

Politechnika Warszawska

WYDZIAŁ ELEKTRONIKI
I TECHNIK INFORMACYJNYCH



Sprawozdanie

Wprowadzenie do sztucznej inteligencji

Ćwiczenie nr. 1

Mikołaj Bańkowski

Numer albumu 310408

prowadzący
Grzegorz Rypeś

Warszawa 2023

Spis treści

1. Treść ćwiczenia	3
2. Pochodne cząstkowe funkcji	4
3. Minima i maksima funkcji	5
4. Czym jest gradient	10
5. Jak punkt startowy wpływa na wynik	11
6. Jak wartość kroku uczącego wpływa na proces optymalizacji.....	13
7. Jak można zwiększyć precyzję znalezionych ekstremów.....	22

1. Treść ćwiczenia

Proszę znaleźć minima oraz maksima funkcji $f(x, y) = \frac{x \cdot y}{e^{(x^2 + 0.5x + y^2)}}$ wykorzystując algorytm spadku wzdłuż gradientu SGD (omówiony na wykładzie). Wzór na pochodne cząstkowe należy wyprowadzić ręcznie na kartce i jej skan/zdjęcie załączyć do raportu.

Odpowiedzieć na pytania:

- 1) Czym jest gradient $\nabla f(x, y)$?
- 2) Jak punkt startowy algorytmu wpływa na wynik?
- 3) Jak wartość kroku uczącego wpływa na proces optymalizacji?
- 4) Jak można zwiększyć precyzję znalezionych ekstremów?

Odpowiedzi na pytania należy poprzeć eksperymentami. Wyniki eksperymentów, odpowiedzi oraz wnioski należy zamieścić w raporcie

Uwagi:

Należy tę funkcję sobie zwizualizować korzystając przykładowo z:

<https://c3d.libretexts.org/CalcPlot3D/index.html> lub <https://www.wolframalpha.com/>.

2. Pochodne cząstkowe funkcji

$$f(x, y) = \frac{xy}{e^{(x^2 + 0.5x + y^2)}}$$

$$\frac{\partial f}{\partial x} = \frac{y e^{(x^2 + 0.5x + y^2)} - xy(2x + 0.5) e^{(x^2 + 0.5x + y^2)}}{e^{(x^2 + 0.5x + y^2)^2}}$$

$$= y \frac{(1 - 2x^2 - 0.5x)}{e^{(x^2 + 0.5x + y^2)}} = \frac{y(-2x^2 - 0.5x + 1)}{e^{(x^2 + 0.5x + y^2)}} \cdot \frac{e^{(-x^2 - 0.5x - y^2)}}{e^{(-x^2 - 0.5x - y^2)}}$$

$$= (-2x^2 - 0.5x + 1) y e^{(-x^2 - 0.5x - y^2)}$$

$$\frac{\partial f}{\partial y} = \frac{e^{(x^2 + 0.5x + y^2)} - xy \cdot 2y \cdot e^{(x^2 + 0.5x + y^2)}}{e^{(x^2 + 0.5x + y^2)^2}}$$

$$= \frac{x - 2xy^2}{e^{(x^2 + 0.5x + y^2)}} = \frac{x - 2xy^2}{e^{(x^2 + 0.5x + y^2)}} \cdot \frac{e^{(-x^2 - 0.5x - y^2)}}{e^{(-x^2 - 0.5x - y^2)}}$$

$$= (x - 2xy^2) e^{(-x^2 - 0.5x - y^2)}$$

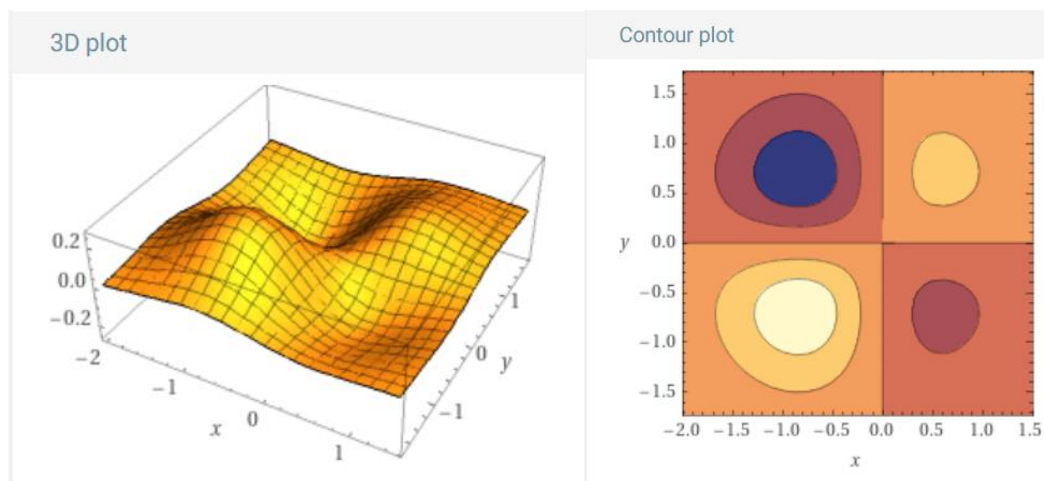
$$\frac{\partial f}{\partial x} = (-2x^2 - 0.5x + 1) y e^{(-x^2 - 0.5x - y^2)}$$

$$\frac{\partial f}{\partial y} = (x - 2xy^2) e^{(-x^2 - 0.5x - y^2)}$$

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] - \text{gradient}$$

3. Minima i maksima funkcji

Do wstępnej analizy badanej funkcji i jej wizualizacji, wykorzystano narzędzie [WolframAlpha](#)



Na podstawie analizy poniższych wykresów funkcji, można zauważyć, że funkcja posiada dwa minima i dwa maksima.

