Politechnika Warszawska

WYDZIAŁ ELEKTRONIKI
I TECHNIK INFORMACYJNYCH

Sprawozdanie

Wprowadzenie do sztucznej inteligencji

Ćwiczenie nr. 1

Mikołaj Bańkowski

Numer albumu 310408

prowadzący
Grzegorz Rypeść

Warszawa 2023

Spis treści

1. Treść ćwiczenia 3

2. Pochodne cząstkowe funkcji 4

3. Minima i maksima funkcji 5

4. Czym jest gradient 14

5. Jak punkt startowy wpływa wynik 15

6. Jak wartość kroku uczącego wpływa na proces optymalizacji 19

7. Jak można zwiększyć precyzję znalezionych ekstremów 27

1. Treść ćwiczenia

Proszę znaleźć minima oraz maksima funkcji $f(x,y) = \frac{9 \cdot x \cdot y}{e^{(x^2+0.5x+y^2)}}$ 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 podeprzeć 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

$$f(x,y) = \frac{9xy}{e^{(x^2+0.5x+y^2)}}$$

$$\frac{\partial f}{\partial x} = \frac{9y e^{(x^2+0.5x+y^2)} - 9xy(2x+0.5)\, e^{(x^2+0.5x+y^2)}}{e^{(x^2+0.5x+y^2)^2}}$$

$$= \frac{9y(1-2x^2-0.5x)}{e^{(x^2+0.5x+y^2)}} = \frac{9y(-2x^2-0.5x+1)}{e^{(x^2+0.5x+y^2)}} \cdot \frac{e^{(-x^2-0.5x-y^2)}}{e^{(-x^2-0.5x-y^2)}}$$

$$= (-2x^2-0.5x+1) \cdot e^{(-x^2-0.5x-y^2)} \cdot 9y$$

$$\frac{\partial f}{\partial y} = \frac{9x e^{(x^2+0.5x+y^2)} - 9xy \cdot 2y \cdot e^{(x^2+0.5x+y^2)}}{e^{(x^2+0.5x+y^2)^2}}$$

$$= \frac{9x - 2xy^2 \cdot 9}{e^{(x^2+0.5x+y^2)}} = \frac{(x-2xy^2)\cdot 9}{e^{(x^2+0.5x+y^2)}} \cdot \frac{e^{(-x^2-0.5x-y^2)}}{e^{(-x^2-0.5x-y^2)}}$$

$$= (x-2xy^2)\, e^{(-x^2-0.5x-y^2)} \cdot 9$$

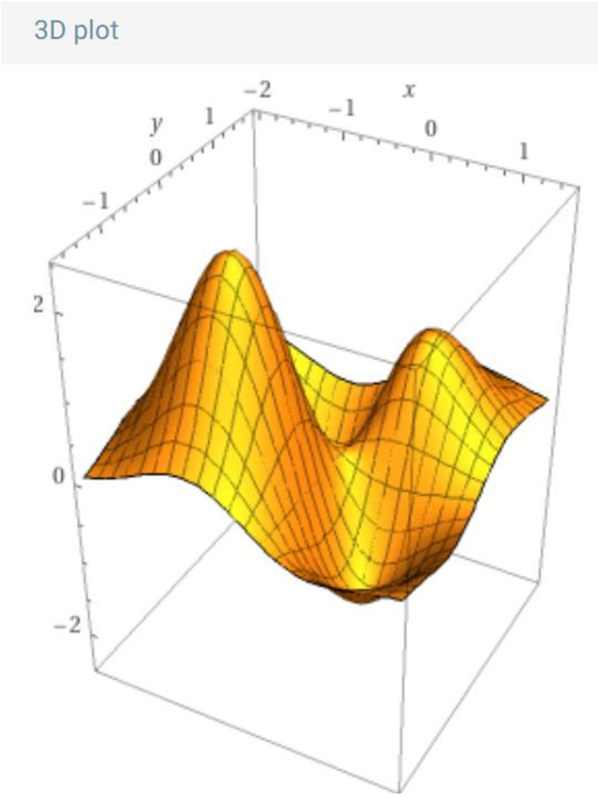
$$\frac{\partial f}{\partial x} = (-2x^2-0.5x+1)\, y\, e^{(-x^2-0.5x-y^2)} \cdot 9$$

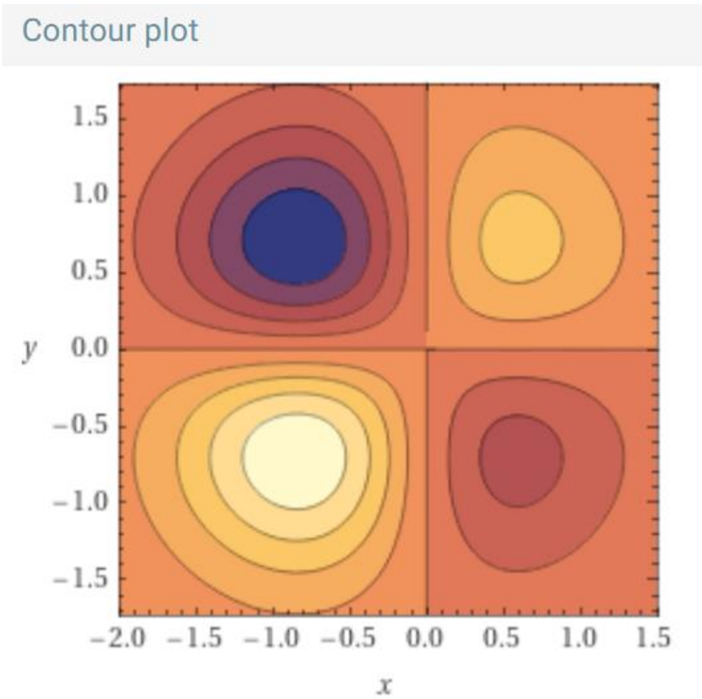
$$\frac{\partial f}{\partial y} = (x-2xy^2)\, e^{(-x^2-0.5x-y^2)} \cdot 9$$

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

3. Minima i maksima funkcji

Do wstępnej analizy badanej funkcji i jej wizualizacji, wykorzystano narzędzie [WolframAplha]





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)} ≈ -1.19715 dla (x, y) ≈ (-0.84307, 0.707107)

min{f(x,y)} ≈ -2.43687 dla (x, y) ≈ (0.59307, -0.707107)

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

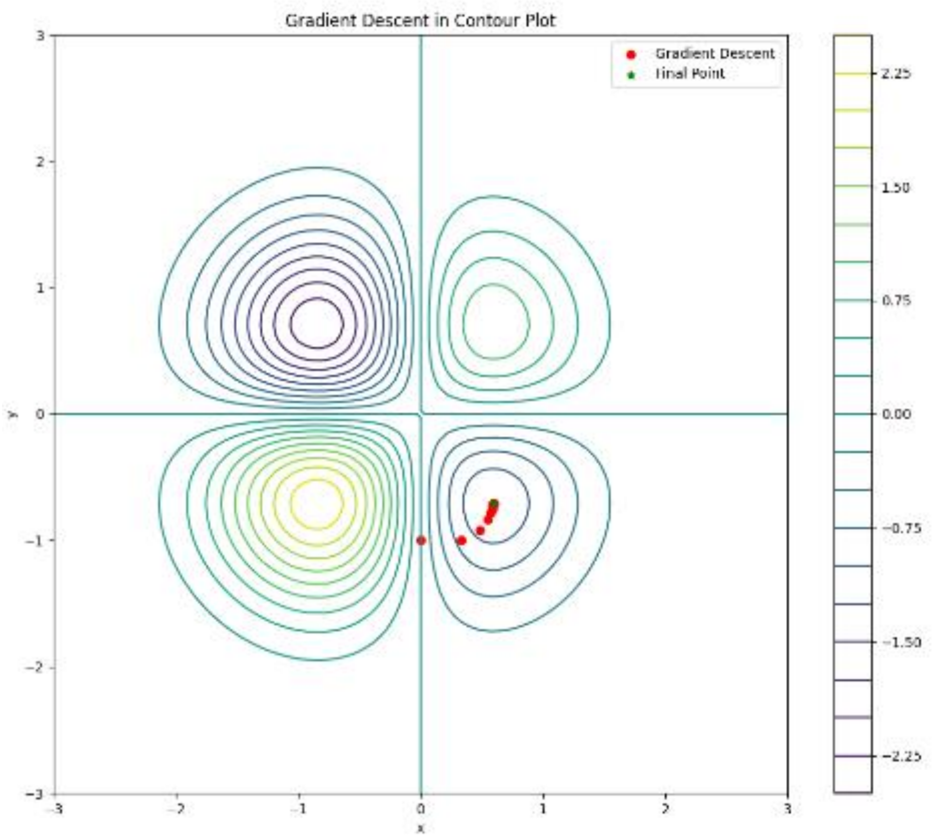
max{f(x,y)} ≈ 1.19715 dla (x, y) ≈ (0.59307, 0.707107)

