

Sieci neuronowe

Sprawozdanie nr 1 Perceptron prosty i Adaline

Przemysław Pietrzak, 238083
Środa, 14:15

Spis treści

1	Plan eksperymentów	2
2	Wpływ hiperparameterów na szybkość uczenia perceptronu prostego	3
2.1	Badanie wpływu wartości wag początkowych na szybkość uczenia perceptronu prostego	3
2.2	Badanie wpływu współczynnika uczenia na szybkość uczenia perceptronu prostego	4
2.3	Badanie wpływu funkcji aktywacji na szybkość uczenia perceptronu prostego . .	5
3	Wpływ hiperparameterów na szybkość uczenia Adaline	7
3.1	Badanie wpływu wartości wag początkowych na szybkość uczenia Adaline	7
3.2	Badanie wpływu współczynnika uczenia na szybkość uczenia Adaline	9
4	Porównanie perceptronu prostego i Adaline	10
5	Wnioski końcowe	10

1 Plan eksperymentów

Celem przeprowadzanych badań było omówienie właściwości i zachowania metod uczenia realizowanych przez pojedynczy neuron oraz zbadanie wpływu hiperparametrów na szybkość uczenia neuronu zarówno dla perceptronu prostego jak i Adaline.

W trakcie badań zbadano wpływ różnych hiperparametrów na proces uczenia perceptronu prostego i Adaline. Podstawową miarą szybkości procesu trenowania jest liczba epok, które były wymagane do ustalenia odpowiednich wag. W przypadku perceptronu prostego jest to liczba epok, po których wagi na połączeniach wejściu neuronu przestają być aktualizowane. W przypadku neuronu Adaline warunkiem kończącym trening jest osiągnięcie błędu średniokwadratowego mniejszego od zadanej wcześniej wartości. W celu przyspieszenia badań wprowadzono także ograniczenie polegające na przerwaniu treningu po upływie 1000 epok. Wszystkie otrzymane wyniki są wartościami uśrednionymi uzyskanymi w skutek wielokrotnego uruchomienia algorytmu.

Badania perceptronu prostego zostały przeprowadzone zarówno dla unipolarnej jak i bipolarnej funkcji przejścia. W przypadku Adaline, ze względu na jego charakterystykę, wykorzystano tylko bipolarną funkcję przejścia. Uczenie neuronu zostało przeprowadzone na zbiorze treningowym przedstawiającym bramkę AND. Postać wykorzystywanych danych prezentuje się następująco:

x1	x2	y
1	1	1
1	0	0
0	1	0
0	0	0

Tabela 1: Zbiór danych dla unipolarnej funkcji aktywacji

x1	x2	y
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1

Tabela 2: Zbiór danych dla bipolarnej funkcji aktywacji

W powyższych tabelach para **(x1, x2)** oznacza dane wejściowe, natomiast kolumna **y** zawiera wartości oczekiwane.

2 Wpływ hiperparameterów na szybkość uczenia perceptronu prostego

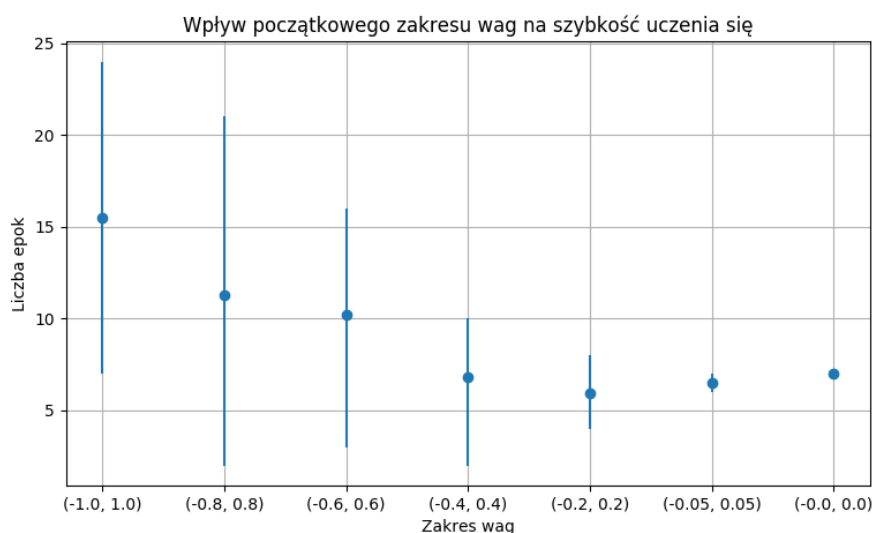
2.1 Badanie wpływu wartości wag początkowych na szybkość uczenia perceptronu prostego

