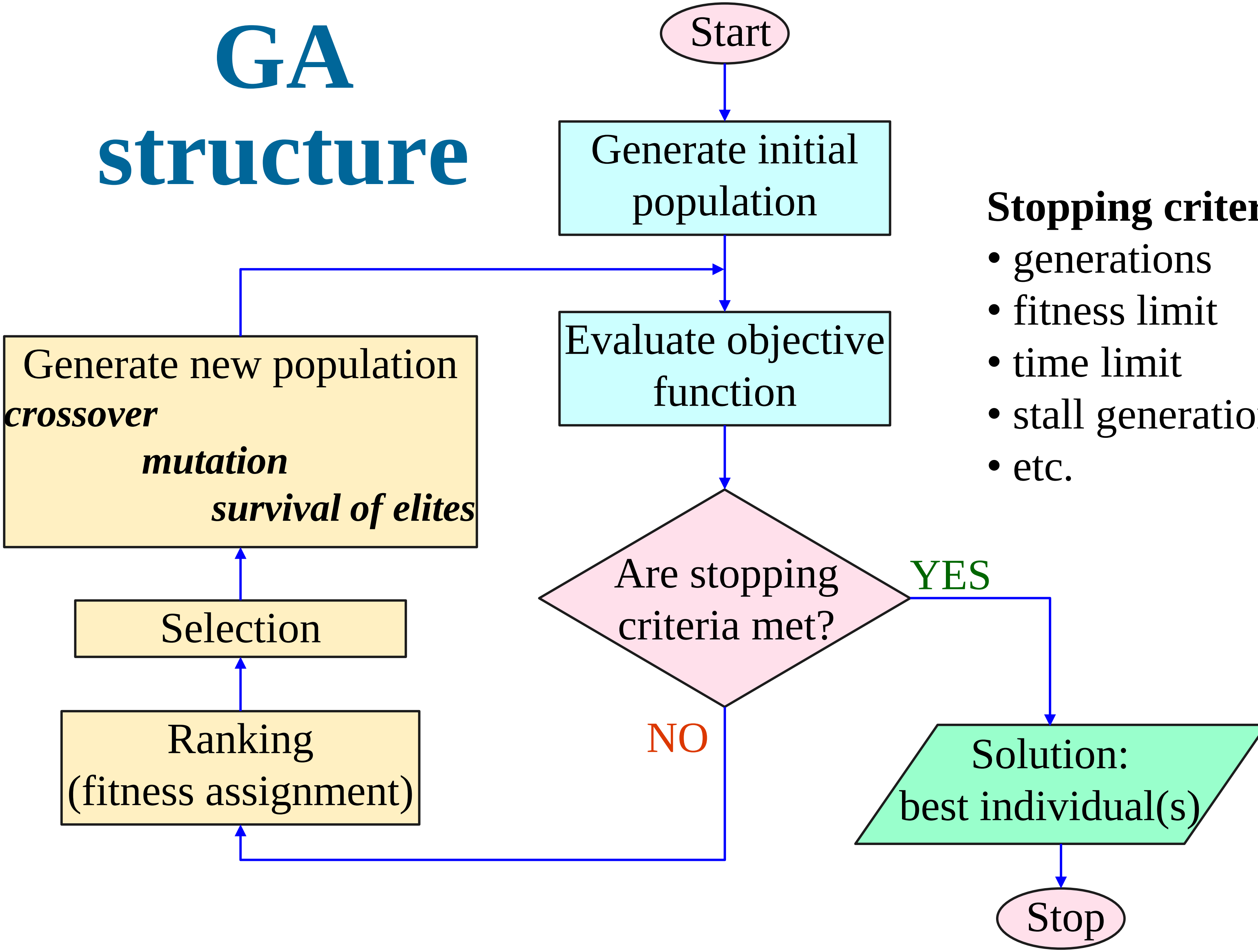
GENETIC ALGORITHMS (GA)

Structure and operation

GA structure



Stopping criteria:

- generations
- fitness limit
- time limit
- stall generations/time
- etc.

Initializing the population

- Usually, the population initialization is done stochastically (randomly)
- Some heuristic individuals (promising individuals) can be introduced into the initial population

- The initial population should consist of a large variety of individuals
- Population size: moderate 50 - 500 individuals

- Population size should grow linearly with the size of an individual (number of genes in a chromosome)