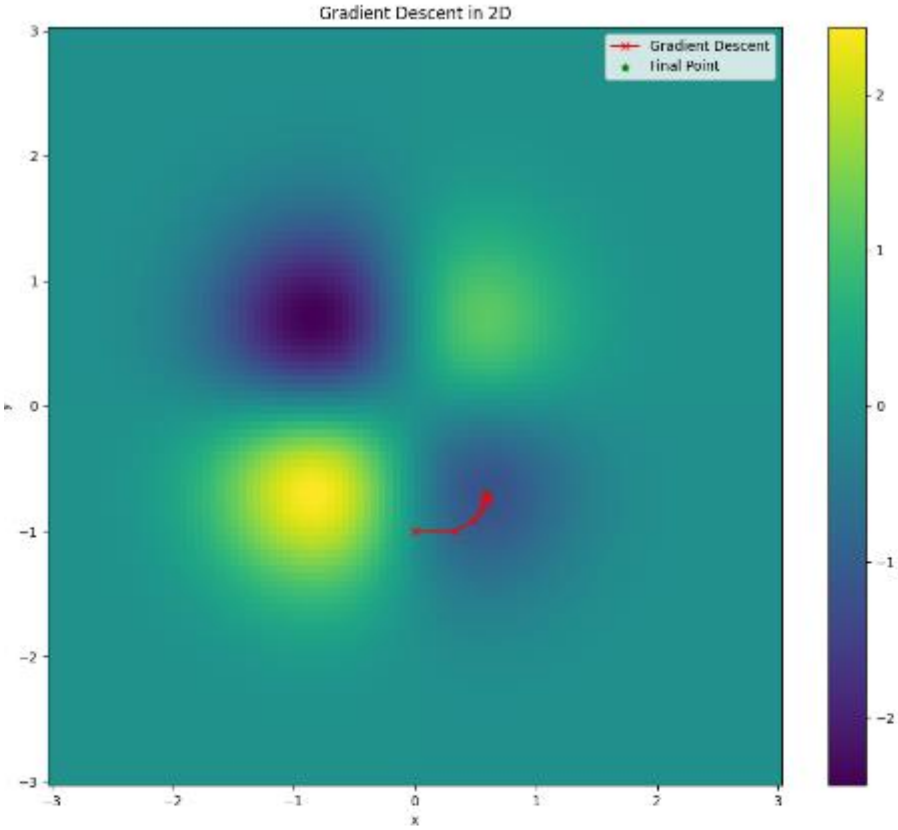
max{f(x,y)} ≈ 2.43687 dla (x, y) ≈ (-0.84307, -0.707107)

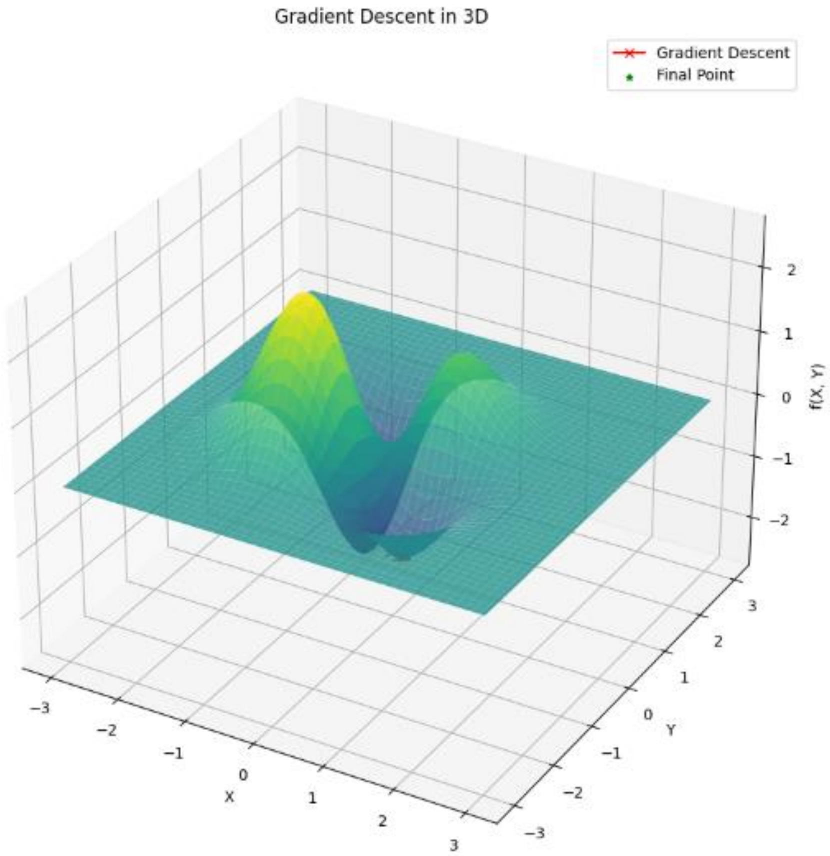
Minima funkcji znalezione za pomocą algorytm spadku wzdłuż gradientu SGD

Dla danych wejściowych:

Krok uczący	Punkt startowy	Punkt końcowy	Znalezione minimum
0.1	x = 0 y = -1	x = 0.593070323659 y = -0.707108068995	-1.197146871898

