

# Chapter 9 Genetic Algorithm

Associate Professor Yachai Limpiyakorn, Ph.D.

## Natural Selection

- Charles Darwin ได้อธิบายการสืบทอดของสิ่งมีชีวิตด้วยกฎ **Evolution through Natural Selection** :-
  - สิ่งมีชีวิตมีแนวโน้มที่จะสืบทอดลักษณะพิเศษให้ลูกหลาน
  - ธรรมชาติจะผลิตสิ่งมีชีวิตที่มีลักษณะพิเศษแตกต่างไปจากเดิม
  - สิ่งมีชีวิตที่เหมาะสมที่สุด (fittest) หรือที่มีลักษณะพิเศษที่ธรรมชาติพอใจมากที่สุด มีแนวโน้มที่จะมีลูกหลานมากกว่าตัวที่ไม่เหมาะสม ดังนั้น ประชากรจะโน้มเอียงไปทางตัวที่เหมาะสม
  - การเปลี่ยนแปลงจะสะสมไปเรื่อยและเกิด species ใหม่ที่เหมาะสมกับสภาพแวดล้อม เมื่อเวลาผ่านไปนานๆ

## Principles of Genetics

ในทางชีววิทยา เซลล์แต่ละเซลล์ในพืชชั้นสูงและสัตว์ประกอบด้วยนิวเคลียส (nucleus) 1 เซลล์ แต่ละนิวเคลียสประกอบด้วยโครโมโซม (chromosome) จำนวนหนึ่ง โดยโครโมโซมจะอยู่กันเป็นคู่มาจากพ่อและแม่อย่างละเส้น โครโมโซมแต่ละเส้นจะมียีน (gene) เป็นตัวกำหนดลักษณะถ่ายทอดทางพันธุกรรมของสิ่งมีชีวิต ในขณะที่มีการจับคู่กันของโครโมโซมอาจเกิด การไขว้เปลี่ยน (crossover) ซึ่งเป็นการที่ยีนจากโครโมโซมพ่อแม่สลับเปลี่ยนกัน ทำให้เกิดโครโมโซมใหม่ขึ้น 2 คู่ และในขณะที่เซลล์แบ่งตัวจะเกิดกระบวนการ คัดลอกโครโมโซม (chromosome copying) ซึ่งบางครั้งจะมีการเปลี่ยนแปลงของยีนที่มาจากยีนพ่อแม่ เกิดเป็นยีนที่ไม่เคยมีมาก่อน เราเรียกการเกิดยีนลักษณะนี้ว่า การกลายพันธุ์ (mutation)

เราอาจกล่าวได้ว่า การคัดเลือกโดยธรรมชาติ (Natural Selection) เกิดจากการเปลี่ยนแปลงทางพันธุกรรมอันเป็นผลมาจาก การไขว้เปลี่ยน (crossover) ของลักษณะทางพันธุกรรม และ การกลายพันธุ์ (mutation)

2110773-9 2/2567

2

## Genetic Algorithm (GA)

- Goldberg and Holland, 1988
- GA has its core idea from Darwin's theory of natural evolution "survival of the fittest"
- one of random-based *evolutionary algorithms* (EAs)
- search-based optimization technique
- optimization → "how to find the best value for k that maximizes the performance of KNN classifier?"
- in order to find a solution, random changes applied to the current solutions to generate new ones.

## How GA works

- GA works on a population consisting of some solutions
- population size is the number of solutions
- each solution is called individual/ hypothesis
- each individual solution has a chromosome
- chromosome is represented as a set of parameters (features) that defines the individual
- each chromosome has a set of genes
- each gene is represented by somehow such as a string of 0s and 1s

2110773-9 2/2567

5

## Hypothesis/ Individual Representation

Hypothesis: IF (Haircolor =black  $\vee$  red)  $\wedge$  (Eyecolor = dark) THEN Sunburn = negative