Illustration

De Jong function – 2 variables

$$x = [x_1,\ x_2];$$

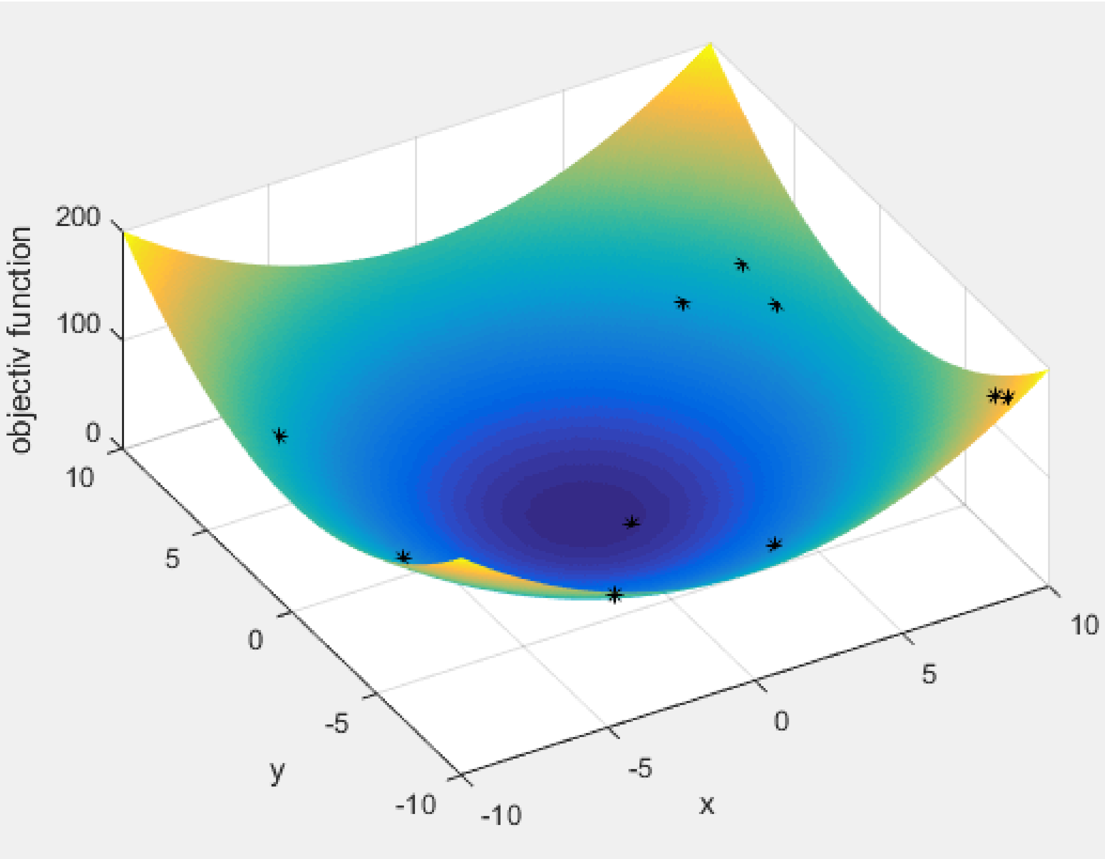
$$F(x_1,\ x_2) = x_1^2 + x_2^2;$$

$$D = [-10,10]\ \text{x}\ [-10,10]$$

25 randomly generated individuals;

we will consider the first 10

x_1	x_2
6.2945	**5.1548**
8.1158	**4.8626**
-7.4603	**-2.1555**
8.2675	**3.1096**
2.6472	**-6.5763**
-8.0492	**4.1209**
-4.4300	**-9.3633**
0.9376	**-1.0500**
9.1501	**-9.0766**
9.2978	**-8.0574**
-6.8477	6.4692
9.4119	3.8966
9.1433	-3.6580
-0.2925	9.0044
6.0056	-9.3111
-7.1623	-1.2251
-1.5648	-2.3688
8.3147	5.3103
5.8441	5.9040
9.1898	-6.2625
3.1148	-0.2047
-9.2858	-1.0883
6.9826	2.9263
8.6799	4.1873
3.5747	5.0937



Ranking

✓ The objective function provides an absolute measure of the performance of each individual.

- The direct use of the objective function for the hierarchy of individuals (ranking) can create difficulties in differentiating, in order to select, individuals with close values of the objective function:

  - premature convergence of the algorithm
  - loss of population diversity.

Ranking

✓ An hierarchy (ranking) of individuals can be used according to the value of the objective function.

- The **best** (most fit) individual has the lowest value of the objective function

- The **weakest** (less fit) individual has the highest value of the objective function

Evaluation of the objective function

	x_1	x_2	f
I1	6.29	5.15	66.09
I2	8.12	4.86	89.55
I3	-7.46	-2.16	60.32
I4	8.27	3.11	78.06
I5	2.65	-6.58	50.32
I6	-8.05	4.12	81.78
I7	-4.43	-9.36	107.23
I8	0.94	-1.05	**1.99** ← Best individual
I9	9.15	-9.08	**166.17** ← Weakest individual
I10	9.30	-8.06	151.45

I8, I5, I3 – neighboring individuals in terms of adequacy
distance between I8 and I5 is |1.99 – 50.32| = 48.33
distance between I5 and I3 is |50.32 – 60.32| = 10.00

Much different distances between neighboring individuals from the POV of the value of the objective function - it can create difficulties in the selection

Ranking – cont

- Generally, **a scaling function** is used that performs a domain transformation:
  - the value of the objective function is transformed into another domain more suitable for the selection function.

- Basically, there are two categories of scaling functions:
  - scaling by using explicitly the values of the objective function (linear scaling, sigma truncation scaling, power function scaling)
  - scaling based on the **rank** of all individuals (rank-based)
    - linear
    - nonlinear

Ranking individuals using a rank-based scaling function

- The population is **ordered according to the value of the objective function**, each individual receiving a position (*Pos*)
- The **fitness assigned** to each individual depends only on its **position** in the individuals rank and not on the actual objective value.
- Subsequently, each individual will receive a **selection probability** used for recombination, a probability that depends on the own value of the fitness and on the values of the fitness of all other individuals.

