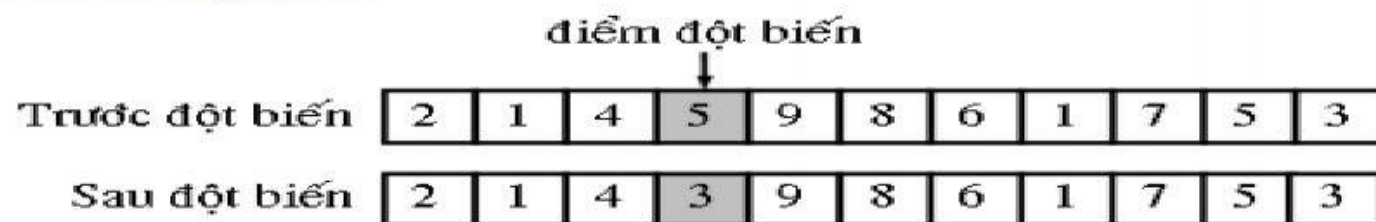
- Toán tử lai ghép:
  - ❖ Lấy giá trị ngẫu nhiên  $p \in [0,1]$  để chọn cá thể lai ghép, cá thể có độ thích nghi cao có xác suất lựa chọn nhiều hơn
  - ❖ Sau khi lựa chọn một cặp cá thể cha mẹ, hoán vị các nhiễm sắc thể tại vị trí ngẫu nhiên với xác suất  $p_c$
- Toán tử lai ghép có xu hướng kéo quần thể về phía các cá thể có độ thích nghi cao  $\Rightarrow$  cục bộ địa phương

# PHÉP ĐỘT BIẾN

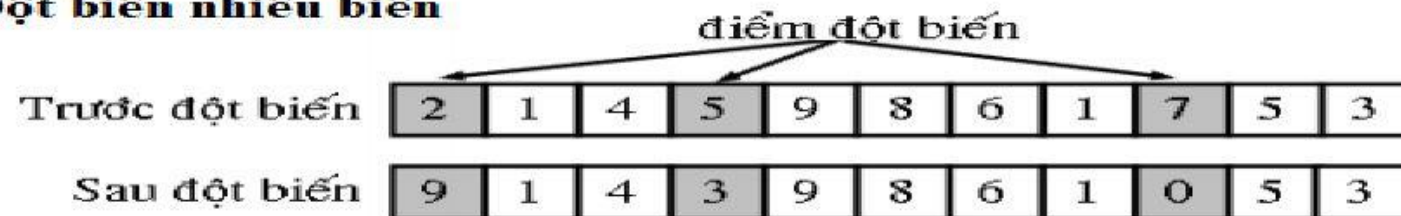
- Phép đột biến thay đổi ngẫu nhiên một hoặc nhiều gen của một cá thể để làm tăng sự đa dạng về cấu trúc trong quần thể.
- Đột biến chỉ xảy ra với xác suất  $PM$  thấp. Các phương pháp đột biến:

Đột biến một biến và đột biến nhiều biến

## Đột biến một biến



## Đột biến nhiều biến





# PHÉP ĐỘT BIẾN

- Toán tử đột biến:
  - Giúp lời giải có thể nhảy ra khỏi các cực trị địa phương
  - Với mỗi cá thể trong quần thể, thực hiện đột biến với xác suất  $p_m$  tại một vị trí ngẫu nhiên (thông thường  $p_m \ll 0.1$ )

0010001



0011001

# HÀM MỤC TIÊU

- Độ tốt của một cá thể: Là giá trị của cá thể cho một vấn đề bài toán cụ thể. Để xác định được độ tốt của các cá thể ta cần một hàm để làm việc này. Hàm này gọi là Hàm mục tiêu .
- Hàm mục tiêu
  - ❖ Dùng để đánh giá độ tốt của một lời giải hoặc cá thể.
  - ❖ Hàm mục tiêu nhận vào tham số là gen của một cá thể và trả ra một số thực.
  - ❖ Tùy theo giá trị của số thực này mà ta biết được độ tốt của cá thể đó

# HÀM MỤC TIÊU

- Độ thích nghi của các cá thể (fitness)
  - ❖ Là khả năng cá thể đó được chọn lọc vào thế hệ sau hoặc là được chọn lọc cho việc lai ghép để tạo ra cá thể con .
  - ❖ Vì độ thích nghi là một xác suất để cá thể được chọn nên người ta thường ánh xạ độ thích nghi vào đoạn  $[0,1]$  (độ thích nghi chuẩn) Dùng để đánh giá độ tốt của một lời giải hoặc cá thể.

$$F(a_i) = \frac{F(a_i)}{\sum_{j=1}^N F(a_j)} \quad i = 1, 2, \dots, N$$

# GIẢI QUYẾT BÀI TOÁN

- Hãy tìm nghiệm của phương trình:

$$X^2 - 14x + 49 = 0.$$

- Hãy sử dụng giải thuật di truyền để mô tả và giải bài toán?

