Sztuczna inteligencja i inżynieria wiedzy – laboratorium  
  
Sprawozdanie: Algorytmy genetyczne  
  
  
Kajetan Pynka, 254495

Spis treści

[Wstępne założenia implementacyjne 3](#_Toc98878002)

[Selekcja turniejowa vs selekcja ruletkowa 4](#_Toc98878003)

[Krzyżowanie 8](#_Toc98878004)

[Mutacja 13](#_Toc98878005)

[Liczba osobników 20](#_Toc98878006)

[Liczba pokoleń 25](#_Toc98878007)

[Próby wyznaczenia rozwiązania optymalnego 30](#_Toc98878008)

# Wstępne założenia implementacyjne

W ramach badań przeprowadzonych w kolejnych rozdziałach sprawozdania modyfikowane były następujące parametry:

* Rozmiar populacji
* Prawdopodobieństwo mutacji
* Prawdopodobieństwo krzyżowania
* Liczba pokoleń
* Rozmiar turnieju

Dla każdej sekcji badań (i każdego źródła danych) stosowałem te same parametry początkowe:

* Liczba pokoleń = 500
* Rozmiar populacji = 10
* Domyślnie selekcja turniejowa
* Rozmiar turnieju = 30% rozmiaru populacji (tutaj 3)
* Prawdopodobieństwo mutacji = 10%
* Prawdopodobieństwo krzyżowania = 60%

Zdaje sobie sprawę, że dla różnych źródeł danych te ustawienia mogą być albo odpowiednie albo nie, natomiast skupiłem się na porównaniu zmian poszczególnych parametrów (chciałem zapewnić wszędzie równe warunki startowe by móc kontrolować pojedynczą zmienną).

W ramach operatora krzyżowania zastosowałem krzyżowanie jednopunktowe.

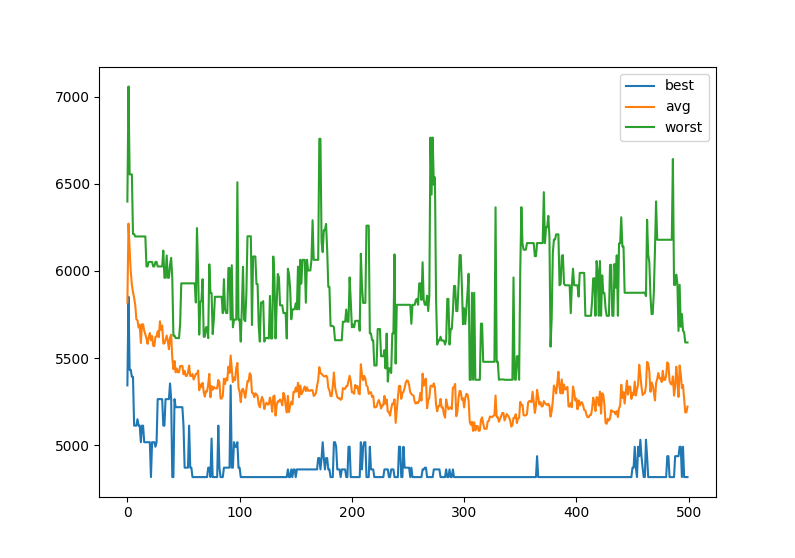
W ramach operatora mutacji prawdopodobieństwo określa szansę na zmutowanie każdego pojedynczego genu (nie mam osobnego parametru na to czy osobnik powinien zmutować).

Następnie przedstawię wyniki przeprowadzonych badań. Best, worst, avg oraz std dotyczą ostatniego pokolenia algorytmu genetycznego (z jednej strony best powinien raczej dotyczyć najlepszej możliwej wartości ale z drugiej strony branie worst na samym początku algorytmu gdy jest kosmicznie duży uważam za bez sensu).

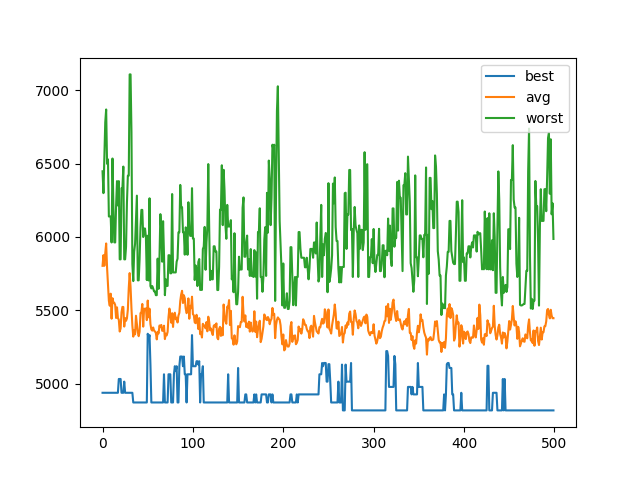
# Selekcja turniejowa vs selekcja ruletkowa

XXX/001 – turniej, XXX/002 - ruletka

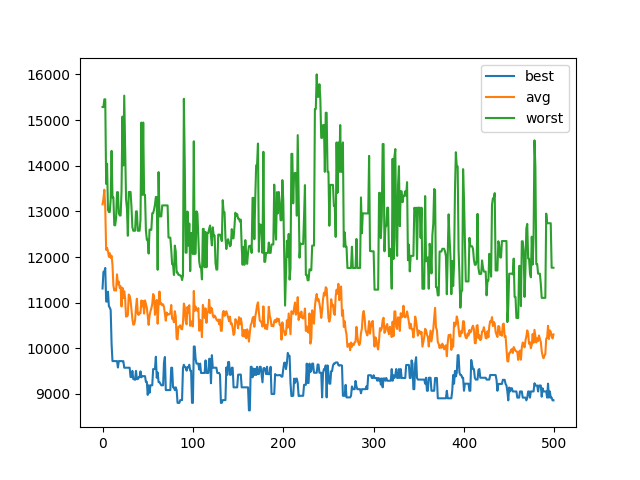
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| instancja | Alg. Ewolucyjny [10x] | | | | Metoda losowa [N] | | | |
|  | best | worst | avg | std | best | worst | avg | std |
| EASY/001 | 4818 | 5590 | 5222.5 | 232.4 | 4818 | 11730 | 7842.7 | 1051.4 |
| EASY/002 | 4818 | 11730 | 7842.7 | 348.1 | 4818 | 11730 | 7848.9 | 1054.5 |
| FLAT/001 | 8860 | 11765 | 10305.5 | 936.6 | 8940 | 25425 | 18087.3 | 2223.6 |
| FLAT/002 | 8605 | 12860 | 10107.5 | 1289.4 | 9245 | 24820 | 18093.2 | 2225.2 |
| HARD/001 | 15914 | 18831 | 17704.9 | 1066 | 15055 | 39553 | 26456.4 | 3175.2 |
| HARD/002 | 15707 | 20016 | 17779.9 | 1423.1 | 15588 | 39183 | 26460.2 | 3185.6 |



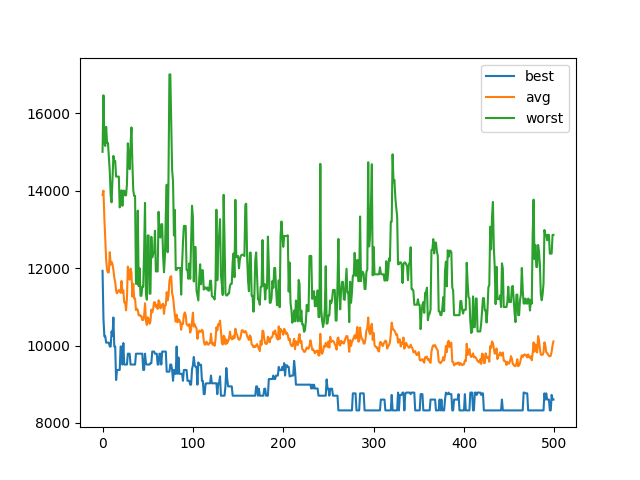
*EASY/001*



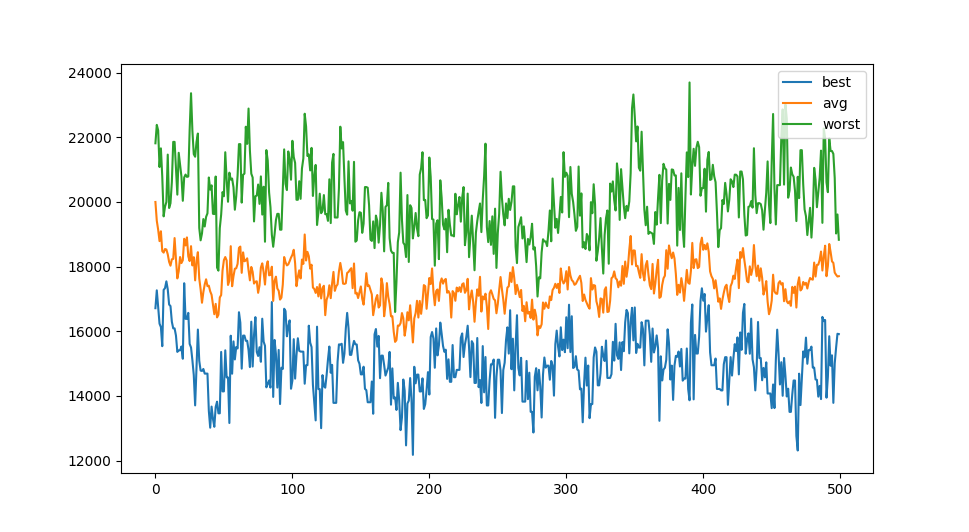
EASY/002



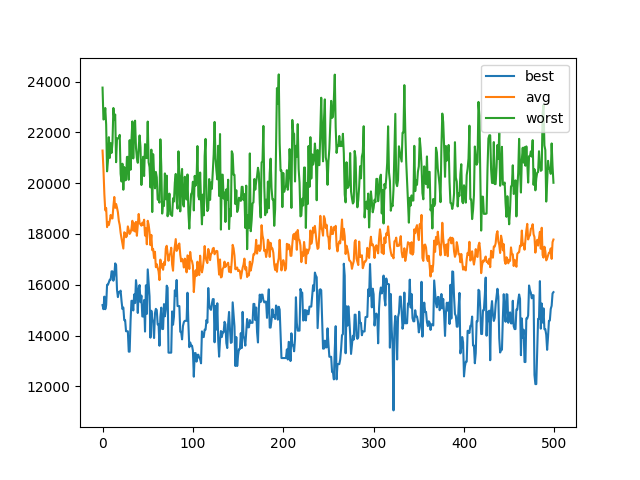
FLAT/001



FLAT/002



HARD/001



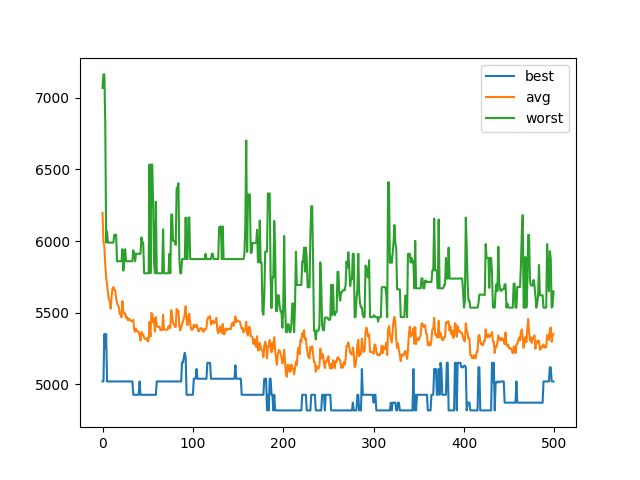
HARD/002

**WNIOSKI:** Ogólnie oba podejścia wydają się spełniać swoje zadanie. W przypadku obu operatorów bardzo istotne jest odpowiednie ustalenie ich parametrów: w przypadku turnieju – jego rozmiaru, w przypadku ruletki – „skalowania prawdopodobieństwa”. Rozmiar turnieju wydaje się być najbardziej zależnym od rozmiaru populacji (pytanie jak dobrać odpowiednią wartość: stałoliczbowa czy proporcjonalna?). W przypadku ruletki musimy odpowiednio skalować prawdopodobieństwa tzn. jeśli zejście z wyniku 5000 na 4990 mielibyśmy uznać za znaczną poprawę to wartość w postaci suma\_wyników/4990 może nie odbiegać tak bardzo jak byśmy chcieli od wartości suma\_wyników/5000. Można wtedy wykorzystywać funkcje eksponencjalne lub mapować wyniki na inne przedziały liczbowe tak aby różnice były znaczące.

