Algorytm gradientu prostego

Marcin Połosak

Listopad 2023

1 Opis algorytmu

Algorytm gradientu prostego służy do wyznaczania minimum funkcji. Wykorzystuje gradient w celu wskazania kierunku malenia x który trzeba obrać w celu dotarcia do minimum.

x_0 - punkt początkowy
α - "learning rate"
d_k - gradient funkcji celu w punkcie x_k

$$x_{k+1} = x_k - \alpha d_k$$

Algorytm powtarza opisane działanie aż do wyznaczonego warunku stopu. Jest nim chociażby przekroczenie limitu iteracji narzuconego przez użytkownika. Drugim warunkiem jest uzyskanie wyniku odpowiadającego użytkownikowi tj. takiego który spełnia równanie:

$$||\nabla f(x_k)|| \leqslant \epsilon$$

Algorytm nie jest zbieżny dla każdej α, a w swoich eksperymentach przetestuje dla jakich wartości jest w stanie wyznaczyć minimum funkcji.

2 Eksperymenty

W swoich eksperymentach za cel obrałem dwie rzeczy. Pierwszą było sprawdzenie jaka jest granica parametru learning rate przy której algorytm traci swoje właściwości i przestaje być zbieżny.

W tym celu dla każdej kombinacji funkcji celu $\alpha = (1, 10, 100)$ starałem się znaleźć granicę zbierzności dobierając różne wartości.

Drugim eksperymentem była próba oszacowania parametru learning rate przy którym algorytm działał najlepiej i potrzebował wykonać najmniej obliczeń. W tym celu na interesujących mnie przedziałach obierałem 100 kolejnych bliskich wartości parametru learning rate.

W celu uzyskania czytelnych wyników badałem za równo liczbę potrzebnych do wykonania kroków jak i czas działania algorytmu. Umożliwiło mi to uzyskanie obrazu pewnej tendencji widocznej na wykresie.

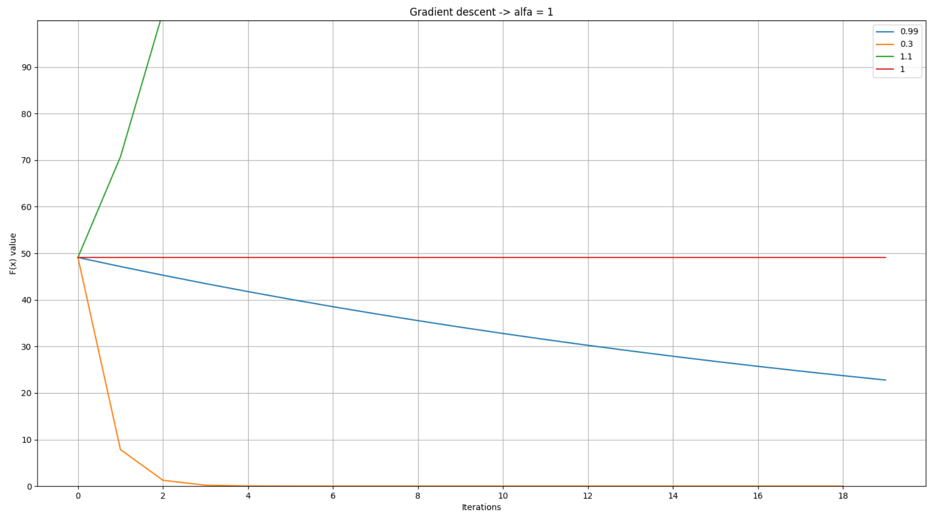
Zadana funkcja:

q(x) = $\sum_{i=1}^{n} \alpha^{\frac{i-1}{n-1}} x_i^2$, n = 10, α = {1, 10, 100}

3 Rezultaty

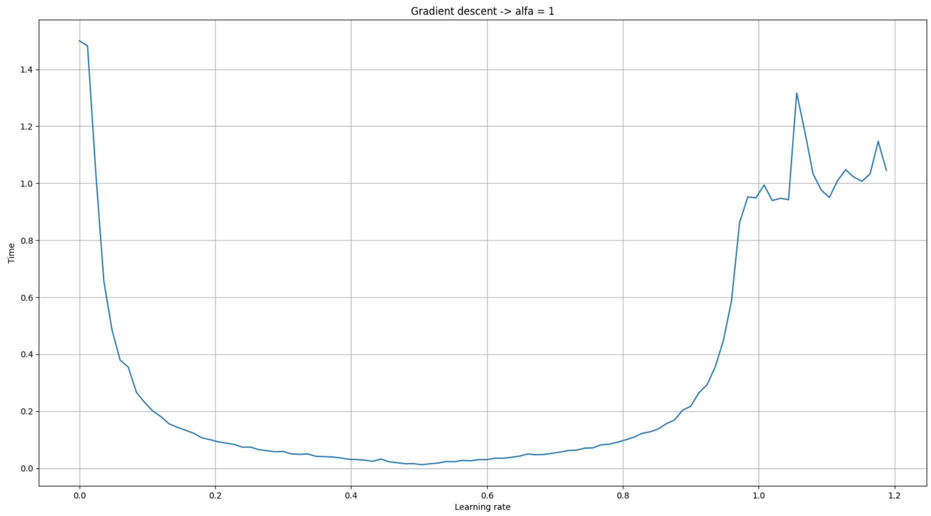
3.1 $\alpha = 1$

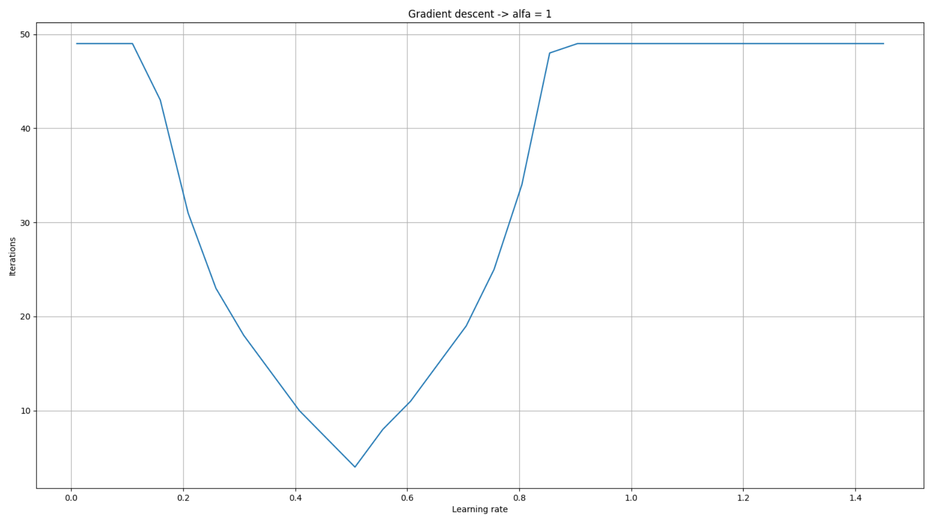
3.1.1 Zbieżność:



Można zauważyć, że granicą zbieżności w przypadku $\alpha = 1$ jest wartość parametru learning rate równa 1. Dla wartości nieznacznie większych widoczny jest znaczny wzrost wartości wyniku wraz z każdą następną iteracją.