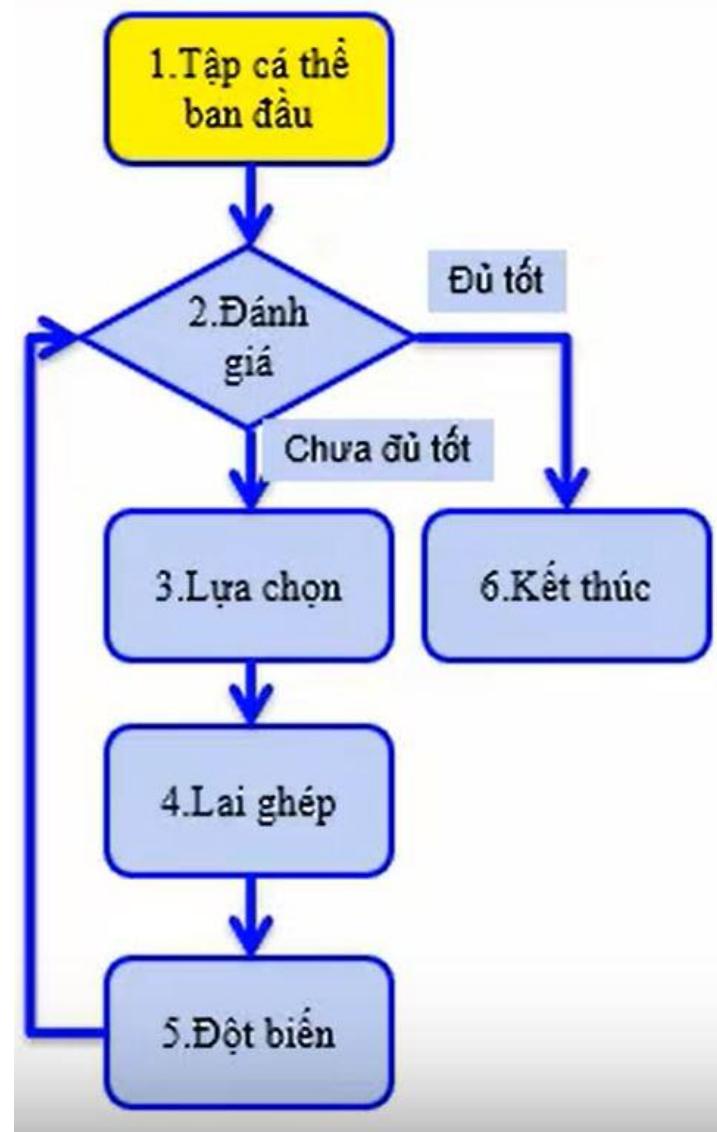
# GIẢI QUYẾT BÀI TOÁN

## ❖ Bước 1: Tạo tập các thể ban đầu

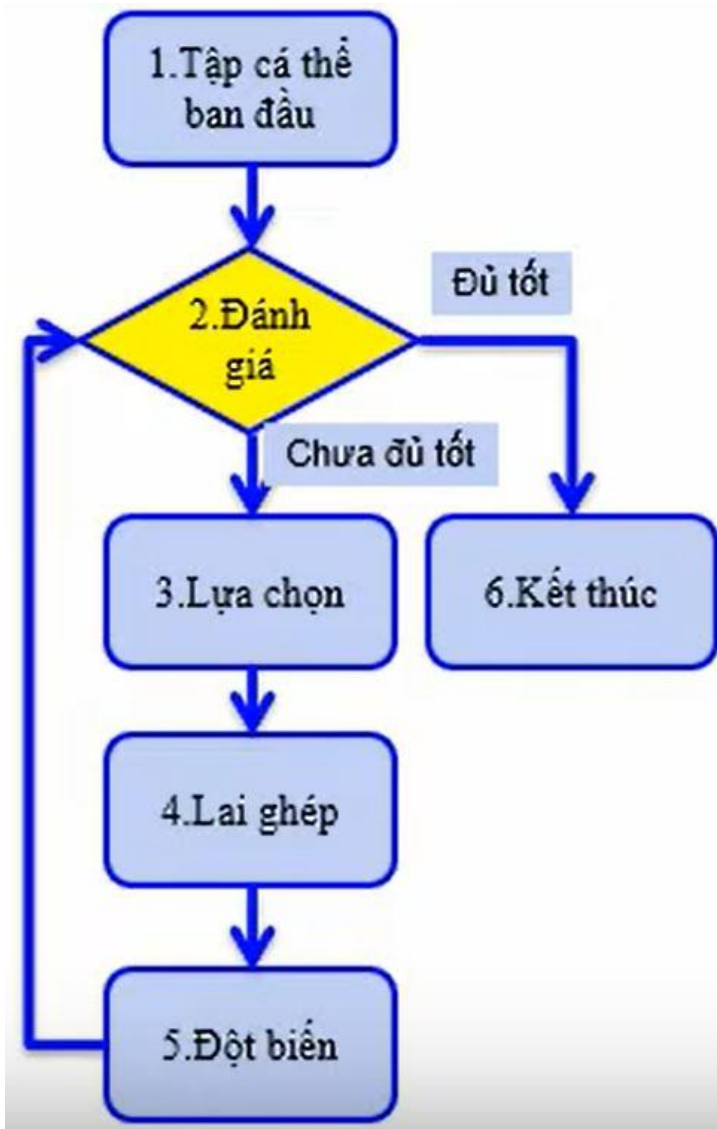
- Thiết kế bộ gen là các số trong hệ nhị phân. Phương án nghiệm của bài toán là số khi đổi qua cơ số 10 của bộ gen
- Tạo ngẫu nhiên tập cá thể ban đầu:

Thập phân (Decimal)	Nhị phân (Binary)
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

**Ví dụ:** 0000, 1011,  
1100, 0101, 0010

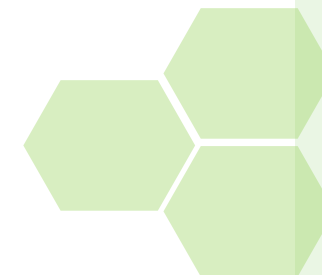


# GIẢI QUYẾT BÀI TOÁN



## ❖ Bước 2: Đánh giá

- Giải sử chọn hàm đánh giá:
  - $f = |x^2 - 14x + 49|$
  - $f$  càng nhỏ thể hiện độ thích nghi càng cao
- Tính các giá trị của  $f$  đối với tất cả các cá thể
  - $F(0000)=49, f(1011)=16, f(1100)=25,$   
 $f(0101)=4, f(0011)=25.$
- Sắp xếp theo trình tự độ thích nghi từ ca đến thấp.
  - $f(0101)=4, f(1011)=16, f(1100)=25$
  - $f(0010)=25, f(0000)=49.$
- Giả sử tiêu chí kết thúc là  $f < 1$