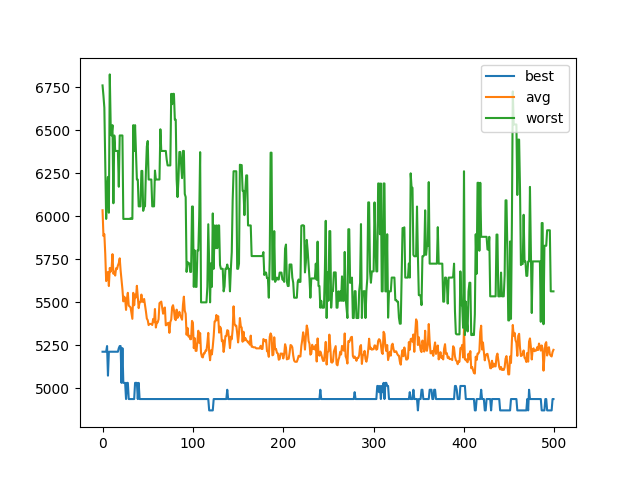
# Krzyżowanie

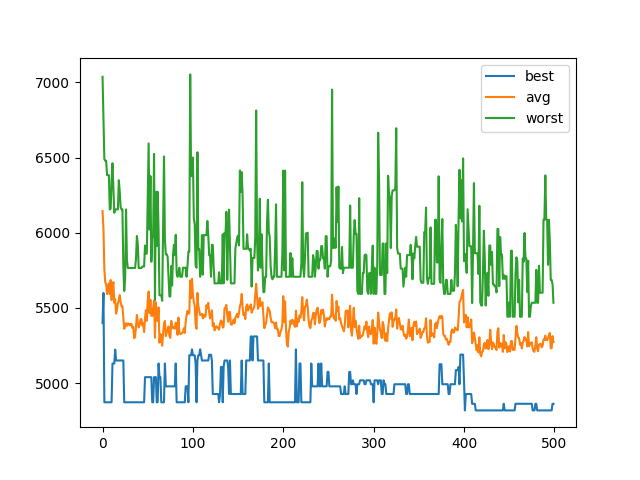
XXX/003 – szansa na krzyżowanie zerowa, XXX/004 – szansa 50%, XXX/005 – szansa 100%

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| instancja | Alg. Ewolucyjny [10x] | | | | Metoda losowa [N] | | | |
|  | best | worst | avg | std | best | worst | avg | std |
| EASY/003 | 5020 | 5649 | 5352.5 | 180 | 4818 | 11730 | 7847.3 | 1051.7 |
| EASY/004 | 4938 | 5564 | 5224.1 | 198.1 | 4818 | 11730 | 7851 | 1055.6 |
| EASY/005 | 4862 | 5534 | 5272.8 | 205.9 | 4818 | 11730 | 7851.4 | 1048.8 |
| FLAT/003 | 8945 | 12665 | 10729 | 1182.2 | 9130 | 24975 | 18091.1 | 2220.1 |
| FLAT/004 | 8930 | 13475 | 10522.5 | 1386.8 | 9295 | 25035 | 18099.6 | 2230.8 |
| FLAT/005 | 9505 | 11065 | 10187.5 | 641.7 | 9025 | 25345 | 18080.7 | 2234.2 |
| HARD/003 | 15021 | 19297 | 17571.8 | 1491.5 | 14357 | 38772 | 26460.4 | 3170.4 |
| HARD/004 | 15744 | 19353 | 17217.4 | 1223.9 | 14133 | 38660 | 26459.3 | 3150.8 |
| HARD/005 | 14308 | 19118 | 17549.6 | 1332.4 | 14702 | 39527 | 26478 | 3166.7 |

