

^xOptimal Solution
best //



유전알고리즘

(Genetic Algorithm; GA)

유전자 알고리즘?

❖ 유전알고리즘의 원리/진화의 원리

생물학적 진화

- 부모의 선택
- 번식 과 자손의 돌연변이

진화의 주요 목표

- 부모보다 생물학적으로 더 나은 자손을 번식시키는 것
- 유전 알고리즘은 주로 다윈의 자연 선택에 의한 진화론을 기반으로 하며, 동일한 시뮬레이션을 시도함

진화의 원리

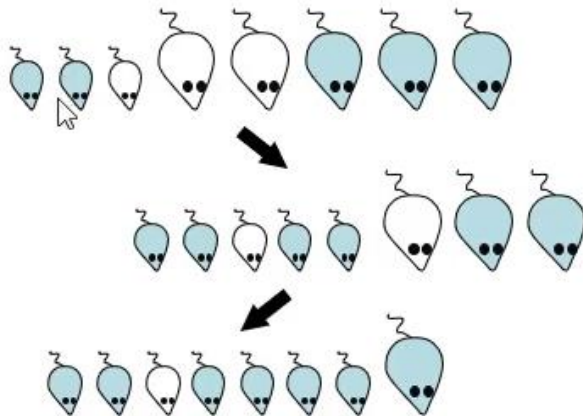
- 집단에서 부모로 최고(적응력이 높은)의 개인을 선택
- 부모의 교차
 - 자녀를 번식하고 낳아 세대를 확장
- 유전자가 돌연변이라는 오류가 발생하는 번식 과정 거침
- 자녀들은 다시 자손을 번식
- 이 과정이 계속되어 더 건강한 세대로 이어짐



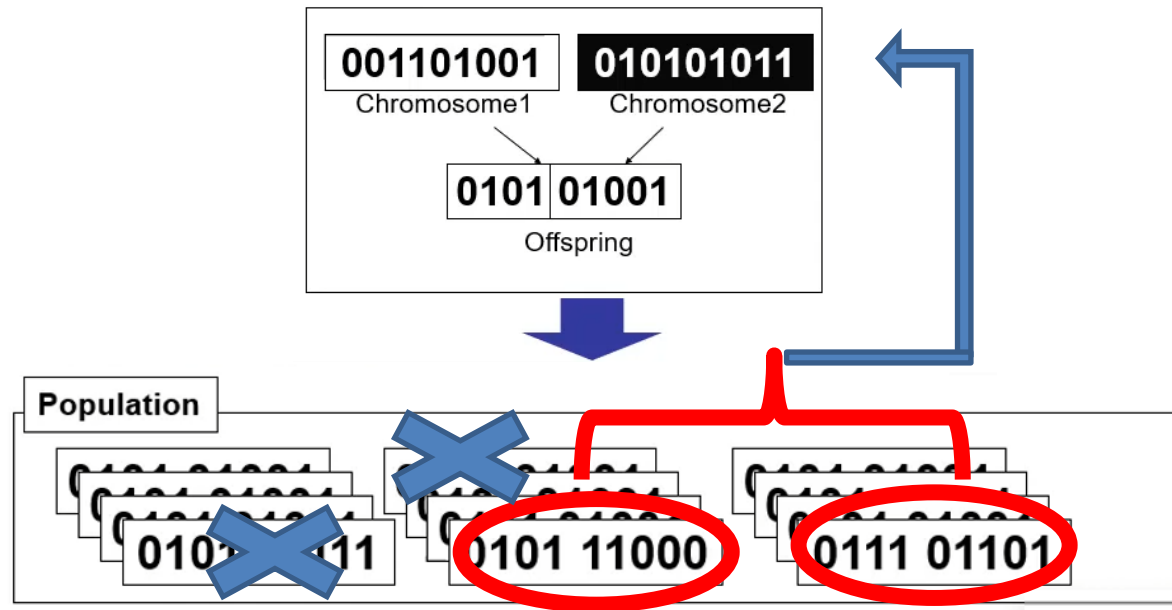
이 이론을 최적화 문제, 기능 선택, 고전적인 배낭 문제 등
여러 공학적 문제에 적용하여 훌륭한 해를 줄 수 있음

❖ 진화의 예시

- 실제 진화과정/ 유전알고리즘에서의 진화과정의 비교



색이 검을수록/몸집이 작을수록
생존에 유리하다



유전 알고리즘

다윈의 진화론으로부터 창안된 해 탐색 알고리즘

- 자연계에 있어서 생물의 유전(Genetics)과 진화(Evolution)의 메커니즘을 공학적으로 모델화
- 1975년에 John Holland가 저서 "Adaptation on Natural and Artificial System"에 처음 소개
- 적자생존의 개념을 최적화 문제를 해결하는데 적용
 - 개체군(population)중에서 환경에 대한 적합도(fitness)가 높은 개체일수록 재생산할 수 있게 되며, 개체군은 환경에 적응을 할 수 있게 됨

유전 알고리즘의 원리

- 원시 인간의 유전체(가정)
 - 염색체는 0, 1의 6개의 유전자로 구성한다고 가정

| | Color | Speed | Intelligence | Fitness |
|---|---------|--------|--------------|---------|
|  | 1 White | Medium | Dumb | 40 |
|  | 2 Black | Slow | Dumb | 43 |
|  | 3 White | Slow | Very dumb | 22 |
|  | 4 Black | Fast | Dumb | 71 |
|  | 5 White | Medium | Very smart | 53 |

White/Medium/Very Dumb

1 0 1 / 1 1 0

0 0 0 / 0 0 0

Black/Fast/Very Smart

- 각 유전자의 배열 구성에 따라 인간의 신체 특징은 달라짐
 - 피부색
 - 첫 1비트 (흑인(0)/백인(1)) : 흑인이 생존에 유리하다
 - 주행속도
 - 다음 2비트(빠름(00)/보통(01)/느림(10)) : 빠를수록 적자
 - 지능
 - 다음 3비트(매우 똑똑/(000)/ ... /Very Dumb(110))
 - 적합도
 - 신체특징을 종합하여 평가

- 상위 두 개체를 교배하여 후손을 낳으면?
- 하위 두 개체를 교배하여 후손을 낳으면?

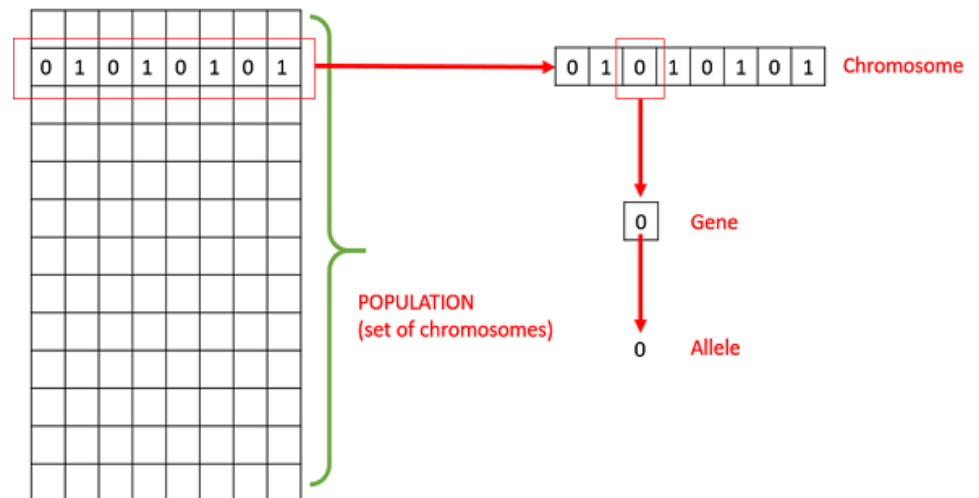
유전알고리즘의 기본 용어(1)

염색체(Chromosome)

- 개체라고도 하며 알고리즘의 핵심 주체
- 문제의 최적해는 가장 좋은 염색체를 찾는 것
- 여러 개의 유전자(gene)로 구성되며 문제마다 달라짐
 - 어떻게 구성하느냐에 따라 알고리즘의 성능이 좌우됨
- 자연계의 염색체가 가변적인 것과 달리, 대부분의 경우 고정 길이를 가진다.
- 표현방법 : 이진 표현, k진수 표현, 그레이 코드, 실수 표현, 트리 표현