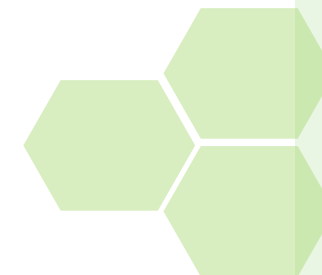
# GIẢI QUYẾT BÀI TOÁN



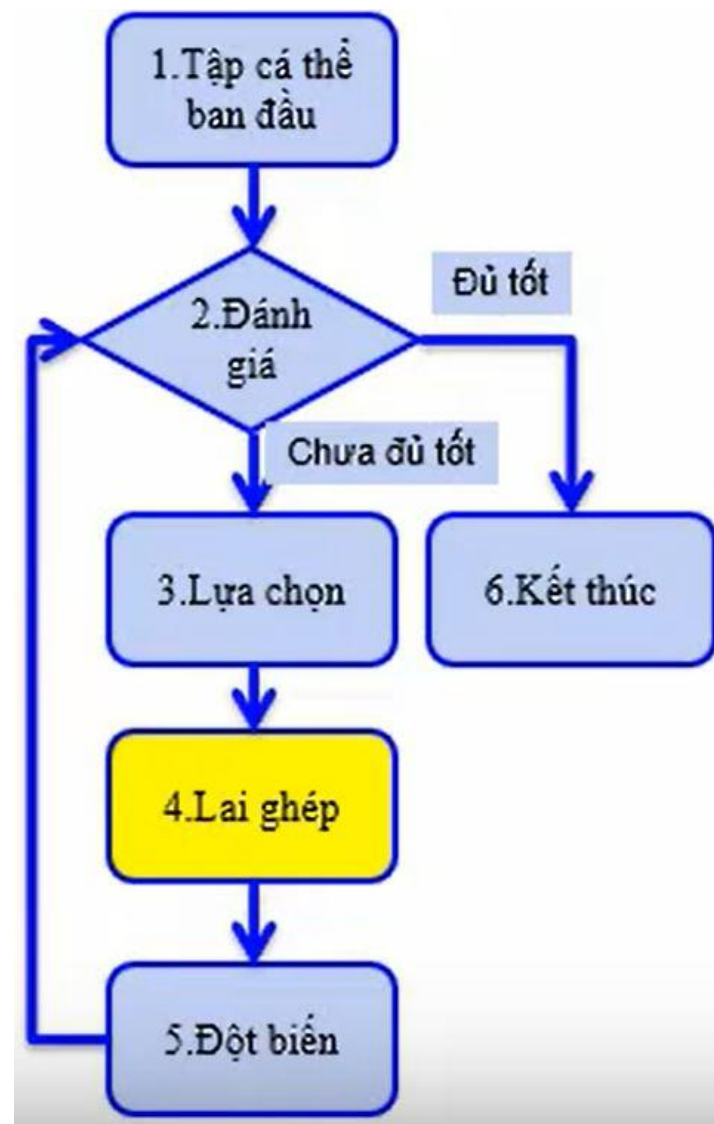
## ❖ Bước 3: Lựa chọn

- Có thể loại bỏ các nghiệm có độ thích nghi thấp, nhân bản cá thể có độ thích nghi cao.

- $0101 \rightarrow f = 4$
- $0101 \rightarrow f = 4 \leftarrow \text{Nhân bản}$
- $1011 \rightarrow f = 16$
- $1100 \rightarrow f = 25$
- $0010 \rightarrow f = 25$
- $0000 \rightarrow f = 49 \leftarrow \text{Loại bỏ}$

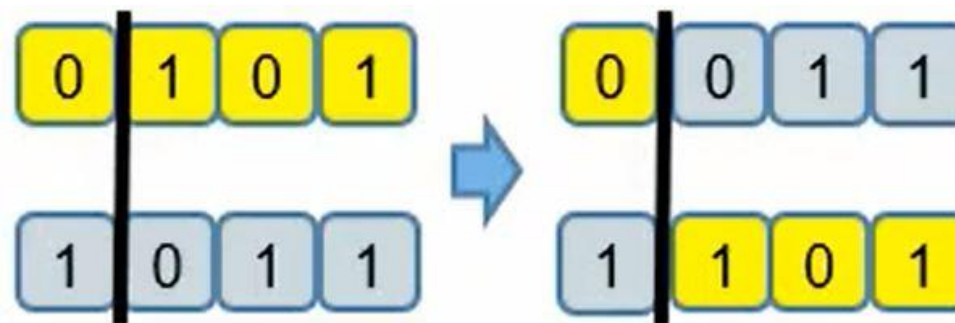


# GIẢI QUYẾT BÀI TOÁN

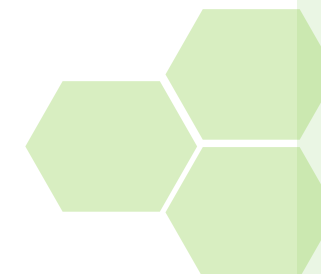


## ❖ Bước 4: Lai ghép

- Ví dụ ta có thể lai ghép 1 điểm đối với nghiệm 0101 và 1011 tại vị trí  $k=1$  như sau:



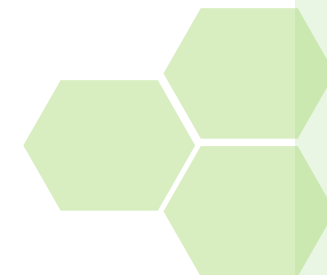
- Ta được nghiệm mới là
  - 0101
  - 0011
  - 1100
  - 0010



# GIẢI QUYẾT BÀI TOÁN

## ❖ Bước 5: Đột biến

- Thay đổi giá trị 1 thành 0 và 0 thành 1
- Ví dụ có thể các thẻ **1100** được chọn và đột biến thành **0100** hoặc **0101** đột biến thành **0111**
- Việc lặp đi lặp lại quá trình đánh giá, lựa chọn, lai ghép và đột biến sẽ tính được cá thể tốt nhất là 0111, đó chính là nghiệm của bài toán.



***Thank you !***