Minima funkcji znajdują się w II i IV ćwiartce układu współrzędnych. Są to kolejno

$$\min\{f(x,y)\} \approx -0.270763 \text{ dla } (x, y) \approx (-0.84307, 0.707107)$$

$$\min\{f(x,y)\} \approx -0.133016 \text{ dla } (x, y) \approx (0.59307, -0.707107)$$

Maxima funkcji znajdują się w I i III ćwiartce układu współrzędnych. Są to kolejno

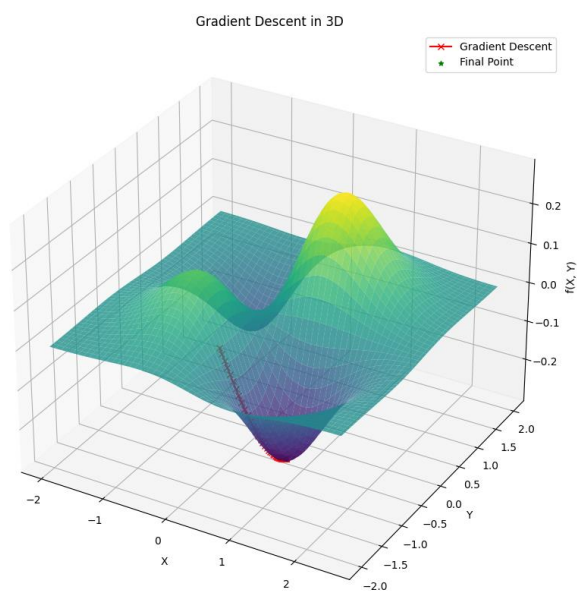
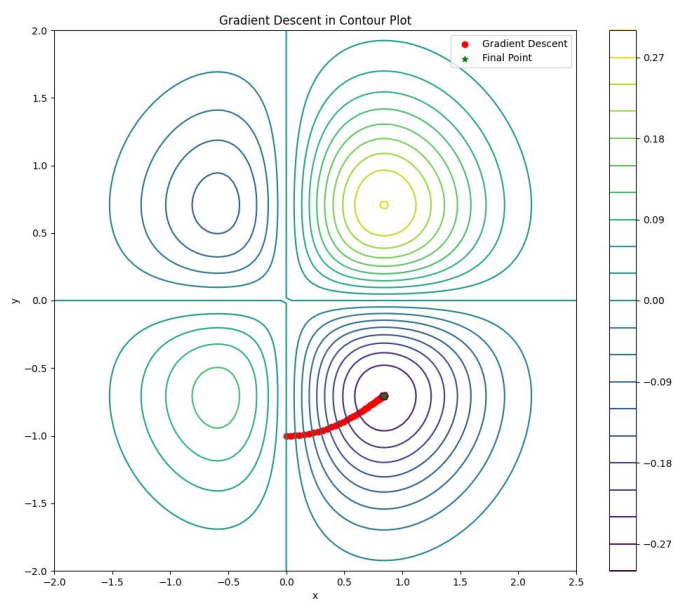
$$\max\{f(x,y)\} \approx 0.133016 \text{ dla } (x, y) \approx (0.59307, 0.707107)$$

$$\max\{f(x,y)\} \approx 0.270763 \text{ dla } (x, y) \approx (-0.84307, -0.707107)$$

Minima funkcji znalezione za pomocą algorytm spadku wzdłuż gradientu SGD

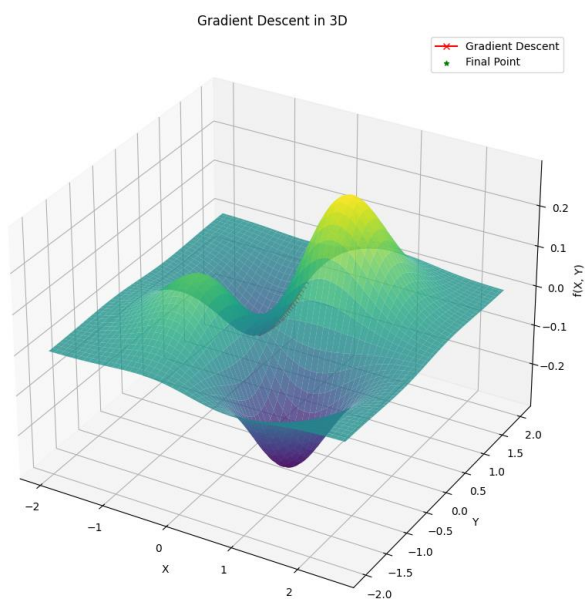
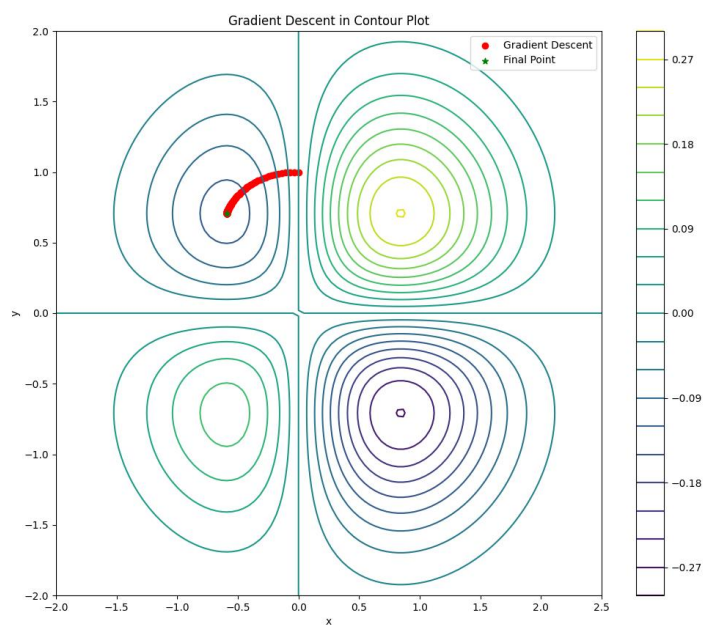
Dla danych wejściowych:

Krok uczenia	Punkt startowy	Znalezione minimum
0.1	$x = 0, y = -1$	-0.270763128698



Dla danych wejściowych

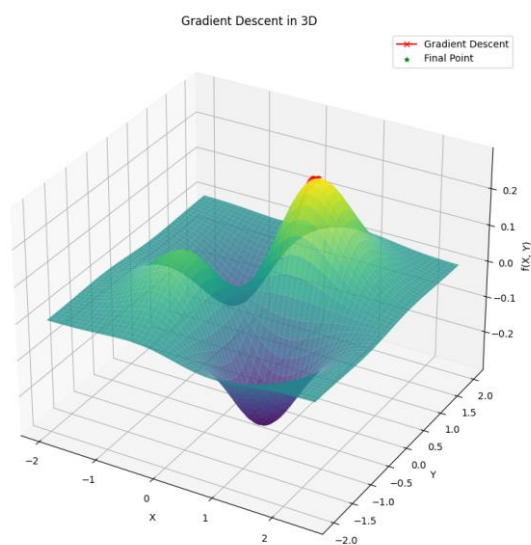
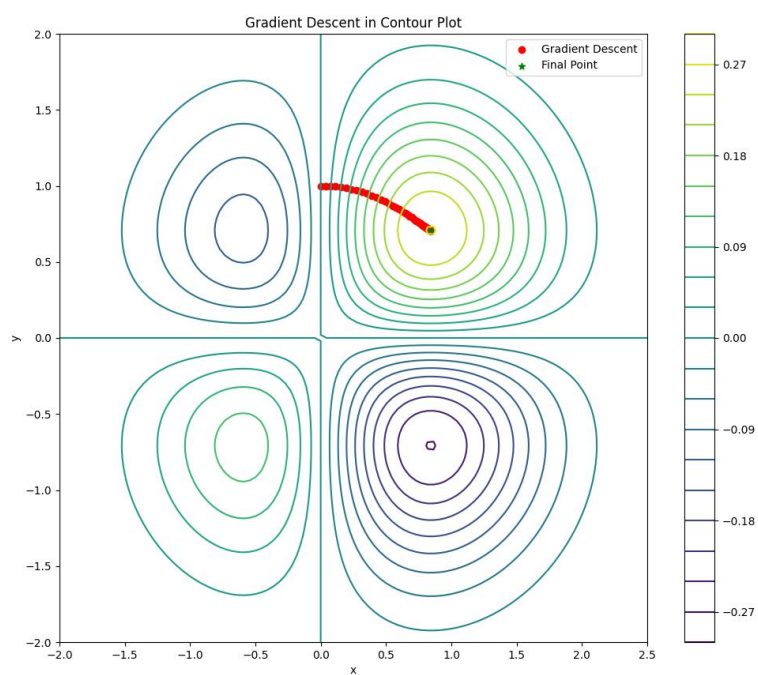
Krok uczący	Punkt startowy	Znalezione minimum
0.1	$x = 0$, $y = 1$	-0.133016318358



Maxima funkcji znalezione za pomocą algorytm spadku wzdłuż gradientu SGD

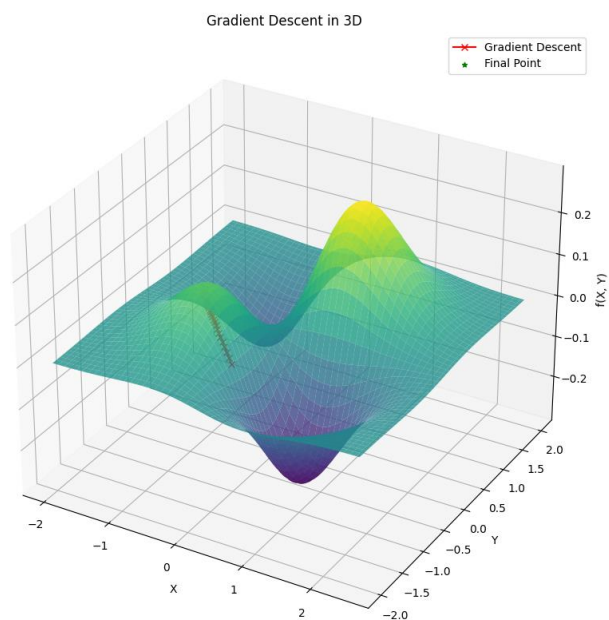
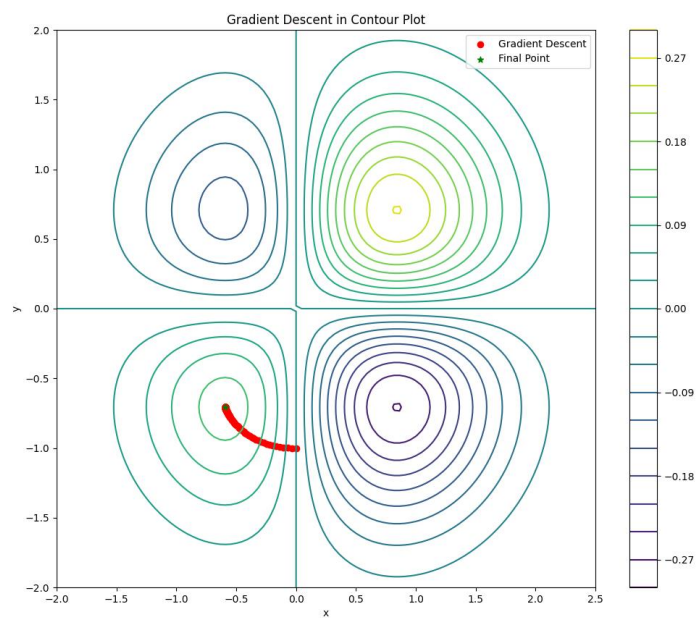
Dla danych wejściowych