Rank-based scaling function – fitness assignment

$$Fit(Pos) = 2 - SP + 2(SP - 1)\frac{Pos - 1}{Nind - 1}$$

Pos - position; $Nind$ – number of individuals

SP - selective preasure $SP \in [1; 2]$

For a minimization problem:

- **For the fittest individual:** ***Pos = Nind***

  the highest fitness value

- **Least fit individual has *Pos = 1***

  the smallest fitness value

Position assignment

	x_1	x_2	f	*Pos*	
I1	6.29	5.15	66.09	7	
I2	8.12	4.86	89.55	4	
I3	-7.46	-2.16	60.32	8	
I4	8.27	3.11	78.06	6	
I5	2.65	-6.58	50.32	9	
I6	-8.05	4.12	81.78	5	
I7	-4.43	-9.36	107.23	3	
I8	0.94	-1.05	1.99	10	← the fittest individual
I9	9.15	-9.08	166.17	1	← least fit individual
I10	9.30	-8.06	151.45	2	

Scaling function

$$Fit(Pos) = 2 - SP + 2(SP - 1)\frac{Pos - 1}{Nind - 1}$$

Pos - position; $Nind$ – number of individuals

SP - selective preasure $SP \in [1; 2]$

- **Increasing the selective pressure** focuses the search on the best performers, **exploiting** the best solutions
  - Premature convergence of the search, loss of genetic diversity, reduction of the exploration capacity in the state space
- **Reducing the selective pressure** can lead to uniformity of the selection, the search becoming less efficient
  - increases the **exploration** capacity, including more chromosomes in the selection process

A balance of exploration – exploitation must be maintained

Fitness assignment

$$Fit(Pos) = 2 - SP + 2(SP - 1)\frac{Pos - 1}{Nind - 1}, \qquad Nind = 10$$

$$SP = 1.5$$

	x_1	x_2	f	Pos	Fit	
I1	6.29	5.15	66.09	7.00	1.17	
I2	8.12	4.86	89.55	4.00	0.83	
I3	-7.46	-2.16	60.32	8.00	1.28	
I4	8.27	3.11	78.06	6.00	1.06	
I5	2.65	-6.58	50.32	9.00	1.39	
I6	-8.05	4.12	81.78	5.00	0.94	
I7	-4.43	-9.36	107.23	3.00	0.72	
I8	0.94	-1.05	1.99	10.00	1.50	← the fittest individual
I9	9.15	-9.08	166.17	1.00	0.50	← least fit individual
I10	9.30	-8.06	151.45	2.00	0.61	

Selection

- Determining the intermediate population containing the parents who will be subjected to the genetic operators of recombination and mutation - selecting the individuals that will produce offspring.

- Methods:
  - **stochastic** – in the selection process individuals are introduced **randomly**, generally by using selection probabilities that depend on the fitness value
    - Individuals with a higher fitness are more likely to be selected, so their number of offspring may be higher than those with a lower fitness
    - **Roulette wheel selection** (stochastic sampling with replacement); Tournament selection; Stochastic uniform selection.
  - **deterministic** – individuals with a high fitness are always selected instead of those with a lower fitness: selection by truncation

- **Survival of elites**: survival of the fittest individuals in the current generation by their direct migration in the next generation

Roulette wheel selection

- Stochastic algorithm
- For each individual I_j, a probability of selection is calculated :

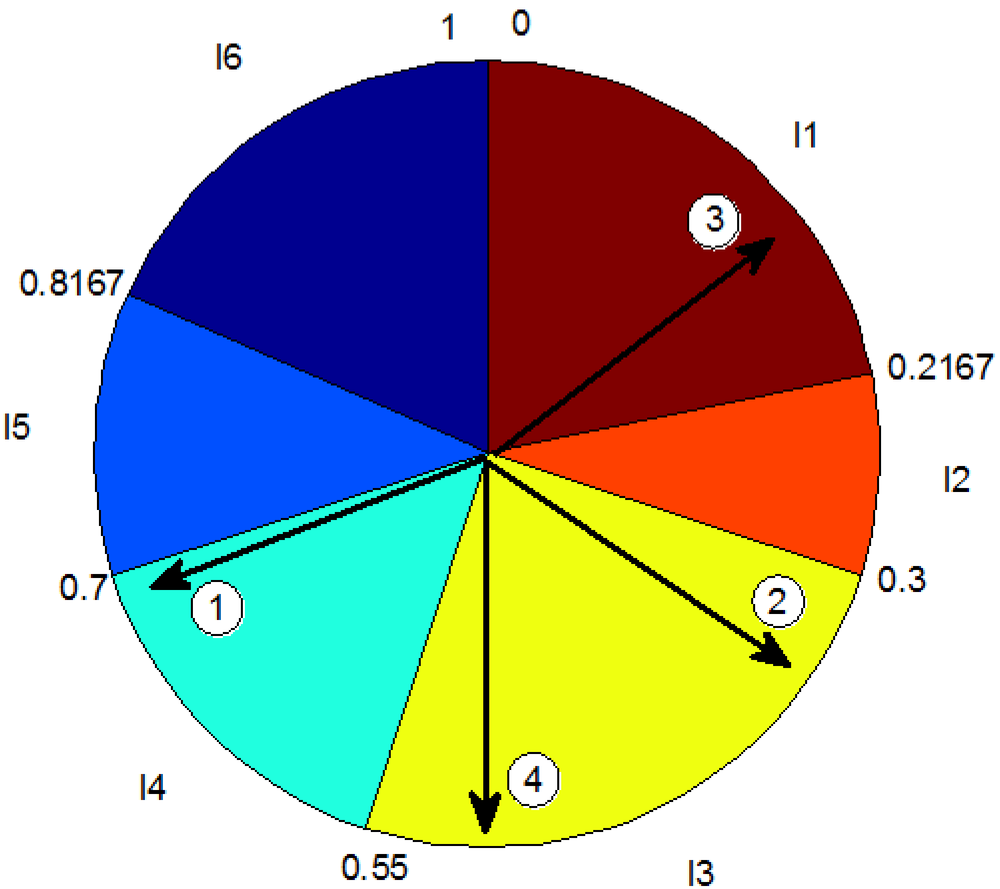
$$Prob(I_j) = \frac{Fit(I_j)}{\sum_{j=1}^{N} Fit(I_j)}$$

