

Autor:

Kacper Górka

Prowadzący:

mgr inż. Rypeś Grzegorz

RAPORT DO LABORATORIUM NR 2

STRATEGIA EWOLUCYJNA $\mu + \lambda$

WPROWADZENIE DO SZTUCZNEJ INTELIGENCJI

1 Treść ćwiczenia

1.1 Zadania do wykonania

Proszę znaleźć minima oraz maksima funkcji $f(x, y) = \frac{9xy}{e^{x^2+0.5x+y^2}}$ wykorzystując strategię ewolucyjną $\mu + \lambda$ (ES, ang. Evolution Strategy). Strategia ma dokonywać mutacji osobnika za pomocą dodania do niego szumu Gaussowskiego. Krzyżowanie ma być dokonane za pomocą interpolacji, tzn. wynikiem krzyżowania osobników $o1$ i $o2$ jest osobnik $a \cdot o1 + (1-a) \cdot o2$, gdzie a jest zmienną losową z rozkładu jednostajnego na przedziale $[0;1]$. Proszę samemu zaproponować strategię selekcji i eliminacji. Osobnik jest tutaj dwuwymiarowym wektorem, niech będzie on reprezentowany przez numpy array, albo torch tensor.

1.2 Pytania

- Co jest przestrzenią poszukiwań?
- Czy jest potrzebny gradient?
- Jak odchylenie standardowe w szumie mutacji (σ) wpływa na wyniki (wartości 0.01, 0.1, 1, 10)?
- Jak liczba rodziców μ i dzieci λ wpływa na proces optymalizacji (wartości $\{1, 1\}$, $\{1, 16\}$, $\{16, 1\}$, $\{16, 16\}$, $\{128, 512\}$)
- Proszę punkt startowy ES (dla $\mu, \lambda = (128, 512)$, $\sigma = 0.1$) oraz algorytmu SGD z poprzedniego zadania ustawić na punkt (10,10) i porównać wyniki. Co i dlaczego można zaobserwować?
- Kiedy ES jest lepszym rozwiązaniem niż SGD?
- Który algorytm optymalizacji jest bardziej złożony obliczeniowo?

2 Wstęp teoretyczny

2.1 Podstawowe pojęcia i złożenia

Co jest przestrzenią poszukiwań?

Przestrzenią poszukiwań to przestrzeń startowa populacji. Najlepiej aby obejmowała rozwiązanie optymalne.

Czy jest potrzebny gradient?

Algorytm ewolucyjny do znalezienia minimów lub maksimów nie potrzebuje znajomości gradientu funkcji, gdyż jego analiza opiera się na porównywaniu wartości osobników.

3 Wyniki

3.1 Odnalezione minima oraz maksima

Minima:

```
Minimum 1: [-0.83945554  0.71161454] with value: -2.4367150361270657
Error x = 0.003614461068629393  y = 0.0045075427033229865  z = 0.00015496387293412184
Minimum 2: [ 0.60363924 -0.72274749] with value: -1.1962441527116605
Error x = 0.010569239025720156  y = -0.015640490955485364  z = 0.0009058472883394231
```

Maksima:

```
Maksimum 1: [0.59407807 0.69858348] with value: 1.1969693004366246
Error x = 0.001008067330290685  y = -0.008523518329582314  z = -0.00018069956337529725
Maksimum 2: [-0.86025504 -0.70217493] with value: 2.4355305984719284
Error x = -0.017185037997487163  y = 0.004932072486147354  z = -0.0013394015280714378
Maksimum 3: [-0.8472394 -0.6901613] with value: 2.43538568564859
Error x = -0.004169398641570177  y = 0.016945696171297264  z = -0.0014843143514098323
```

Błąd bezwzględny dla wartości nie przekracza $2e-3$.