개체집단(Population) 또는 개체군

- 염색체의 집단
- 후보해(Candidate Solution)들의 집합을 말함



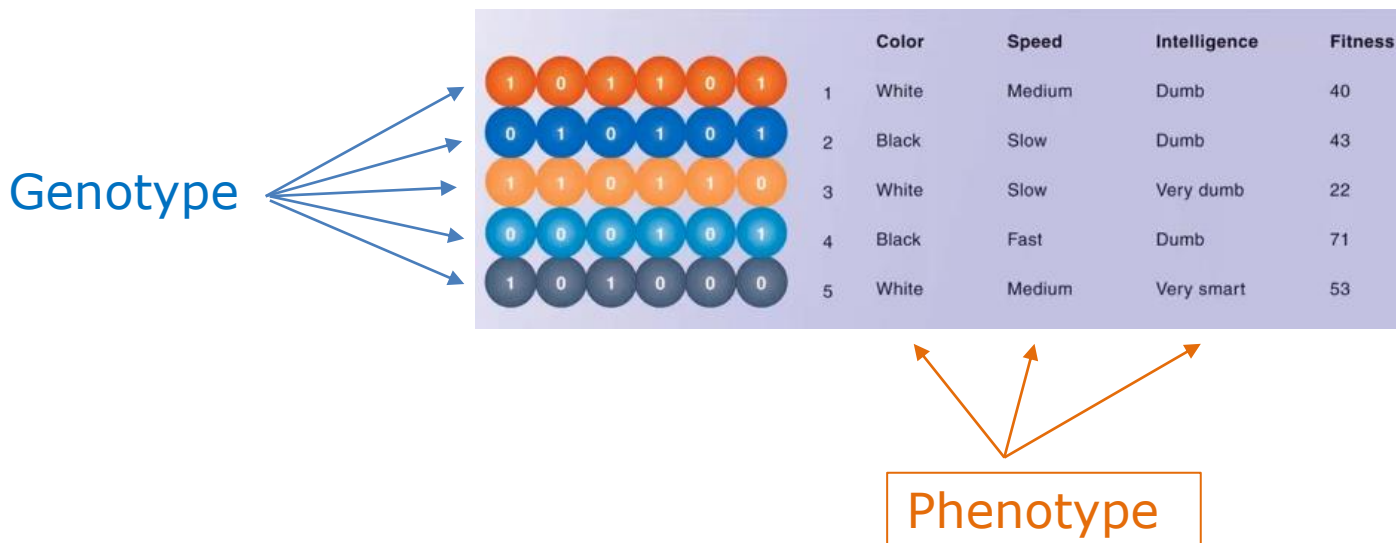
유전 알고리즘의 기본 용어(2)

Genotype(유전자형)

- 계산 공간의 모집단
- 계산 공간에서 솔루션은 컴퓨팅 시스템을 사용하여 쉽게 이해하고 조작할 수 있는 방식으로 표현됨

Phenotype(표현형)

- 솔루션이 실제 상황에서 표현되는 방식으로 표현됨
- 실제 솔루션 공간

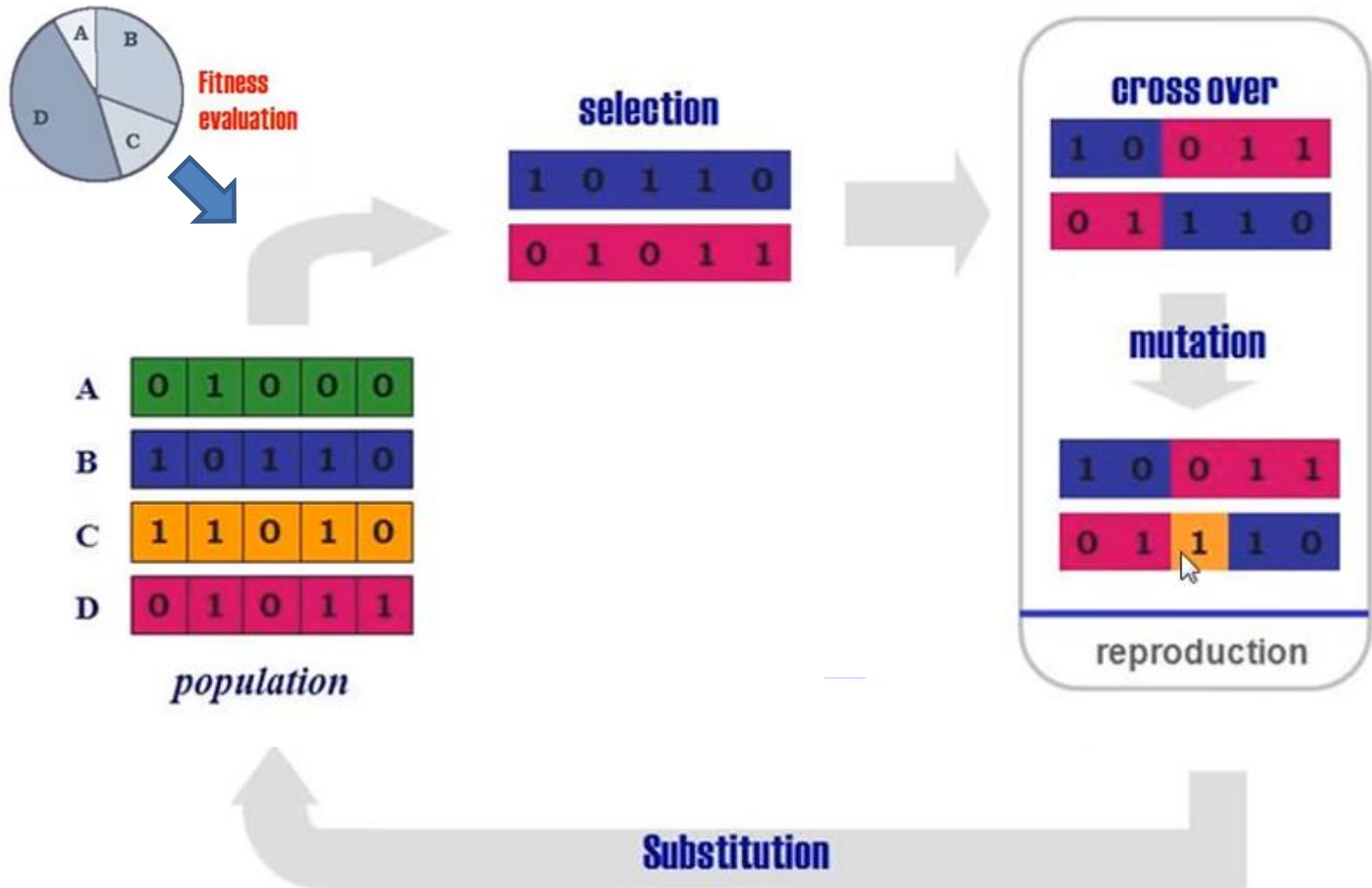


유전알고리즘의 기본 용어(3)

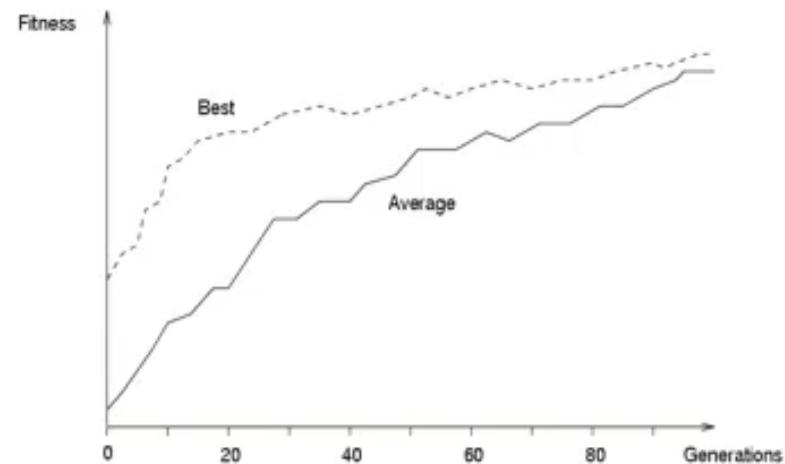
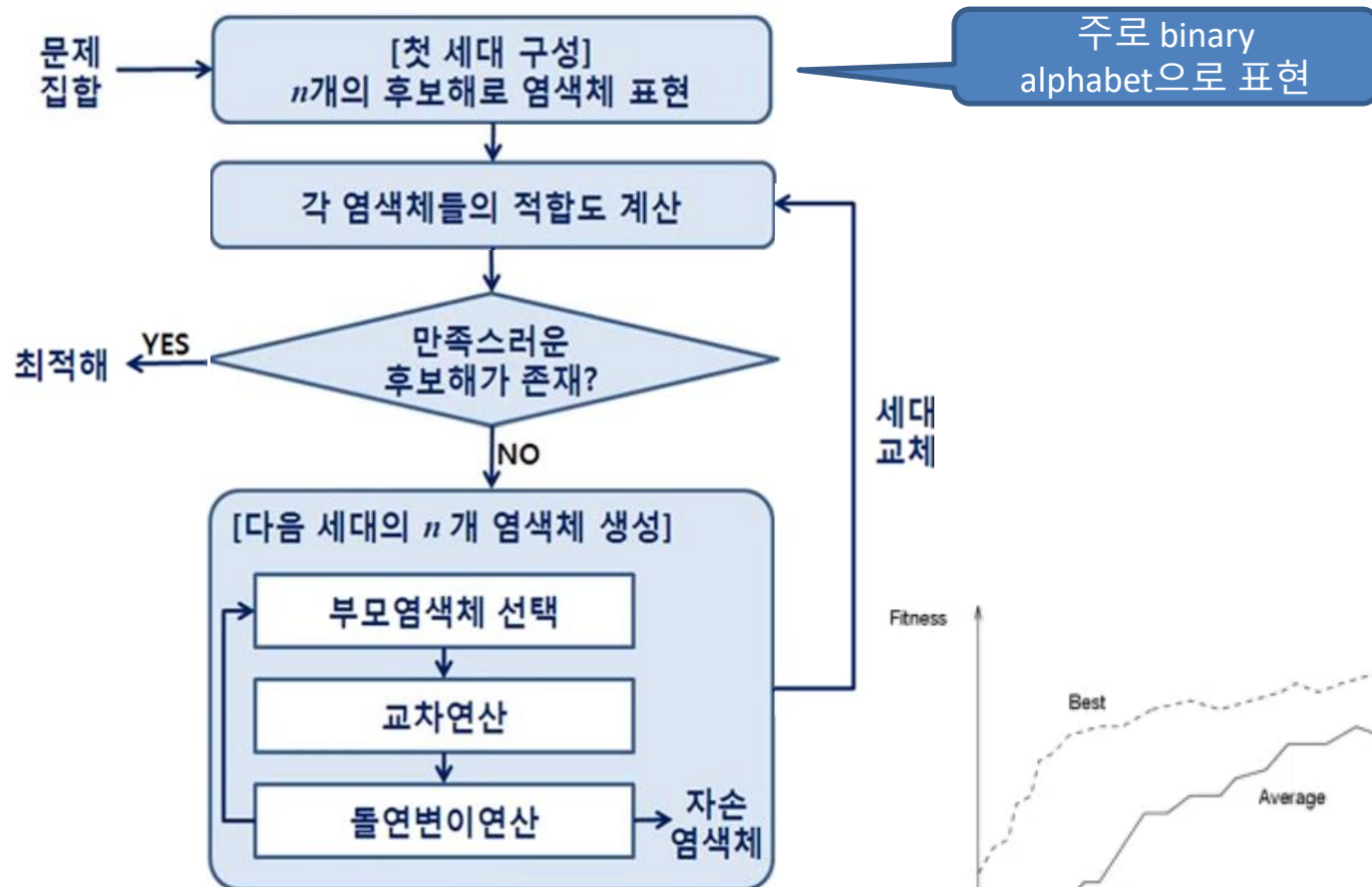
평가 함수(Evaluation function) / 적합도 함수(Fitness function)