Krok uczący	Punkt startowy	Znalezione maximum
0.1	$x = 0, y = 1$	0.270763128698



Dla danych wejściowych

Krok uczący	Punkt startowy	Znalezione maximum
0.1	$x = 0$, $y = -1$	0.133016318358



4. Czym jest gradient

Gradient to wektor, który w danym punkcie przestrzeni wskazuje kierunek i wartość najszybszego wzrostu funkcji skalarnej w tym punkcie. W jego skład wchodzi pochodne cząstkowe funkcji względem każdej ze zmiennych niezależnych.

W kontekście optymalizacji funkcji celu, gradient jest kluczowym narzędziem, które pomaga algorytmom optymalizacyjnym znajdować minimum lub maksimum funkcji. Dzięki niemu możemy określić kierunek najszybszego wzrostu (gdzie wartość funkcji rośnie najszybciej) oraz przeciwnie - kierunek najszybszego spadku (gdzie wartość funkcji maleje najszybciej).

Algorytm spadku wzdłuż gradientu SGD używa gradientu do aktualizacji parametrów modelu w celu minimalizacji funkcji kosztu. Algorytmy te iteracyjnie poruszają się w kierunku przeciwnym do gradientu, aż osiągną zbieżność do lokalnego minimum lub maksimum

5. Jak punkt startowy wpływa wynik

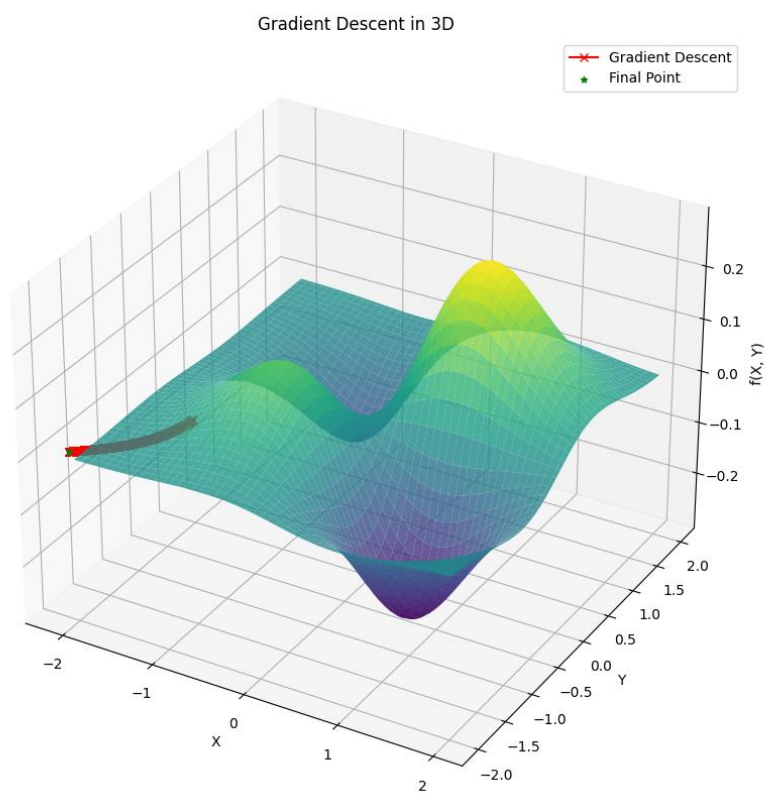
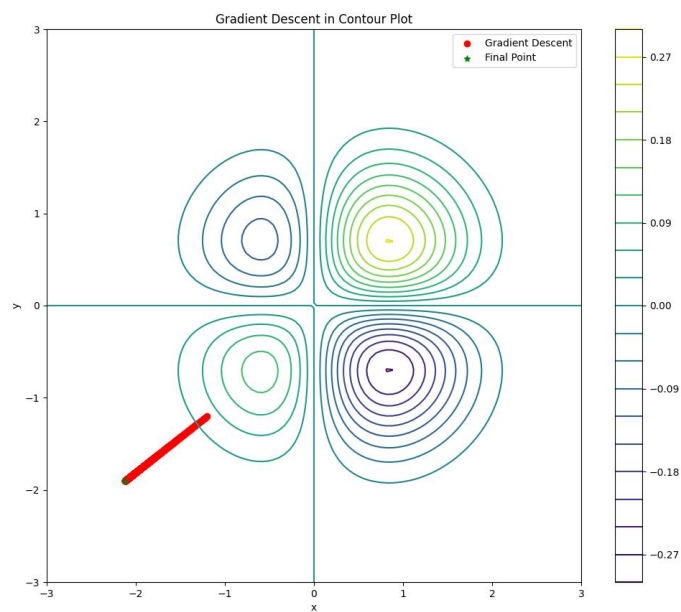
Szukanie minimum funkcji za pomocą SGD w zależności od punktu startowego