- The individuals are mapped to contiguous segments of a line (or slice in a circle), such that each individual's segment (slice) is proportional to its fitness or with its selection probability
- A random number in [0; 1] is generated and the individual whose segment (slice) spans the random number is selected.
- The process is repeated until the desired number of individuals is obtained (called mating population)
- This technique is analogous to a roulette wheel with each slice proportional in size to the fitness

The probability of each individual being selected for mating depends on its fitness normalized by the total fitness of the population.

Roulette wheel selection - illustration

Individ	*I1*	*I2*	*I3*	*I4*	*I5*	*I6*
Fit	1,3	0,5	1,5	0,9	0,7	1,1
Prob	0,2167	0,0833	0,2500	0,1500	0,1167	0,1833



We want to select 4 parents as part in the mating population.

For this, 4 numbers are randomly generated in the interval [0; 1]:

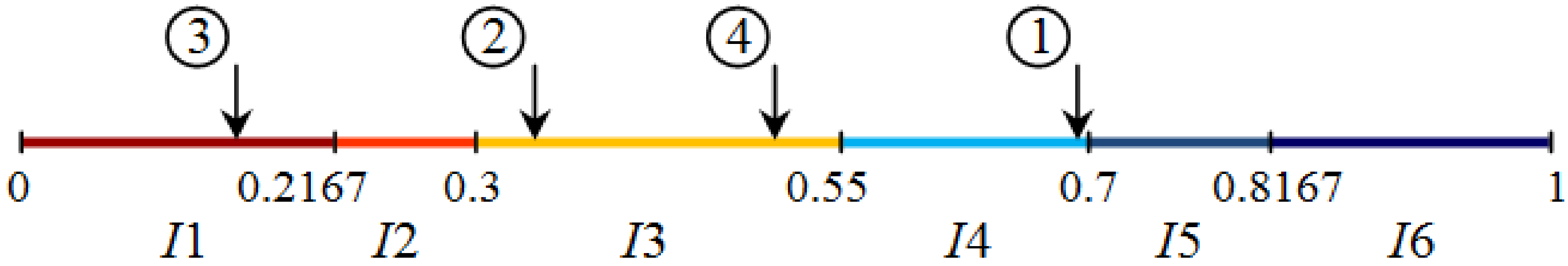
0.69; 0.34; 0.15; 0.5.

Accordingly, the 4 individuals selected as parents are:

I4; I3; I1; I3

Roulette wheel selection - illustration

Individ	*I1*	*I2*	*I3*	*I4*	*I5*	*I6*
Fit	1,3	0,5	1,5	0,9	0,7	1,1
Prob	0,2167	0,0833	0,2500	0,1500	0,1167	0,1833



We want to select 4 parents as part in the mating population.

For this, 4 numbers are randomly generated in the interval [0; 1]:

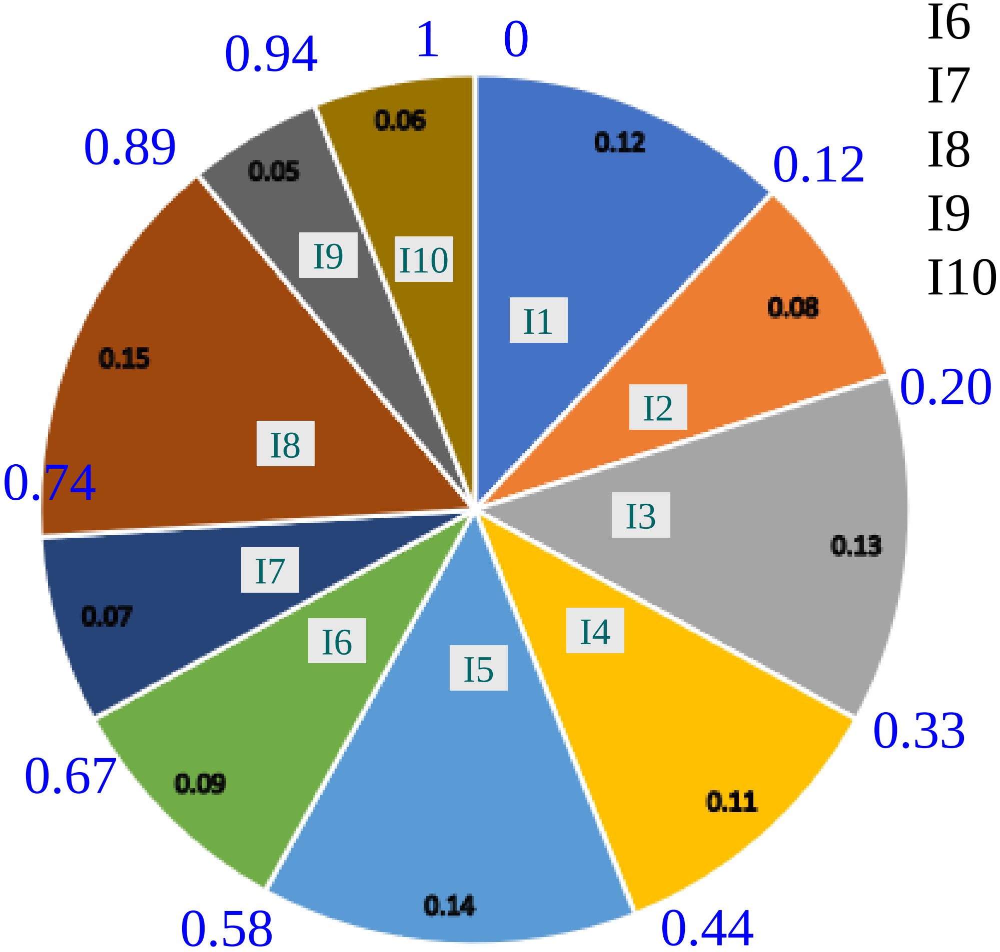
0.69; 0.34; 0.15; 0.5.