	Haircolor	Eyecolor	Sunburn
<u>Bit String:</u>	110	10	01

Hypothesis: IF Haircolor =blonde THEN Sunburn = positive

	Haircolor	Eyecolor	Sunburn
<u>Bit String:</u>	001	11	10

**Note:** Eyecolor = 11  $\rightarrow$  Don't care

2110773-9 2/2567

6

## Genetic Operators

◆ การไขว้เปลี่ยน (Crossover) เป็นการสร้างสายอักขระลูกหลาน 2 สายจากสายอักขระพ่อแม่ 2 สาย โดยการคัดลอกบิตจากสายอักขระพ่อแม่ตามตำแหน่งที่กำหนดโดยหน้ากากไขว้เปลี่ยน (Crossover Mask) การไขว้เปลี่ยนทำได้หลายวิธีดังตัวอย่างที่แสดงข้างล่าง อาทิ การไขว้เปลี่ยน 1 ตำแหน่ง (Single-point Crossover) การไขว้เปลี่ยน 2 ตำแหน่ง (Two-point crossover) การไขว้เปลี่ยนยูนิฟอร์ม (Uniform Crossover)

◆ การกลายพันธุ์ (Mutation) เป็นการสร้างการเปลี่ยนแปลงต่อสายอักขระลูกหลาน โดยสุ่มเลือกเปลี่ยนค่าบิตใดบิตหนึ่งดังแสดงในตัวอย่างข้างล่าง

2110773-9 2/2567

7

Parent Strings	Crossover Mask	Offsprings
<i>Single-point crossover:</i>		
11101001000	11111000000	11101010101
00001010101		00001001000
<i>Two-point crossover:</i>		
11101001000	00111110000	11001011000
00001010101		00101000101
<i>Uniform crossover:</i>		
11101001000	10011010011	10001000100
00001010101		01101011001
<i>Point mutation:</i>		
11101001000		11101011000

8

2110773-9 2/2567

## Application: Knapsack Problem

- Prepare a knapsack that can only hold a certain weight
- A number of items to choose
- Each item has its own weight and value e.g. food, water, tent
- Goal: select the items that maximize the total value you carry without exceeding the knapsack's weight capacity

2110773-9 2/2567

9

## How GA works

1. **Representation (Chromosomes):** Each possible combination of items is represented as a "chromosome" (a bit string). For example, if there are 5 items, a chromosome could be "10110". A '1' means the item is included in the knapsack, and a '0' means it's not.
2. **Population:** Start with a population of random chromosomes (i.e., random combinations of items).
3. **Fitness Function:** The total value of the items in the knapsack \*if\* the total weight is within the capacity. If the weight exceeds the capacity, the fitness can be penalized (e.g., set to 0 or a negative value).
4. **Selection:** Chromosomes with higher fitness have a greater chance of being selected for reproduction. This mimics natural selection where the "fittest" individuals are more likely to pass on their genes.
5. **Crossover:** Selected chromosomes are paired up, and their genetic material is combined to create offspring. For example, two chromosomes, "10110" and "01101" might swap parts of them to create "10101" and "01110".
6. **Mutation:** To introduce diversity, some bits in the offspring chromosomes are randomly flipped e.g., a '1' becomes a '0', or vice versa).
7. **Repeat:** Steps 3-6 are repeated for a number of generations. With each generation, the population tends to evolve towards better solutions (higher total value in the knapsack without exceeding the weight limit).