Algorytm spadku wzdłuż gradientu został uruchomiony 20 razy w celu znalezienia minimum funkcji, punkt startowy x i punkt startowy y, za każdym razem były wybierane losowo z przedziału $[-2,2]$

Tabela przedstawia rezultat eksperymentu

	Learning Rate	Initial X	Initial Y	Iterations	Minimum
0	0.1	-1.768787	0.563134	232.0	-0.133016
1	0.1	-0.251901	-0.792047	117.0	-0.270763
2	0.1	-0.954328	-1.771894	1001.0	0.000417
3	0.1	-1.157402	-1.161601	1001.0	0.000422
4	0.1	0.807193	0.111890	92.0	-0.270763
5	0.1	-1.714941	-0.554264	1001.0	0.000375
6	0.1	-0.616258	1.108441	173.0	-0.133016
7	0.1	-0.175266	-0.491809	116.0	-0.270763
8	0.1	1.122022	0.009465	105.0	-0.270763
9	0.1	-1.743933	-0.228734	426.0	-0.133016
10	0.1	0.989390	0.744971	1001.0	0.000407
11	0.1	-0.542471	-0.800831	151.0	-0.270763
12	0.1	-1.609157	1.493078	273.0	-0.133016
13	0.1	-1.742897	0.213615	252.0	-0.133016
14	0.1	-0.823052	-0.064359	177.0	-0.133016
15	0.1	-0.100665	-1.814429	183.0	-0.270763
16	0.1	-1.138644	0.060765	184.0	-0.133016
17	0.1	1.321954	-0.032247	113.0	-0.270763
18	0.1	1.335133	-1.909179	146.0	-0.270763
19	0.1	0.994294	0.084405	99.0	-0.270763

Dla punktu startowego $x = -1.2$, $y = -1.2$, krok uczenia = 0.1



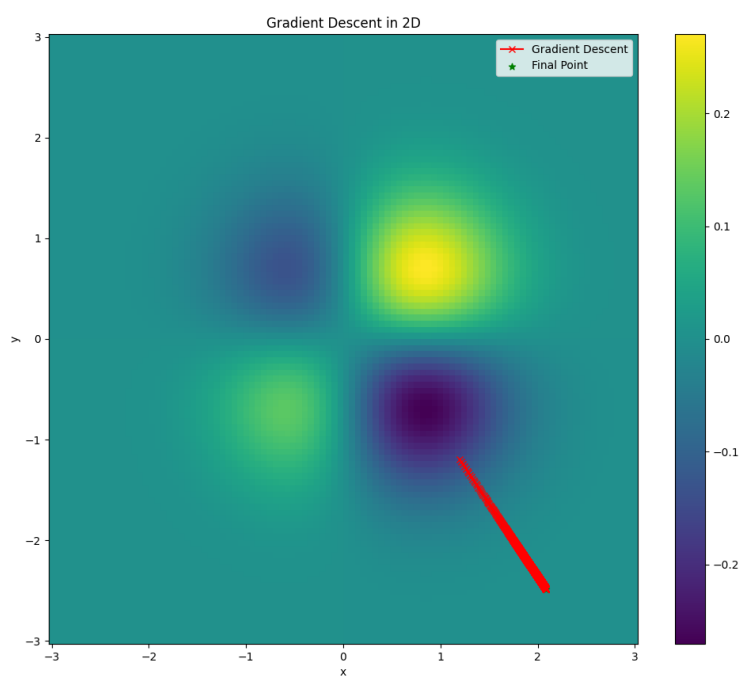
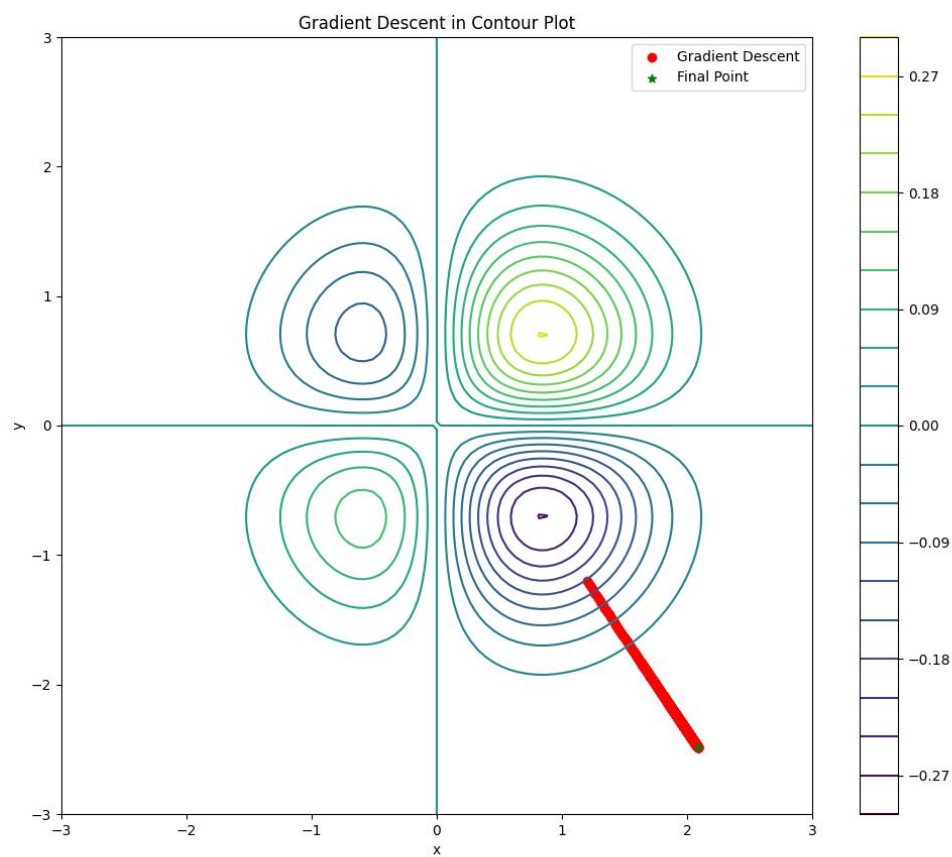
Szukanie maximum funkcji za pomocą SGD w zależności od punktu startowego