Badanie ma na celu udowodnienie wpływu wartości wag początkowych na działanie perceptronu prostego. Badania przeprowadzone zostały dla unipolarnej funkcji aktywacji oraz unipolarnego zbioru uczącego. Wykorzystano współczynnik uczenia $\alpha = 0.05$.

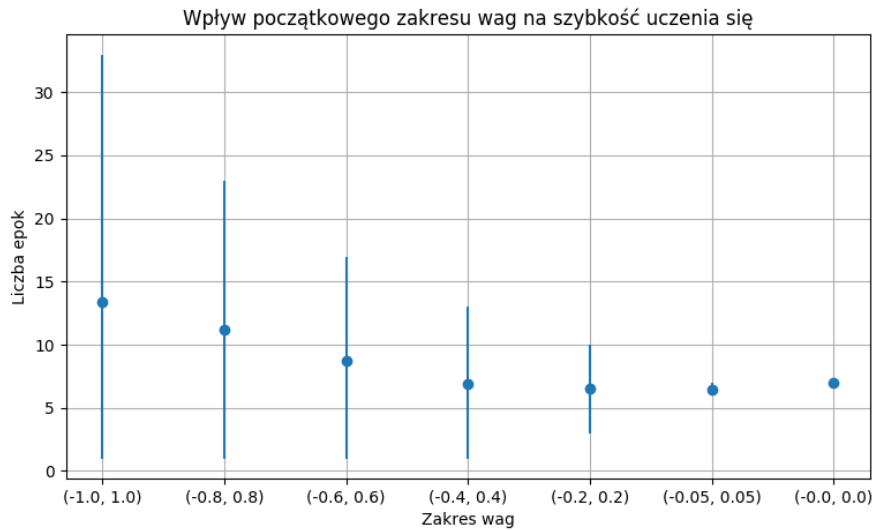
Zbadano natomiast następujące przedziały:

- $(-1, 1)$
- $(-0.8, 0.8)$
- $(-0.6, 0.6)$
- $(-0.4, 0.4)$
- $(-0.2, 0.2)$
- $(-0.05, 0.05)$
- $(0, 0)$

Tak jak zostało to wcześniej wspomniane, prezentowane wyniki są wartościami uśrednionymi, hline uzyskanymi w skutek wielokrotnego uruchomienia algorytmu i prezentują się one następująco.



Rysunek 1: Wyniki badań uzyskane w skutek 10 uruchomień



Rysunek 2: Wyniki badań uzyskane w skutek 100 uruchomień

W przeprowadzonych badaniach można zauważyć wzrost liczby epok wymaganych do dobrania odpowiednich wag, wraz ze wzrostem wielkości przedziału losowanych wag. Wynika to z faktu, że w przypadku dużych przedziałów wag mogą one wylosować skrajnie różne wartości. Duża różnica między początkową, a optymalną wartością wagi prowadzi, przy stałym współczynniku uczenia do wzrostu wymaganej liczby epok, a co za tym idzie czasu wymaganego na ukończenie treningu. Wraz ze spadkiem wymaganej liczby epok, zmniejsza się także odchylenie standardowe, oznacza to, że w przypadku małego przedziału wag wyniki są bardziej zróżnoważone i powtarzalne.

