

Podstawy Sztucznej Inteligencji

Laboratorium

Algorytmy genetyczne

Przygotował:
dr inż. Piotr Urbanek

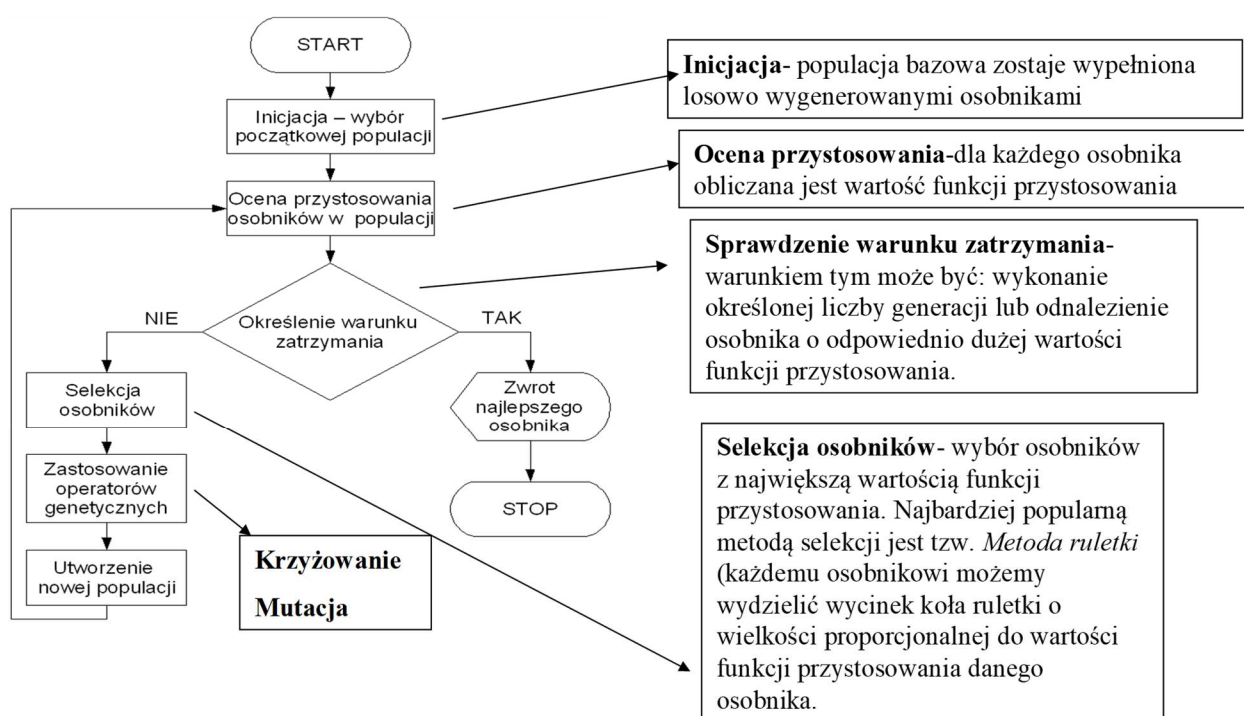
1. Teoria.

Definicja, czyli co to jest algorytm genetyczny?

Algorytmy genetyczne to metody rozwiązywania problemów, głównie zagadnień optymalizacyjnych, wzorowane na naturalnej ewolucji. Są to procedury przeszukiwania zbioru rozwiązań oparte na mechanizmach doboru naturalnego i dziedziczenia, korzystających z ewolucyjnej zasady przeżycia osobników najlepiej przystosowanych.

Ideę algorytmów ewolucyjnych przedstawił John Holland na przełomie lat sześćdziesiątych i siedemdziesiątych XX w.

Schemat blokowy działania AG



Cechy charakterystyczne AG

- Reprezentacja osobnika jako łańcucha binarnego (geny to 0 lub 1, świadczy o kodowaniu problemu/zadania na poziomie genotypu) o stałej długości (ilości bitów).
- Operatorami genetycznymi są: proste krzyżowanie (ang.crossover) i prosta mutacja (ang.mutation), operacje te są prowadzone na poziomie genotypu.
- Stałe prawdopodobieństwo operatorów genetycznych.
- Parametrem algorytmu jest prawdopodobieństwo zajścia krzyżowania.
- Osobnik jest przekształcany z genotypu do fenotypu i dopiero oceniany (to istotna cecha, gdyż mała zmiana na poziomie genotypu może spowodować duże zmiany na poziomie fenotypu).