- 염색체가 얼마나 Goal과 가까운지를 평가함
 - 유전자의 품질을 결정하여 우수 유전자를 선별
 - 유전자가 유효한 해인지를 판단하는 함수
 - 계산하기에 충분히 빨라야 함
- 평가 함수 결과가 기준 값을 넘으면 알고리즘 종료

유전알고리즘의 흐름도1(예제)



유전알고리즘의 흐름도2



유전 알고리즘의 연산자(유전과정표현)

선택 연산(Selection)

- 품질 비례 룰렛 휠 선택
- 토너먼트 선택

교차 연산(Cross-Over)

- 교배라고도 함.
- 1점 교차, 다점 교차, 균등 교차, 싸이클 교차, 순서 교차, PMX(Partially matched Crossover), 산술적 교차, 휴리스틱 교차, 간선 재결합 등

변이 연산(Mutation)

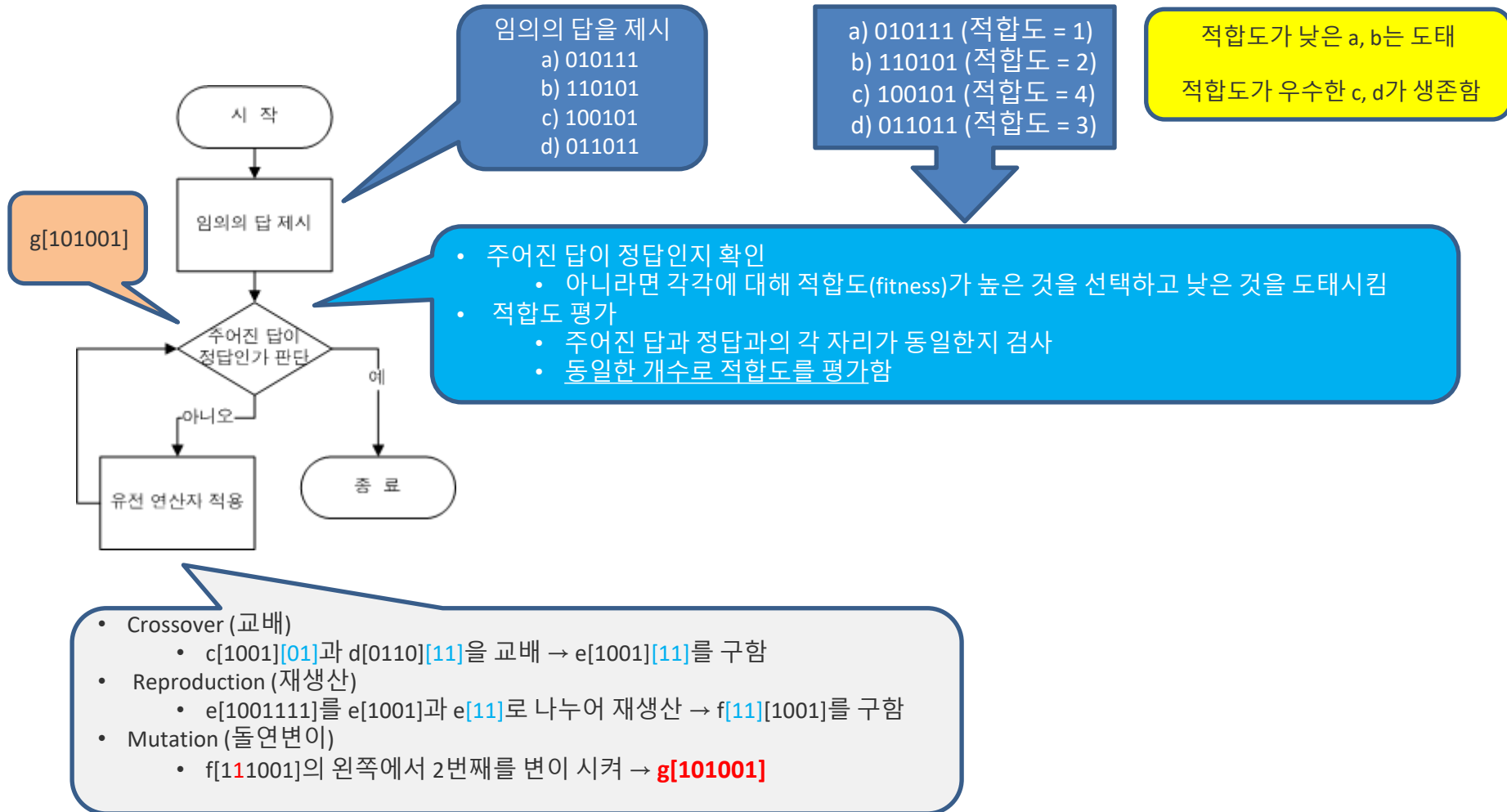
- 부모 해에 없는 속성을 도입하여 탐색 공간을 넓히려는 목적을 가진 연산
- 지역 최적해에 빠지는 것을 완화시켜 준다.

대치 연산(reproduction; 재생산, 번식)

- 품질이 나쁜 유전자를 품질이 좋은 유전자로 바꾸는 연산
- 유전자의 수를 일정하게 유지하기 위한 연산
- 일반적으로 가장 우수한 해는 대치되지 않음(엘리트주의; Elitism)

유전 알고리즘의 예

- ❖ 문제(상대방의 숫자 알아내기)
 - ❖ 0과 1로 이루어진 6자리 수 맞추어 보자
 - ❖ **정답 : 101001**
 - ❖ 랜덤의 경우 $2^6 = 64$ 번, 평균 32번의 추측 필요함



유전알고리즘의 연산

초기 염색체 생성

- 초기에는 이전 염색체가 존재하지 않기 때문에 자손을 생성할 수가 없음
- 초기 염색체를 생성하는 연산을 별도로 정의
- 가장 많이 이용되는 방법은 어떠한 규칙도 없이 단순히 임의의 값으로 염색체를 생성

적합도 계산

- 염색체에 표현된 정보를 기반으로 적합도를 계산하는 연산
- 해결하고자 하는 문제에 매우 종속
- 문제가 매우 복잡하여 계산이 가능하지 않을 경우, 근사치를 계산할 수도 있음

염색체 선택

- 훌륭한 부모 선택은 GA의 수렴 속도에 매우 중요
- 적합도가 높은 개체 위주로 선발하며 룰렛 휠 선택 등 확률적인 개념으로 선택
- 단순히 적합도가 가장 높은 두 개의 염색체를 선택하는 것은 다양성 훼손으로 지양함

자손 생성

- 선택된 두 개의 부모 염색체 들을 교배(Crossover) 하여 자손 염색체(offspring)를 생성
- 초기 염색체를 생성하는 연산을 별도로 정의하며 가장 많이 이용되는 방법은 어떠한 규칙도 없이 단순히 임의의 값으로 염색체를 생성

돌연변이

- 새롭게 생성된 염색체에 확률적으로 돌연변이가 발생하도록 함
- 일반적으로 0.1%, 0.05% 등의 아주 낮은 확률로 돌연변이가 발생하도록 설정
- 지역 최적점에 빠지는 문제를 해결하기 위해 사용되며 Reverse, exchange 등 매우 다양

핵심

염색체 선택연산 1

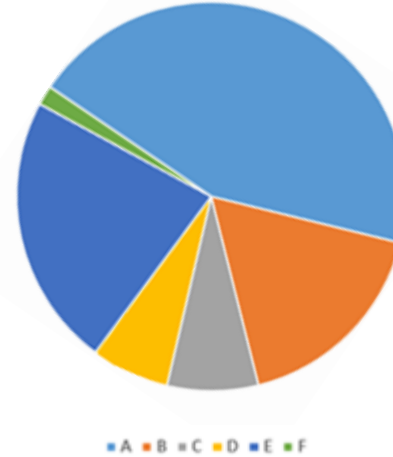
Fitness Proportionate Selection(적합도비례 선택)