2110773-9 2/2567

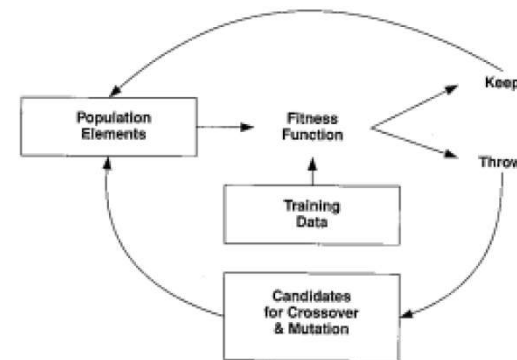
10

## Knapsack is about optimization problem

- "Find the values of 'x<sub>i</sub>' (which items to include or exclude) that *maximize* the total value (the objective function) while ensuring that the total weight does not exceed the knapsack's capacity (the constraint)."
- Given item 1: Value = 6, Weight = 2  
Item 2: Value = 10, Weight = 3  
Item 3: Value = 12, Weight = 4  
capacity of W=5  
Equation: Maximize  $6X_1 + 10X_2 + 12X_3$   
Subject to:  $2X_1 + 3X_2 + 4X_3 \leq 5$  where  $X_1, X_2$ , and  $X_3$  can be either 0 or 1
- Optimal solution is to include items 1 and 2 ( $X_1 = 1, X_2 = 1, X_3 = 0$ ), giving a total value of 16 and a total weight of 5

2110773-9 2/2567

11



2110773-9 2/2567

12

## GA Overview

การเรียนรู้จำลอง  
วิวัฒนาการของ  
สิ่งมีชีวิต  
(Biological  
Evolution)

## GA Parameters

- Fitness function for ranking candidate patterns
- Stopping Criteria:
  - Maximum of hypotheses' fitness values  $\geq$  fitness threshold
  - Max fitness does not change after many generations
  - Reach fixed number of iterations of learning process
- Size of population  $p$  to be maintained
- Ratio of population  $r$  to be replaced at each generation
- Mutation rate  $m$

2110773-9 2/2567

13

GA (Fitness, Fitness\_threshold, p, r, m)

- ◆ Initialize:  $P \leftarrow p$  random hypotheses
- ◆ Evaluate: for each  $h$  in  $P$  compute  $Fitness(h)$
- ◆ While  $[max_i Fitness(h_i)] < Fitness\_threshold$ 
  1. *Select*: Probabilistically select  $(1-r)p$  individuals of  $P$  to add to  $P_s$ 
$$Prob(h_j) = Fitness(h_j) / \sum_j Fitness(h_j) ; j=1,...,p$$
  2. *Crossover*: Probabilistically select  $rp/2$  pairs of hypotheses from  $P_s$ .  
For each pair  $\langle h_1, h_2 \rangle$ , produce two offspring by applying Crossover operator. Add all offspring to  $P_s$ .
  3. *Mutate*: Invert a randomly selected bit in  $mp$  random individuals of  $P_s$ .
  4. *Update*:  $P \leftarrow P_s$ .
  5. *Evaluate*: for each  $h$  in  $P$  compute  $Fitness(h)$
- ◆ Return the hypothesis from  $P$  that has the highest fitness.

2110773-9 2/2567

14

15

## Fitness Function and Selection

โดยทั่วไป โครโมโซมจะถูกเลือกแบบสุ่มเพื่อสร้างประชากรรุ่นใหม่ด้วยค่าความน่าจะเป็นตามสมการที่แสดงใน GA Algorithm ข้างต้น ซึ่งวิธีการเลือกดังกล่าว เรียกว่า Fitness Proportionate selection หรือ Roulette wheel selection เราจะสังเกตได้ว่า ฟังก์ชันค่าความเหมาะสม (fitness function) จะเป็นตัวกำหนดความน่าจะเป็นที่โครโมโซมจะอยู่รอดในประชากรรุ่น (generation) ถัดไป ซึ่งโครโมโซมที่มีค่าความเหมาะสมสูงจะมีโอกาสอยู่รอดมากกว่าโครโมโซมเส้นอื่นๆ แต่ก็ไม่ได้หมายความว่า โครโมโซมที่มีค่าความเหมาะสมสูงสุดจะถูกเลือกทุกครั้ง ขึ้นอยู่กับการสุ่มค่า

2110773-9 2/2567

*Crowding* เป็นปรากฏการณ์ที่สมาชิกบางตัวในประชากร ซึ่งมีค่าความเหมาะสมสูง ถูก reproduce อย่างรวดเร็ว ทำให้ความหลากหลาย (diversity) ของประชากรลดลง การแก้ปัญหา *Crowding* สามารถทำได้โดย