Operatory genetyczne

Krzyżowanie.

Polega na wymianie materiału genetycznego pomiędzy losowo dobranymi parami osobników wybranych podczas selekcji. W wyniku krzyżowania powstają nowe chromosomy, które wejdą w skład kolejnej populacji (pokolenia). Okazuje się, że chromosomy powstałe w wyniku krzyżowania często są lepiej przystosowane (mają większą wartość funkcji przystosowania) niż ich “rodzice”.

Krzyżowanie proste

Przy klasycznym rozmieszczeniu genów w chromosomie jest to operacja stosunkowo prosta. Jeśli mamy dwa chromosomy to wybieramy (losowo) miejsce, w którym je rozcinamy (oba w tym samym miejscu). Punkt rozcięcia to inaczej **punkt krzyżowania**. Po rozcięciu wymieniamy odcięte części. W ten sposób powstają dwa nowe chromosomy

Selekcja

Polega na wyborze z bieżącej populacji najlepiej przystosowanych osobników, których materiał genetyczny zostanie poddany operacji krzyżowania i przekazany osobnikom następnej populacji.

Kryterium wyboru jest wartość funkcji przystosowania.

Mutacja (zachodząca zwykle z niewielkim prawdopodobieństwem)

Polega na zmianie wartości losowo wybranego genu. Zadaniem operatora mutacji jest zapewnienie zmienności chromosomów (np. niedopuszczenie do powstania całej populacji identycznych osobników) i tym samym stworzenie możliwości wyjścia procedury optymalizacji z maksimów lokalnych funkcji przystosowania.

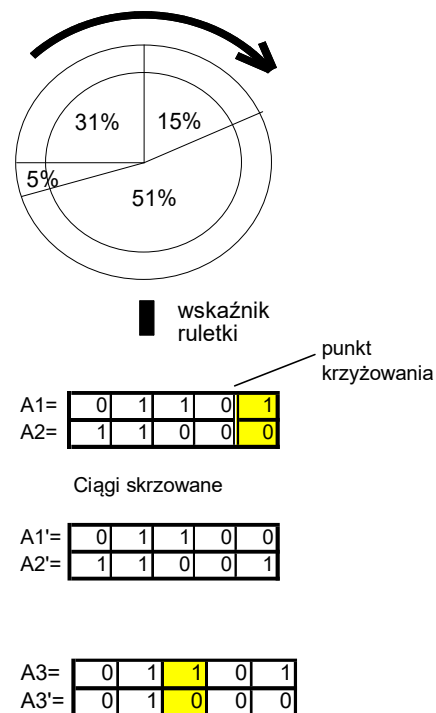
Elementarny algorytm genetyczny

Elementarny algorytm genetyczny jest skonstruowany z trzech następujących operacji: Reprodukacja, Krzyżowanie, Mutacja

Reprodukacja to proces, w którym indywidualne ciągi kodowe zostaną powielone w stosunku zależnym od wartości, jakie przybiera dla nich funkcja celu f (funkcja przystosowania). Jest to inaczej pewien miernik zysku, który chcemy zmaksymalizować.

Krzyżowanie (proste) to proces polegający na losowym kojarzeniu ciągów z puli rodzicielskiej w pary, losowy wybór **punktu krzyżowania** ciągów oraz zamiany wszystkich znaków na prawo od punktu krzyżowania

Mutacja polega na sporadycznej zmianie elementu ciągu kodowego.



Symulacja odręczna działania algorytmu genetycznego

| Nr ciągu | Populacja początkowa (wygenerowana losowo) | Wartość x | $f(x)=x^2$ | Wskaźnik przystosowania | Oczekiwana liczba kopii | Liczba kopii wygenerowanych wg reguły ruletki |
|----------|---|-----------|------------|-------------------------|-------------------------|---|
| 1 | 0 1 1 0 1 | 13 | 169 | 0,14 | 0,6 | 1 |
| 2 | 1 1 0 0 0 | 24 | 576 | 0,49 | 2,0 | 2 |
| 3 | 0 1 0 0 0 | 8 | 64 | 0,06 | 0,2 | 0 |
| 4 | 1 0 0 1 1 | 19 | 361 | 0,31 | 1,2 | 1 |
| | | | | | | |
| Suma | | | 1170 | 1,00 | | |
| Średnia | | | 293 | 0,25 | | |
| Maksimum | | | 576 | 0,49 | | |