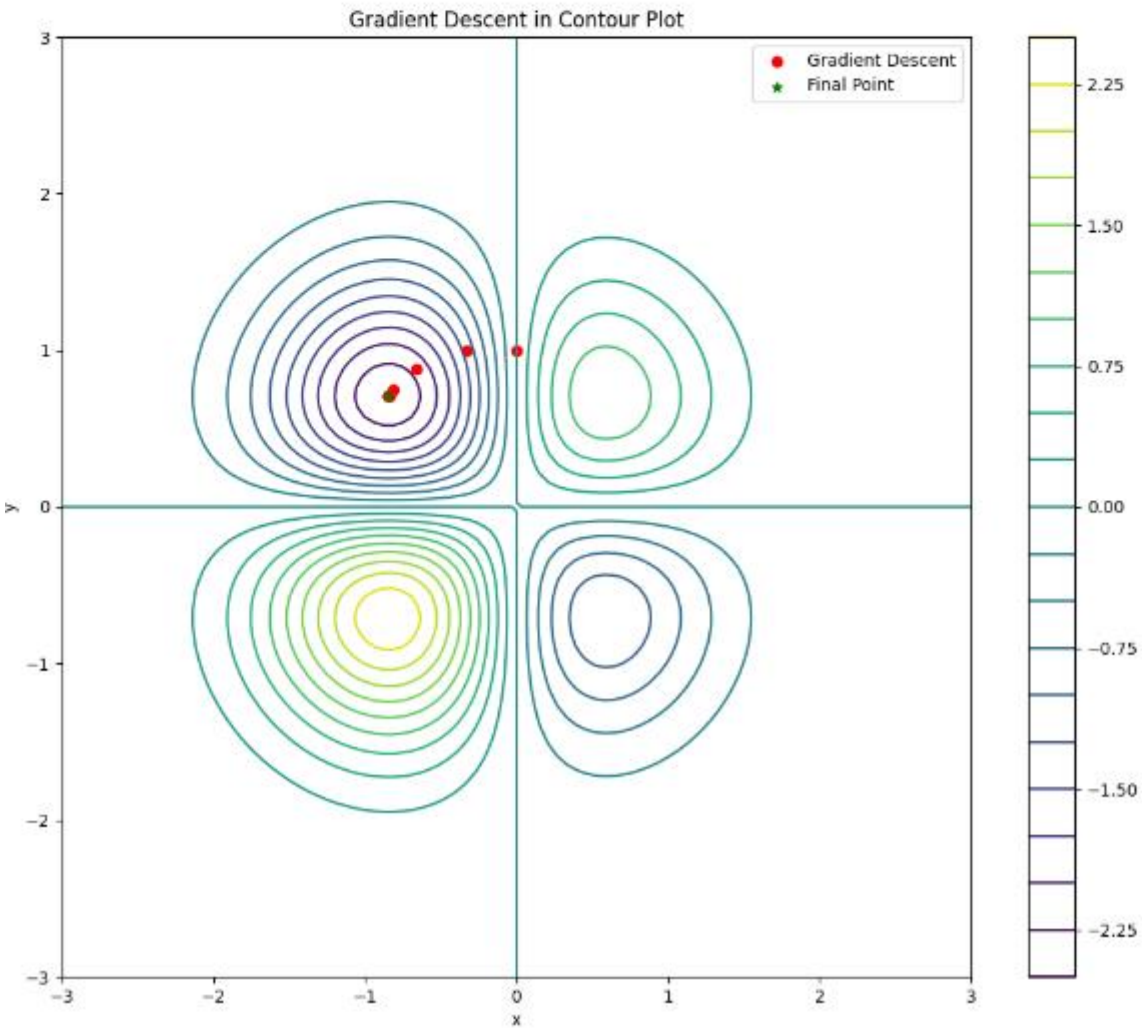


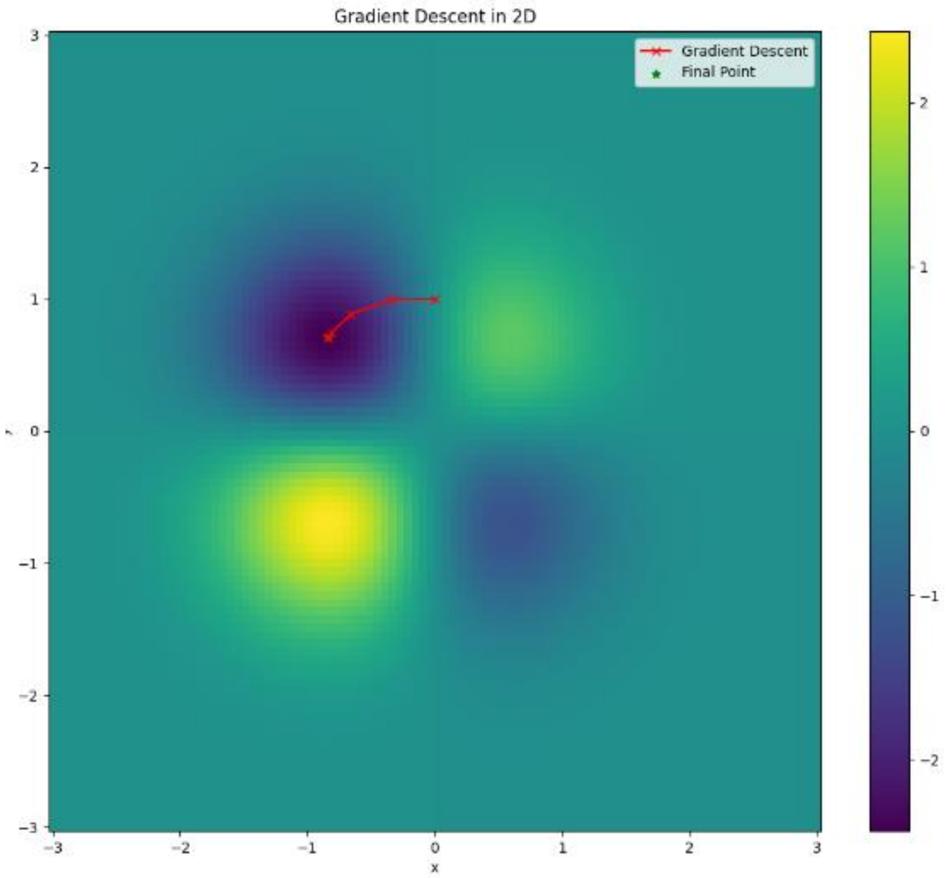

Gradient Descent in 2D
Gradient Descent
Final Point
x
y

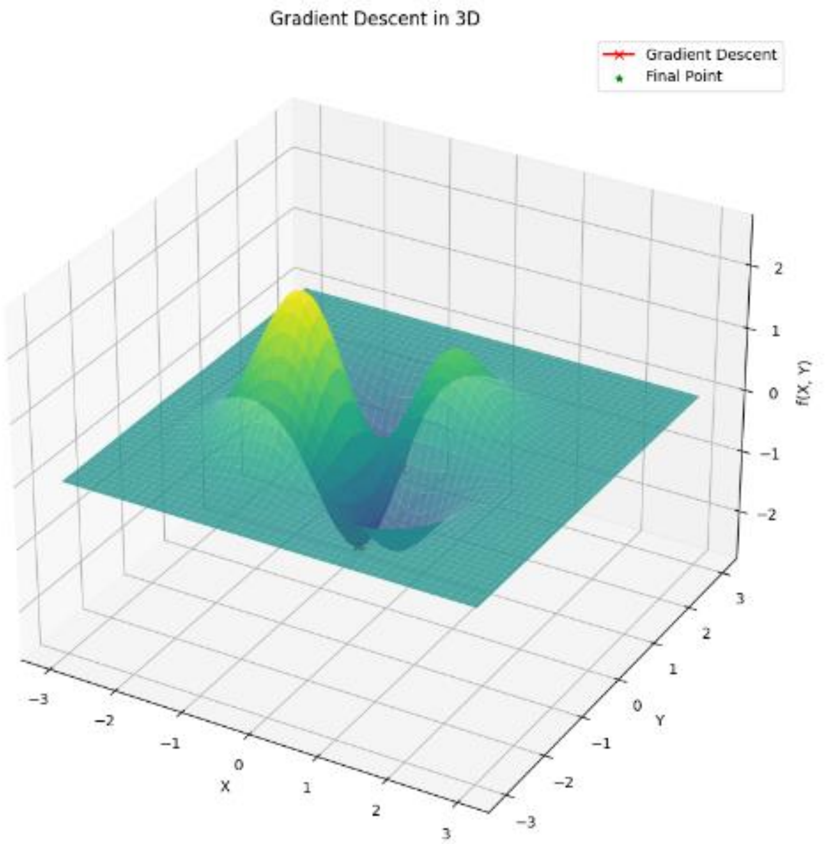

Gradient Descent in 3D
Gradient Descent
Final Point
X
Y
f(X, Y)

Dla danych wejściowych

Krok uczący	Punkt startowy	Punkt końcowy	Znalezione minimum
0.1	x = 0 y = 1	x = -0.843070208719 y = 0.707106781187	-2.436868161523



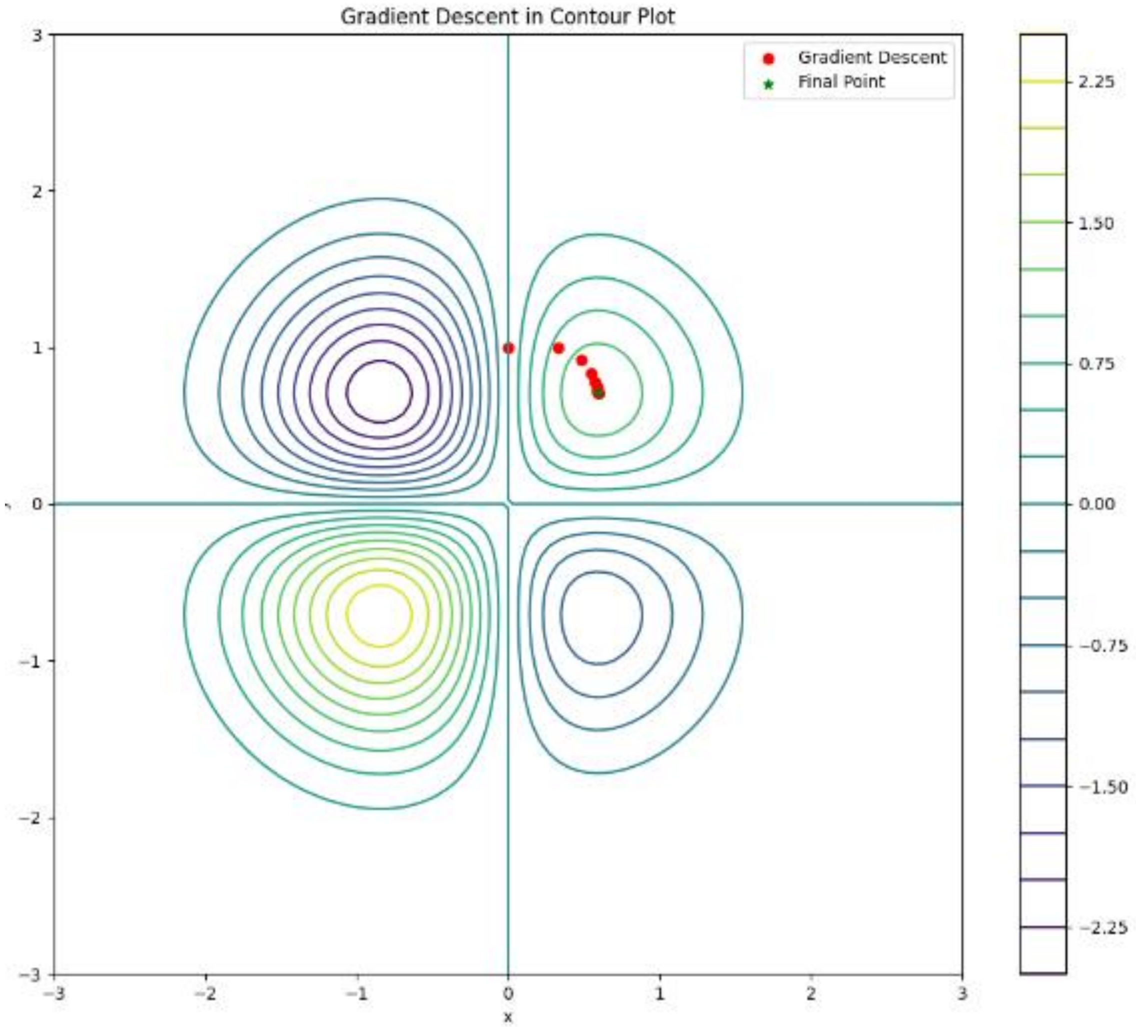

Gradient Descent in 2D
Gradient Descent
Final Point
x
y

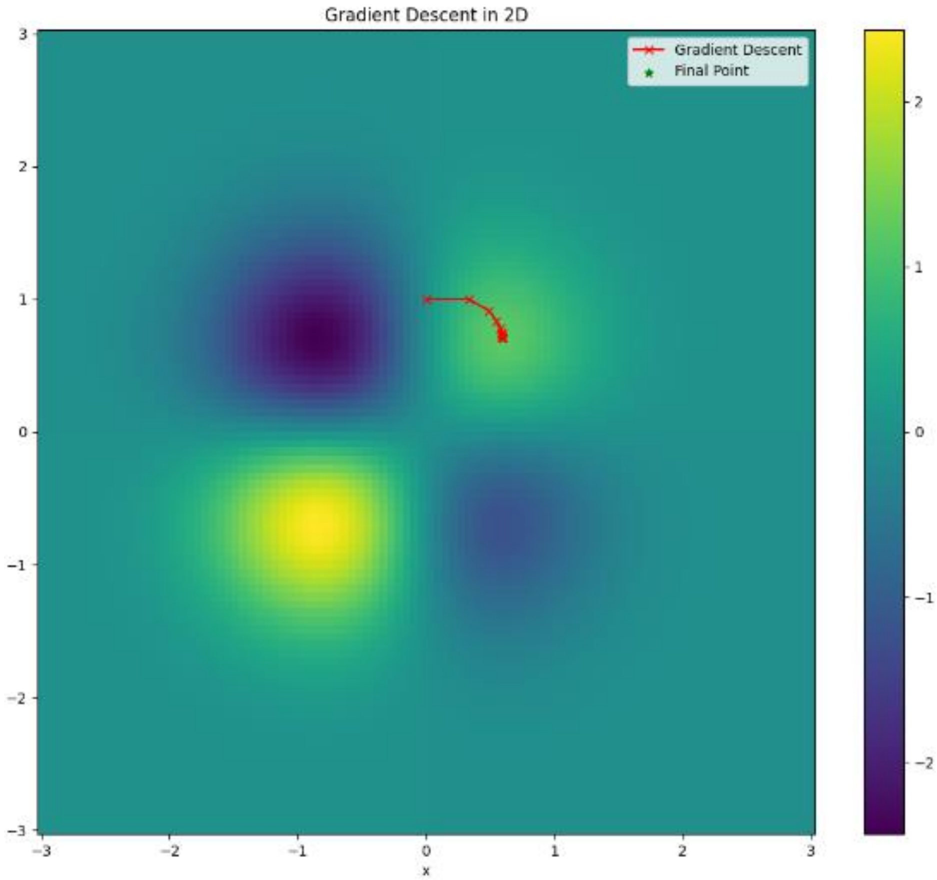

Gradient Descent in 3D
Gradient Descent
Final Point
X
Y
f(X, Y)