- 가장 일반적인 부모 선택 방법
- Roulette Wheel Selection (룰렛 휠 선택), Tournament Selection 등이 있음

룰렛 휠 선택

| Chromosome | Fitness Value |
|------------|---------------|
| A | 8.2 |
| B | 3.2 |
| C | 1.4 |
| D | 1.2 |
| E | 4.2 |
| F | 0.3 |

적합도에
비례하는 룰렛



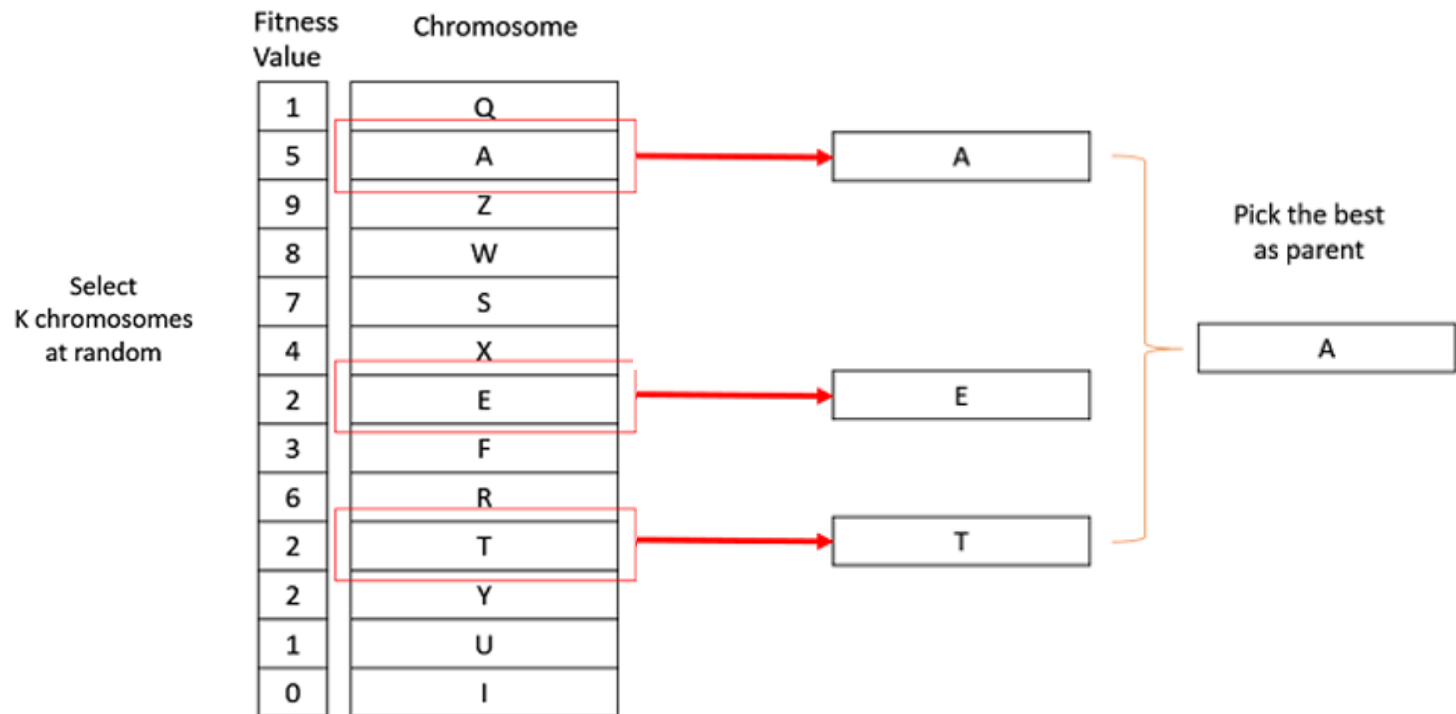
선택절차

- Sum을 계산 = fitnesses의 합계.
- 0과 Sum 사이의 난수 생성(S)
- 모집단의 맨 위에서 시작하여 $P < \text{Sum}$ 까지 부분 합계 P에 적합도를 계속 더함
- P가 S를 초과하는 염색체를 선택
 - 예를 들어 Sum= 18.5, 난수 S=12.1이면 c가 선택됨

염색체 선택연산2

- Tournament Selection

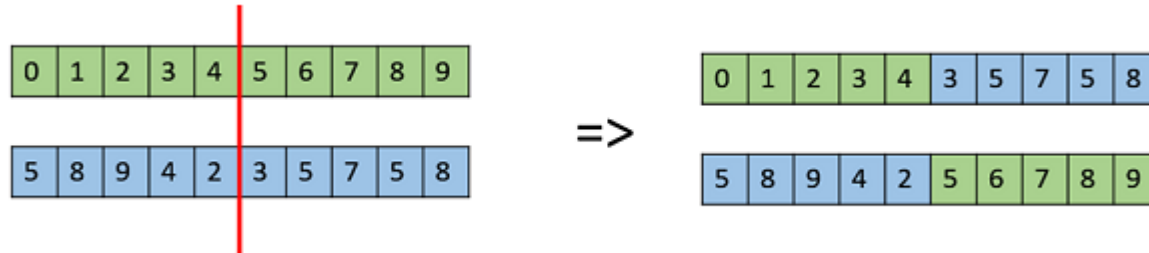
- K-Way 토너먼트 선택에서 우리는 무작위로 K 명의 개인을 선택하고 이들 중 최고를 선택



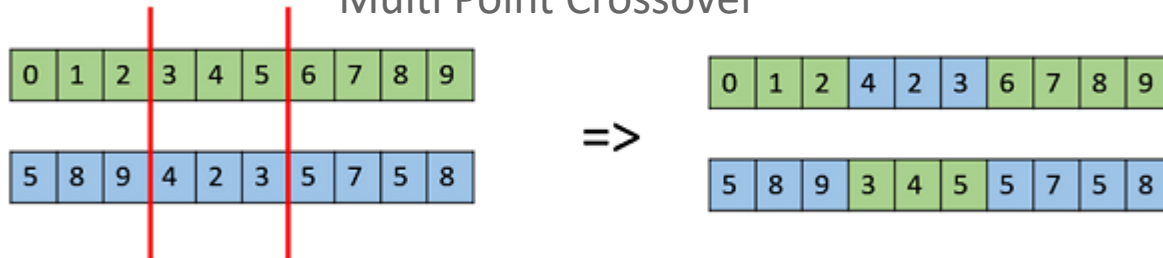
자손생성 연산 1 (Crossover; 교차)

- 교차 (혹은 교배)는 생식 및 생물학적 교차와 유사
- 하나 이상의 부모가 선택되고 부모의 유전 물질을 사용하여 하나 이상의 자손이 생성됨
- 교차 메커니즘의 종류
 - One Point Crossover, Multi Point Crossover, Uniform Crossover

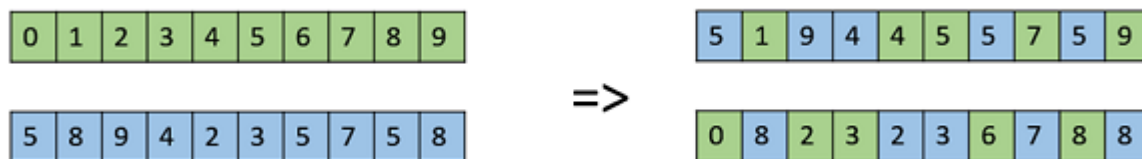
One Point Crossover



Multi Point Crossover



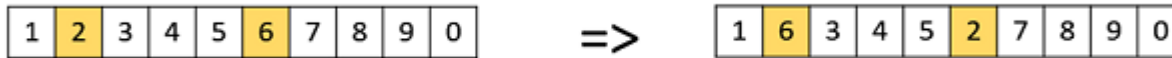
Uniform Crossover



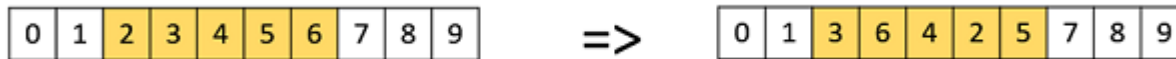
자손생성 연산 2 (Mutation; 돌연변이)

- 지역 최적점에 빠지는 문제를 해결하기 위해 새롭게 생성된 염색체에 확률적으로 돌연변이가 발생하도록 함.
 - 일반적으로 0.1%, 0.05% 등의 아주 낮은 확률로 돌연변이가 발생하도록 설정하며
 - 염색체에서 돌연변이를 발생시키는 연산은 매우 다양

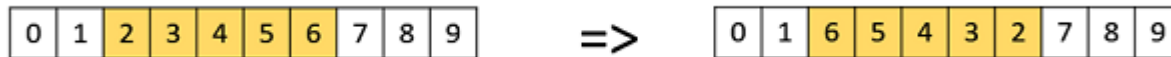
Swap Mutation(exchange; 교환)