3.1.2 Oszacowanie wartości najlepszej:

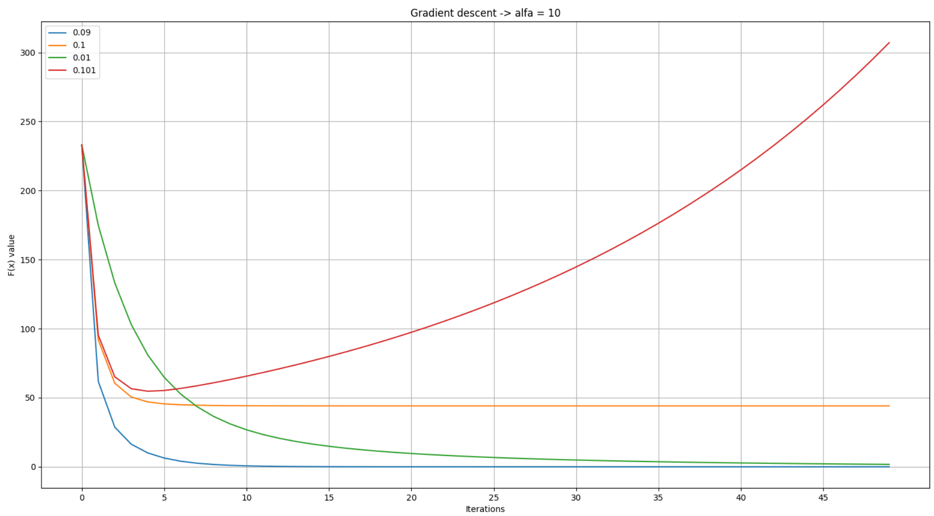




Na podstawie otrzymanych wyników można oszacować, że wartość parametru dla którego otrzymamy wynik najszybciej przy $\alpha = 1$ to około 0.5.

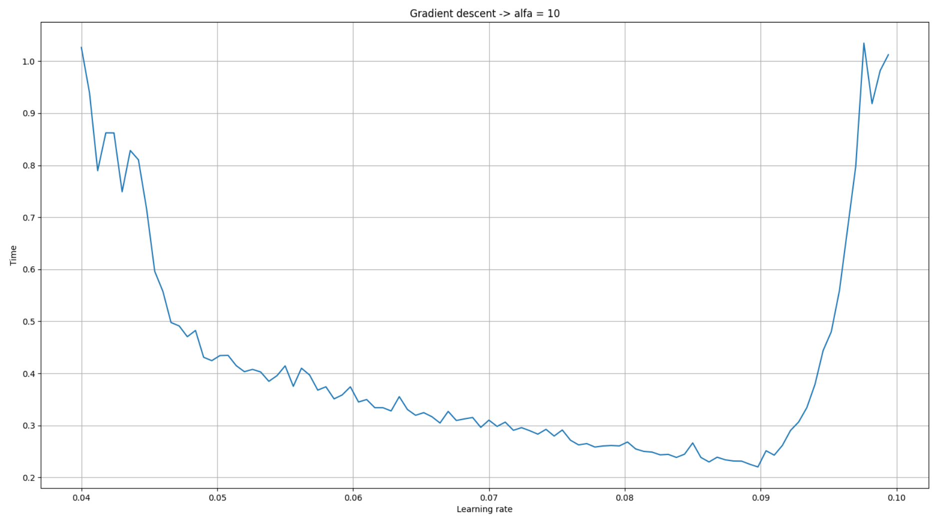
3.2 $\alpha = 10$

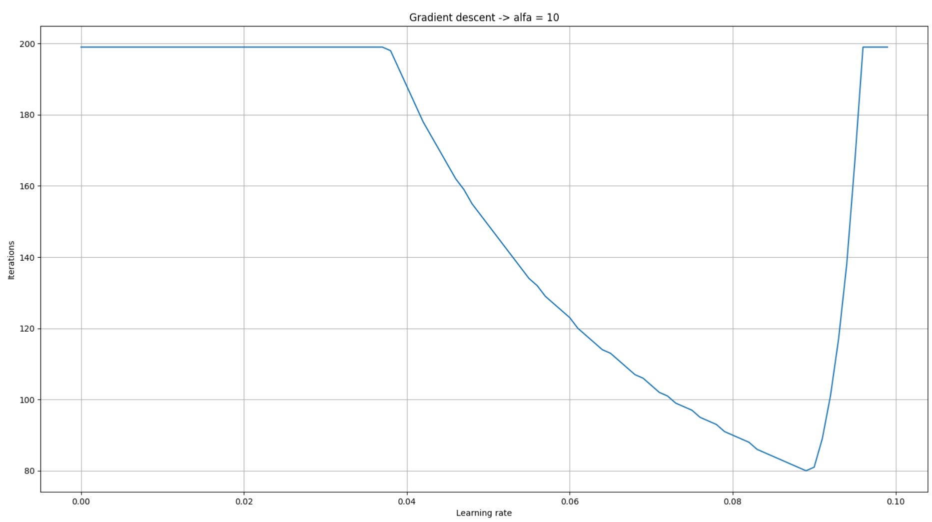
3.2.1 Zbieżność:



W przypadku $\alpha = 10$ wartość parametru z granicy zbieżności krztałtuje się na poziomie 0.1.
Na podstawie wykresu można zauważyć, że nawet niewielka nadwyżka o 0.001 w wartości parametru learning rate powoduje całkowitą utratę właściwości algorytmu.

3.2.2 Oszacowanie wartości najlepszej:

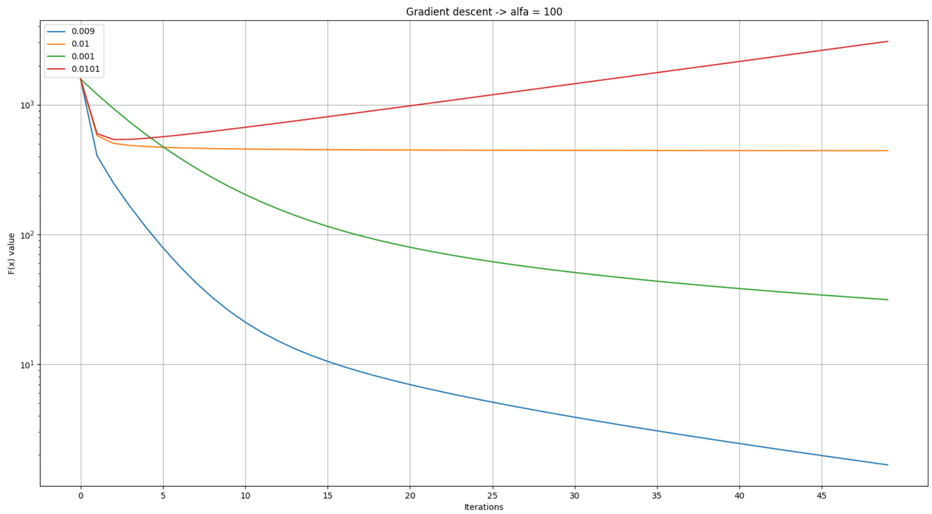




W tym przypadku wyznaczenie wartości "najlepszej" staje się już trudnejsze jednakże można oszacować, że wynosi ona około 0.09.