Accordingly, the 4 individuals selected as parents are:

I4; I3; I1; I3

Selection probability and roulette wheel

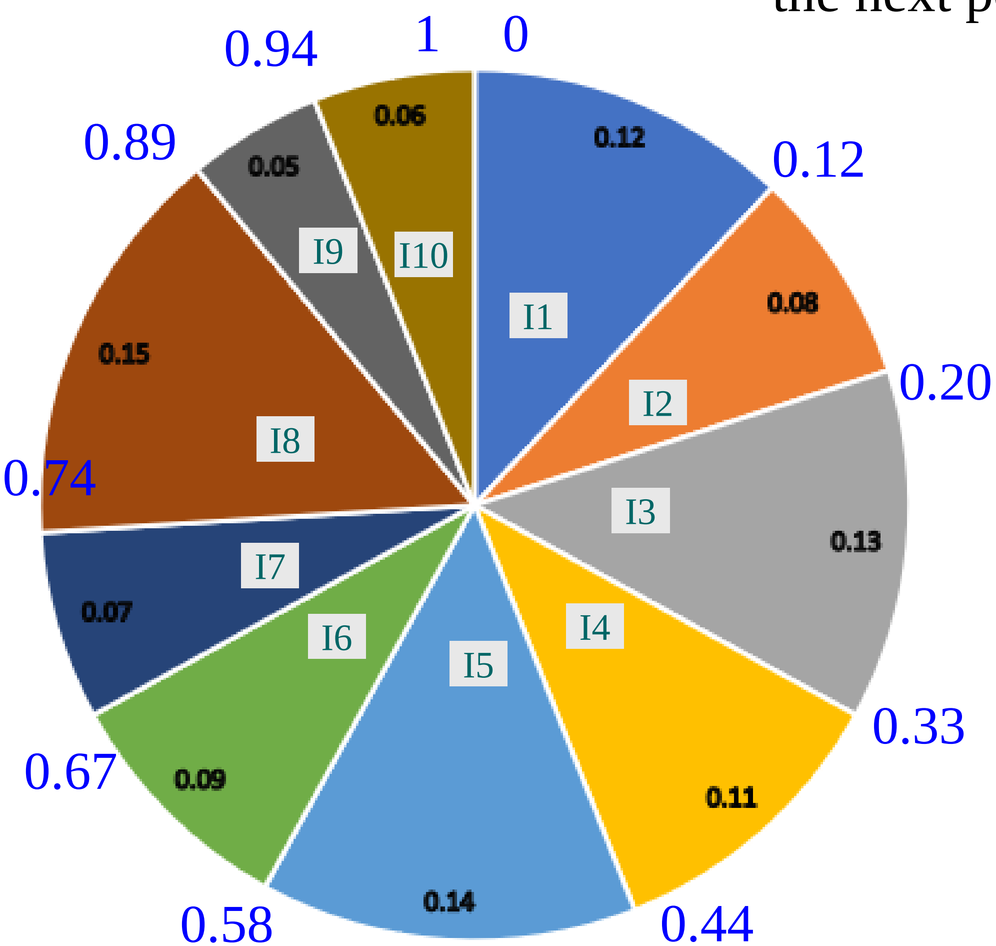
	x_1	x_2	f	*Pos*	*Fit*	*Prob*
I1	6.29	5.15	66.09	7.00	1.17	0.12
I2	8.12	4.86	89.55	4.00	0.83	0.08
I3	-7.46	-2.16	60.32	8.00	1.28	0.13
I4	8.27	3.11	78.06	6.00	1.06	0.11
I5	2.65	-6.58	50.32	9.00	1.39	0.14
I6	-8.05	4.12	81.78	5.00	0.94	0.09
I7	-4.43	-9.36	107.23	3.00	0.72	0.07
I8	0.94	-1.05	1.99	10.00	1.50	0.15
I9	9.15	-9.08	166.17	1.00	0.50	0.05
I10	9.30	-8.06	151.45	2.00	0.61	0.06



Roulette wheel selection

Randomly generate 10 values in the range [0; 1] and select accordingly 10 individuals as parents.

Each of the two neighbors may constitute a pair of parents for the generation of two offspring for the next population.