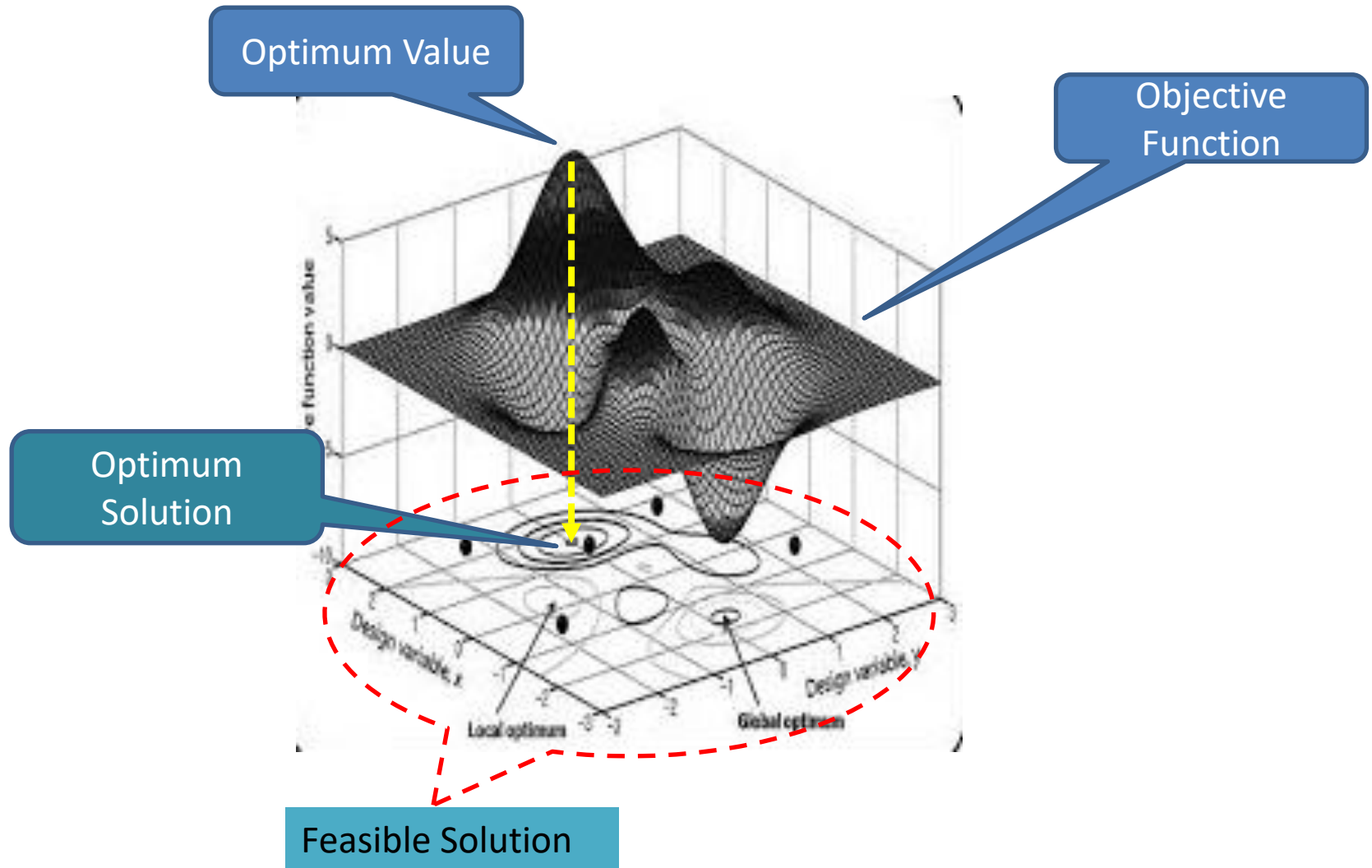
Scramble Mutation(random; 섞기)



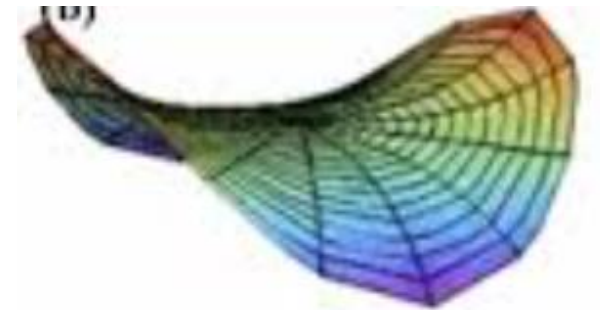
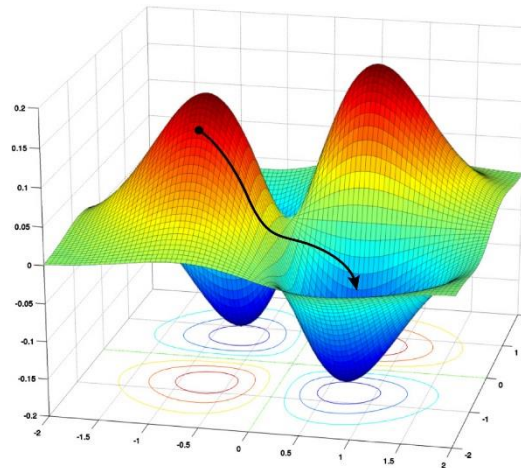
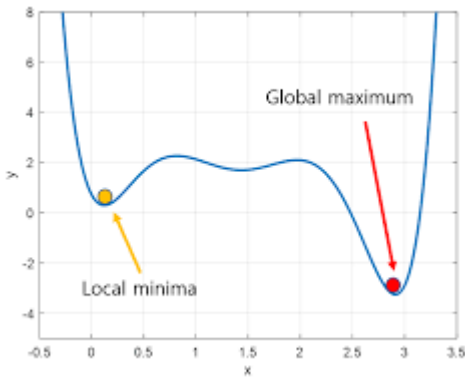
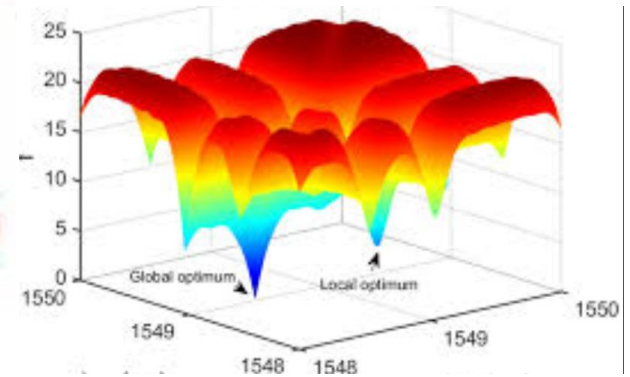
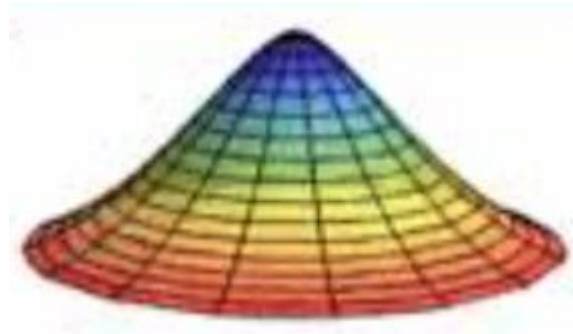
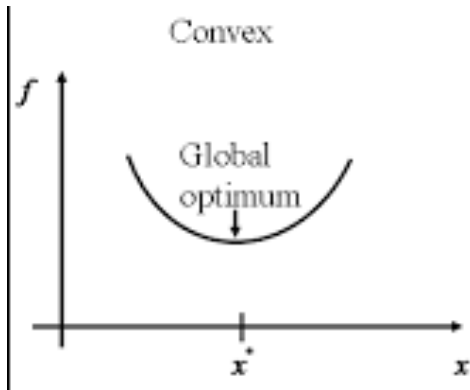
Reverse Mutation (역배열)