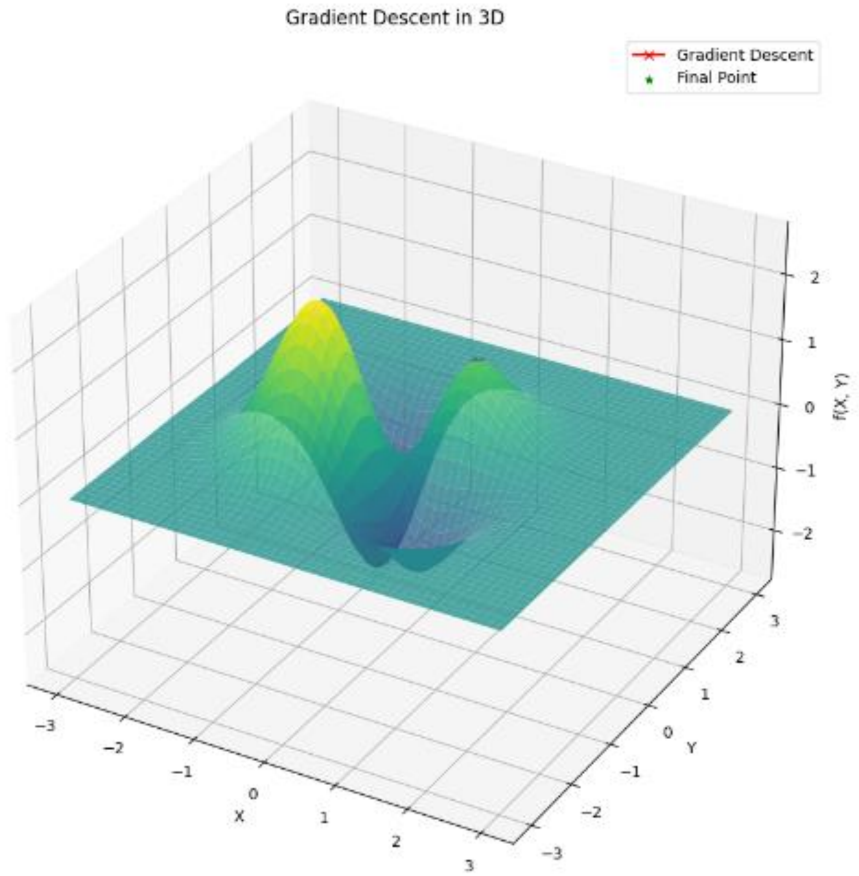
Maxima funkcji znalezione za pomocą algorytm spadku wzdłuż gradientu SGD

Dla danych wejściowych

Krok uczący	Punkt startowy	Punkt końcowy	Znalezione maximun
0.1	x = 0 y = 1	x = 0.593070323659 y = 0.707108068995	1.197146871898