2.2 Badanie wpływu współczynnika uczenia na szybkość uczenia perceptronu prostego

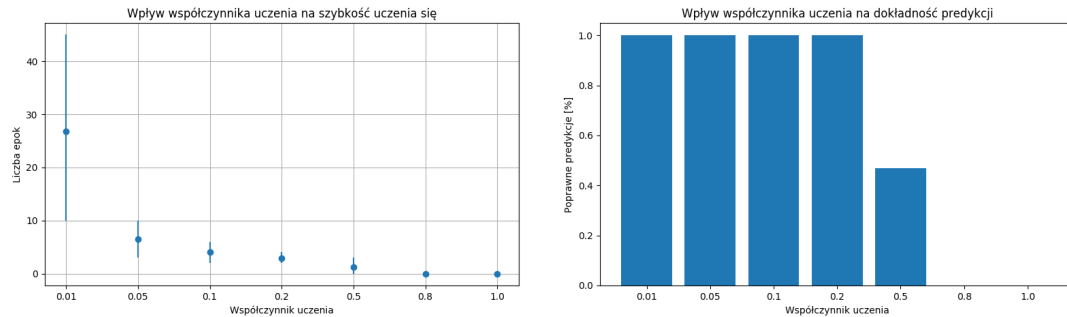
Badanie ma na celu zweryfikowanie wpływu wartości współczynnika uczenia na działanie perceptronu prostego. Badania przeprowadzone zostały dla unipolarnej funkcji aktywacji oraz unipolarnego zbioru uczącego. Przedział wag do badania został wybrany na podstawie poprzednich obserwacji i ma wartość $[-0.2, 0.2]$.

Badania przeprowadzono na następujących wartościach współczynnika.

- 0.01
- 0.05
- 0.1
- 0.2
- 0.5
- 0.8

- 1.0

Wynikiem przeprowadzonych badań są następujące wykresy:



(a) Liczba epok w zależności od współczynnika uczenia

(b) Procent prawidłowych predykcji w zależności od współczynnika uczenia

Rysunek 3: Wpływ współczynnika uczenia na szybkość i jakość treningu

Na pierwszy rzut oka widoczny jest spadek wymaganej liczby epok wraz ze wzrostem współczynnika uczenia. Jednak obserwacja ta nie jest całkowicie poprawna, ze względu na to, że współczynnik uczenia o wartości większej niż 0.5, uniemożliwiał perceptronowi odnalezienie odpowiednich wag. Co widoczne jest na drugim wykresie przedstawiającym procent poprawnych predykcji. Dodatkowo wraz ze wzrostem współczynnika uczenia można zauważyć zmniejszanie się odchylenia standardowego. Dla współczynnika o wartości **0.01** występowały duże wahania pomiędzy kolejnymi testami. Najlepsze wyniki uzyskiwały współczynnik o wartości **0.2**, który otrzymał bezbłędny wynik przy najmniejszej średniej liczbie epok, dodatkowo małe odchylenie standardowe wskazuje na to, że podobne wyniki otrzymane zostały w każdym z stu uruchomień algorytmu. Podsumowując, zbyt mały współczynnik uczenia wydłuża czas potrzebny na przeprowadzenie treningu dodatkowo powodując dużą różnorodność w liczbie epok między kolejnymi uruchomieniami algorytmu. Zbyt duży współczynnik natomiast całkowicie uniemożliwia przeprowadzenie udanego treningu.

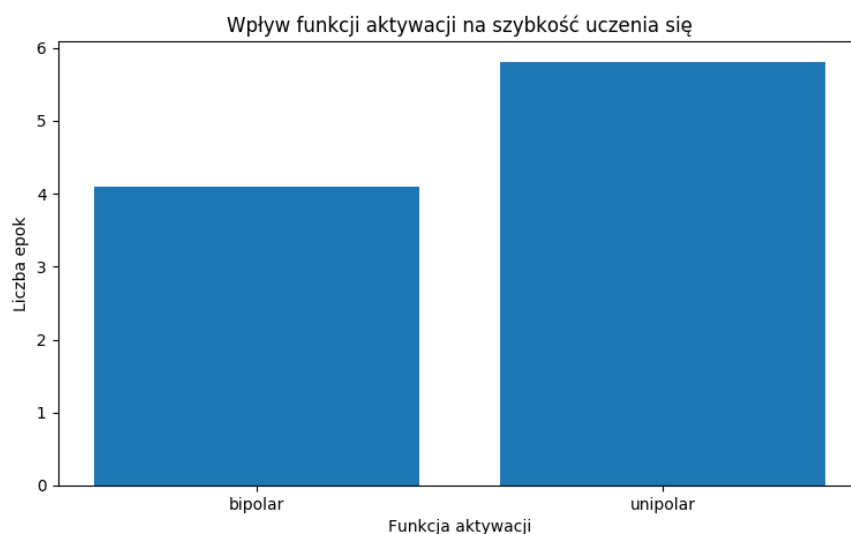
2.3 Badanie wpływu funkcji aktywacji na szybkość uczenia perceptronu prostego