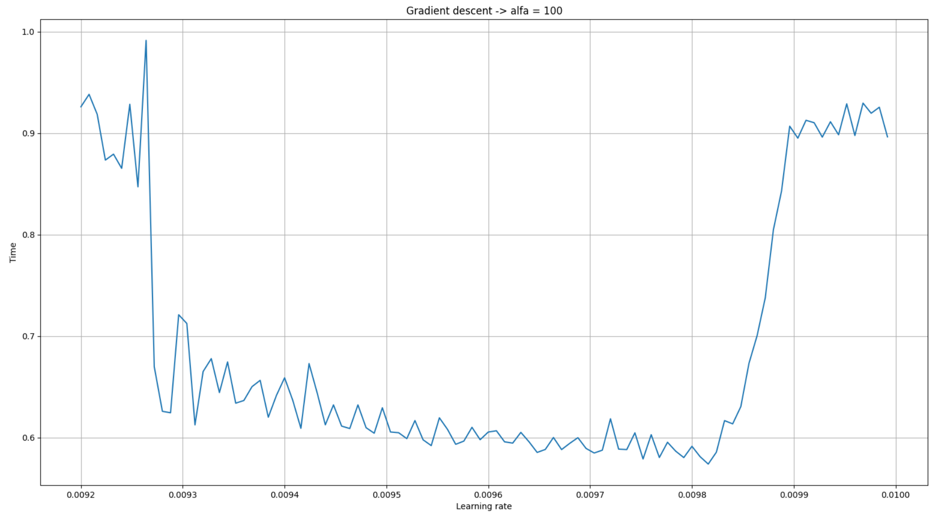
3.3 $\alpha = 100$

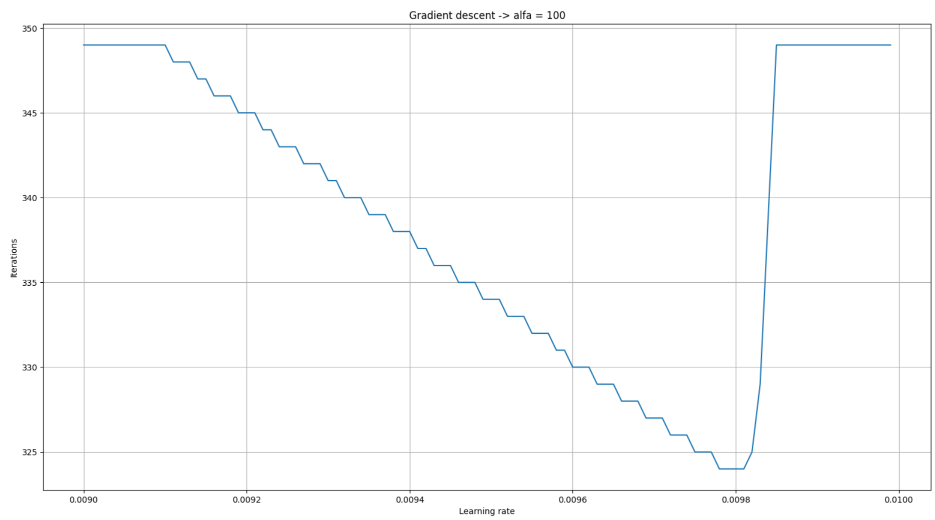
3.3.1 Zbieżność:



Dla tego przypadku konieczne było już użycie skali logarytmicznej ponieważ operujemy na znacznie większych wartościach. Wartość granicy zbieżności wynosi 0.001.

3.3.2 Oszacowanie wartości najlepszej:





W przypadku $\alpha = 100$ uzyskanie wykresu, który mógłby pomóc w oszacowaniu odpowiedniej wartości wymagało działania na odpowiednio wąskim przedziale jednocześnie określając maksymalną liczba iteracji na bardzo wysoką.
W tym przypadku dużo bardziej pomocny jest wykres liczby wykonanych iteracji od parametru learning rate. Pokazuje on, że wartości najlepszej należy się spodziewać w okolicy 0.0098.

4 Wnioski

Udało się wykorzystać stworzony algorytm do przeprowadzenia eksperymentów.

W pierwszym eksperymencie można zauważyć zależność, że dla zadanej funkcji celu wraz z dziesięciokrotnym zwiększeniem parametru α również dziesięciokrotnie maleje granica zbieżności.

W drugim eksperymencie można dostrzec istotę parametru learning rate. W każdej z badanych funkcji widoczna jest podobna charakterystyka algorytmu. Idąc od 0 zwiększanie parametru learning rate zmniejsza liczbę potrzebnych obliczeń w celu odnalezienie minimum.

Istnieje jednak wartość graniczna przy której każde kolejne zwiększenie wartości gwałtownie zwiększa liczbę potrzebnych iteracji. Co ciekawe, można zauważyć, że dla większego parametru α punkt najlepszy przesuwa się coraz bliżej punktu granicznego.

Przeprowadzone eksperymenty obrazują istotę właściwego dobrania parametru learning rate, co jest dużym wyzwaniem przy skomplikowanych funkcjach.