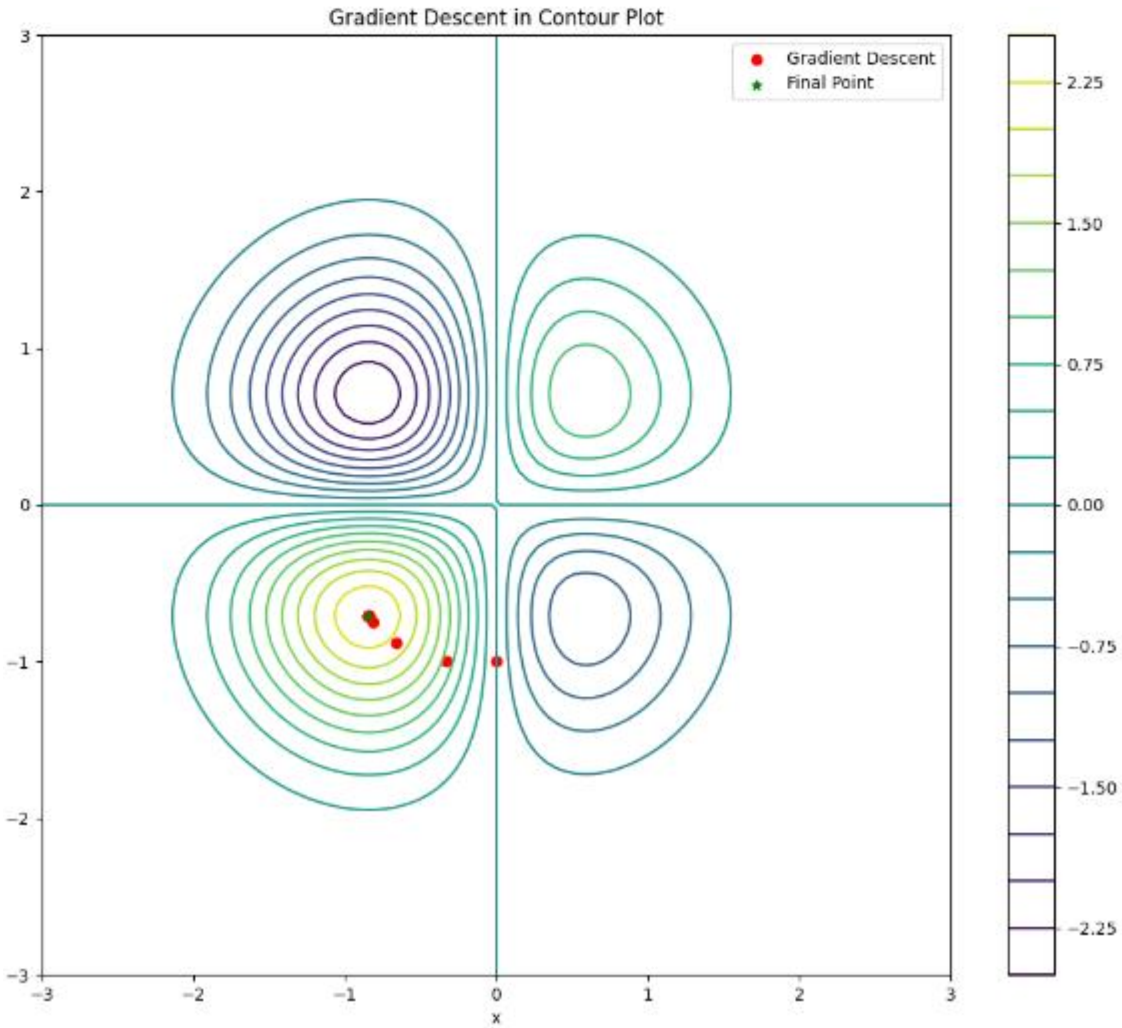


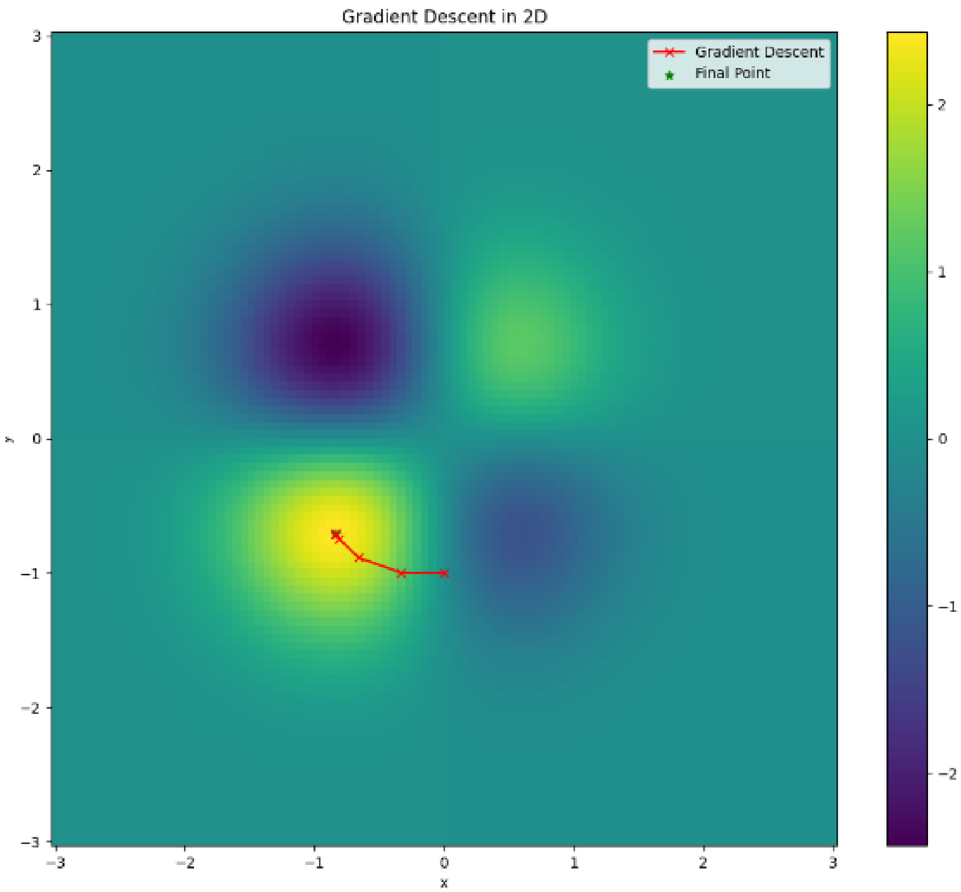

Gradient Descent in 2D
Gradient Descent
Final Point
x

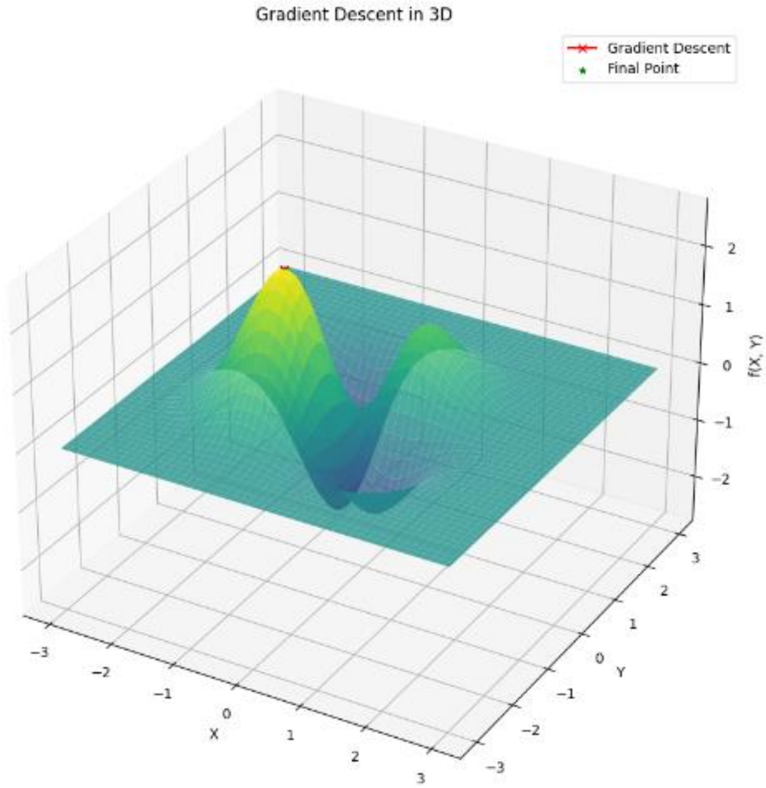

Gradient Descent in 3D
Gradient Descent
Final Point
X
Y
f(X, Y)

Dla danych wejściowych

Krok uczący	Punkt startowy	Punkt końcowy	Znalezione maximum
0.1	x = 0 y = -1	x = -0.843070208719 y = -0.707106781187	2.436868161523




Gradient Descent in 2D
Gradient Descent
Final Point
x
y


Gradient Descent in 3D
Gradient Descent
Final Point
X
Y
f(X, Y)

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 wchodzą 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

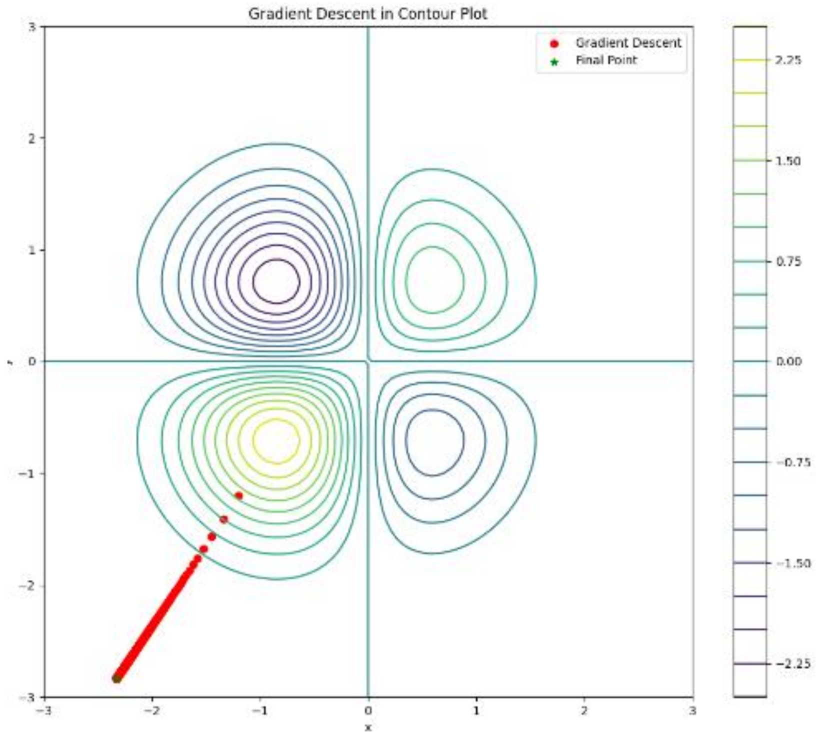
Szukanie minimum funkcji za pomocą SGD w zależności od punktu startowego

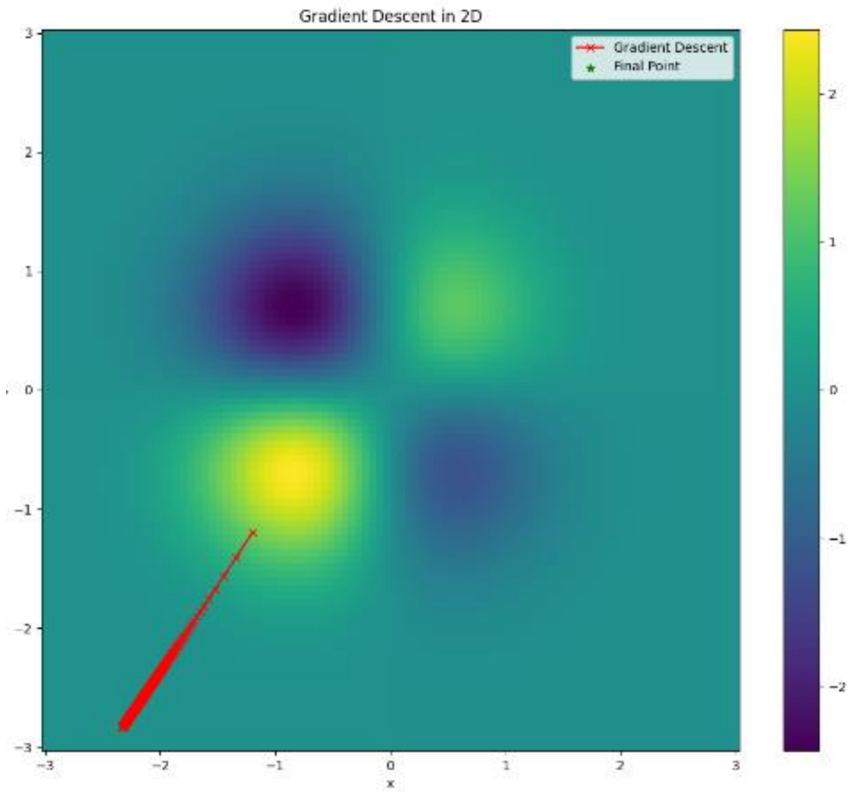
Algorytm spadku wzdłuż gradientu został uruchomiony 20 razy w celu znalezienia minimum funkcji, punkt startowy x i punkt startowy y, za każdym razem były wybierane losowo z przedziału [-2,2]S

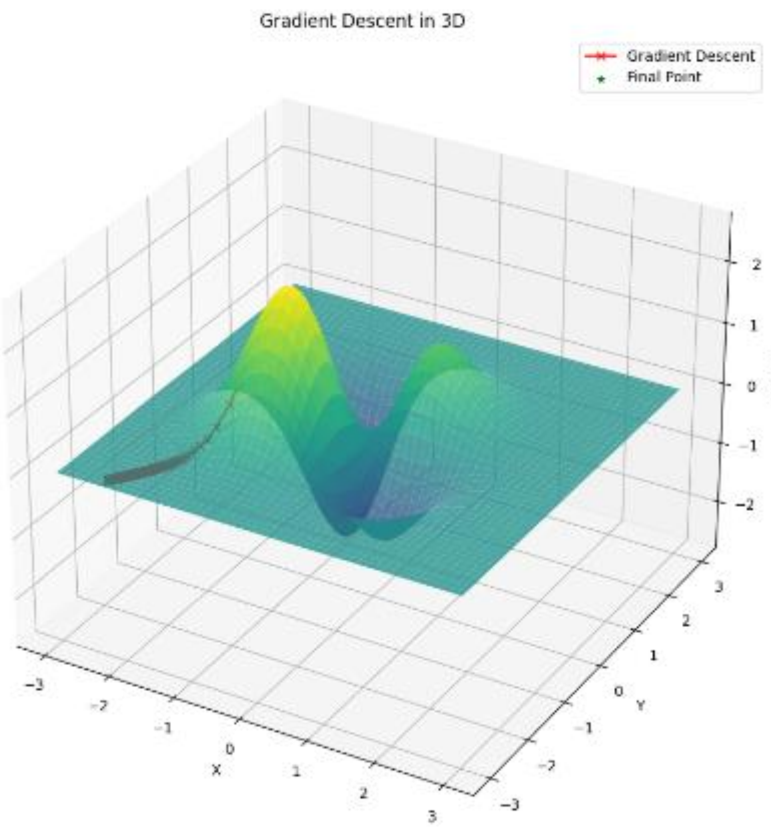
Tabela przedstawia rezultat eksperymentu