최적화



Objective Function의 종류



Unimodal function(단봉형함수)를 찾아보자

변이와 Global Optimum

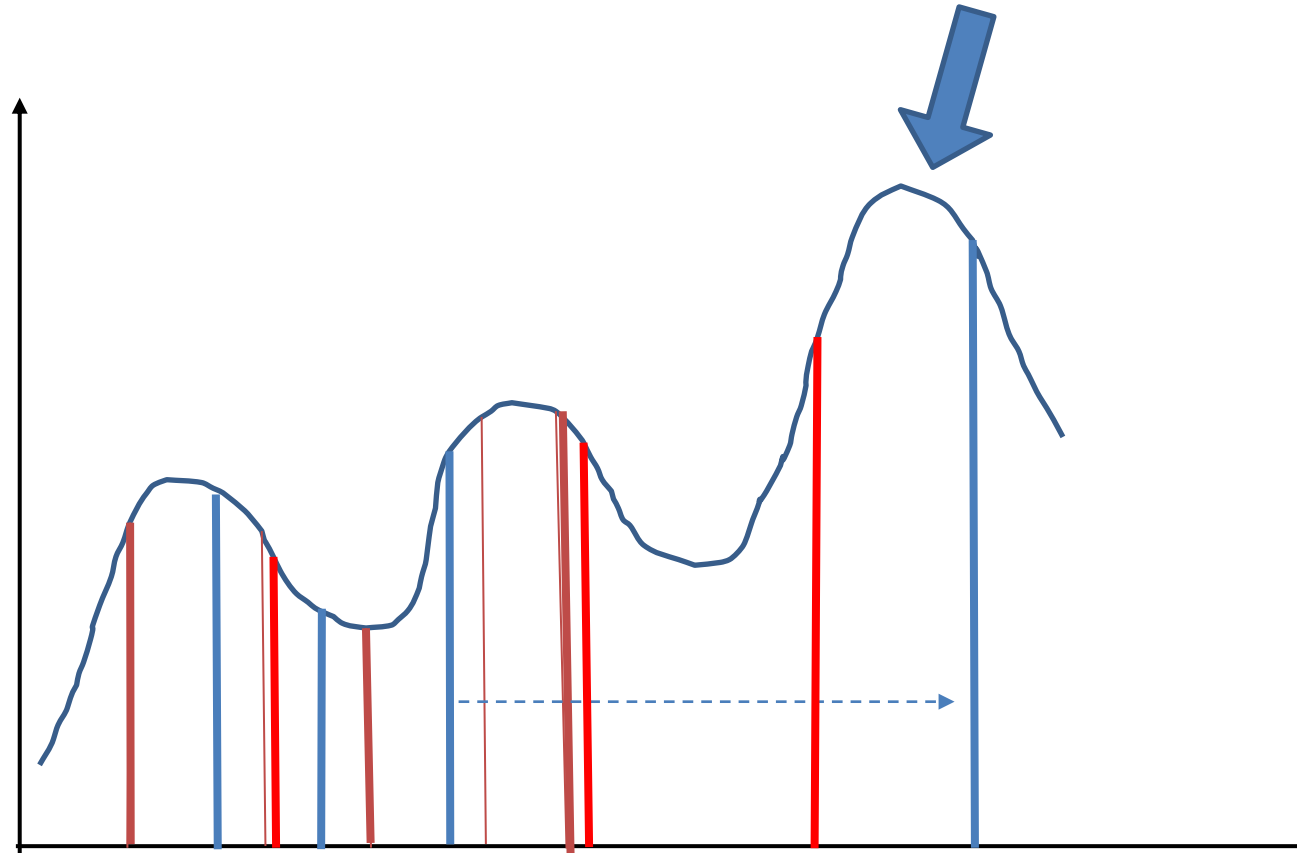
1세대 population

→ 1세대 부모선택

→ 2세대 자손생성

→ 변이발생

→ 3세대 자손생성

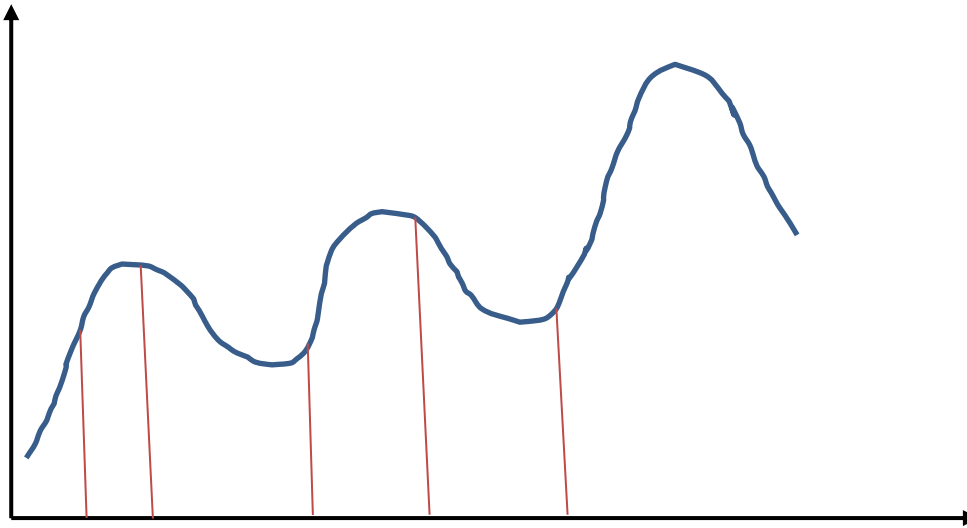


Global Optimum을 놓칠
가능성이 매우 높다

Global Optimum을
탐색할 가능성이 생겼다

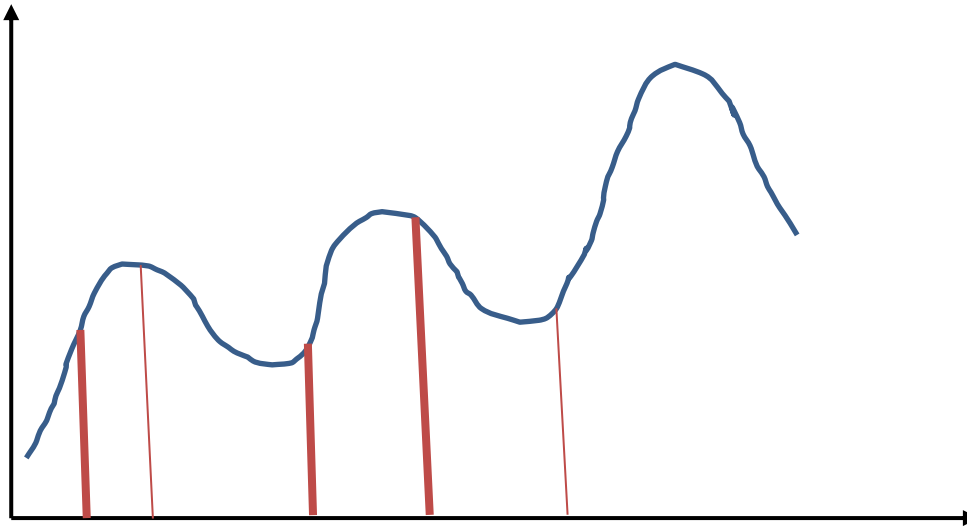
변이와 Gloval Optimum

- 1세대 population



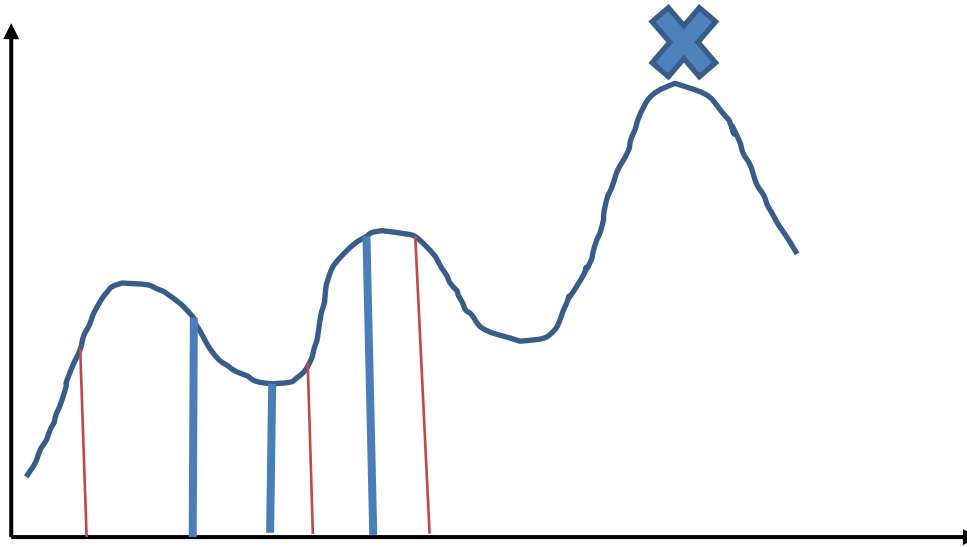
변이와 Gloval Optimum

- 1세대 부모선택



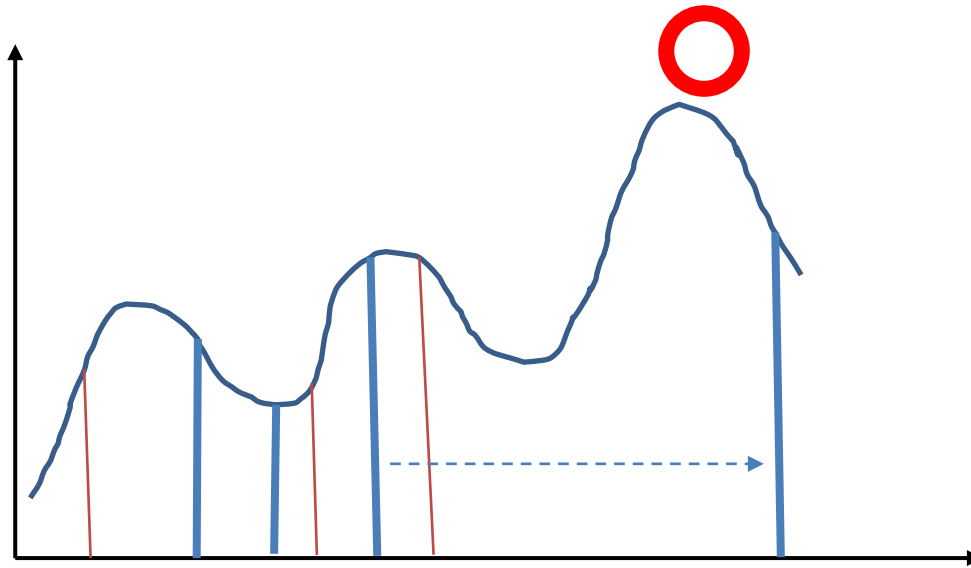
변이와 Gloval Optimum

- 2세대 자손생성



변이와 Gloval Optimum

- 변이발생



유전 알고리즘의 종료조건

종료조건(Termination Condition)이란?

- 초기에 GA는 반복 작업마다 더 나은 솔루션을 제공하면서 매우 빠르게 진행
- 그러나 후반 단계에서 개선이 매우 적은 경우가 대부분
- GA 실행이 언제 종료 될지를 결정하는 데 중요

종료 조건 의 유형

- 반복에 대한 모집단의 개선이 없는 경우.
- 목적 함수 값이 미리 정의 된 특정 값에 도달했을 때.
- 절대 세대에 도달했을 때.