| Pula rodzicielska po reprodukcji | Partner (wybrany losowo) | Punkt krzyżowania (wybrany losowo) | Nowa populacja | Wartość x | $f(x)=x^2$ | Wskaźnik przystosowania | Liczba kopii wygenerowanych wg reguły ruletki |
|----------------------------------|--------------------------|------------------------------------|----------------|-----------|------------|-------------------------|---|
| 0 1 1 0 1 | 2 | 4 | 0 1 1 0 0 | 12 | 144 | 0,33 | 0 |
| 1 1 0 0 0 | 1 | 4 | 1 1 0 0 1 | 25 | 625 | 1,43 | 1 |
| 1 1 0 0 0 | 4 | 2 | 1 1 0 1 1 | 27 | 729 | 1,66 | 2 |
| 1 0 0 1 1 | 3 | 2 | 1 0 0 0 0 | 16 | 256 | 0,58 | 1 |
| | | | | | | | |
| Suma | | | | | 1754 | 1 | |
| Średnia | | | | | 439 | 0,25 | |
| Maksimum | | | | | 729 | 0,42 | |

Podsumowanie

Różnice pomiędzy AG a konwencjonalnymi technikami optymalizacji:

- Operowanie na ciągach kodowych,
- Działanie na populacjach, a nie na pojedynczych punktach,
- Poszukiwanie metodą próbkowania („ślepe”),
- Losowe reguły wyboru.

Zadanie 1.

Napisać program znajdujący metodą algorytmu genetycznego maksimum funkcji

$$f(x) = 0.2\sqrt{x} + 2 \sin(2\pi \cdot 0.02 \cdot x) + 5, \text{ gdzie } x \in \langle 0; 255 \rangle.$$

Zbadać wpływ prawdopodobieństwa krzyżowania jednopunktowego ***pk*** oraz prawdopodobieństwa wystąpienia mutacji ***pm*** na szybkość

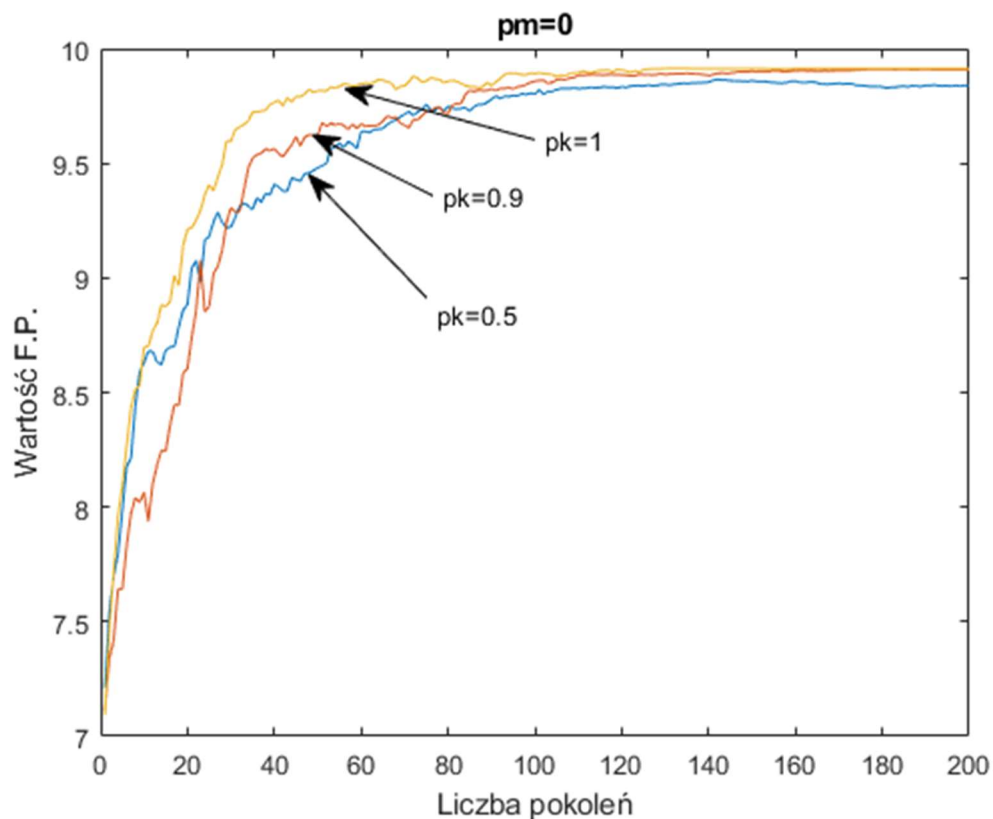
zbieżności wyników optymalizacji. Przyjąć następujące wartości pm i pk . Obliczenia przeprowadzić dla liczby chromosomów = 200.