	Learning Rate	Initial X	Initial Y	Iterations	Minimum
0	0.1	0.809649	1.133588	1001.0	0.000285
1	0.1	1.828542	0.707124	1001.0	0.000297
2	0.1	-0.661918	0.422566	10.0	-2.436868
3	0.1	-1.355129	0.734948	11.0	-2.436868
4	0.1	1.999433	1.464029	1001.0	0.000276
5	0.1	0.851391	-0.111590	21.0	-1.197147
6	0.1	1.674150	1.554615	1001.0	0.000279
7	0.1	-0.288446	0.136323	11.0	-2.436868
8	0.1	-0.918366	-1.313571	1001.0	0.000276
9	0.1	1.527317	0.586425	1001.0	-0.000299
10	0.1	-0.652445	0.897647	9.0	-2.436868
11	0.1	0.742125	-0.258380	20.0	-1.197147
12	0.1	-0.912279	0.454214	9.0	-2.436868
13	0.1	0.594690	0.066216	21.0	-1.197147
14	0.1	0.208278	-1.175930	23.0	-1.197147
15	0.1	0.415269	-0.187363	20.0	-1.197147
16	0.1	-0.683947	-0.411024	12.0	-2.436868
17	0.1	-0.063826	0.318984	11.0	-2.436868
18	0.1	1.104275	1.567129	1001.0	0.000284
19	0.1	1.687121	0.271001	51.0	-1.197147

Krok uczący	Punkt startowy	Punkt końcowy	Znalezione minimum
0.1	x = 1.2 y = -1.2	x = -2.331571045911 y = -2.835618989202	0.000267833261



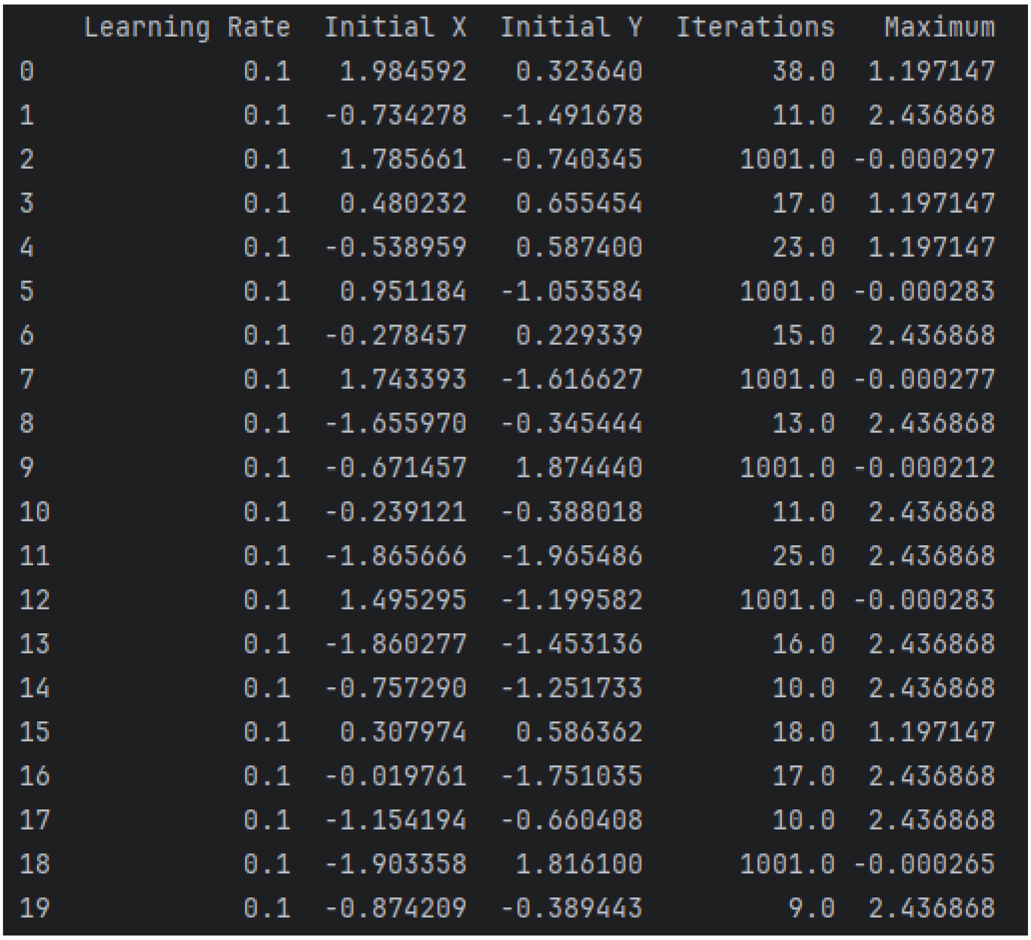



Gradient Descent in 3D
Gradient Descent
Final Point
f(X, Y)
X
Y
−3
−2
−1
0
1
2
3

Szukanie maximum funkcji za pomocą SGD w zależności od punktu startowego

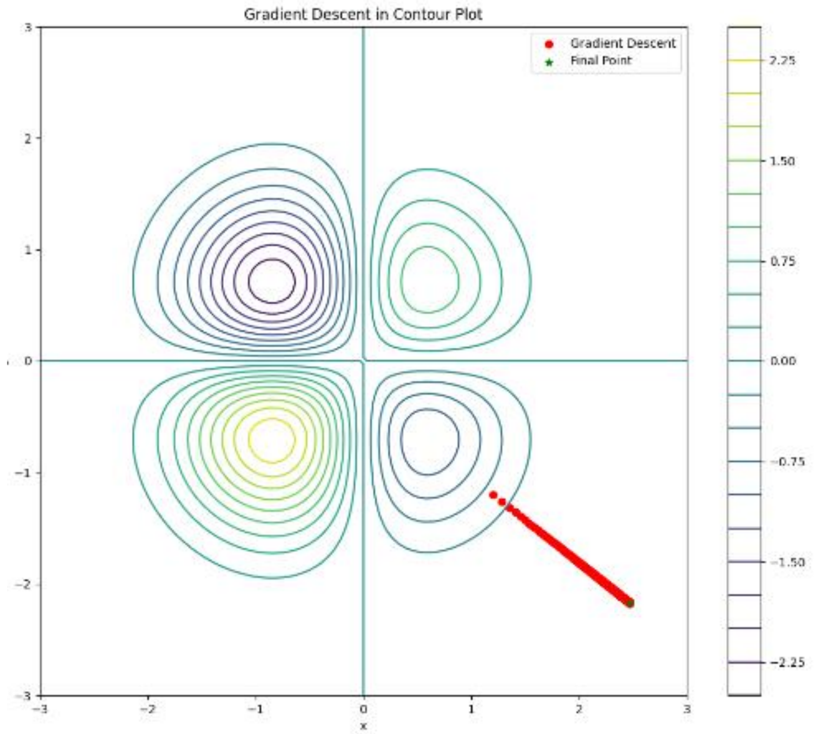
Algorytm spadku wzdłuż gradientu został uruchomiony 20 razy w celu znalezienia maximum funkcji, punkt startowy x i punkt startowy y, za każdym razem były wybierane losowo z przedziału [-2,2]

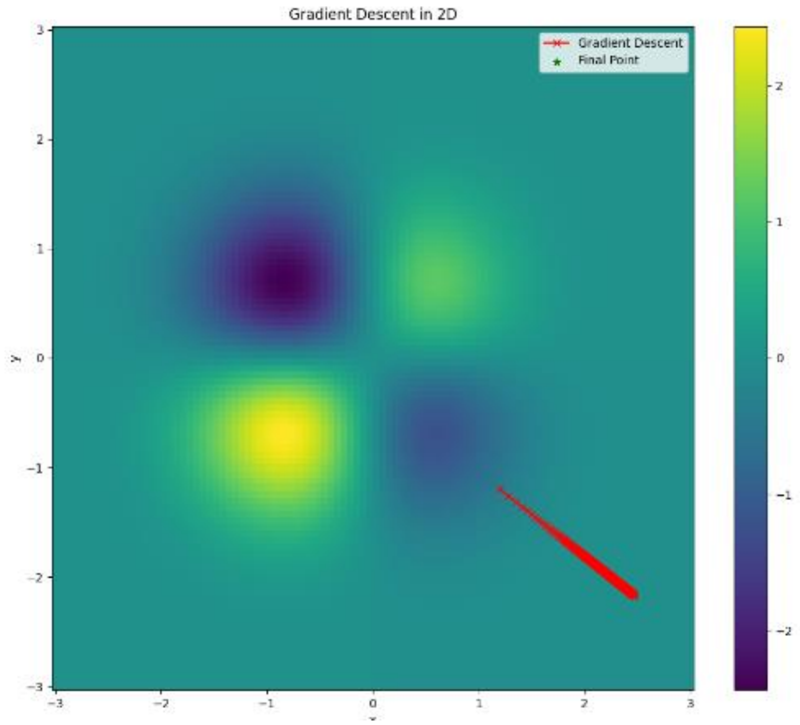
Tabela przedstawia rezultat eksperymentu