Decyzję o aktywacji neuronu można przeprowadzić na podstawie wielu funkcji. Badanie skupiło się na zaobserwowaniu zmian w czasie trwania treningu przy różnych funkcjach aktywacji. W przeprowadzonych badaniach wykorzystano współczynnik uczenia równy 0.1 oraz początkowy przedział wag wynoszący $[-0.2, 0.2]$.

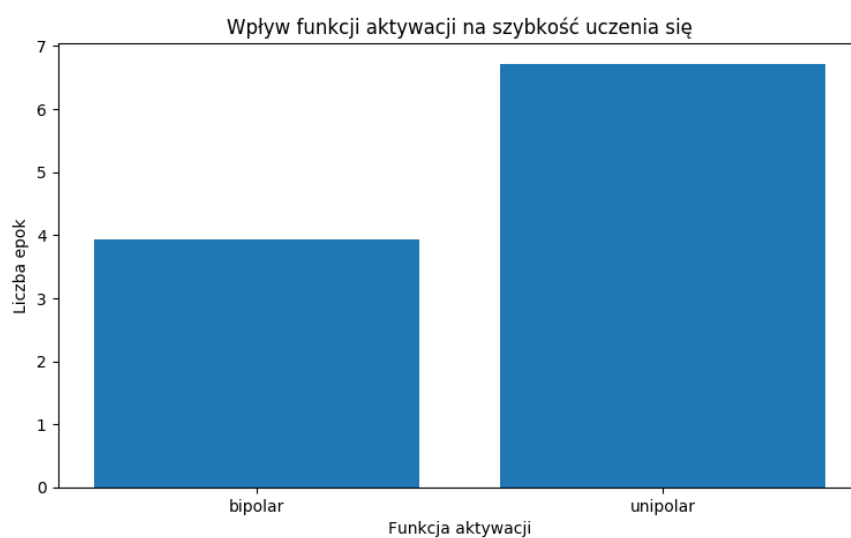
Badania przeprowadzono na następujących funkcjach aktywacji:

- Funkcja bipolarna
- Funkcja unipolarna

Po przeprowadzeniu pomiarów otrzymano następujące wyniki:



Rysunek 4: Wyniki badań uzyskane w skutek 10 uruchomień



Rysunek 5: Wyniki badań uzyskane w skutek 100 uruchomień

Jak można zauważyć funkcja bipolarna uzyskuje znacznie lepsze wyniki niż funkcja unipolarna. Wpływ na otrzymane wyniki ma jednak to, że w obu przypadkach użyto tej samej wartości współczynnika uczenia. W przypadku funkcji unipolarnej błąd, który używany jest do poprawiania wag może przyjąć wartości ze zbioru $[-1, 0, 1]$. Jednak użycie funkcji bipolarnej powoduje, że

zbiór możliwych błędów ma postać $[-2, 0, 2]$. Wagi aktualizowane są więc o większe wartości, co umożliwia szybsze dobranie optymalnych wag.