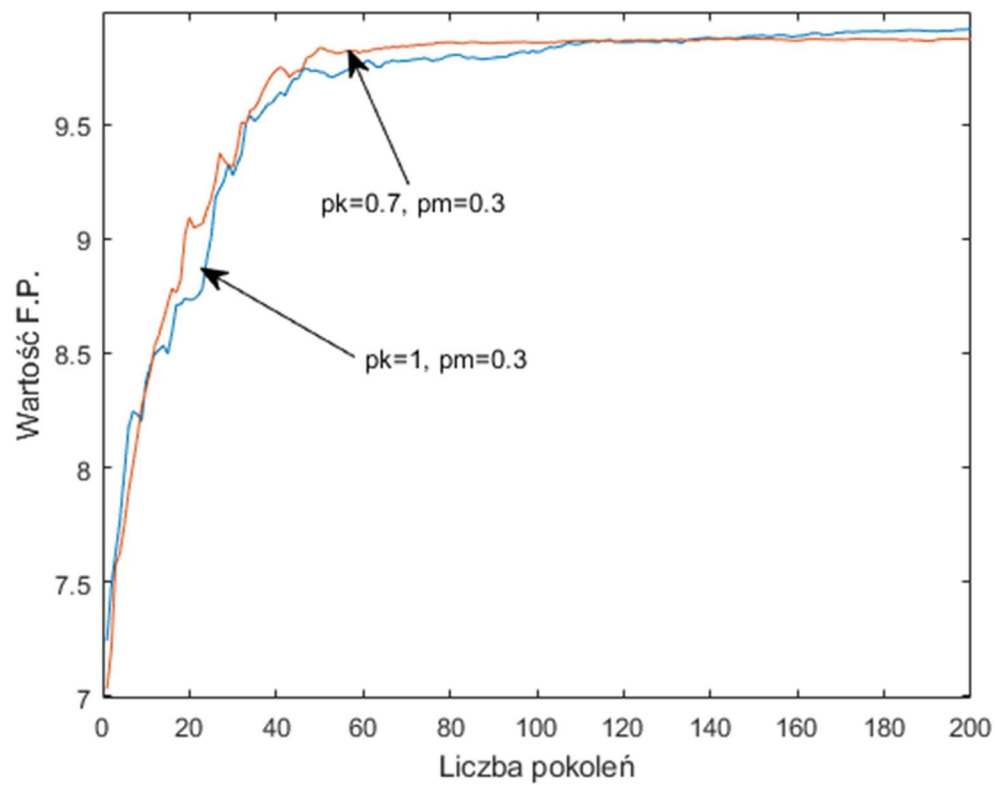
Średnia wartość funkcji przystosowania

| $pm \backslash pk$ | 0,5 | 0,6 | 0,7 | 0,8 | 1 |
|--------------------|-----|-----|-----|-----|---|
| 0 | | | | | |
| 0,01 | | | | | |
| 0,06 | | | | | |
| 0,1 | | | | | |
| 0,2 | | | | | |
| 0,3 | | | | | |
| 0,5 | | | | | |