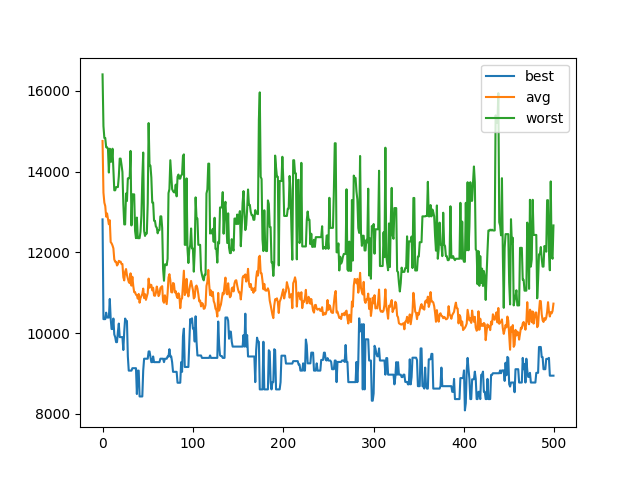


EASY/003

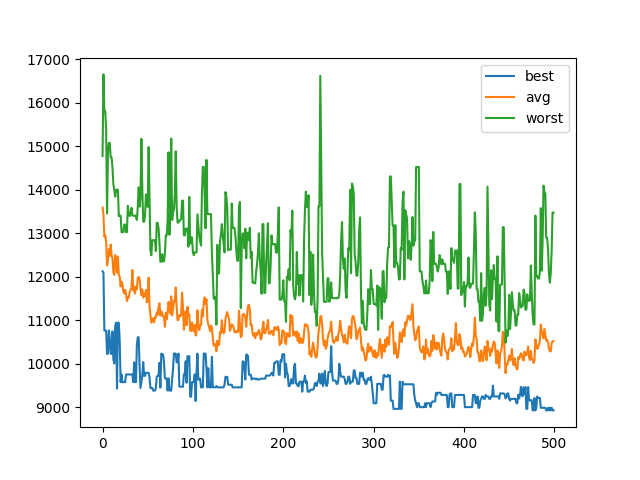


EASY/004

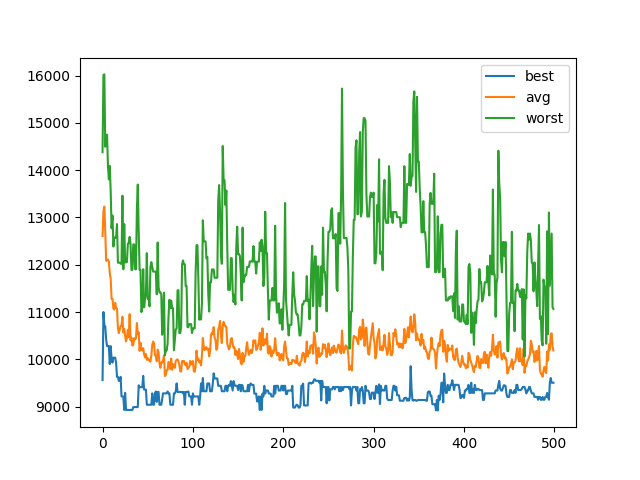
EASY/005



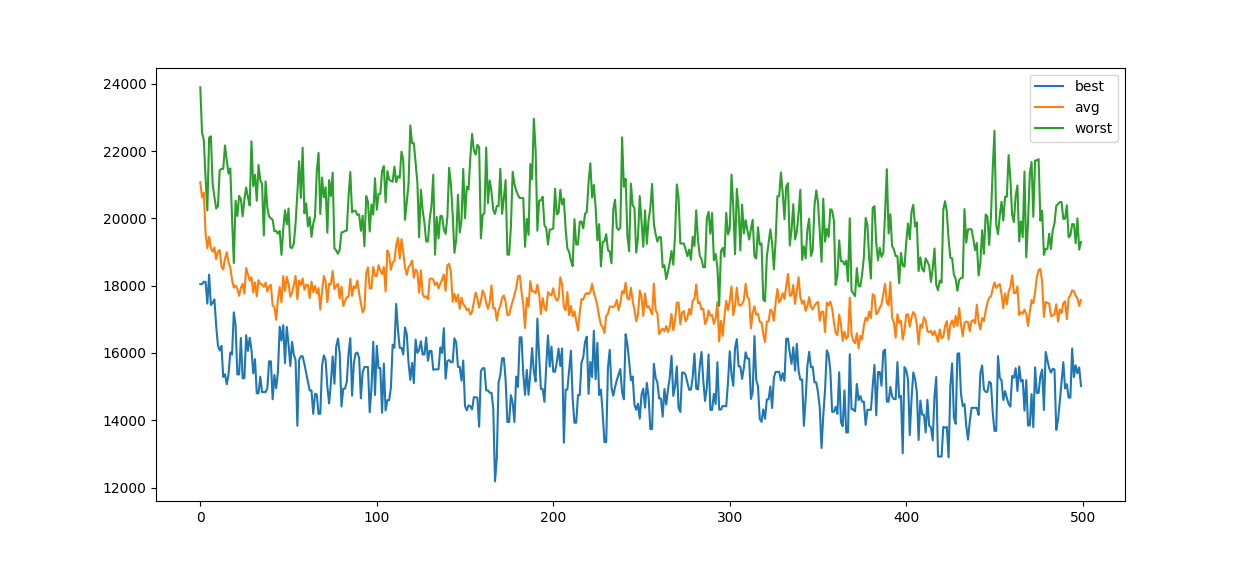
FLAT/003



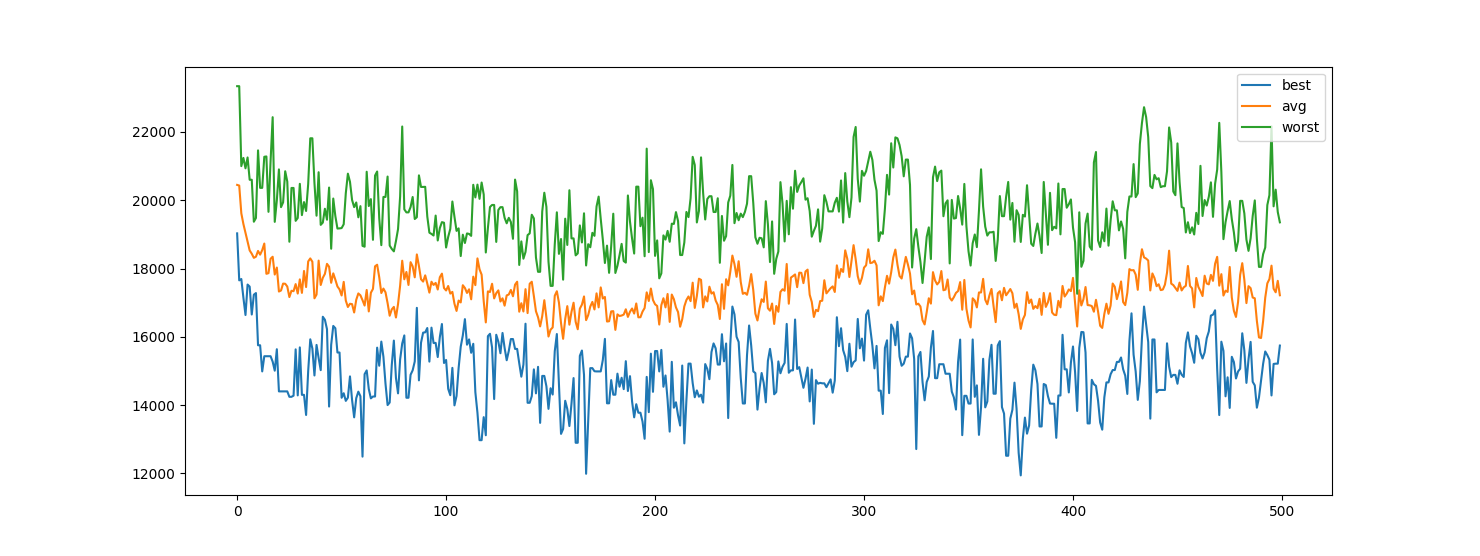
FLAT/004



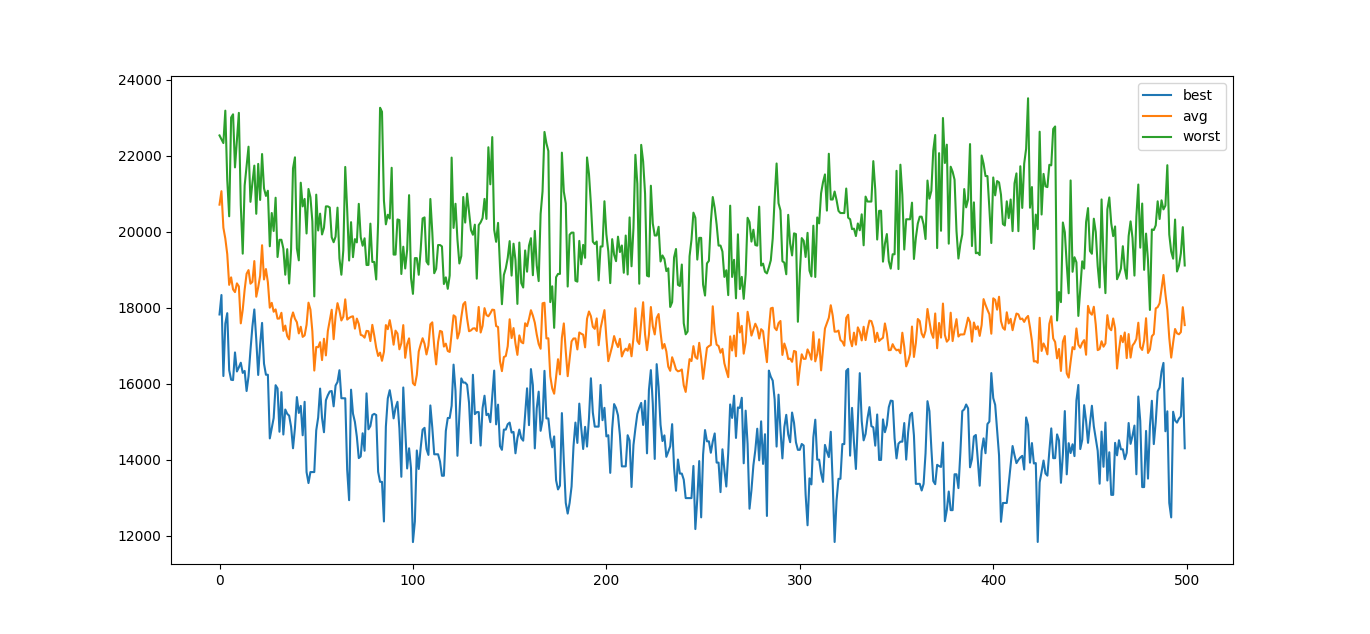
FLAT/005



HARD/003



HARD/004



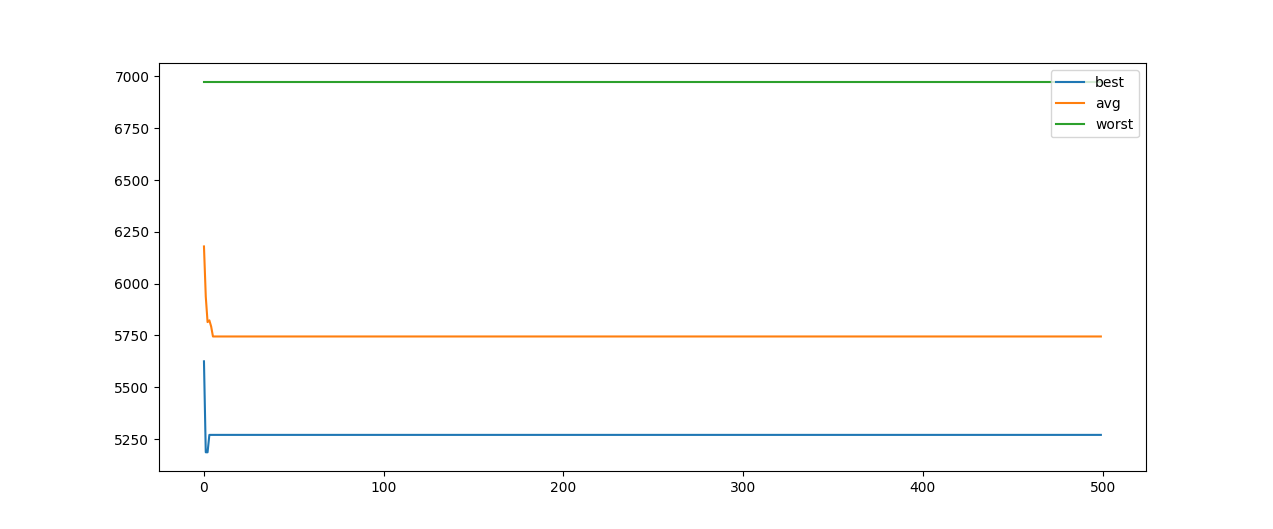
HARD/005

**WNIOSKI:** Krzyżowanie jest ściśle powiązane z mutacją i oba operatory na swój sposób modyfikują genotyp. W przypadku krzyżowania nie są to jednak tak losowe zmiany (wymieniamy bowiem części rodziców a nie zmieniamy samych genów). Odpowiednio duże krzyżowanie wydaje się wprowadzać stabilizację wykresu a z drugiej strony powoduje odpowiednią wymianę genów. Bez tego operatora polegać możemy jedynie na mutacji, która jest bardziej nieprzewidywalna (w zależności od rozważanego problemu może to być na plus i na minus).