Algorytm spadku wzdłuż gradientu został uruchomiony 20 razy w celu znalezienia maximum funkcji, punkt startowy x i punkt startowy y, za każdym razem były wybierane losowo z przedziału $[-2,2]$

Tabela przedstawia rezultat eksperymentu

	Learning Rate	Initial X	Initial Y	Iterations	Maximum
0	0.1	1.024225	-1.001710	1001.0	-0.000394
1	0.1	-0.762723	-1.210794	181.0	0.133016
2	0.1	-0.280136	-1.040307	170.0	0.133016
3	0.1	-1.320266	-1.174747	197.0	0.133016
4	0.1	-1.571060	0.937275	1001.0	-0.000422
5	0.1	1.337109	-0.727291	1001.0	-0.000422
6	0.1	1.218906	-1.384417	1001.0	-0.000387
7	0.1	0.757619	1.185667	94.0	0.270763
8	0.1	-0.636234	0.014182	173.0	0.133016
9	0.1	0.703060	-1.273920	905.0	0.133016
10	0.1	-0.874418	-0.780728	144.0	0.133016
11	0.1	-0.241188	-1.786898	244.0	0.133016
12	0.1	-0.102690	-1.069366	175.0	0.133016
13	0.1	-0.831516	1.345213	1001.0	-0.000436
14	0.1	-1.465600	0.256568	282.0	0.133016
15	0.1	-1.009114	0.521751	271.0	0.133016
16	0.1	-1.153855	-1.285370	197.0	0.133016
17	0.1	0.911322	1.312824	98.0	0.270763
18	0.1	1.315761	1.065801	108.0	0.270763
19	0.1	-0.566555	0.259225	181.0	0.133016

Dla punktu startowego $x = -1.2$, $y = -1.2$, krok uczenia = 0.1



Wybór różnych punktów startowych może prowadzić do różnych lokalnych minimów/maksimów

Punkt startowy blisko minimum może przyspieszyć zbieżność, ale nie zawsze jest łatwy do znalezienia.

Wybór punktu startowego może wpływać na liczbę iteracji wymaganych do zbieżności algorytmu.

Dobranie nieodpowiedniego punktu startowego może doprowadzić, że algorytm

6. Jak wartość kroku uczącego wpływa na proces optymalizacji

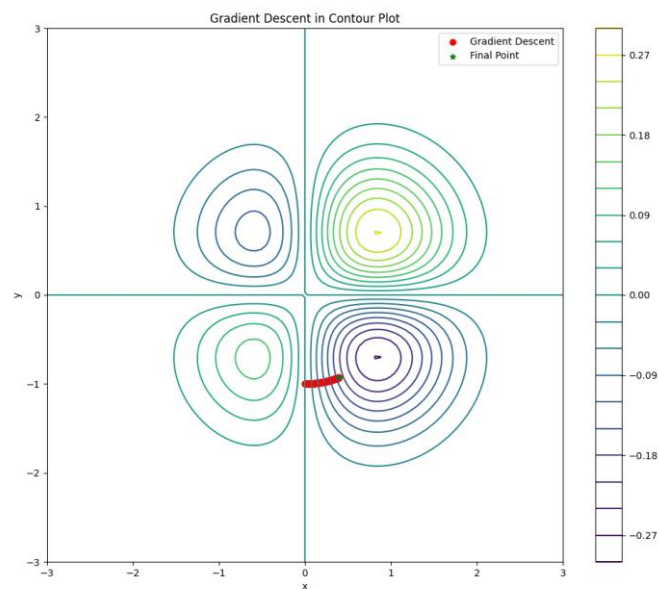
Szukanie minimum funkcji za pomocą SGD w zależności od kroku uczącego

Algorytm spadku wzdłuż gradientu został uruchomiony 7 razy, z różnym krokiem uczącym w celu znalezienia minimum funkcji, dla punktu startowego $x = 0$, $y = -1$ oraz $x = 0$, $y = 1$