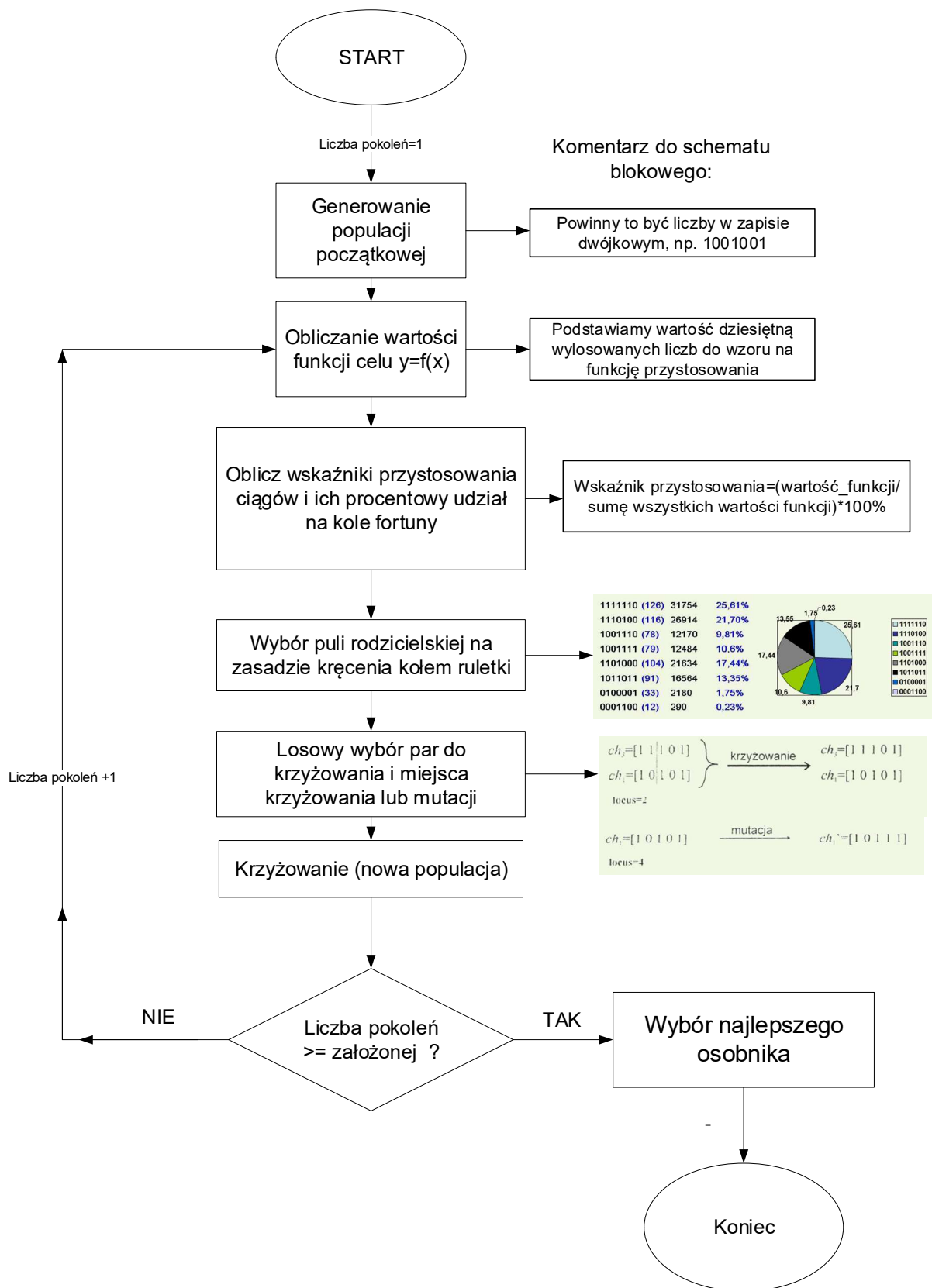
Dla poszczególnych wartości pm i pk narysować na jednym wykresie zależności wartości funkcji przystosowania fp od numeru pokolenia ($fp=f(nr_pokolenia)$).