0.6557	I6	O1
0.0357	I1	O2
0.8491	I8	O3
0.9340	I9	O4
0.6787	I7	O5
0.7577	I8	O6
0.7431	I8	O7
0.3922	I4	O8
0.6555	I6	O9
0.4911	I5	O10

Recombination

- Recombination produces new individuals in combining the information contained in two or more parents (parents - mating population)
  - by combining the variable values of the parents
- **Methods for recombination:**
  - discrete (an exchange of variable values between the individuals)
  - for real valued variables:
    - **intermediate recombination**
    - line recombination
    - extended line recombination
  - for binary valued variables (crossover):
    - **single-point / double point / multi-point crossover**
    - uniform crossover
    - shuffle crossover

Intermediate recombination

$$Var_j^O = a_j Var_j^{P1} + (1 - a_j) Var_j^{P2}, \quad j = 1, 2, \ldots, Nvar$$

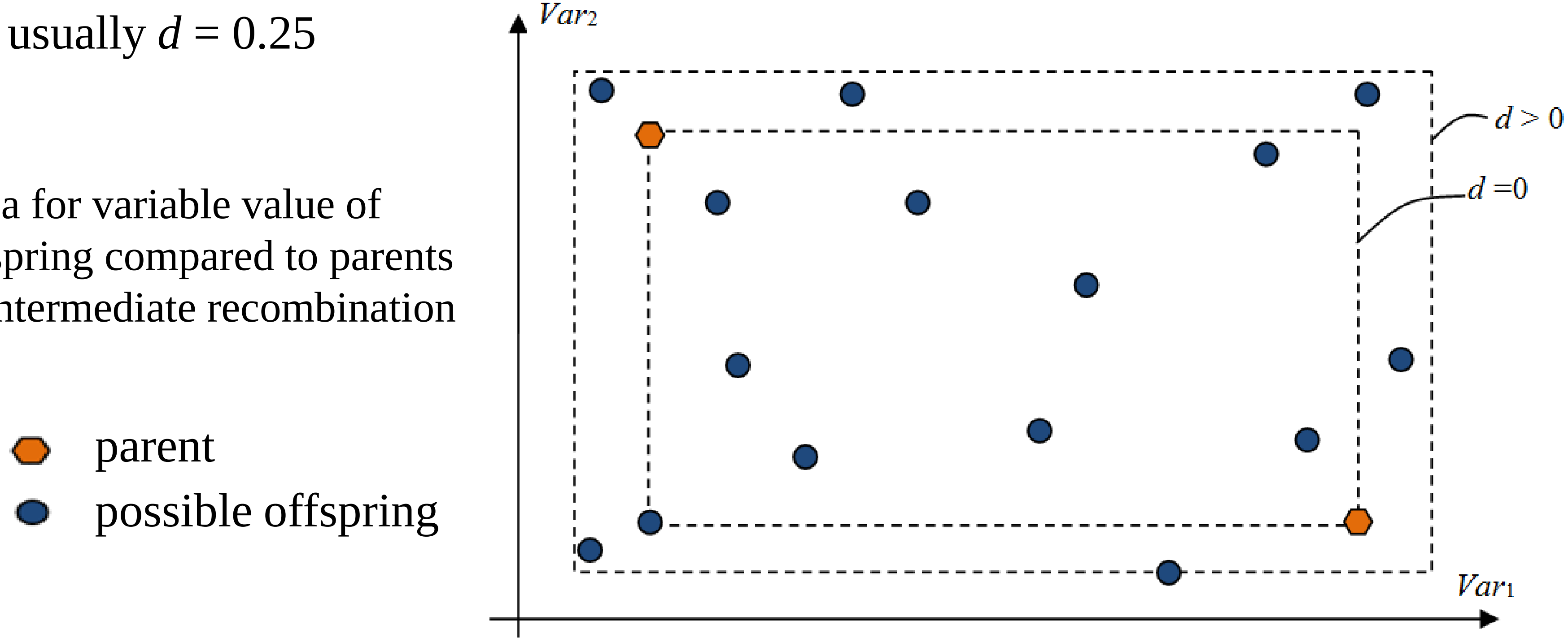
Var_j^O - Represent the j^{th} variable of the offspring

Var_j^{P1}, Var_j^{P2} - Represent the j^{th} variable of the 1st , respectively 2nd parent

a – scaling factor, chosen uniformly at random over an interval [-d, 1+d] for each variable anew

- usually d = 0.25

Area for variable value of offspring compared to parents in intermediate recombination



Intermediate recombination – illustration

$$Var_j^O = a_j Var_j^{P1} + (1 - a_j) Var_j^{P2}, \quad j = 1, 2, ..., Nvar$$

$P1\ (I6)$: -8.05 4.12

$P2\ (I1)$: 6.29 5.15

Assume d = 0.25.

For the 1st offspring $O1$, the scaling factors was randomly chosen as

a_1 = 1.1, for the 1st variable and a_2 = 0.7, for the 2nd variable

$$Var_1^{O_1} = 1.1 \cdot (-8.05) + (1 - 1.1) \cdot 6.29 = -9.484$$

$$Var_2^{O_1} = 0.7 \cdot 4.12 + (1 - 0.7) \cdot 5.15 = 4.429$$

The offspring $O1$:

$O1$: -9.484 4.429

Intermediate recombination – illustration

$$Var_j^O = a_j Var_j^{P1} + \left(1 - a_j\right) Var_j^{P2}, \quad j = 1, 2, \ldots, Nvar$$

$$P1\,(I6): \quad -8.05 \quad 4.12$$

$$P2\,(I1): \quad 6.29 \quad 5.15$$

For the 2nd offspring $O2$, the scaling factors was randomly chosen as
$a_1 = 0.2$, for the 1st variable and $a_2 = -0.15$, for the 2nd variable

$$Var_1^{O_2} = 0.2 \cdot (-8.05) + (1 - 0.2) \cdot 6.29 = 3.422$$

$$Var_2^{O_2} = (-0.15) \cdot 4.12 + (1 + 0.15) \cdot 5.15 = 5.30$$

The offspring O2:

$$O2: \quad 3.422 \quad 5.30$$

Mutation

- By mutation individuals are randomly altered
  - These variations (mutation steps) are mostly small. They will be applied to the variables of the individuals with a low probability (mutation probability or mutation rate)