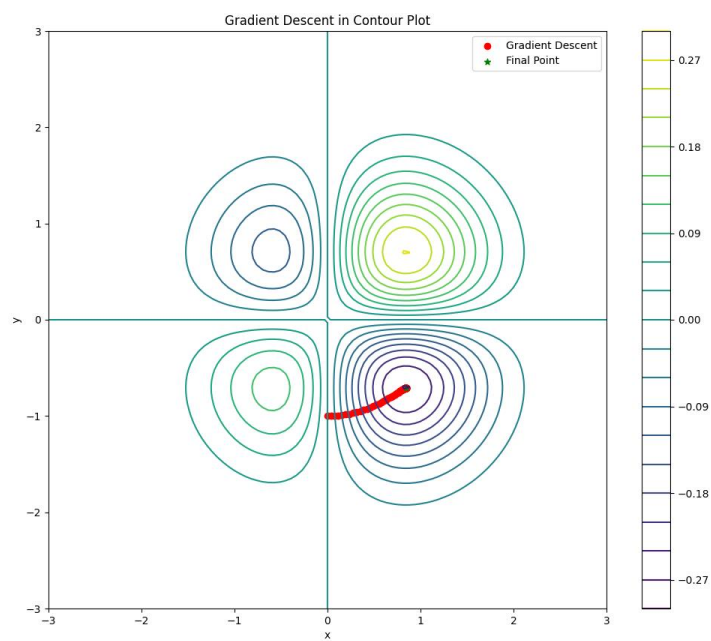
	Learning Rate	Initial X	Initial Y	Iterations	Minimum
0	0.0001	0.0	1.0	1001.0	-1.302985e-02
1	0.0010	0.0	1.0	1001.0	-8.848966e-02
2	0.0100	0.0	1.0	1001.0	-1.330141e-01
3	0.1000	0.0	1.0	172.0	-1.330163e-01
4	1.0000	0.0	1.0	18.0	-1.330163e-01
5	10.0000	0.0	1.0	4.0	-2.852196e-07
6	100.0000	0.0	1.0	3.0	-0.000000e+00

	Learning Rate	Initial X	Initial Y	Iterations	Minimum
0	0.0001	0.0	-1.0	1001.0	-0.014022
1	0.0010	0.0	-1.0	1001.0	-0.160115
2	0.0100	0.0	-1.0	1001.0	-0.270763
3	0.1000	0.0	-1.0	111.0	-0.270763
4	1.0000	0.0	-1.0	9.0	-0.270763
5	10.0000	0.0	-1.0	1001.0	0.000001
6	100.0000	0.0	-1.0	3.0	-0.000000

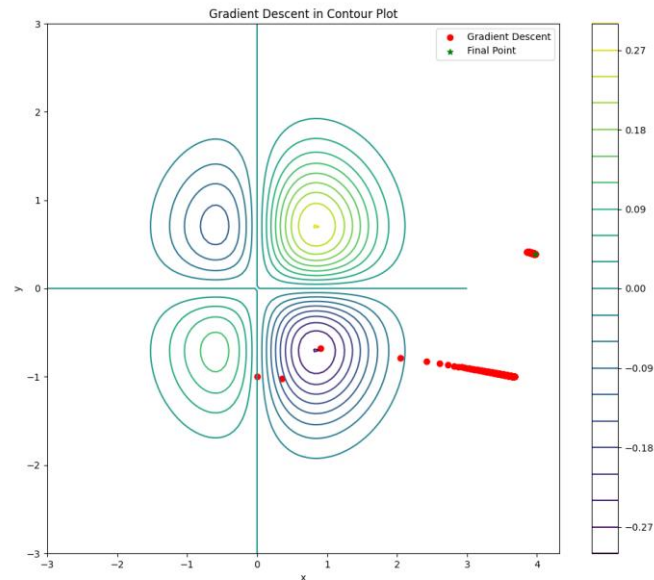
Dla punktu startowego $x = 0$, $y = -1$, krok uczenia = 0.001



Dla punktu startowego $x = 0$, $y = -1$, krok uczenia = 0.1



Dla punktu startowego $x = 0$, $y = -1$, krok uczenia = 10



Szukanie maximum funkcji za pomocą SGD w zależności od kroku uczącego

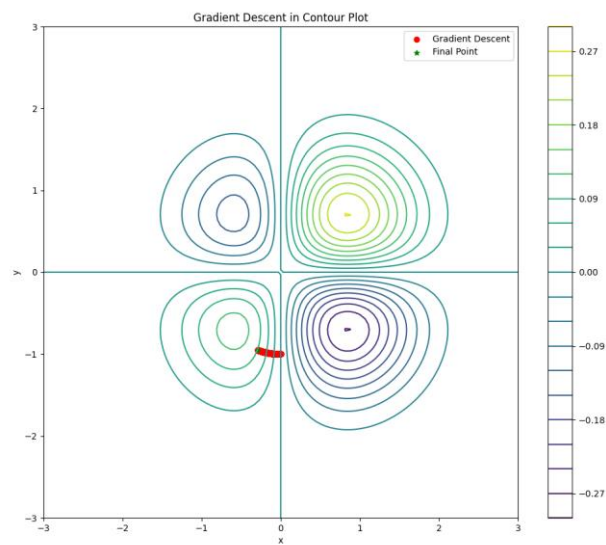
Algorytm spadku wzdłuż gradientu został uruchomiony 7 razy, z różnym krokiem uczącym w celu znalezienia maximum funkcji, dla punktu startowego $x = 0$, $y = -1$ oraz $x = 0$, $y = 1$