Przykładowe wykresy dla różnych wartości pm , pk mogą wyglądać następująco:





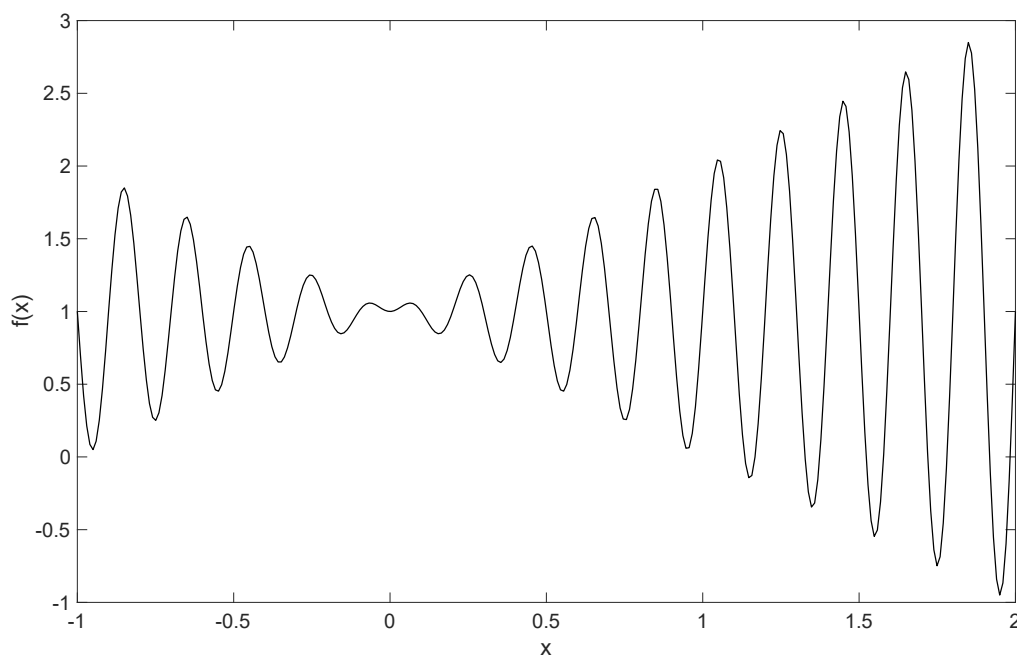
Otrzymane wyniki skomentować. Zadanie powtórzyć dla liczby chromosomów = 50.



Schemat działania algorytmu genetycznego

Reprezentacja zmiennoprzecinkowa Algorytmu Genetycznego.

W poprzednim zadaniu znalezienia maksimum funkcji za pomocą AG w każdym chromosomie można było zakodować jedynie liczby całkowite. W przypadku konieczności dokładnego wyznaczenia maksimum funkcji $f(x) = x \cdot \sin(10\pi x) + 1$ z dokładnością np. 6 miejsc po przecinku należy zastosować kodowanie zmiennoprzecinkowe liczb binarnych.



Wykres funkcji $f(x) = x \cdot \sin(10\pi x) + 1$

Do reprezentacji wartości rzeczywistych zmiennej x można użyć wektora binarnego, którego długość zależy od żądanej dokładności obliczenia rozwiązania. Dla użytego przykładu wiemy, że:

- Dziedzina funkcji x w przedziale $[-1;2]$ ma długość 3, gdyż $(2 - (-1)) = 3$.
- Liczba bitów chromosomów zależy od podziału funkcji na podprzedziały. Jeśli chcemy mieć dokładność 6 cyfr po przecinku (1×10^6 elementów), to całą dziedzinę należy podzielić na 3×10^6 elementów. W efekcie narzuca to zastosowanie chromosomów o długości 22 bitów, gdyż: $2^{21} \leq 3 \times 10^6 \leq 2^{22}$.

- Aby odwzorować łańcuch binarny w liczbę rzeczywistą z zakresu $[-1;2]$ należy przeliczyć ją wg wzoru:

$$x = -1.0 + \frac{3 \cdot x'}{2^{22} - 1}$$

gdzie -1.0 jest lewym krańcem przedziału, 3 jest długością przedziału a x' jest wartością zawartą w chromosomie w postaci binarnej.

Przykładowo:

Wartość binarna chromosomu

[1 1 0 1 0 0 1 1 1 0 0 1 0 0 1 0 1 0 0 0 0 1] wynosi 1.4794.

Zadanie 2:

Uzupełnij wykonany w zadaniu 1 program o możliwość wyznaczenia maksimum funkcji

$f(x) = 0.2\sqrt{x} + 2 \sin(2\pi \cdot 0.02 \cdot x) + 5$ w dziedzinie liczb rzeczywistych. Liczbę potrzebnych podprzedziałów proszę ustalić samodzielnie.