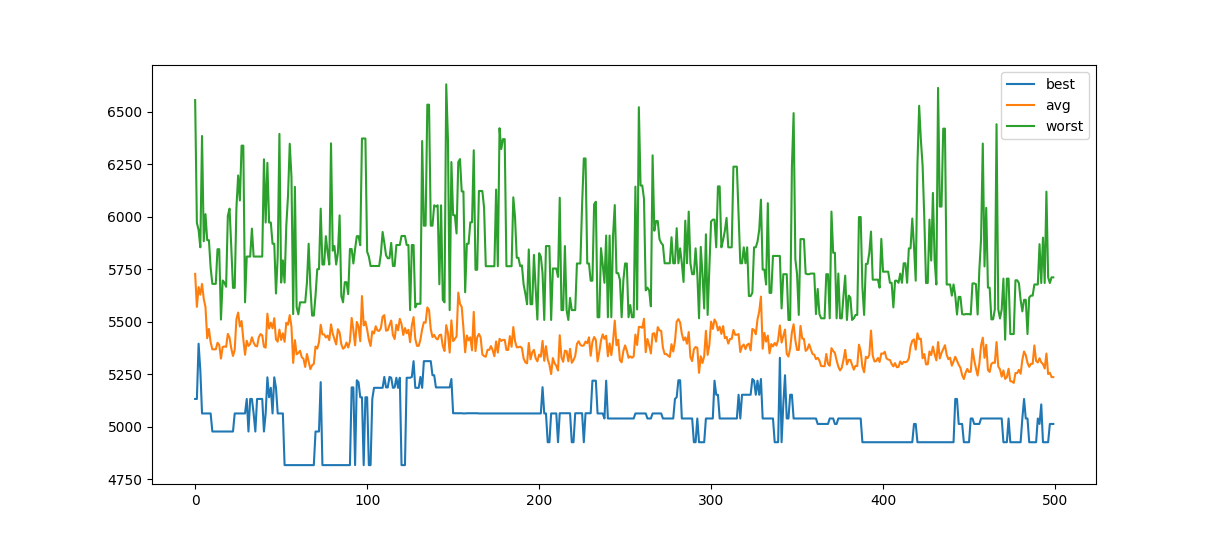
# Mutacja

XXX/006 – szansa 0%, XXX/007 – szansa 10%, XXX/008 – 50%, XXX/009 – szansa 100%

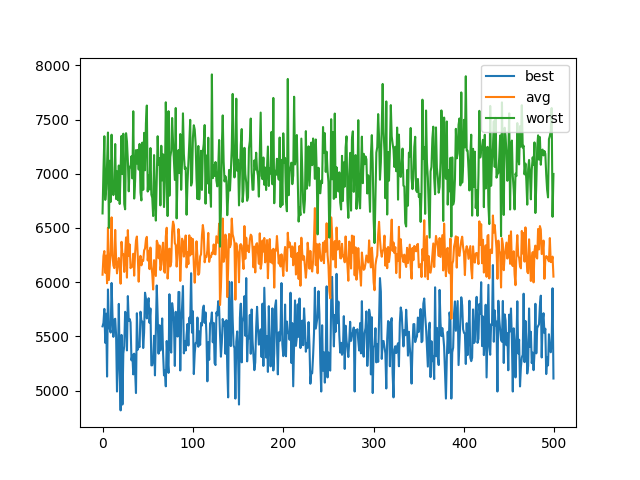
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| instancja | Alg. Ewolucyjny [10x] | | | | Metoda losowa [N] | | | |
|  | best | worst | avg | std | best | worst | avg | std |
| EASY/006 | 5270 | 6973 | 5744.9 | 455.1 | 4818 | 11730 | 7850 | 1057.87 |
| EASY/007 | 5014 | 5712 | 5236.8 | 200.4 | 4862 | 11730 | 7849.3 | 1052.2 |
| EASY/008 | 5112 | 6998 | 6051.6 | 645 | 4862 | 11730 | 7856.6 | 1047.6 |
| EASY/009 | 5930 | 8688 | 7281.6 | 787 | 4862 | 11730 | 7854.9 | 1052.2 |
| FLAT/006 | 10830 | 16275 | 14301.5 | 1536.7 | 9145 | 24880 | 18090.9 | 2222.8 |
| FLAT/007 | 9700 | 13570 | 10892 | 1068.8 | 8845 | 25085 | 18104 | 2236.8 |
| FLAT/008 | 11225 | 15275 | 13473.5 | 1580.6 | 9185 | 25200 | 18081.7 | 2233.5 |
| FLAT/009 | 11495 | 18280 | 15637.5 | 1695 | 8820 | 25215 | 18084.6 | 2230.9 |
| HARD/006 | 15761 | 23092 | 20664 | 2233.3 | 15169 | 39825 | 26450 | 3160.4 |
| HARD/007 | 14866 | 21281 | 17640.1 | 2122.4 | 14823 | 39186 | 26437.3 | 3185.1 |
| HARD/008 | 15744 | 19353 | 17217.4 | 1223.9 | 14133 | 38660 | 26459.3 | 3150.8 |
| HARD/009 | 21725 | 30622 | 25980.7 | 2548.2 | 14469 | 39169 | 26461 | 3156.8 |