1. เปลี่ยนวิธีการเลือกความเหมาะสม (Altering fitness selection) โดยใช้ Tournament Selection หรือ Rank Selection
2. ใช้วิธีการ Fitness sharing หมายถึง ลดค่าความเหมาะสมของสมาชิกในประชากรรุ่นหนึ่งที่มีความคล้ายกัน (the fitness of a member is reduced by the presence of similar members)

2110773-9 2/2567

16

Tournament selection – เลือกสมาชิก 1 คู่จากประชากรรุ่นปัจจุบันแบบสุ่ม แล้วจึงสุ่มเลือกระหว่างสมาชิกคู่นั้น โดยสมาชิกตัวที่มีค่าความเหมาะสมสูงกว่าจะถูกสุ่มเลือกด้วยค่าความน่าจะเป็นที่กำหนดไว้ล่วงหน้า  $p$  ส่วนสมาชิกตัวที่มีค่าความเหมาะสมต่ำกว่าจะถูกสุ่มเลือกด้วยความน่าจะเป็น  $(1-p)$

Rank selection – เป็นวิธีที่ใช้ควบคุมการเลือกโครโมโซมโดยไม่สนใจค่าความเหมาะสมของโครโมโซมว่ามีค่าเท่าไร แต่จะใช้ค่าความเหมาะสมเพียงแค่จัดลำดับเรียงโครโมโซมตามค่าความเหมาะสมที่มีค่าสูงสุดจนถึงต่ำสุด จากนั้น กำหนดค่าคงที่  $p$  เป็นความน่าจะเป็นที่โครโมโซมลำดับที่ 1 จะถูกเลือก และเป็นความน่าจะเป็นที่โครโมโซมลำดับที่ 2 จะถูกเลือกเมื่อโครโมโซมลำดับที่ 1 ไม่ถูกเลือก และเป็นความน่าจะเป็นที่โครโมโซมลำดับที่ 3 จะถูกเลือกเมื่อลำดับที่ 1 และ 2 ไม่ถูกเลือก เรื่อยไปจนกระทั่งถึงลำดับสุดท้ายซึ่งจะถูกเลือกเมื่อลำดับก่อนหน้าไม่ถูกเลือกเลย ดังนั้น ความน่าจะเป็นที่โครโมโซมลำดับที่  $r$  จะถูกเลือก เท่ากับ  $p(1-p)^{r-1}$ ;  $r = 1, 2, 3, \dots$

Hypothesis	Fitness( $h_i$ )	$P(h_i)$	Rank	Probability of Rank( $h_i$ )
$h_1$	4	0.4	1	$=p = 0.667$
$h_2$	3	0.3	2	$=p(1-p) = 0.667 \times 0.333 = 0.222$
$h_3$	2	0.2	3	$=p(1-p)^2 = 0.667(0.333)^2 = 0.074$
$h_4$	1	0.1	4	$=p(1-p)^3 = 0.667(0.333)^3 = 0.025$
$h_5$	0	0	5	$=p(1-p)^4 = 0.667(0.333)^4 = 0.012$

## Example GA Application

ตัวอย่างชุดข้อมูลสอนการจำแนกประเภทลักษณะลูกค้าที่สนใจ Life Insurance Promotion

Training Instance	Income Range	Life Insurance Promotion	Credit Card Insurance	Sex	Age
1	30-40K	Yes	Yes	Male	30-39
2	30-40K	Yes	No	Female	40-49
3	50-60K	Yes	No	Female	30-39
4	20-30K	No	No	Female	50-59
5	20-30K	No	No	Male	20-29
6	30-40K	No	No	Male	40-49

## Fitness function

$$F(E_i) = N / (M+1)$$

where

$N$  คือ จำนวนตัวอย่างในชุดข้อมูลสอนซึ่งอยู่ในคลาสเดียวกับ  $E_i$  ที่มีค่าคุณลักษณะตรงกัน

$M$  คือ จำนวนตัวอย่างในชุดข้อมูลสอนซึ่งไม่อยู่ในคลาสเดียวกับ  $E_i$  ที่มีค่าคุณลักษณะตรงกัน

หมายเหตุ เพื่อป้องกันการหารด้วยค่าศูนย์ จึงบวกหนึ่งที่ตัวหารในฟังก์ชันความเหมาะสม