3 Wpływ hiperparameterów na szybkość uczenia Adaline

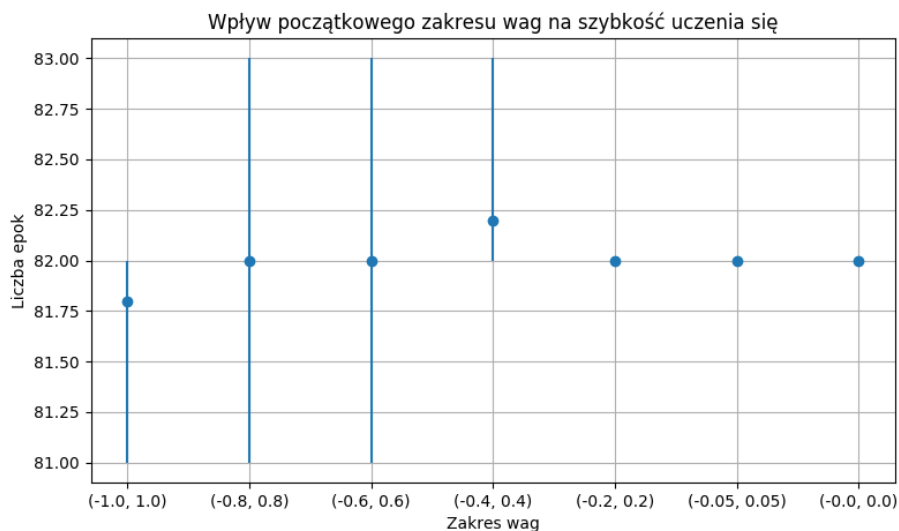
3.1 Badanie wpływu wartości wag początkowych na szybkość uczenia Adaline

Badanie ma na celu zweryfikowanie wpływu wartości wag początkowych na działanie Adaline. Badania przeprowadzone zostały dla unipolarnej funkcji aktywacji oraz unipolarnego zbioru uczącego. Wykorzystano współczynnik uczenia $a = 0.05$, zakres dopuszczalnego błędu natomiast został ustawiony na 0.08.

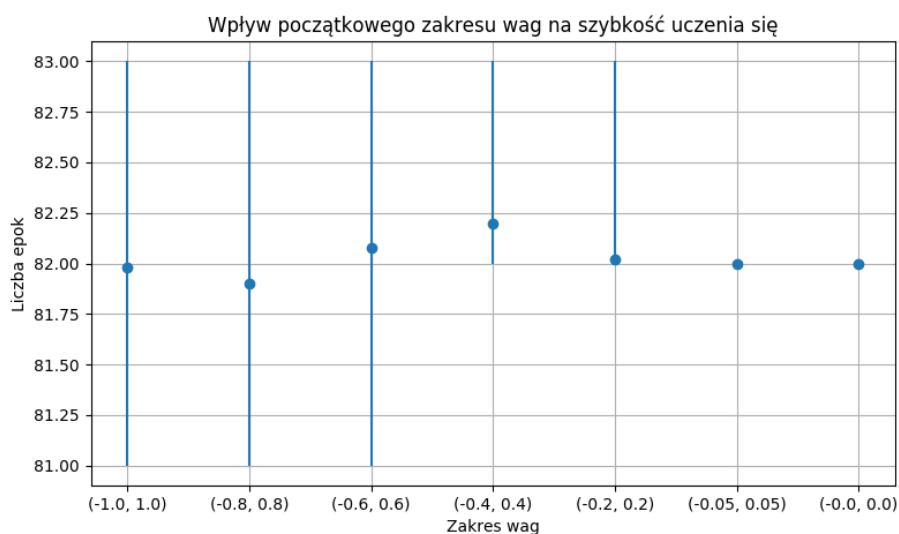
Zbadano natomiast następujące przedziały:

- $(-1, 1)$
- $(-0.8, 0.8)$
- $(-0.6, 0.6)$
- $(-0.4, 0.4)$
- $(-0.2, 0.2)$
- $(-0.05, 0.05)$
- $(0, 0)$

Tak jak zostało to wcześniej wspomniane, prezentowane wyniki są wartościami uśrednionymi, hline uzyskanymi w skutek wielokrotnego uruchomienia algorytmu i prezentują się one następująco.



Rysunek 6: Wyniki badań uzyskane w skutek 10 uruchomień



Rysunek 7: Wyniki badań uzyskane w skutek 100 uruchomień

Można zauważyć, że wyniki całkowicie różnią się od tych uzyskanych przy wykorzystaniu perceptronu prostego. W przypadku Adaline, średnia liczba epok wymaganych do wytrenowania modelu jest praktycznie stała, z niewielkimi odchyłami. Widać jednak, że w przypadku mniejszych przedziałów zmniejsza się różnorodność uzyskanych wyników, jednoznacznie wskazując na to zaobserwowane odchylenie standardowe. Wyniki te spowodowane są zasadą działania neuronu Adaline, który zawsze zmierza do ustalenia tych samych optymalnych wag, niezależnie od

wartości wag początkowych. Dodatkowo, w przypadku Adaline, wagi poprawiane są nie o błąd otrzymany w wyniku nałożenia funkcji aktywacji, a całkowite pobudzenie neuronu. Umożliwia to bardziej dokładne aktualizacje wartości wag, co prowadzi do szybszego znalezienia rozwiązania.