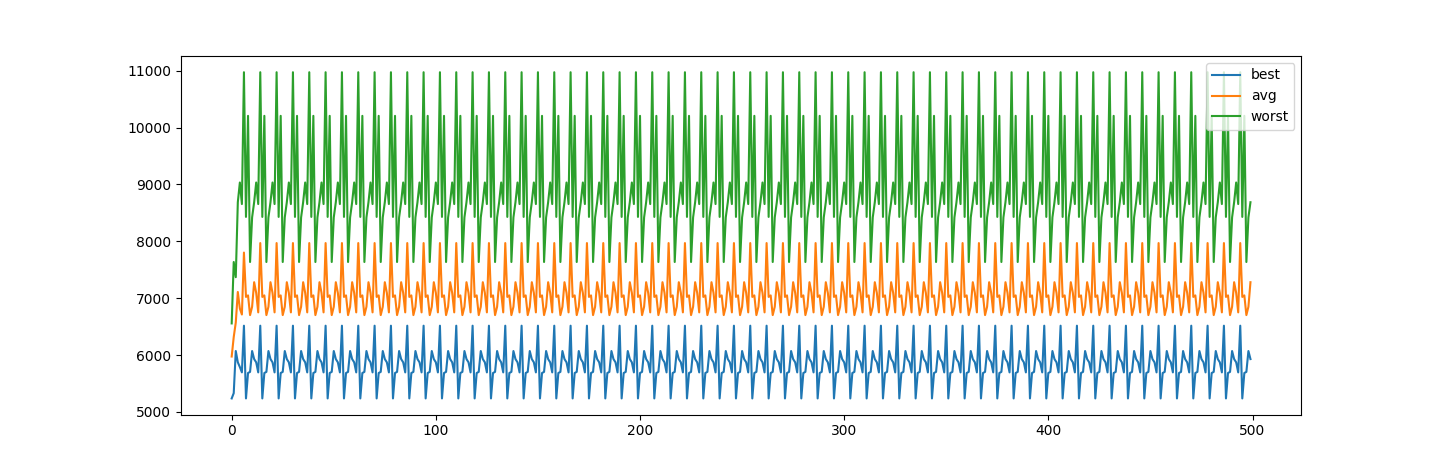
EASY/006



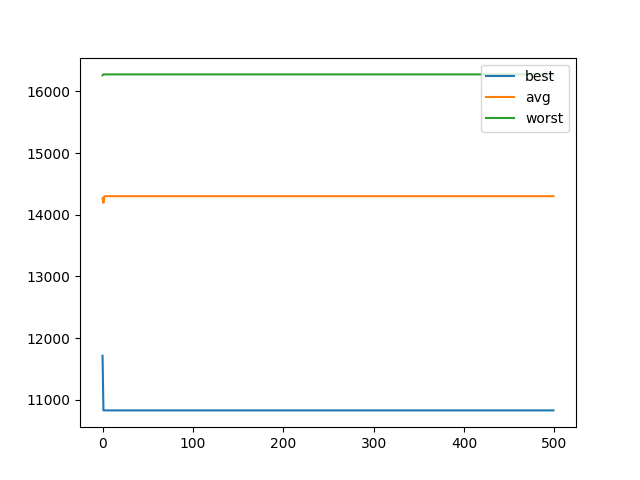
EASY/007



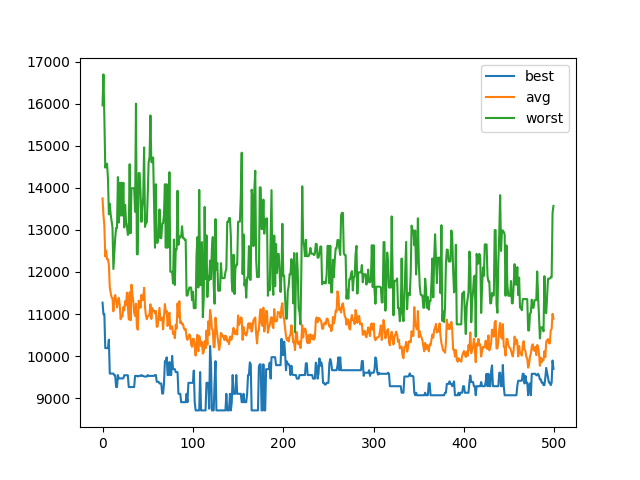
EASY/008



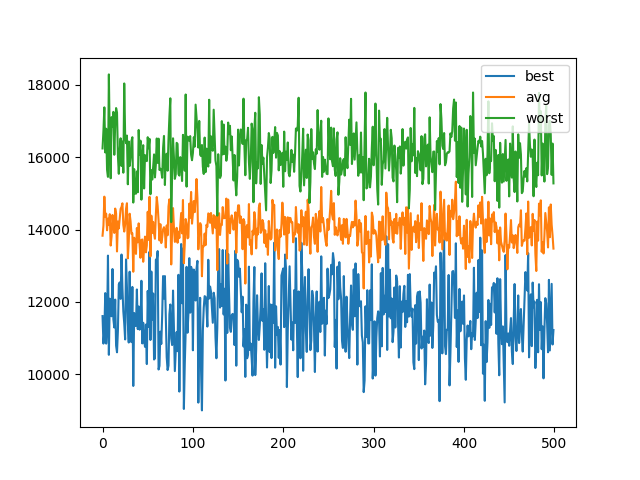
EASY/009



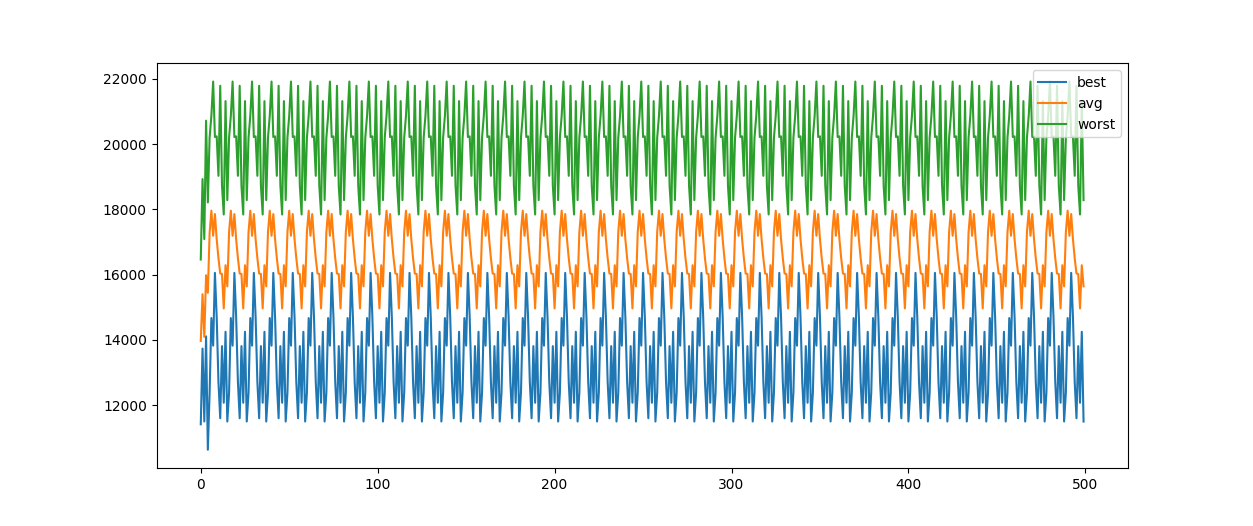
FLAT/006



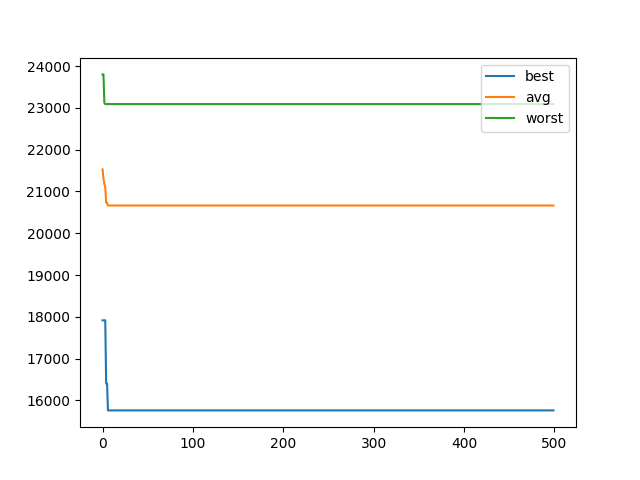
FLAT/007



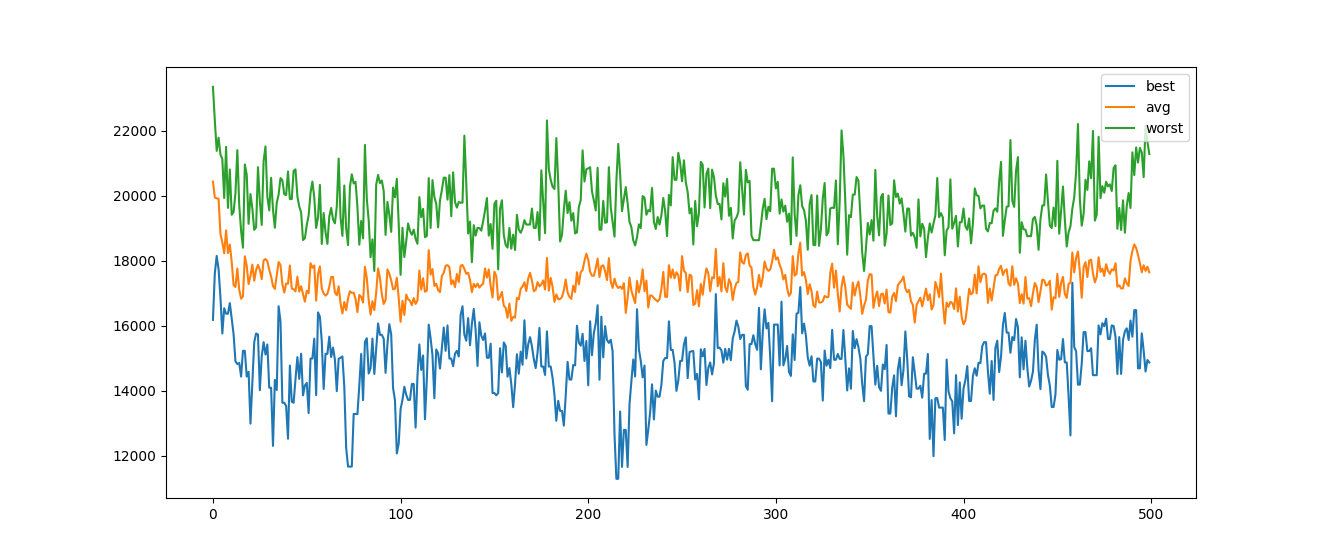
FLAT/008



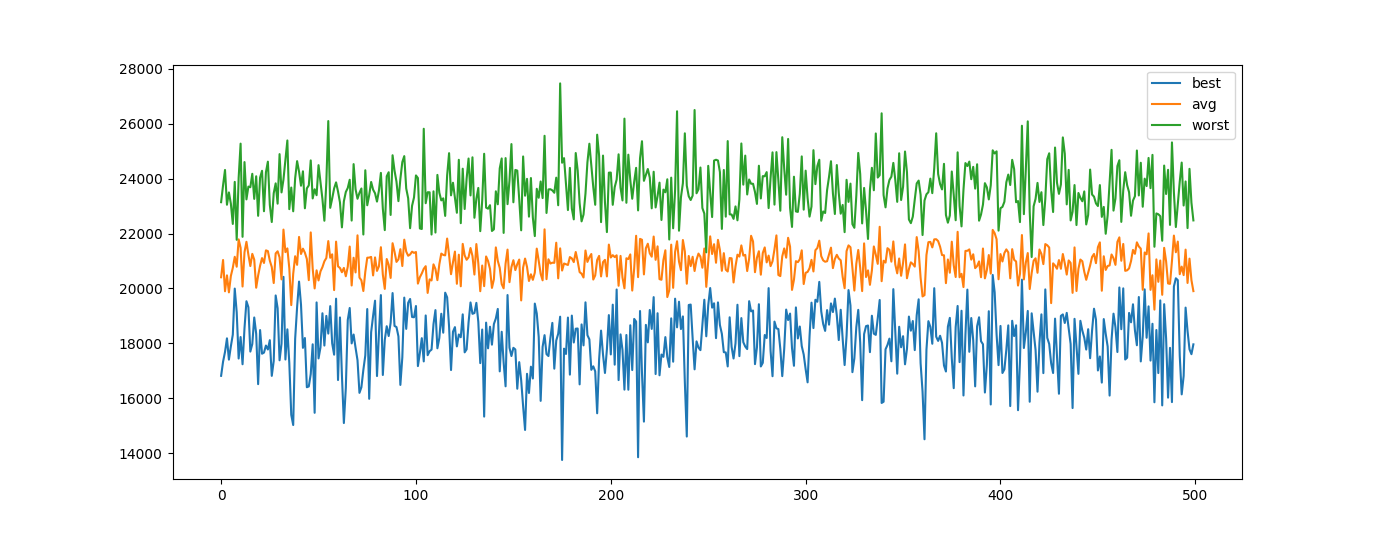
FLAT/009



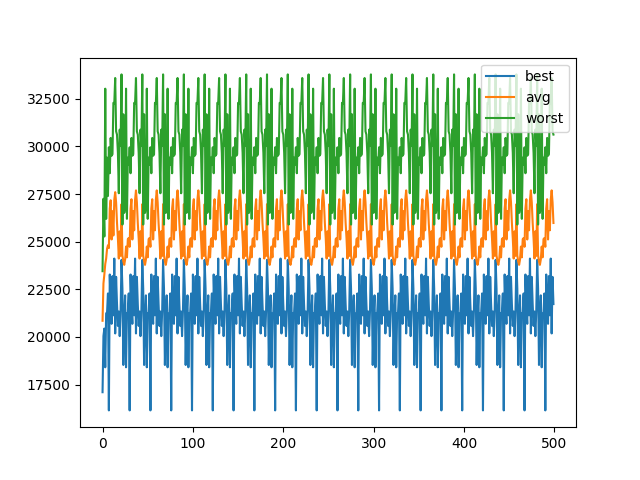
HARD/006



HARD/007



HARD/008



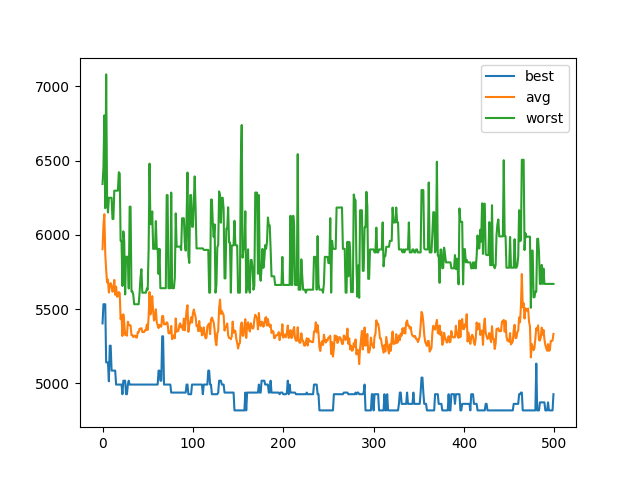
HARD/009

**WNIOSKI:** Ze względu na moją implementację operatora mutacji wykresy dla 0% oraz 100% szans mutacji mają sens. Gdy mutacja ma 0% szans to osobniki do pewnego momentu się krzyżują (przez pierwsze parę pokoleń), a następnie wszystkie zatrzymują się gdy mają ten sam genotyp (nie da się już wyjść z tej sytuacji). Dla 100% szans każdy gen u każdego osobnika zostanie zmieniony (każda maszyna zmieni pozycję o 1, co po chwili oczywiście wróci do początku i tak nastąpi cykl). W moim przypadku szansa między 5-15% zapewnia najlepsze rezultaty (pozwala uciec z najlepszego rozwiązania i eksplorować inne a z drugiej strony nie wprowadza chaosu). Szansa powyżej 20% oznacza, że średnio w każdym osobniku nastąpi mutacja genu co doprowadza do chaosu.

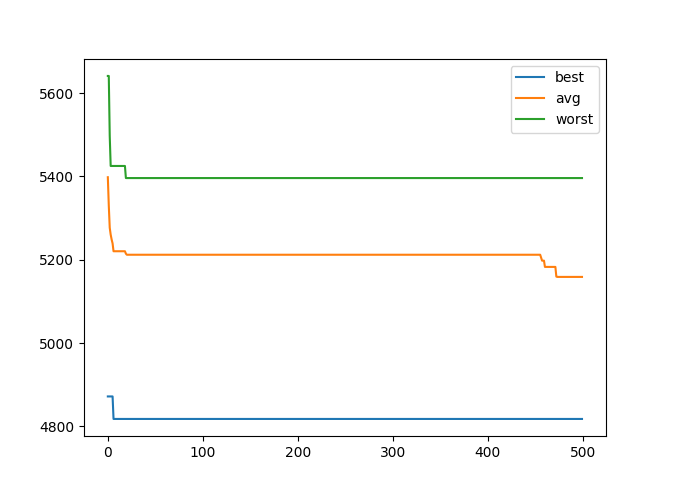
# Liczba osobników

XXX/010 – 10 osobników, XXX/011 – 50 osobników, XXX/012 – 200 osobników

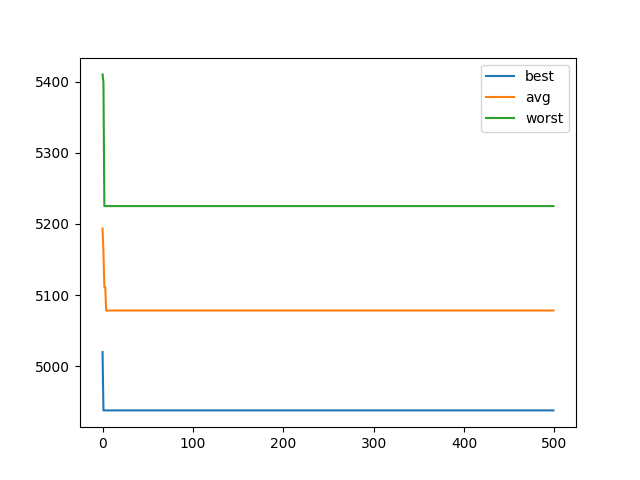
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| instancja | Alg. Ewolucyjny [10x] | | | | Metoda losowa [N] | | | |
|  | best | worst | avg | std | best | worst | avg | std |
| EASY/010 | 4927 | 5670 | 5333 | 250.3 | 4818 | 11730 | 7851.3 | 1053.8 |
| EASY/011 | 4818 | 5396 | 5158.8 | 194 | 4818 | 11730 | 7848.1 | 1051.7 |
| EASY/012 | 4938 | 5225 | 5078.3 | 108.9 | 4818 | 11730 | 7849.2 | 1052.9 |
| FLAT/010 | 9405 | 11205 | 10341.5 | 478.5 | 9290 | 24945 | 18093.6 | 2217.6 |
| FLAT/011 | 8445 | 9485 | 9033.5 | 360.1 | 8770 | 25195 | 18094 | 2225.4 |
| FLAT/012 | 8115 | 8930 | 8669.5 | 305.1 | 8625 | 25475 | 18090.5 | 2224.1 |
| HARD/010 | 15759 | 20147 | 17355.4 | 1238.9 | 14256 | 39321 | 26461 | 3165.7 |
| HARD/011 | 10439 | 14046 | 12163.3 | 1092.4 | 13429 | 41082 | 26451 | 3166.7 |
| HARD/012 | 11319 | 13779 | 12314.6 | 681.5 | 12330 | 41110 | 26449.7 | 3163.5 |