3.2 Porównanie rezultatów dla różnych wartości sigma oraz mu+lambda

ZMIENNE SIGMA

```
Results for sigma = 0.001:
Minimum 1: [-0.84280802  0.70735507] with value: -2.4368675754542397
Error x = 0.00026198002328547965 y = 0.00024807029479989584 z = 2.4245457601601572e-06

Results for sigma = 0.1:
Minimum 1: [-0.86097975  0.69615366] with value: -2.4349573546274925
Error x = -0.017909752092305498 y = -0.010953338623209352 z = 0.001912645372507349

Results for sigma = 1:
Minimum 1: [-1.00138417  0.76829109] with value: -2.3225641844816183
Error x = -0.1583141695254493 y = 0.06118408624816973 z = 0.11430581551838159

Results for sigma = 10:
```

Jak odchylenie standardowe w szumie mutacji (σ) wpływa na wyniki (wartości 0.01, 0.1, 1, 10)?

Najlepsze rezultaty otrzymano dla $\sigma = 0.001$. Wynika to ze znajomości funkcji i rozrzuceniu populacji na powierzchni 6 na 6 o środka w układzie współrzędnych. W przypadku szukania rozwiązania na nieznanym funkcji ta metoda nie byłaby skuteczna. Dla zbyt dużego σ nie osiągnięto rozwiązania w pobliżu optimum.

ZMIENNE MU+LAMBDA

```
Results for mu = 1 and lambda = 1:

Results for mu = 1 and lambda = 16:

Results for mu = 16 and lambda = 1:
Minimum 1: [-0.7858971  0.72696973] with value: -2.421177606107779
Error x = 0.05717290494640703 y = 0.019862729765618203 z = 0.015692393892220657

Results for mu = 16 and lambda = 16:
Minimum 1: [-0.83651257  0.72655814] with value: -2.434862649635202
Error x = 0.006557425916124959 y = 0.019451142886243278 z = 0.002007350364797933

Results for mu = 128 and lambda = 512:
Minimum 1: [-0.84416792  0.70318811] with value: -2.436788183759856
Error x = -0.0010979213387500852 y = -0.003918893827491443 z = 8.181624014369504e-05
```

Jak liczba rodziców μ i dzieci λ wpływa na proces optymalizacji (wartości {1, 1}, {1, 16}, {16, 1}, {16, 16}, {128, 512})?

Na podstawie otrzymanych wyników można stwierdzić, że w tym przypadku większa ich liczba przynosi lepsze rezultaty, aczkolwiek nie jest to ogólna prawda. Dla {1, 1} i {1, 16} nie otrzymano wyników w pobliżu minimum.

Proszę punkt startowy ES (dla μ , $\lambda=(128, 512)$, $\sigma=0.1$) oraz algorytmu SGD z poprzedniego zadania ustawić na punkt (10,10) i porównać wyniki. Co i dlaczego można zaobserwować?

```
Minimum 1: [-0.84228887  0.70408086] with value: -2.4368209376844376  
Error x = 0.0007811268753885603  y = -0.0030261366237763765  z = 4.906231556223162e-05
```

Algorytm SGD nie otrzymał rozwiązania gdyż gradient biegnie w przeciwną stronę i rozwiązanie odbiega od minimum. W przypadku algorytmu ewolucyjnego: wymagało one kilku usprawnień. Przede wszystkim należało zastosować zmienną sigmę, która na początku wynosiła 3, a na końcu 0.05 – stąd dobry wynik. Nie jest to jednak wystarczające, gdyż należało również zmienić strategię. Co generację wybierano tylko 10% najlepszych osobników, a resztę losowano. Pozwoliło to na eksplorację terenu przy zachowaniu najlepszego wyniku.

4 Wnioski

Kiedy ES jest lepszym rozwiązaniem niż SGD?

Algorytm spadku gradientu może być lepszy dla problemów gdy gradient jest łatwy do wyznaczenia, a funkcja jest gładka i ciągła.

Algorytm ewolucyjny może być lepszy dla problemów, w których ciężko jest obliczyć gradient lub gdy nie jest on wyznaczalny w pewnych punkt.

Który algorytm optymalizacji jest bardziej złożony obliczeniowo?

Złożoność obliczeniowa obydwu algorytmów silnie zależy od problemu. W przypadku gdy gradient jest trudny do wyznaczenia, to algorytm SGD może być bardziej złożony obliczeniowo. Natomiast gdy do rozwiązania będziemy potrzebowali wielu osobników i ich sposób selekcji również będzie złożony to ES okaże się być bardziej złożony obliczeniowo.