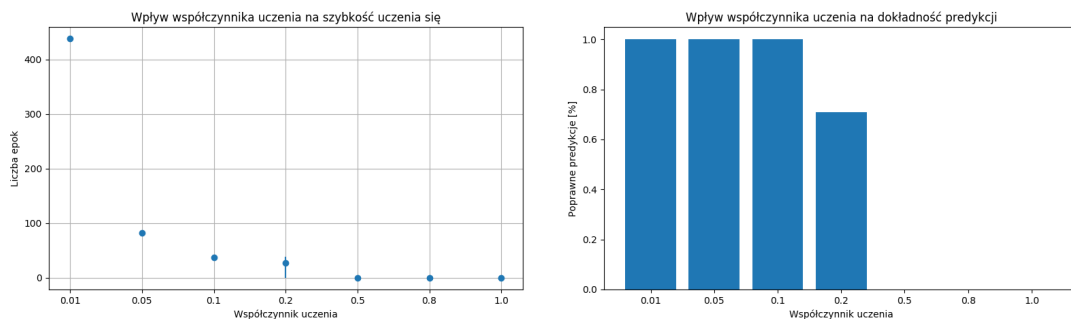
3.2 Badanie wpływu współczynnika uczenia na szybkość uczenia Adaline

Badanie ma na celu określenie wpływu wartości współczynnika uczenia na działanie perceptronu prostego. Badania przeprowadzone zostały dla unipolarnej funkcji aktywacji oraz unipolarnego zbioru uczącego. Przedział wag do badania został wybrany na podstawie poprzednich obserwacji i ma wartość $[-0.2, 0.2]$.

Badania przeprowadzono na następujących wartościach współczynnika.

- 0.01
- 0.05
- 0.1
- 0.2
- 0.5
- 0.8
- 1.0

Wynikiem przeprowadzonych badań są następujące wykresy:



(a) Liczba epok w zależności od współczynnika uczenia

(b) Procent prawidłowych predykcji w zależności od współczynnika uczenia

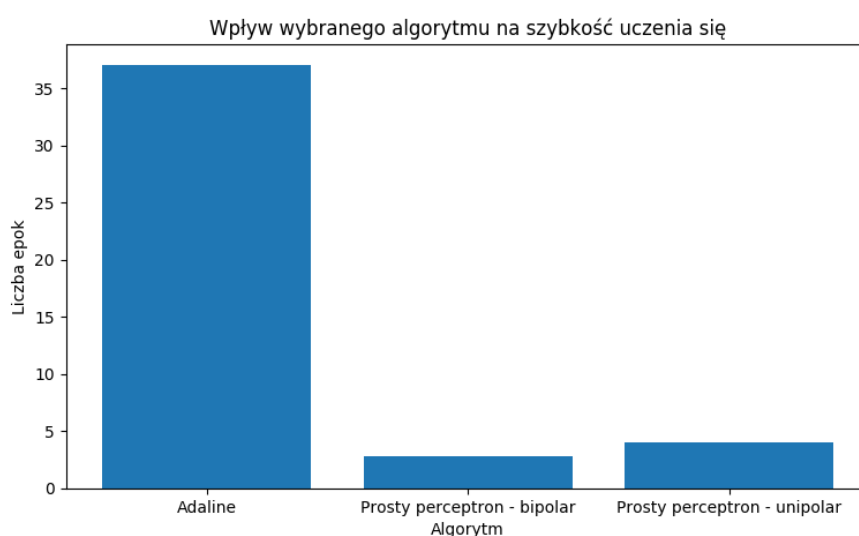
Rysunek 8: Wpływ współczynnika uczenia na szybkość i jakość treningu