- Mutation
  - Real valued mutation
  - Binary mutation
    - for binary valued individuals, mutation means the flipping of variable values, because every variable has only two states
    - the size of the mutation step is always 1

Real valued mutation

- Mutation of real variables means, that randomly created values are added to the variables with a low probability
    - the **probability** of mutating a variable (mutation rate) and the **size of the changes** for each mutated variable (mutation step) must be defined
  - The **probability** of mutating a variable is inversely proportional to the number of variables (dimensions)
    - a mutation rate of 1/*Nvar* is suggested
- The **size** of the mutation step is usually difficult to choose.
    - The optimal step-size depends on the problem considered and may even vary during the optimization process
        - small steps (small mutation steps) are often successful, especially when the individual is already well adapted
        - larger changes (large mutation steps) can, when successful, produce good results much quicker
        - a good mutation operator should often produce small step-sizes with a high probability and large step-sizes with a low probability

http://www.geatbx.com/download/GEATbx_Intro_Algorithmen_v38.pdf

Mutation operator

- Mutation specify how the genetic algorithm makes small random changes in the individuals in the population to create mutation offspring.
- Mutation provides genetic diversity and enables the genetic algorithm to search a broader space

Gaussian mutation

- can be used for unconstrained optimization
- adds a random number taken from a Gaussian distribution with mean 0
- the standard deviation of this distribution is determined by two parameters *scale* and *shrink*

scale - determines the standard deviation at the first generation
shrink - controls how the standard deviation shrinks as generations go by

This way the mutation is stronger at the beginning of the algorithm and "weaker" as the algorithm advances and new generations are created.

Gaussian mutation

In generation k, for the j^{th} variable of the parent subjected to the mutation:

$$Var_j^O = Var_j^P + r_j \cdot d_j \cdot scale_k$$

r_j - randomly chosen from a normal Gaussian distribution with mean 0

d_j - represents the range in which this variable takes values, that is

$$d_j = Var_{j,\max} - Var_{j,\min}$$

$$scale_k = scale\left(1 - shrink \cdot \frac{k}{N}\right)$$

N - the maximum number of iterations the genetic algorithm performs (stopping criteria)

Gaussian mutation - illustration

$N = 100$ $k = 50$ $scale = 1, shrink = 1$

$$scale_k = scale\left(1 - shrink \cdot \frac{k}{N}\right) \qquad scale_{50} = 1 \cdot \left(1 - 1 \cdot \frac{50}{100}\right) = 0.5$$

The range for both variables is: [-10; 10] $d_1 = d_2 = 20$

P:

-3	1

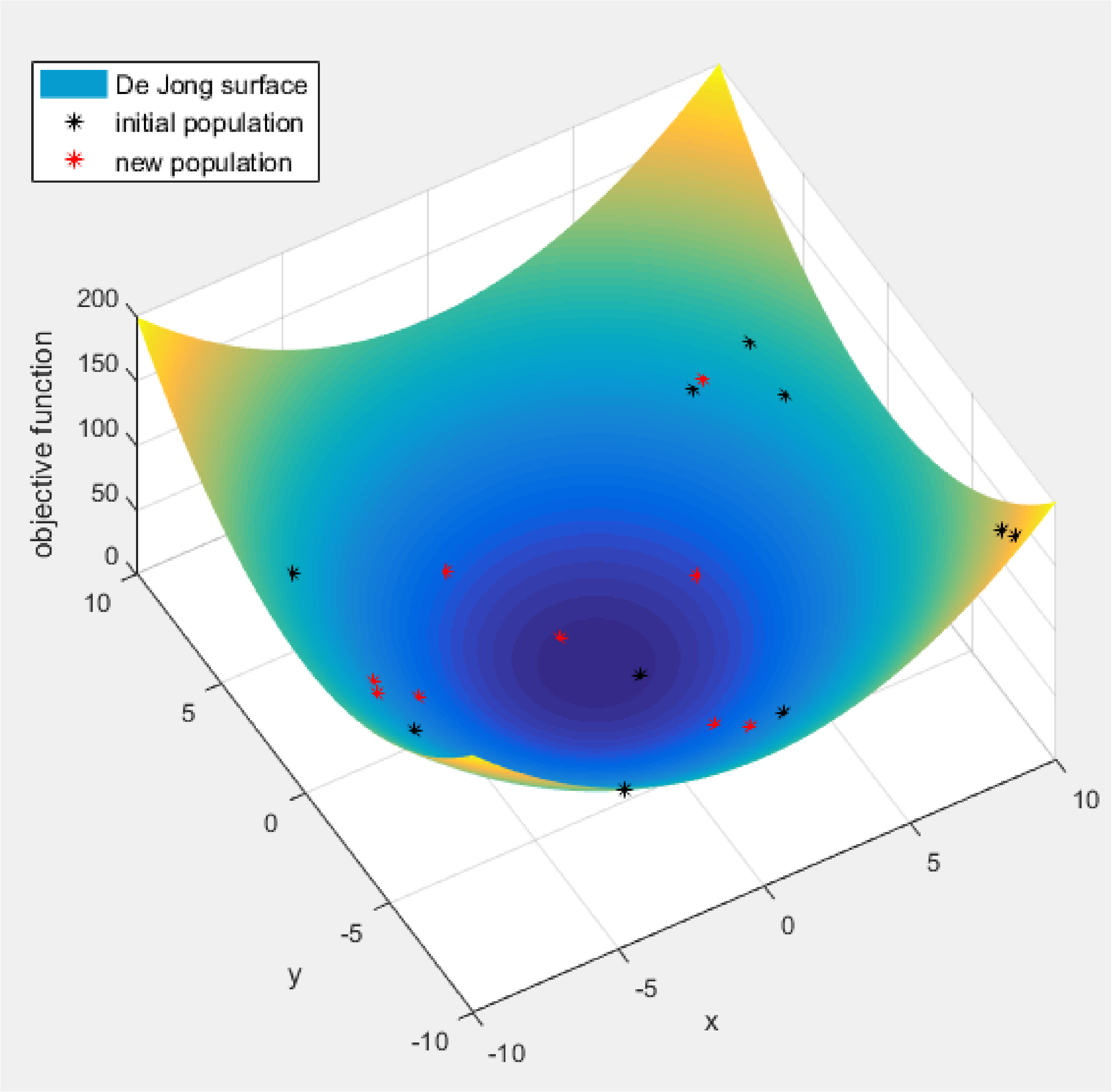
$$Var_j^O = Var_j^P + r_j \cdot d_j \cdot scale_k$$