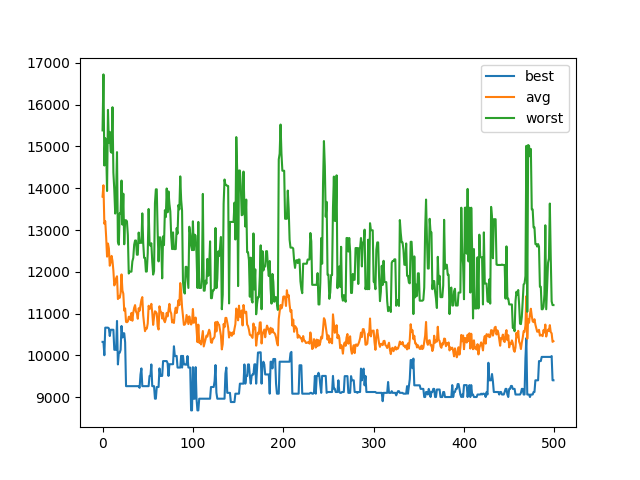
EASY/010



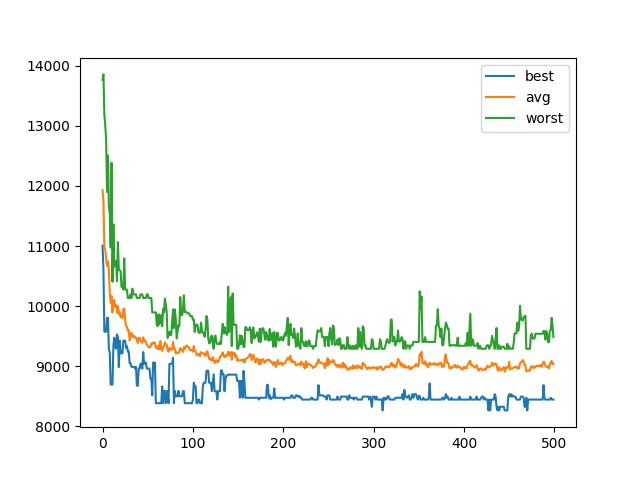
EASY/011



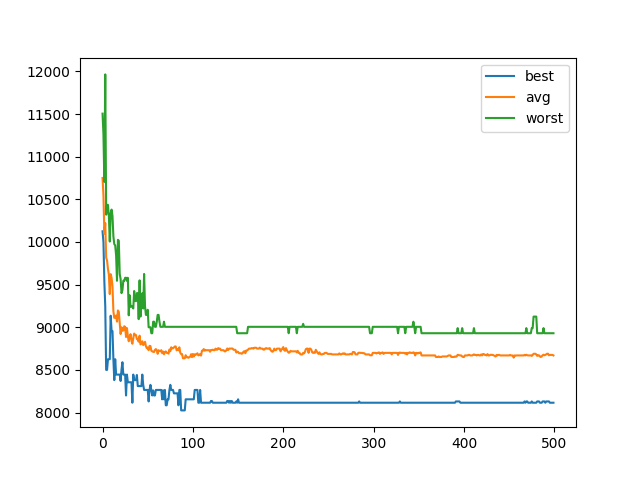
EASY/012



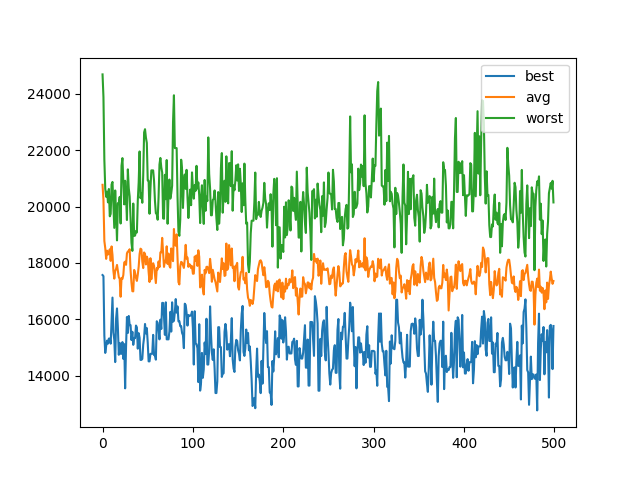
FLAT/010



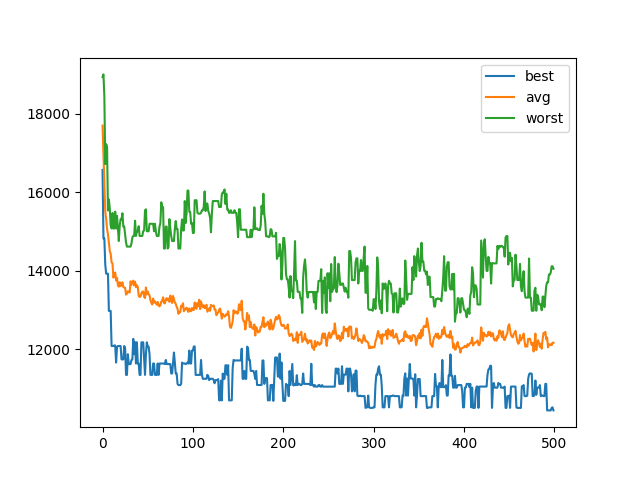
FLAT/011



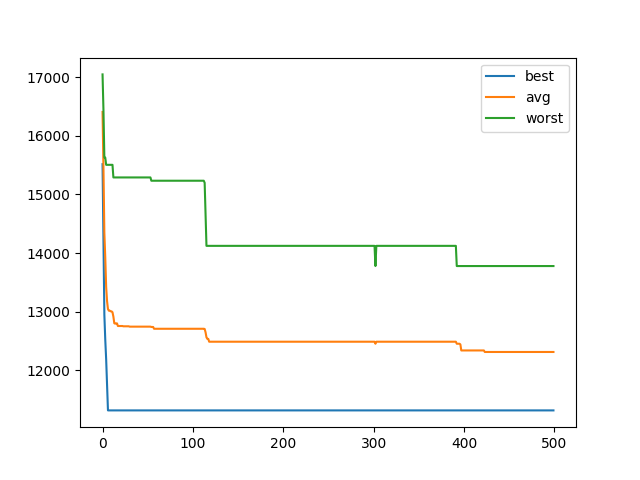
FLAT/012



HARD/010



HARD/011



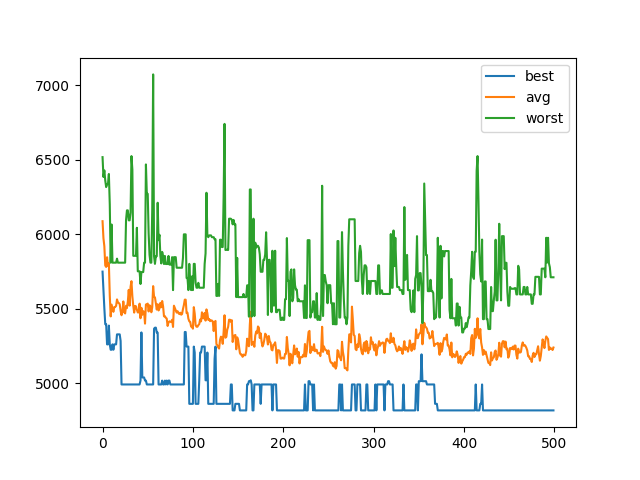
HARD/012

**WNIOSKI:** W zasadzie można stwierdzić, że im więcej osobników tym szybciej zbiegamy do lepszych rozwiązań i potrzeba na to mniej pokoleń (oba parametry są ze sobą powiązane). Wprowadza to jednak oczywiście dodatkowy nakład obliczeniowy, więc należy liczbę osobników dopasować do problemu oraz poszukiwanego rozwiązania.