몇 세대 동안 개체집단의 적합도가 정체될 수 있음.
이럴 경우 정해진 세대수에 도달하면 알고리즘을
종료하고 지금까지의 해집합 중에 최적을 찾음

유전알고리즘의 예제

•다음의 2차 함수에 대해 유전자 알고리즘으로 $0 \leq x \leq 31$ 구간에서 최대값을 찾아보자.

$$f(x) = -x^2 + 38x + 80$$

•한 세대의 후보해 수를 4로 가정

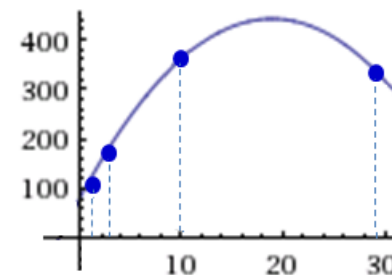
•초기 세대 후보해

-0~31에서 랜덤하게 4개의 후보해인 1, 29, 3, 10을 선택

1세대 후보해의 적합도

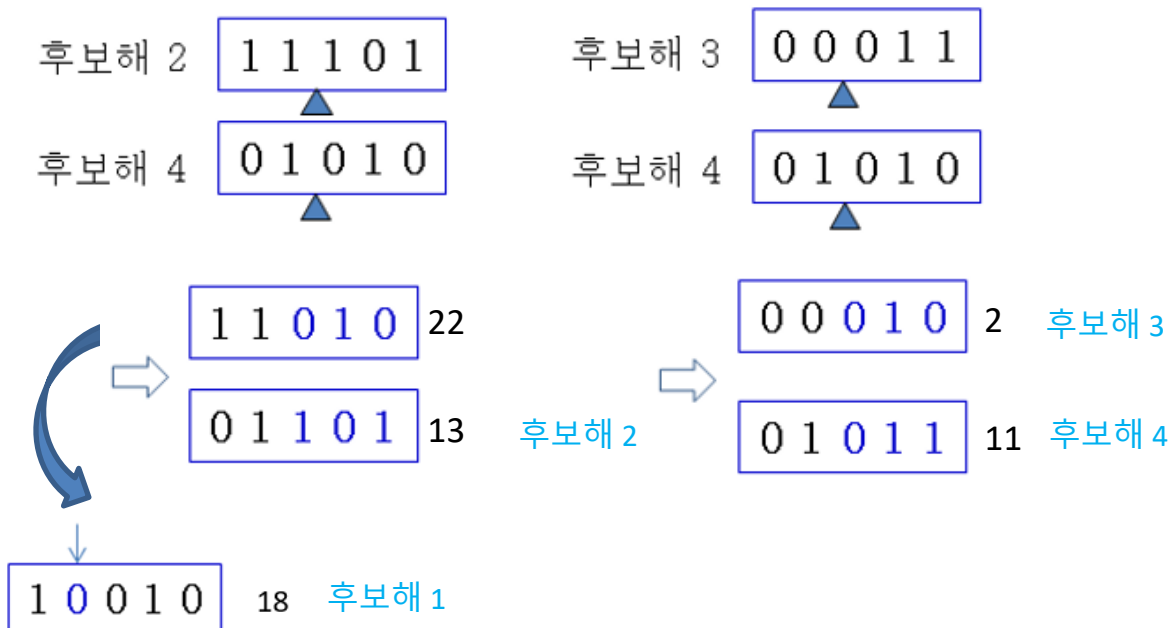
| 후보해 | 2진 표현 | x | 적합도 f(x) | 룰렛구성비율(%) |
|-----|-----------|----|-------------|-----------|
| 1 | 0 0 0 0 1 | 1 | 117 | 12 |
| 2 | 1 1 1 0 1 | 29 | 341 | 34 |
| 3 | 0 0 0 1 1 | 3 | 185 | 18 |
| 4 | 0 1 0 1 0 | 10 | 360 | 36 |
| 계 | | | 1,003 | 100 |
| 평균 | | | 250.75 | |

$f(1) = -(1)^2 + 38(1) + 80 = 117,$
 $f(29) = 341,$
 $f(3) = 185,$
 $f(10) = 360$



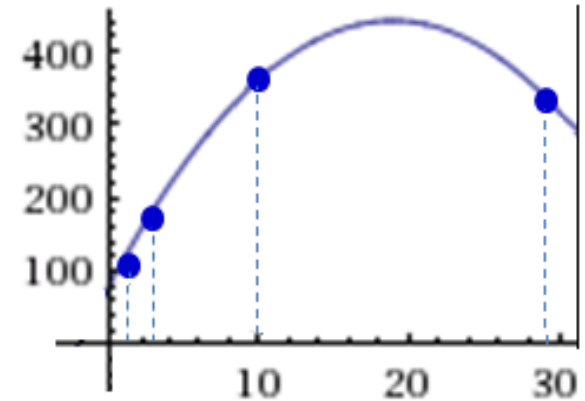
유전알고리즘의 예제(계속)

- 선택 연산
 - 룰렛 휠 선택 방법을 이용하여, 후보해 4는 2번, 후보해 2와 3은 각각 1번 선택, 후보해 1은 0번 선택되었다고 하자
- 교차 연산
 - (후보해 2, 후보해 4), (후보해 3, 후보해 4) 를 1점-교차 연산
- 돌연변이
 - 후보해 1의 왼쪽에서 두 번째 bit가 돌연변이가 되어서 '1'에서 '0'으로 바뀌었다고 가정하자



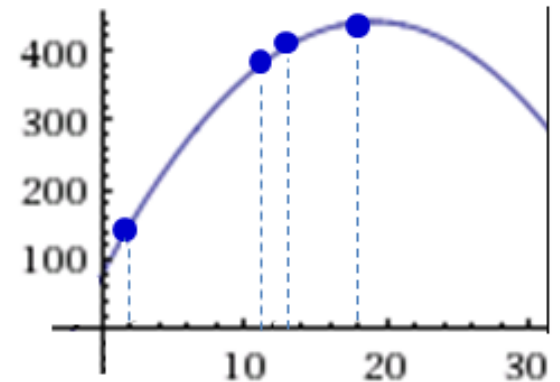
1세대 후보해의 적합도

| 후보해 | 2진 표현 | x | 적합도 f(x) | 룰렛구성비율(%) |
|-----|-----------|----|----------|-----------|
| 1 | 0 0 0 0 1 | 1 | 117 | 12 |
| 2 | 1 1 1 0 1 | 29 | 341 | 34 |
| 3 | 0 0 0 1 1 | 3 | 185 | 18 |
| 4 | 0 1 0 1 0 | 10 | 360 | 36 |
| 계 | | | 1,003 | 100 |
| 평균 | | | 250.75 | |



2세대 후보해의 적합도

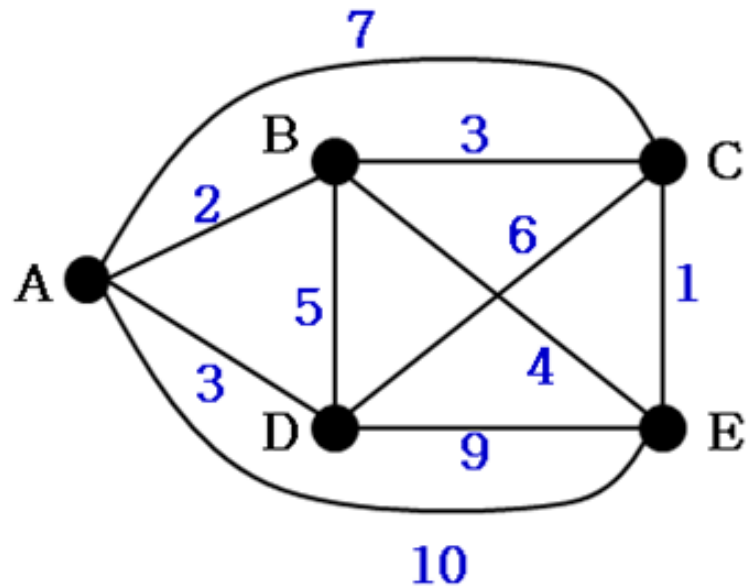
| 후보해 | 2진 표현 | x | 적합도 f(x) | 원반 면적 (%) |
|-----|-----------|----|----------|-----------|
| 1 | 1 0 0 1 0 | 18 | 440 | 32 |
| 2 | 0 1 1 0 1 | 13 | 405 | 29 |
| 3 | 0 0 0 1 0 | 2 | 152 | 11 |
| 4 | 0 1 0 1 1 | 11 | 377 | 27 |
| 계 | | | 1,374 | 100 |
| 평균 | | | 343.5 | |



적합도상승
250.75→343.5

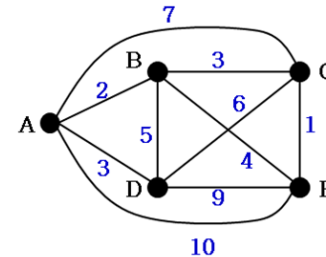
GeneticAlgorithm in TSP

세일즈맨은 A 도시에서 출발하여 4개의 도시(B, C, D, E)를 방문해야 하고, 현재 도시에서 다른 도시로 이동하는데 소모되는 비용은 아래의 표와 같다.