$r_1 = +0.05$ $Var_1^O = -3 + (0.05 \cdot 20 \cdot 0.5) = -2.5$

$r_2 = -0.2$ $Var_2^O = 1 + ((-0.2) \cdot 20 \cdot 0.5) = -1$

O:

-2.5	-1



A new population vs. initial population

New population is created by recombination and mutation

Problem 1

Ind	x_1	x_2
I1	1	-4
I2	0	-5
I3	2	-2
I4	-2	2
I5	-1	1
I6	-3	0.5

Consider the function $f(x_1, x_2) = x_1^2 + 2x_2$, for which it is desired to find the minimum using a genetic algorithm. The initial population is:

a) Determine the values of the fitness function (*Fit*) for the initial population using a rank-based scaling function, with *SP* = 1.5

$$Fit(Pos) = 2 - SP + 2(SP - 1)\frac{Pos - 1}{Nind - 1}$$

Pos - position; *Nind* – number of individuals; *SP* - selective preasure

For the fittest individual (highest fitness value), *Pos* = *Nind*, while for the least fit individual (smallest fitness value) *Pos* = 1

b) The roulette-wheel method is used for the selection. The probability of selection is determined by the next relation. Determine which individuals are selected if the numbers 0.1, 0.5 and 0.8 are randomly generated.

$$Prob(I_j) = \frac{Fit(I_j)}{\sum_{j=1}^{N} Fit(I_j)}$$

c) Using the first two individuals selected above and intermediate recombination determine a possible offspring. $Var_j^O = a_j Var_j^{P1} + (1 - a_j) Var_j^{P2}, \quad j = 1, 2, \ldots, Nvar$

Var_j^O - the j^{th} variable of the offspring

Var_j^{P1}, Var_j^{P2} - the j^{th} variable of the 1st , respectively 2nd parent

a – scaling factor, chosen uniformly at random over the interval [-d, 1+d]. Consider *d* = 0.2

d) Is the resulting offspring better than both parents, than one parent, or is it no better than either parent? How do you justify the answer?

Problem 2

Ind	x_1	x_2
I1	2	-2
I2	5	5
I3	0	-2
I4	-1	6
I5	7	-10
I6	-3	1

Consider the function $f(x_1, x_2) = x_1^2 + 2x_2$, for which it is desired to find the minimum using a genetic algorithm. The current population is:

We want to create a new generation as follows: elitist survival: 1 individual; selection and recombination: 4 individuals; mutation: 1 individual

a) Determine the values of the fitness function (*Fit*) for the current population using a rank-based scaling function, with *SP* = 1.5

$$Fit(Pos) = 2 - SP + 2(SP-1)\frac{Pos-1}{Nind-1}$$

Pos - position; $Nind$ – number of individuals;

SP - selective preasure

For the fittest individual (highest fitness value), *Pos* = *Nind*, while for the least fit individual (smallest fitness value) *Pos* = 1

b) What is the individual in the new generation, $I1^{new}$, due to elitist survival?

c) For the mutation, Gaussian mutation is used: $Var_j^O = Var_j^P + r_j \cdot d_j \cdot scale$

r_j - randomly chosen from a normal Gaussian distribution with mean 0; *scale* = 0.5

d_j - represents the range in which this variable takes values, that is $d_j = Var_{j,\max} - Var_{j,\min}$

The mutation operation is applied to I6 whose both variables lie in the [-10; 10]. For the two variables, the values $r_1 = +0.05$, and $r_2 = -0.2$ were generated randomly.

What is the individual in the new generation, $I6^{new}$, due to mutation?

d) Due to selection and recombination the new individuals are:

$I2^{new}$:	-1.1	3.6	$I4^{new}$:	1.6	-2
$I3^{new}$:	-0.2	-3.2	$I5^{new}$:	-0.5	-2

Who are the six individuals in the new generation?

Perform a comparison between the current generation and the new generation both from the point of view of the best individual and from the point of view of the population as a whole.