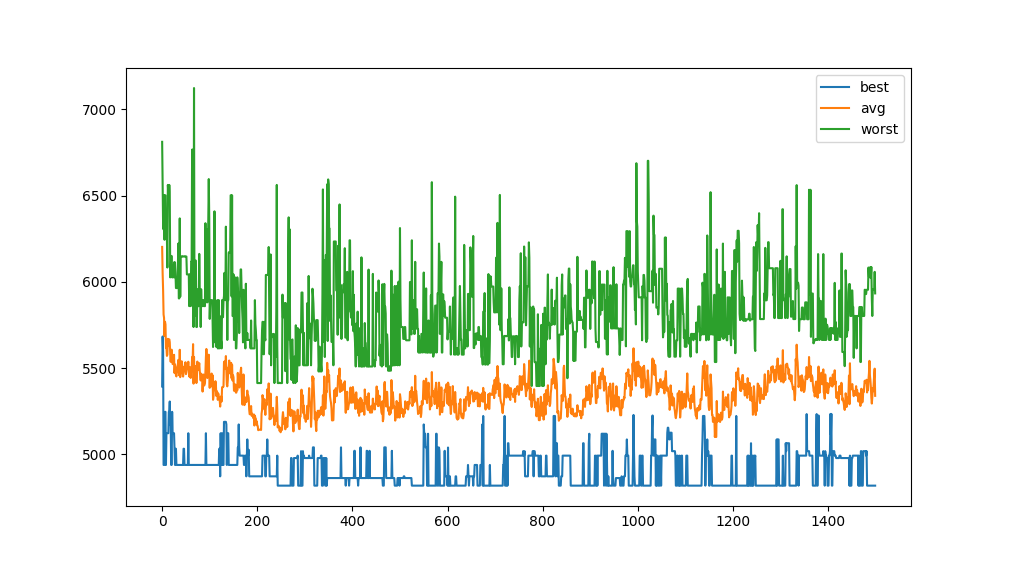
# Liczba pokoleń

XXX/013 – 500 pokoleń, XXX/014 – 1500 pokoleń, XXX/015 – 4000 pokoleń

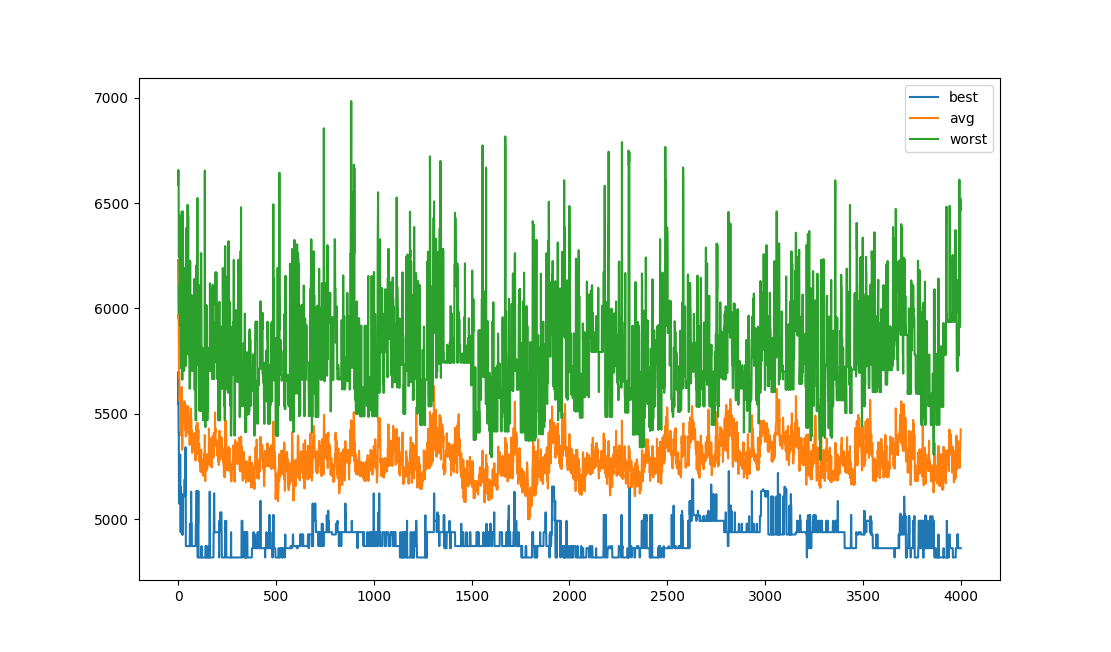
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| instancja | Alg. Ewolucyjny [10x] | | | | Metoda losowa [N] | | | |
|  | best | worst | avg | std | best | worst | avg | std |
| EASY/013 | 4818 | 5711 | 5239.8 | 287.7 | 4818 | 11724 | 7838.7 | 1052.1 |
| EASY/014 | 4818 | 5933 | 5338.8 | 357 | 4818 | 11730 | 7852.2 | 1054.7 |
| EASY/015 | 4862 | 6497 | 5425.5 | 456.9 | 4818 | 11730 | 7848.3 | 1052.7 |
| FLAT/013 | 9590 | 12595 | 10902 | 868.7 | 8820 | 25030 | 18085.3 | 2220.3 |
| FLAT/014 | 9340 | 11585 | 10060.5 | 653.8 | 8130 | 25185 | 18081 | 2224.6 |
| FLAT/015 | 9325 | 11305 | 10081.5 | 572.4 | 8960 | 25205 | 18095.3 | 2222.2 |
| HARD/013 | 16257 | 20003 | 17603.6 | 1122.8 | 14905 | 39920 | 26431.5 | 3172.9 |
| HARD/014 | 16012 | 20930 | 17842.6 | 1471.6 | 13800 | 41481 | 26437.2 | 3166.5 |
| HARD/015 | 15014 | 20530 | 18358.6 | 1659.1 | 12985 | 40362 | 26444.1 | 3164.8 |



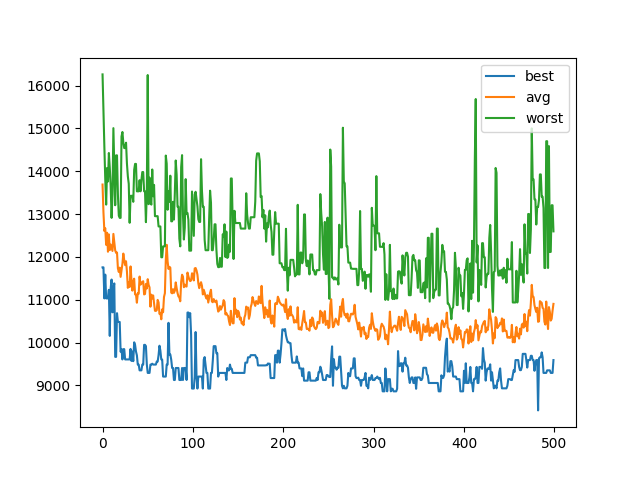
EASY/013



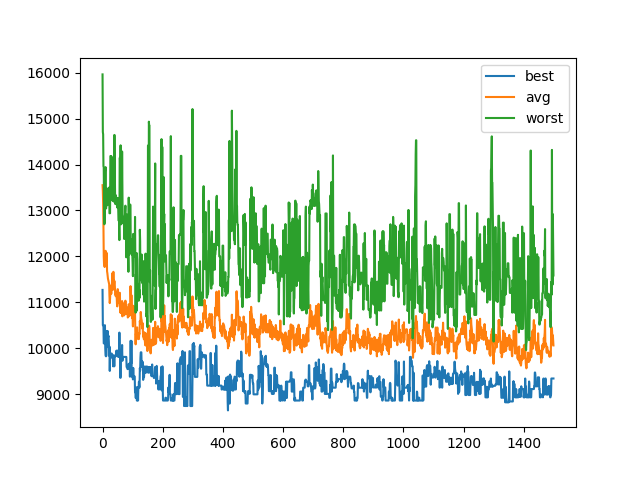
EASY/014



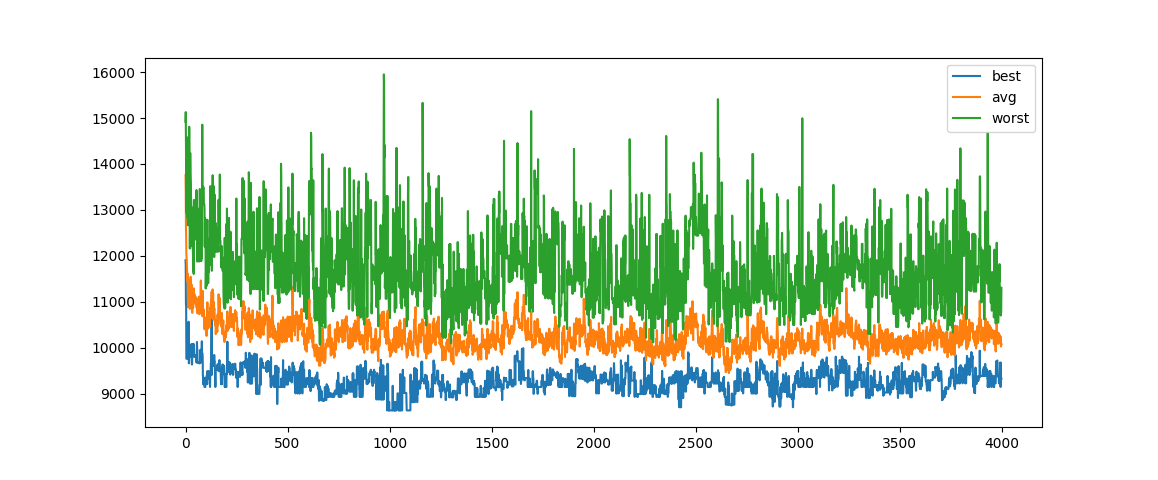
EASY/015



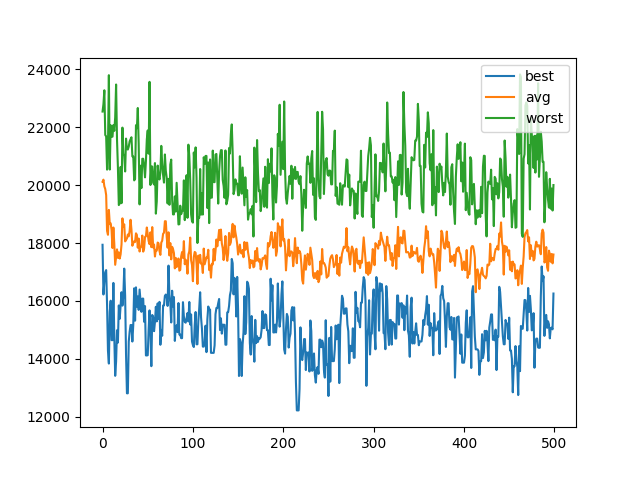
FLAT/013



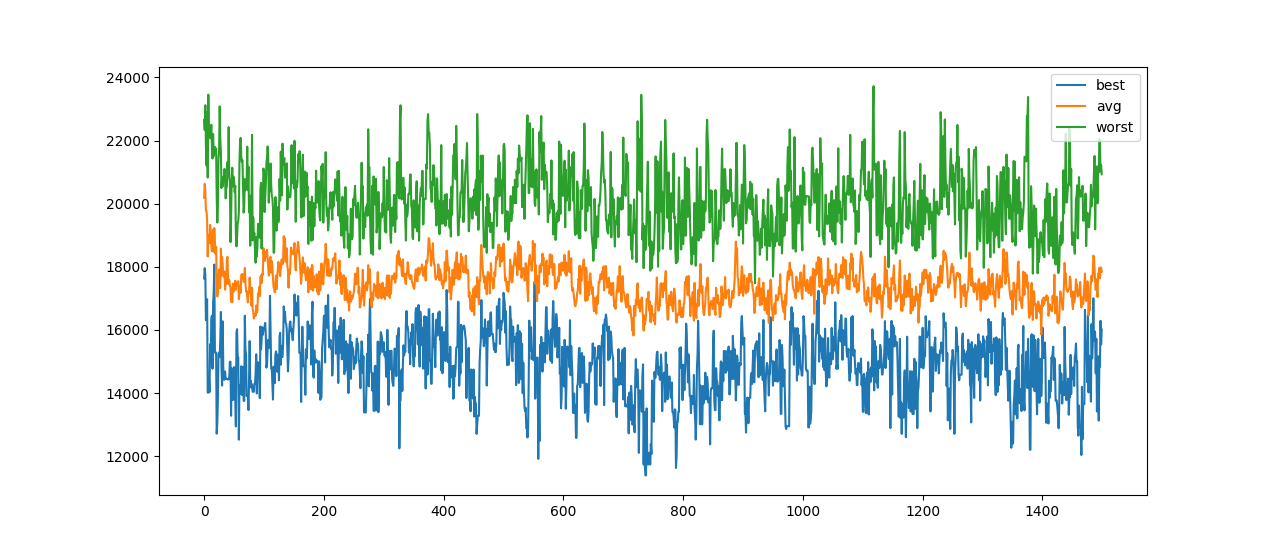
FLAT/014



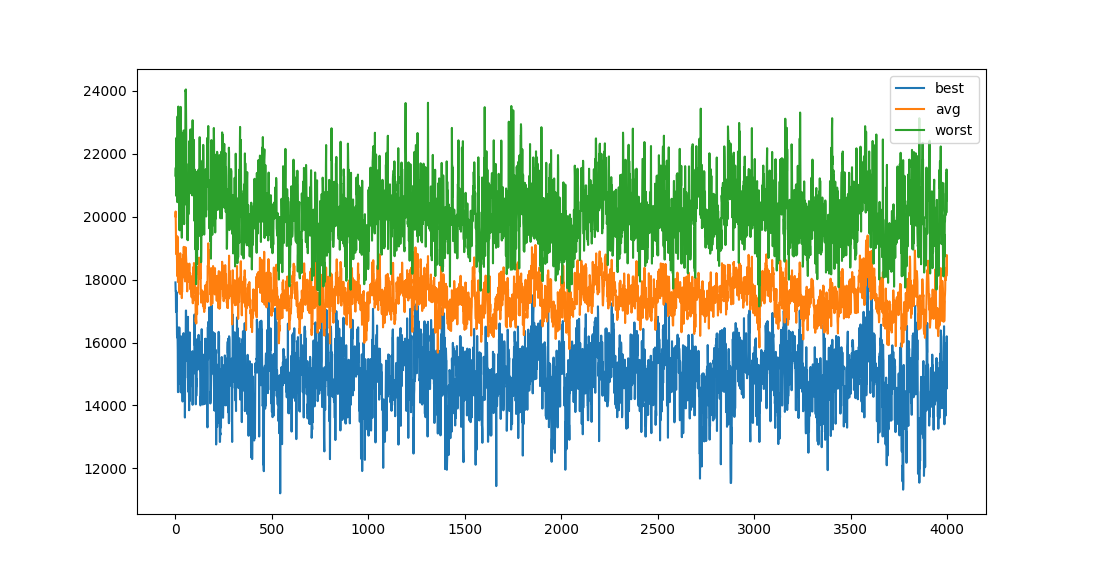
FLAT/015



HARD/013



HARD/014



HARD/015

**WNIOSKI:** Ze względu na zbyt małą populację wyniki nie oddają znaczenia liczby pokoleń. Im większa populacja tym bardziej istotne będzie zwiększenie liczby pokoleń (potrzeba przynajmniej kilkudziesięciu osobników i odpowiedniej mutacji by móc zejść do odpowiedniego poziomu i ewentualnie w późniejszych pokoleniach eksplorować nowe ścieżki).