Wpływ współczynnika uczenia na szybkość uczenia się neuronu Adaline wygląda tak samo jak w przypadku perceptronu prostego. Tak samo jak w przypadku perceptronu prostego, duże wartości współczynnika uczenia uniemożliwiały ukończenie treningu. Można jednak zauważyć, że w przypadku Adaline liczba wymaganych epok dla tej samej wartości współczynnika jest większa. Nieudane treningi przy dużych wartościach współczynnika spowodowane są tym, że algorytm ciągle wprowadzał zmiany do wag, które powodowały, że błąd średniokwadratowy nigdy nie

mógł spaść poniżej zadanej wartości. Oznacza to, że trening nigdy nie mógł się skończyć, a ciągle rosnące wartości błędu prowadziły do otrzymania gigantycznych wag. Efekt ten jest nazywany eksplozją gradientu.

4 Porównanie perceptronu prostego i Adaline

W celu zweryfikowania wyników otrzymanych w poprzednich badaniach, przeprowadzono trening perceptronu prostego oraz Adaline przy takich samych wartościach hiperparametrów. Wybrany współczynnik uczenia się wynosi 0.1, parametr theta ma natomiast wartość 0.5. Wagi początkowe wygenerowane zostały natomiast z przedziału $[-0.2, 0.2]$. W przypadku perceptronu prostego uruchomiono pomiary zarówno dla funkcji unipolarnej jak i bipolarnej.



Rysunek 9: Porównanie wymaganej liczby epok dla różnych algorytmów

Na powyższym wykresie widocznym jest, że obie wersje perceptronu prostego, zarówno ta wykorzystująca unipolarną funkcję aktywacji jak i bipolarną, osiągają znacznie lepsze wyniki średnie niż Adaline. Związane jest to z obserwacjami pojawiającymi się w poprzednich badaniach. W przypadku Adaline wartości wag początkowych nie są tak samo ważne jak w przypadku perceptronu prostego, ponieważ Adaline i tak zawsze zmierza do ustanowienia tych samych wag. Jednocześnie to, że w trakcie uczenia, nie transformujemy wyników predykcji przy użyciu funkcji aktywacji powoduje, że zmiany w wagach mogą być bardzo gwałtowne przez co algorytm wymaga więcej czasu na odpowiednie dopasowanie. Tak jak zauważono w poprzednich badaniach, w przypadku perceptronu prostego, bipolarna funkcja aktywacji osiąga lepsze wyniki niż unipolarna. Nie jest to jednak tak gwałtowna różnica jak w przypadku Adaline.

5 Wnioski końcowe

Zbadano wpływ różnych parametrów na proces uczenia perceptronu prostego jak i neuronu Adaline, we wszystkich przypadkach potwierdzono wpływ parametrów na trening neuronów.

W zależności od współczynnika uczenia, przedziału wag początkowych czy funkcji aktywacji, proces trenowania wydłużał się lub skracał. Badania wykazały także, że znalezienie optymalnych hiperparametrów może znacznie skrócić proces trenowania neuronu.

Badania jasno wskazują na to, że Adaline uczy model bardziej dokładnie. Spowodowane jest to tym, że w przypadku neuronu Adaline, wagi aktualizowane są o rzeczywistą wartość błędu, a nie tak jak w przypadku perceptronu prostego wartość otrzymaną po aktywacji. Także, w przeciwieństwie do Adaline, perceptron prosty nie aktualizuje wag w sytuacji, kiedy model sklasyfikował zadaną wartość poprawnie. Dodatkowo Adaline, niezależnie od wylosowanych wag początkowych, zawsze próbuje dojść do tych samych wag na wejściach do neuronu. Oznacza to, że w jego przypadku dobranie odpowiedniego przedziału wag początkowych nie jest tak bardzo ważne. Bez wątplenia jednak, oba z badanych neuronów są użyteczne tylko w przypadku problemów separowalnych liniowo i jest to ich największą wagą. W przypadku problemów nie-separowalnych liniowo, takich jak przykładowo bramka XOR, wymagane jest użycie bardziej zaawansowanego modelu, np. MultiLayer Perceptron (MLP).