## กฎการคัดเลือกสมาชิกของประชากรรุ่นถัดไป

- สมาชิกที่มีค่าความเหมาะสมน้อยกว่า ค่าขีดแบ่ง (threshold) ที่กำหนดไว้มีค่าเท่ากับ 1 จะถูกนำไป crossover หรือ mutation
- ถ้าสมาชิกทุกตัวมีค่าความเหมาะสมไม่น้อยกว่าค่าขีดแบ่ง จะสุ่มเลือกสมาชิกเพื่อไป crossover หรือ mutation
- ในประชากรทุกรุ่น จะต้องมีสมาชิกที่อยู่ในคลาส “Yes” และ “No” อย่างละ 2 ตัว



กำหนดขนาดประชากรแต่ละรุ่นเท่ากับ 4 ตัว  
Income Range = ? → don't care

สมาชิกประชากรรุ่นแรก

#### ● An Initial Population for Supervised Genetic Learning

Population Element	Income Range	Life Insurance Promotion	Credit Card Insurance	Sex	Age
1	20-30K	No	Yes	Male	30-39
2	30-40K	Yes	No	Female	50-59
3	?	No	No	Male	40-49
4	30-40K	Yes	Yes	Male	40-49

2110773-9 2/2567

21

## ตัวอย่างการคำนวณ $F(E_1)$

- Income Range = 20-30K ตรงกับตัวอย่างสอนที่ 4 และ 5
- Credit Card Insurance = Yes ไม่ตรงกับตัวอย่างสอนใดๆ
- Sex = Male ตรงกับตัวอย่างสอนที่ 5 และ 6
- Age = 30-39 ไม่ตรงกับตัวอย่างสอนใดๆ
- จะได้ว่า ค่า  $N = 2+0+2+0 = 4$
- Income Range = 20-30K ไม่ตรงกับตัวอย่างสอนใดๆ
- Credit Card Insurance = Yes ตรงกับตัวอย่างสอนที่ 1
- Sex = Male ตรงกับตัวอย่างสอนที่ 1
- Age = 30-39 ตรงกับตัวอย่างสอนที่ 1 และ 3
- จะได้ว่า ค่า  $M = 0+1+1+2 = 4$

2110773-9 2/2567

22

การไขว้เปลี่ยนเพื่อสร้างสมาชิกลักษณะใหม่

- $F(E_1) = 4/5 = 0.80$ ,  $F(E_2) = 6/7 = 0.86$ ;  $F(E_1)$  and  $F(E_2) < 1$  → crossover
- $F(E_3) = 6/5 = 1.20$ ,  $F(E_4) = 5/5 = 1.00$ ;  $F(E_3)$  and  $F(E_4) \geq 1$  → keep

Population Element	Income Range	Life Insurance Promotion	Credit Card Insurance	Sex	Age
#1	20-30K	No	Yes	Male	30-39
#2	30-40K	Yes	No	Fem	50-59

Population Element	Income Range	Life Insurance Promotion	Credit Card Insurance	Sex	Age
#2	30-40K	Yes	No	Fem	50-59
#1	20-30K	No	No	Fem	50-59

2110773-9 2/2567

23

GECCO 2025

#GECCO2025 CALLS PROGRAM PAPERS CONFERENCE INFO

# GECCO 2025 @ Málaga (hybrid)

## The Genetic and Evolutionary Computation Conference

### July 14 - 18, 2025

### Why GECCO?

The Genetic and Evolutionary Computation Conference (GECCO) presents the latest high-quality results in genetic and evolutionary computation since 1999. Topics include: genetic algorithms, genetic programming, swarm intelligence, complex systems, evolutionary combinatorial optimization and metaheuristics, evolutionary machine learning, learning for evolutionary computation, evolutionary multiobjective optimization, evolutionary numerical optimization, neuroevolution, real world applications, search-based software engineering, theory, benchmarking, reproducibility, hybrids and more.

THIS IS GECCO CALL FOR SPONSORS

2110773-9 2/2567

24