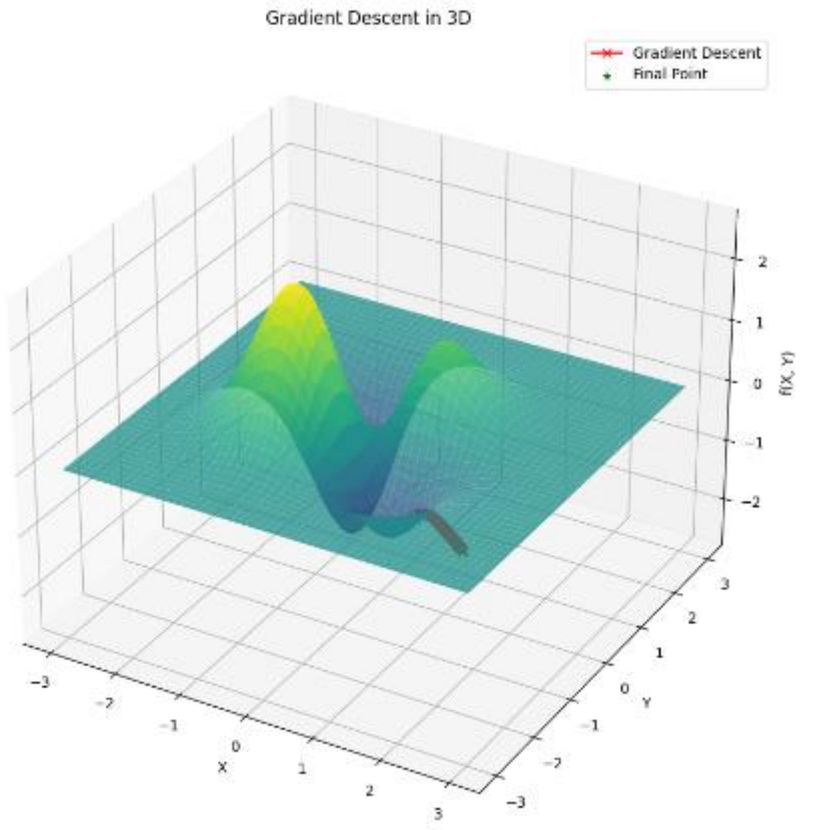


	Learning Rate	Initial X	Initial Y	Iterations	Maximum
0	0.1	1.984592	0.323640	38.0	1.197147
1	0.1	-0.734278	-1.491678	11.0	2.436868
2	0.1	1.785661	-0.740345	1001.0	-0.000297
3	0.1	0.480232	0.655454	17.0	1.197147
4	0.1	-0.538959	0.587400	23.0	1.197147
5	0.1	0.951184	-1.053584	1001.0	-0.000283
6	0.1	-0.278457	0.229339	15.0	2.436868
7	0.1	1.743393	-1.616627	1001.0	-0.000277
8	0.1	-1.655970	-0.345444	13.0	2.436868
9	0.1	-0.671457	1.874440	1001.0	-0.000212
10	0.1	-0.239121	-0.388018	11.0	2.436868
11	0.1	-1.865666	-1.965486	25.0	2.436868
12	0.1	1.495295	-1.199582	1001.0	-0.000283
13	0.1	-1.860277	-1.453136	16.0	2.436868
14	0.1	-0.757290	-1.251733	10.0	2.436868
15	0.1	0.307974	0.586362	18.0	1.197147
16	0.1	-0.019761	-1.751035	17.0	2.436868
17	0.1	-1.154194	-0.660408	10.0	2.436868
18	0.1	-1.903358	1.816100	1001.0	-0.000265
19	0.1	-0.874209	-0.389443	9.0	2.436868

Krok uczący	Punkt startowy	Punkt końcowy	Znalezione maksimum
0.1	x = 1.2 y = -1.2	x = 2.470023196368 y = -2.170725183170	-0.000282576685







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 źle zadziała, zamiast zbliżać się do oczekiwanego rozwiązania będzie się od niego oddalać

6. Jak wartość kroku uczącego wpływa na proces optymalizacji

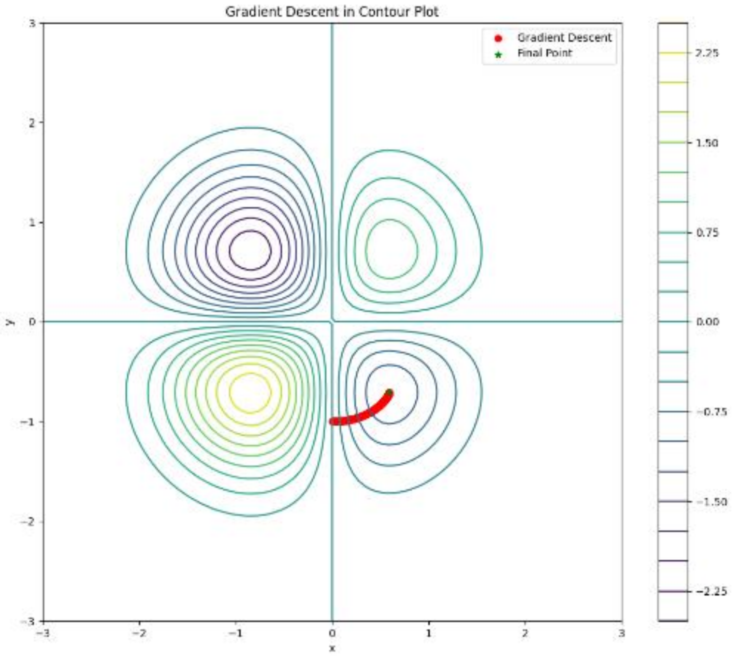
Szukanie minimum funkcji za pomocą SGD w zależności od kroku uczącego

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

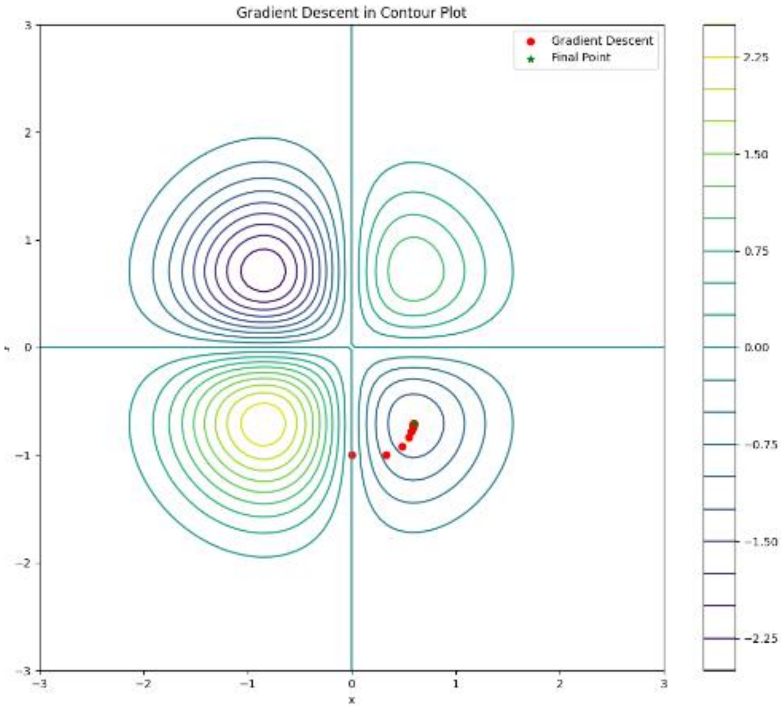
	Learning Rate	Initial X	Initial Y	Iterations	Minimum
0	0.0001	0.0	-1.0	1001.0	-0.748689
1	0.0010	0.0	-1.0	1001.0	-1.197087
2	0.0100	0.0	-1.0	213.0	-1.197147
3	0.1000	0.0	-1.0	22.0	-1.197147
4	1.0000	0.0	-1.0	655.0	-0.000001
5	10.0000	0.0	-1.0	3.0	-0.000000
6	100.0000	0.0	-1.0	3.0	-0.000000

	Learning Rate	Initial X	Initial Y	Iterations	Minimum
0	0.0001	0.0	1.0	1001.0	-1.303784
1	0.0010	0.0	1.0	1001.0	-2.436868
2	0.0100	0.0	1.0	136.0	-2.436868
3	0.1000	0.0	1.0	11.0	-2.436868
4	1.0000	0.0	1.0	1001.0	0.000028
5	10.0000	0.0	1.0	3.0	-0.000000
6	100.0000	0.0	1.0	3.0	-0.000000

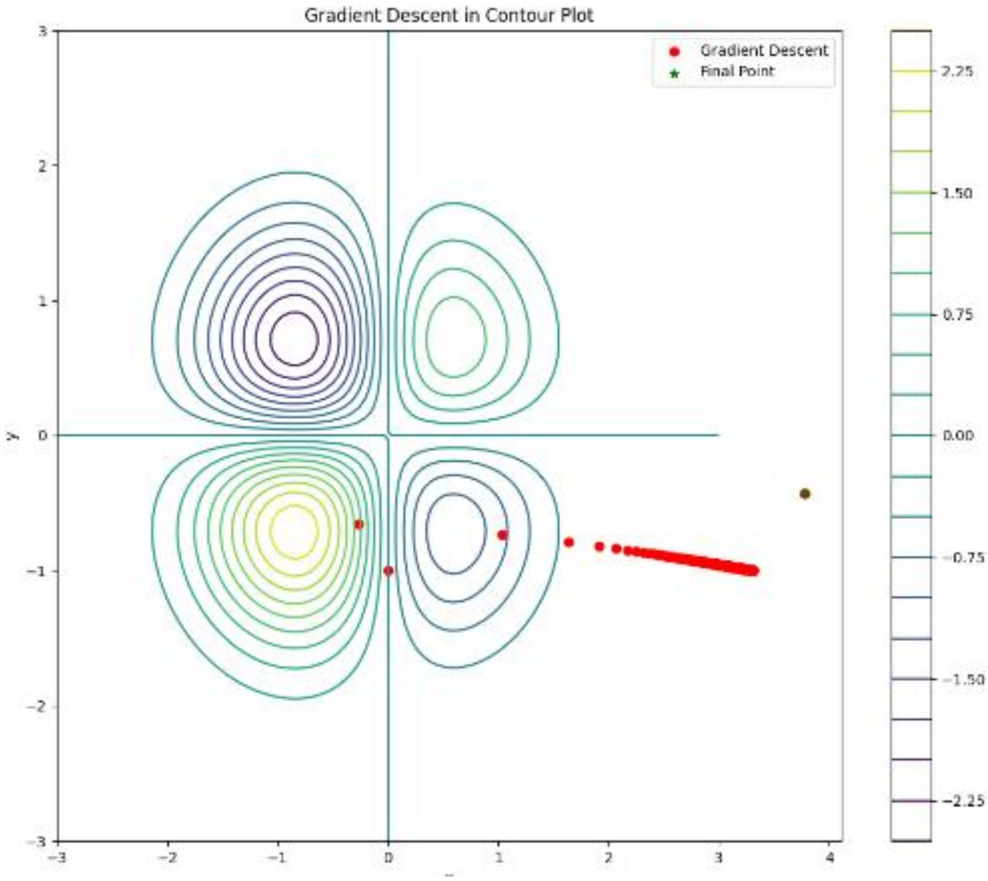
Krok uczący	Punkt startowy	Punkt końcowy	Znalezione minimum
0.001	x = 0 y = -1	x = 0.591573675756 y = -0.711839163416	-1.197086871197