Tabela przedstawia rezultat eksperymentu

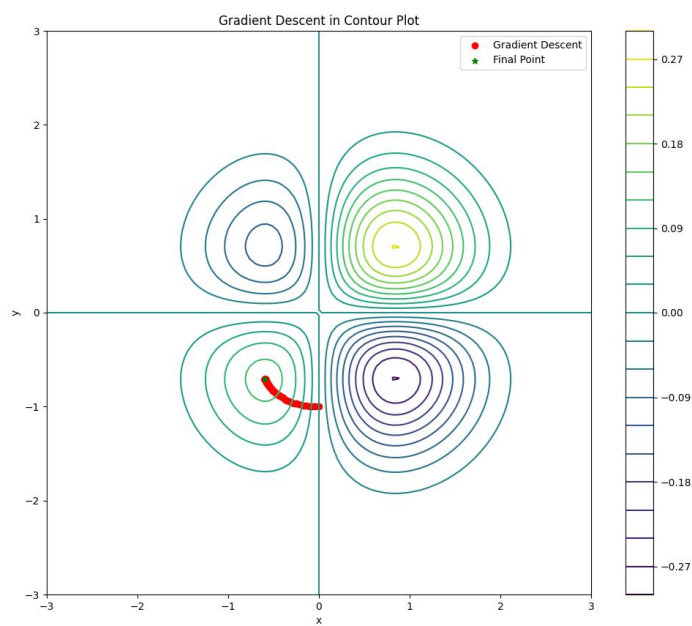
	Learning Rate	Initial X	Initial Y	Iterations	Maximum
0	0.0001	0.0	-1.0	1001.0	1.302985e-02
1	0.0010	0.0	-1.0	1001.0	8.848966e-02
2	0.0100	0.0	-1.0	1001.0	1.330141e-01
3	0.1000	0.0	-1.0	172.0	1.330163e-01
4	1.0000	0.0	-1.0	18.0	1.330163e-01
5	10.0000	0.0	-1.0	4.0	2.852196e-07
6	100.0000	0.0	-1.0	3.0	0.000000e+00

	Learning Rate	Initial X	Initial Y	Iterations	Maximum
0	0.0001	0.0	1.0	1001.0	0.014022
1	0.0010	0.0	1.0	1001.0	0.160115
2	0.0100	0.0	1.0	1001.0	0.270763
3	0.1000	0.0	1.0	111.0	0.270763
4	1.0000	0.0	1.0	9.0	0.270763
5	10.0000	0.0	1.0	1001.0	-0.000001
6	100.0000	0.0	1.0	3.0	0.000000

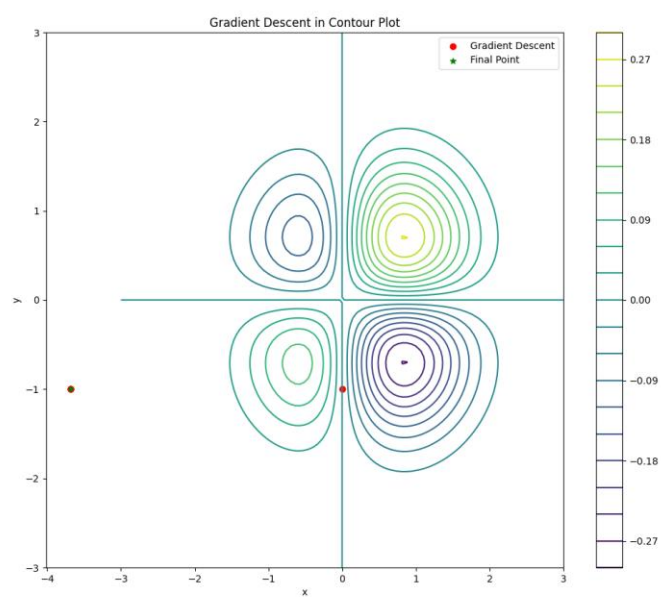
Dla punktu startowego $x = 0$, $y = -1$, krok uczenia = 0.001



Dla punktu startowego $x = 0$, $y = -1$, krok uczenia = 0.1



Dla punktu startowego $x = 0$, $y = -1$, krok uczenia = 10



Wartość kroku uczącego wpływa na szybkość zbieżności algorytmu. Zbyt mała wartość kroku może spowodować wolną zbieżność, podczas gdy zbyt duża może prowadzić do oscylacji lub braku zbieżności.

Zbyt duży krok uczący może sprawić, że algorytm "przeskakuje" minimum, podczas gdy zbyt mały krok może spowolnić zbliżanie się do minimum lub jego nieosiągnięcie

7. Jak można zwiększyć precyzję znalezionych ekstremów

Poprzez dobranie odpowiedniego punktu startowego, im bliżej punkt startowy znajduje się szukanego minimum/maksimum tym większa szansa, że znajdziemy szukane rozwiązanie.

Poprzez dobranie odpowiedniego kroku uczącego, bardzo mały krok uczący wymaga większej liczby iteracji, ale może prowadzić do precyzyjniejszych wyników. Dobór wartości kroku to kompromis pomiędzy szybkością zbieżności a precyzją

Możemy jednak połączyć dobranie odpowiedniego punktu startowego z bardzo małym krokiem. Punkt startowy położony bliżej szukanego rozwiązania zmniejszy nam ilość iteracji, natomiast mniejszy krok uczenia pozwoli nam na otrzymanie dokładniejszego wyniku.