초기세대

- 초기 세대 생성
 - 일정한 규칙 없이 임의로 생성(population 크기는 5를 유지)
- 후보해 집합
 - 여행자 문제의 조건은 시작 도시에서 출발하여 모든 다른 도시를 1번씩만 방문하고 시작 도시로 돌아와야 함
 - 만일 n 개의 도시가 있다면, 후보해의 수는 $(n-1)!$ 이다.
 - 이 문제의 후보해의 수는 시작 도시를 제외한 5개의 도시를 일렬로 나열하는 방법의 수와 같으므로 $5! = 120$ 이다.
- 초기 Population
 - 후보해 중에서 5개 고르면 됨
 - ABCDEA, ACDEBA, AECDBA, AEDCBA, ADCEBA



- 초기 세대 적합도 평가
 - TSP 문제는 거리의 합이 최소를 목표로 함
 - 적합도 함수는 $100 - (\text{거리의 합})$ 으로 가정

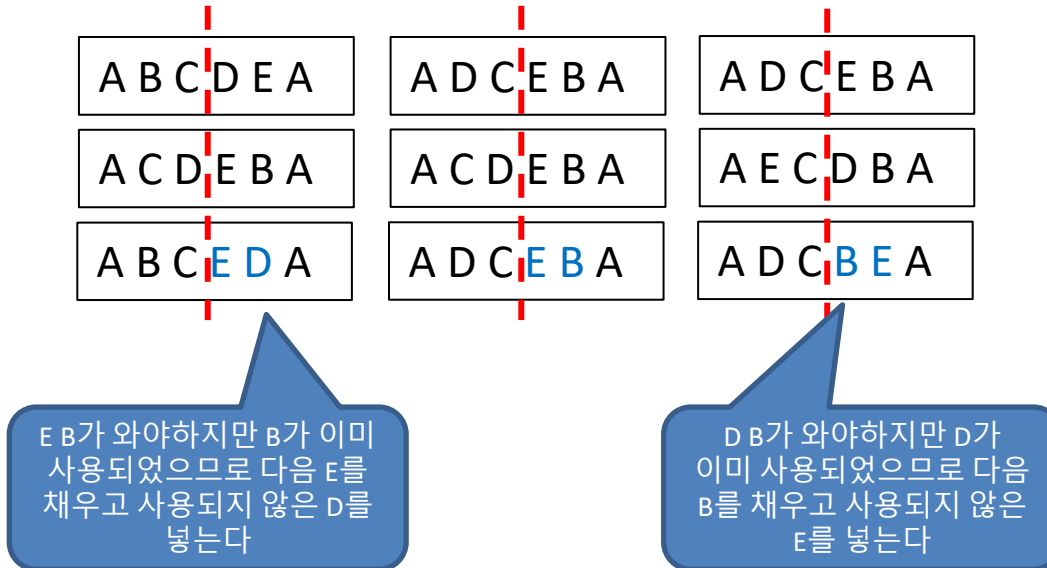
- $\text{Fit}(A B C D E A) = 100 - (2+3+6+9+10) = 100 - 30 = 70$
- $\text{Fit}(A C D E B A) = 100 - (7+6+9+4+2) = 100 - 28 = 72$
- $\text{Fit}(A E C D B A) = 100 - (10+1+6+5+2) = 100 - 24 = 76$
- $\text{Fit}(A E D C B A) = 100 - (10+9+6+3+2) = 100 - 30 = 70$
- $\text{Fit}(A D C E B A) = 100 - (3+6+1+4+2) = 100 - 16 = 84$

- $P(A B C D E A) = 70/372 * 100 = 18.8\%$
- $P(A C D E B A) = 72 / 372 * 100 = 19.3\%$
- $P(A E C D B A) = 76 / 372 * 100 = 20.4\%$
- $P(A E D C B A) = 70/372 * 100 = 18.8\%$
- $P(A D C E B A) = 84/372 * 100 = 22.6\%$

적합도의 합 = 372
평균적합도 = 74.4
최고적합도 = 84

초기 세대 교배

- 부모선택
 - 룰렛으로 두개를 택하여 자손 생성
- 교차연산
 - 두 개의 염색체를 선택하여 교차연산을 통해 다음 세대의 유전체를 생성한다.
 - 1점교배를 해보자
 - 두 염색체의 유전자를 교차



엘리트 주의 및 돌연변이

- 엘리트주의

- 이전 세대에서 적합도가 가장 좋은 염색체를 다음 세대에 그대로 보존하는 것
 - 몇 개를 보존할지는 알고리즘의 설계에 따라 다름
 - 이 예제에서는 **적합도가 가장 높은 상위 2개의 염색체를 보존**
- 초기 세대에서 적합도가 가장 높은 [A D C E B A]와 [A E C D B A]를 다음 세대까지 그대로 보존하고 나머지 3개만 교배하여 생성

- 초기세대

- $\text{Fit}(\text{A B C D E A}) = 100 - (2+3+6+9+10) = 100 - 30 = 70$

- $\text{Fit}(\text{A C D E B A}) = 100 - (7+6+9+4+2) = 100 - 28 = 72$ ★

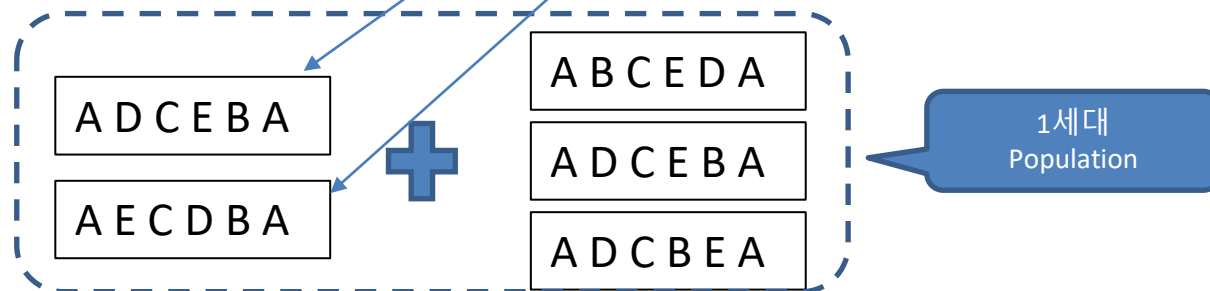
- $\text{Fit}(\text{A E C D B A}) = 100 - (10+1+6+5+2) = 100 - 24 = 76$

- $\text{Fit}(\text{A E D C B A}) = 100 - (10+9+6+3+2) = 100 - 30 = 70$ ★

- $\text{Fit}(\text{A D C E B A}) = 100 - (3+6+1+4+2) = 100 - 16 = 84$

- 돌연변이

- 이 문제에서는 발생하지 않았다고 가정하자



목표 적합도 검사

A D C E B A

$$\bullet \text{Fit}(\text{A D C E B A}) = 100 - (3+6+1+4+2) = 100 - 16 = 84$$

A E C D B A

$$\bullet \text{Fit}(\text{A E C D B A}) = 100 - (10+1+6+5+2) = 100 - 24 = 76$$

A B C E D A

$$\bullet \text{Fit}(\text{A B C E D A}) = 100 - (2+3+1+9+3) = 100 - 18 = 82$$

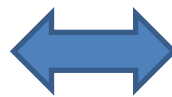
A D C E B A

$$\bullet \text{Fit}(\text{A D C E B A}) = 100 - (3+6+1+4+2) = 100 - 16 = 84$$

A D C B E A

$$\bullet \text{Fit}(\text{A D C B E A}) = 100 - (3+6+3+6+10) = 100 - 28 = 78$$

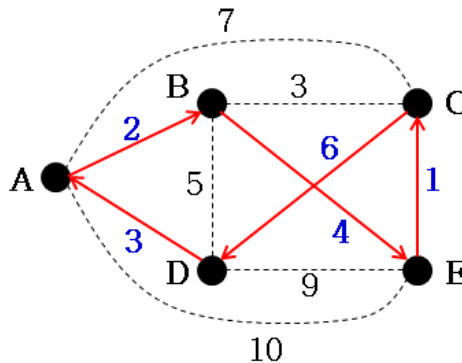
초기적합도의 합 = 372
초기 평균적합도 = 74.4
초기최고적합도 = 84



적합도의 합 = 404
평균적합도 = 80.8
최고적합도 = 84

계속진행여부를
판단해야 함

중단한다면
현재까지의
최고적합도를
해로 하고 종료



비교. Branch&Bound의 해

유전알고리즘의 파라미터

결정사항

- 유전자 알고리즘은 실제로 적지 않은 실험을 요구
- 다음과 같은 파라미터의 값들과 적절한 연산을 선택해야 한다.
 - 모집단 크기
 - 선택 연산
 - 교차 연산과 교차율
 - 돌연변이율
 - repeat-루프의 종료 조건

유전알고리즘의 응용

특징

- 유전자 알고리즘은 문제의 최적해를 알 수 없고, 기존의 어느 알고리즘으로도 해결하기 어려운 경우에,
- 최적해에 가까운 해를 찾는데 매우 적절한 알고리즘이다.
- 유전자 알고리즘이 최적해를 반드시 찾는다라는 보장은 없으나 대부분의 경우 매우 우수한 해를 찾는다.

응용 분야

- 통 채우기
- 작업 스케줄링
- 차량 경로
- 배낭 문제
- 로봇 공학
- 기계 학습 (Machine Learning)
- 신호 처리 (Signal Processing)
- 반도체 설계
- 항공기 디자인
- 통신 네트워크
- 패턴 인식
- 그 외에도 경제, 경영, 환경, 의학, 음악, 군사 등과 같은 다양한 분야에서 최적화 문제를 해결하는데 활용된다