Krok uczący	Punkt startowy	Punkt końcowy	Znalezione minimum
0.1	x = 0 y = -1	x = 0.593070323659 y = -0.707108068995	-1.197146871898



Krok uczący	Punkt startowy	Punkt końcowy	Znalezione minimum
1	x = 0 y = -1	x = 3.772484966940 y = -0.428943649562	-0.000001211866



Szukanie maximum funkcji za pomocą SGD w zależności od kroku uczącego

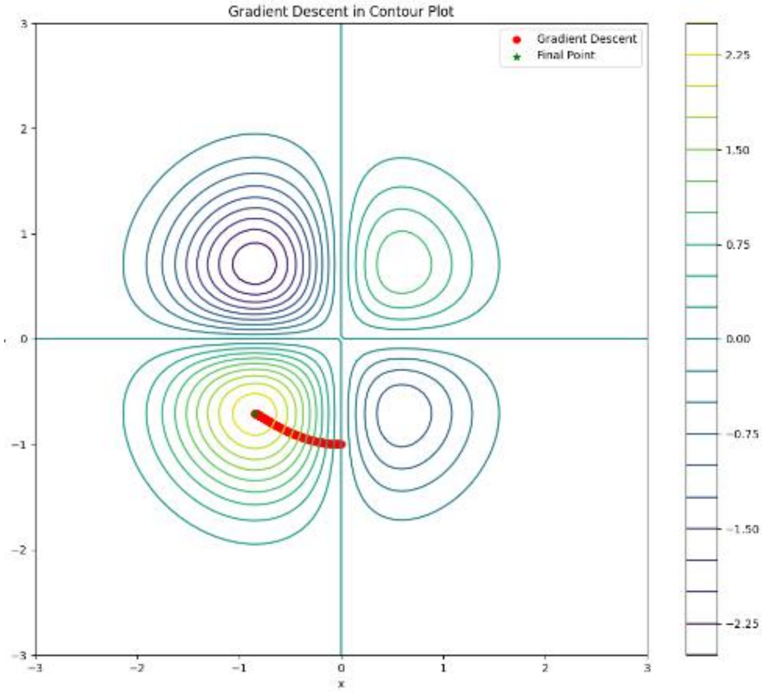
Algorytm spadku wzdłuż gradientu został uruchomiony 7 razy, z różnym krokiem uczącym w celu znalezienia maximum funkcji, dla punktu startowego x = 0, y = -1 oraz x = 0, y = 1

Tabela przedstawia rezultat eksperymentu

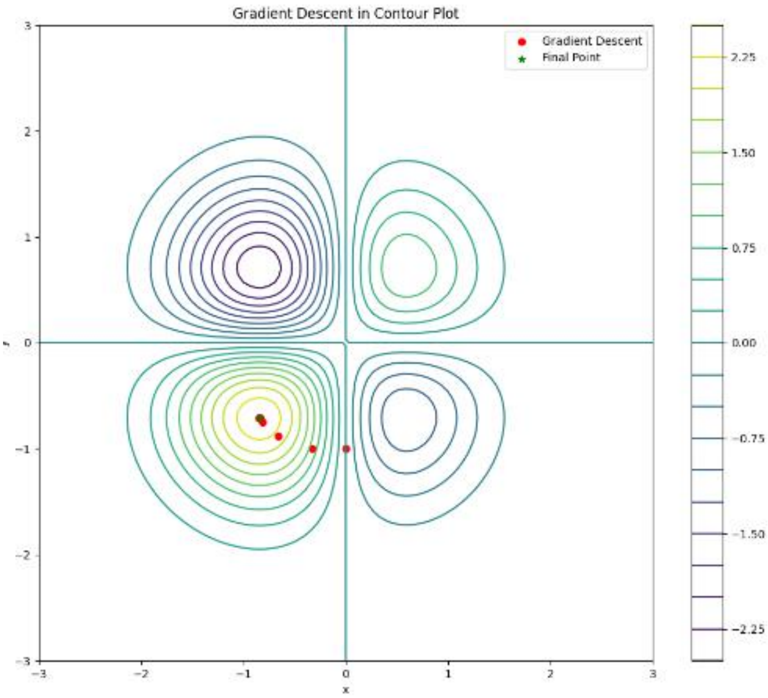
	Learning Rate	Initial X	Initial Y	Iterations	Maximum
0	0.0001	0.0	-1.0	1001.0	1.303784
1	0.0010	0.0	-1.0	1001.0	2.436868
2	0.0100	0.0	-1.0	136.0	2.436868
3	0.1000	0.0	-1.0	11.0	2.436868
4	1.0000	0.0	-1.0	1001.0	-0.000028
5	10.0000	0.0	-1.0	3.0	0.000000
6	100.0000	0.0	-1.0	3.0	0.000000

	Learning Rate	Initial X	Initial Y	Iterations	Maximum
0	0.0001	0.0	1.0	1001.0	0.748689
1	0.0010	0.0	1.0	1001.0	1.197087
2	0.0100	0.0	1.0	213.0	1.197147
3	0.1000	0.0	1.0	22.0	1.197147
4	1.0000	0.0	1.0	655.0	0.000001
5	10.0000	0.0	1.0	3.0	0.000000
6	100.0000	0.0	1.0	3.0	0.000000

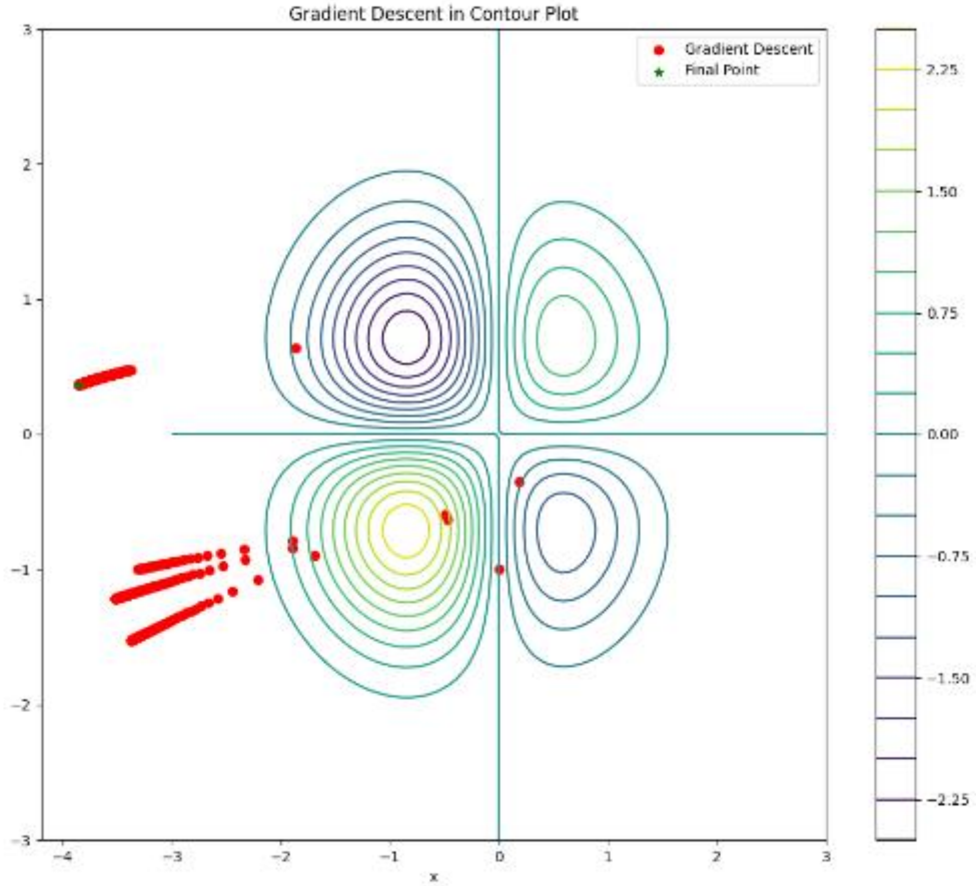
Krok uczący	Punkt startowy	Punkt końcowy	Znalezione maksimum
0.001	x = 1.2 y = -1.2	x = -0.842806526532 y = -0.707163165748	2.436867857116



Krok uczący	Punkt startowy	Punkt końcowy	Znalezione maksimum
0.1	x = 0 y = -1	x = -0.843070208719 y = -0.707106781187	2.436868161523



Krok uczący	Punkt startowy	Punkt końcowy	Znalezione maximum
1	x = 0 y = -1	x = -3.850166167893 y = .369230749634	-0.000027924101



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.