# Próby wyznaczenia rozwiązania optymalnego

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| instancja | Alg. Ewolucyjny [10x] | | | | Metoda losowa [N] | | | |
|  | best | worst | avg | std | best | worst | avg | std |
| SOL/001 | 9464 | 12045 | 10760.2 | 728.1 | 13357 | 42331 | 26451.4 | 3164.2 |
| SOL/002 | 9314 | 11842 | 10803.4 | 673.5 | 14273 | 40692 | 26449.5 | 3163.6 |
| SOL/003 | 9862 | 12542 | 10905.7 | 943.7 | 12619 | 41207 | 26451.5 | 3164.6 |

**PARAMETRY:**

num\_of\_generations = 500

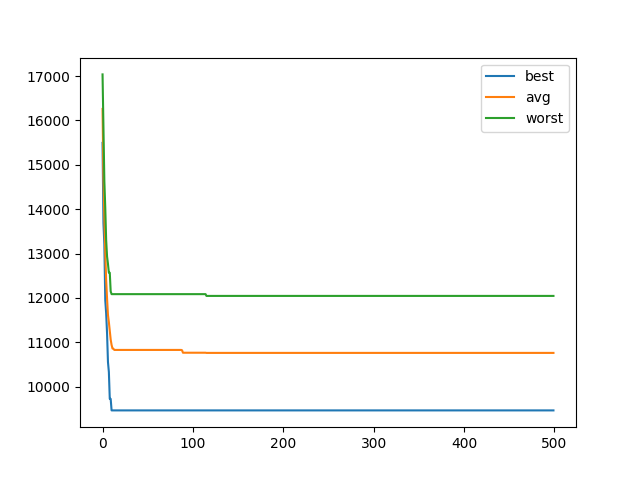
is\_tournament = True

num\_of\_pop = 500

select\_n = 25

mut\_prob = 0.07

cross\_prob = 0.7



HARD\_SOLVING\_1

**PARAMETRY:**

num\_of\_generations = 50

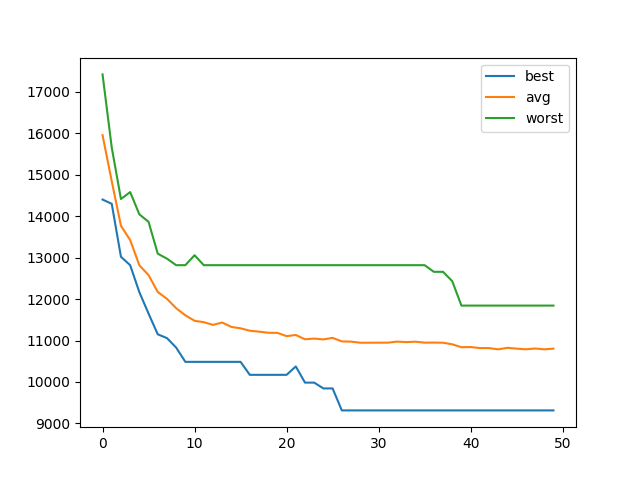
is\_tournament = True

num\_of\_pop = 500

select\_n = 50

mut\_prob = 0.15

cross\_prob = 0.7



HARD\_SOLVING\_2

**PARAMETRY:**

num\_of\_generations = 500

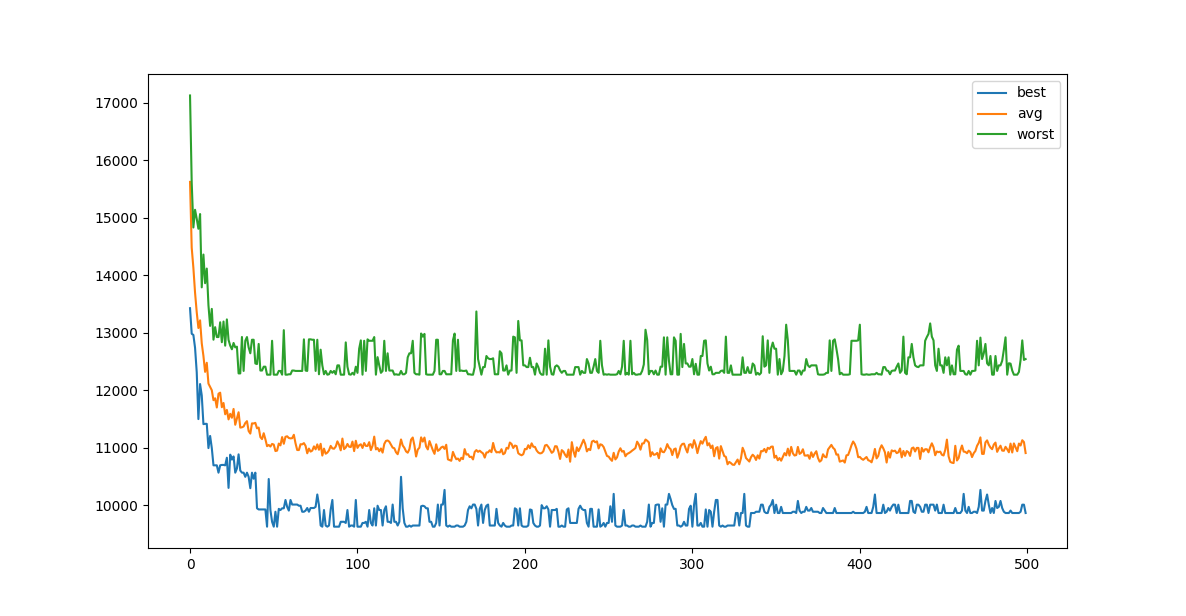
is\_tournament = True

num\_of\_pop = 500

select\_n = 50

mut\_prob = 0.18

cross\_prob = 0.7



HARD\_SOLVING\_3

**PARAMETRY:**

num\_of\_runs = 1

num\_of\_generations = 500

is\_tournament = True

no\_progress\_for = 250

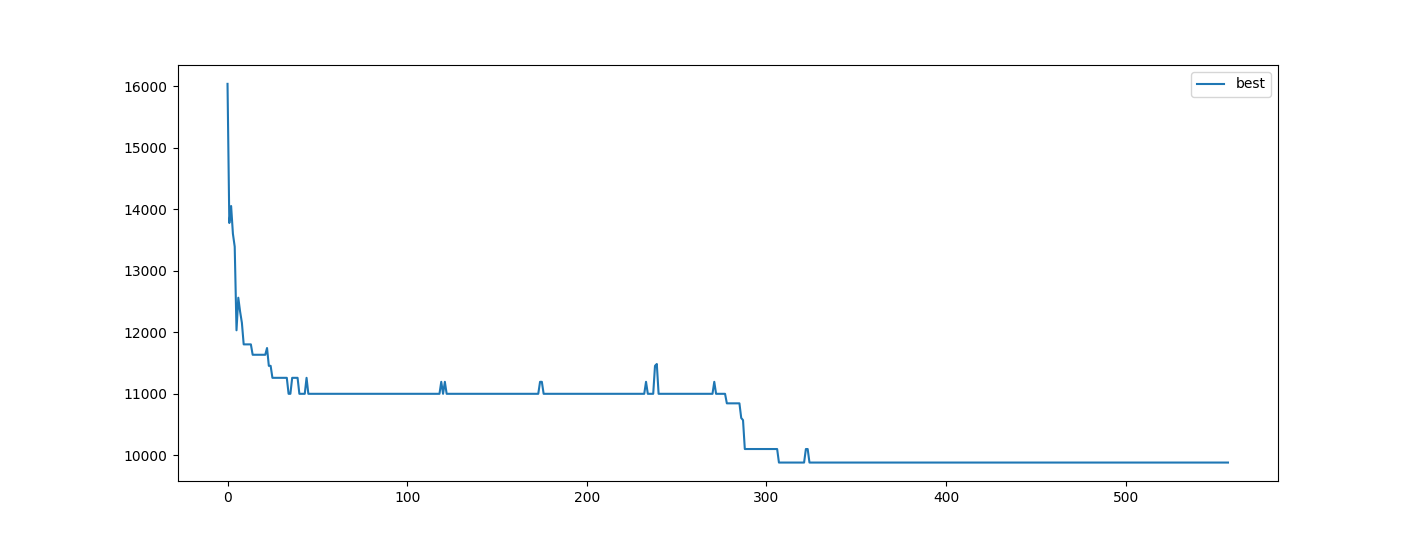
is\_end\_cond\_dynamic = True

num\_of\_pop = 500

select\_n = 50

mut\_prob = 0.15

cross\_prob = 0.75



HARD\_SOLVING\_4

**WNIOSKI:**

HARD\_SOLVING\_1 – za duża liczba pokoleń i za mała mutacja

HARD\_SOLVING\_2 a HARD\_SOLVING\_3 – obie próby uważam za dość dobre, odróżnia je jedynie 3% szansy na mutację a już widać na wykresie wyraźną różnicę. HS2 prawie w ogóle nie eksploruje nowych ścieżek, natomiast udało mu się zbiec do przyzwoitego wyniku. HS3 wiele razy zbiegał do około 9,5tys. a dodatkowo cały czas wychodził i eksplorował

HARD\_SOLVING\_4 – pojedynczy bieg dla dynamicznego warunku stopu, jak widać nawet po 280 pokoleniach udało się znacząco poprawić wynik i utrzymać na tym poziomie

Ogólnie w zasadzie każdy parametr ma duże znaczenie i odpowiedni ich dobór jest zarazem ciężki (bo zależy od wielu czynników: implementacji, danych, mocy obliczeniowej, poszukiwanego rozwiązania) a zarazem bardzo istotny. Najmniejsza zmiana choćby jednego z nich może znacząco wpłynąć na proces poszukiwania rozwiązania a co za tym idzie na przydatność zaimplementowanego algorytmu.