
Sprawozdanie 2

Konrad Gmyrek
247008

Abstract

Poniżej przeprowadzono kilka badań dla perceptronu wielowarstwowego, w których badany jest wpływ różnych hiperparametrów na szybkość uczenia. Głównym celem tych badań było sprawdzenie jak poszczególne hiperparametry wpływają na modele oraz wyrobienie sobie intuicji odnośnie działania prostej sieci neuronowej. Repozytarium kodu: <https://github.com/KonradGmy/Lab02>

Każde badanie zostało powtórzone 10 razy, a wyniki przedstawione w tabelach są uśrednione. Do inicjalizacji wag i biasów jest używana funkcja `numpy.random.normal` (średnia = 0), maksymalna ilość epok wynosi 200, perceptron wielowarstwowy jest badany dla zbioru danych MNIST.

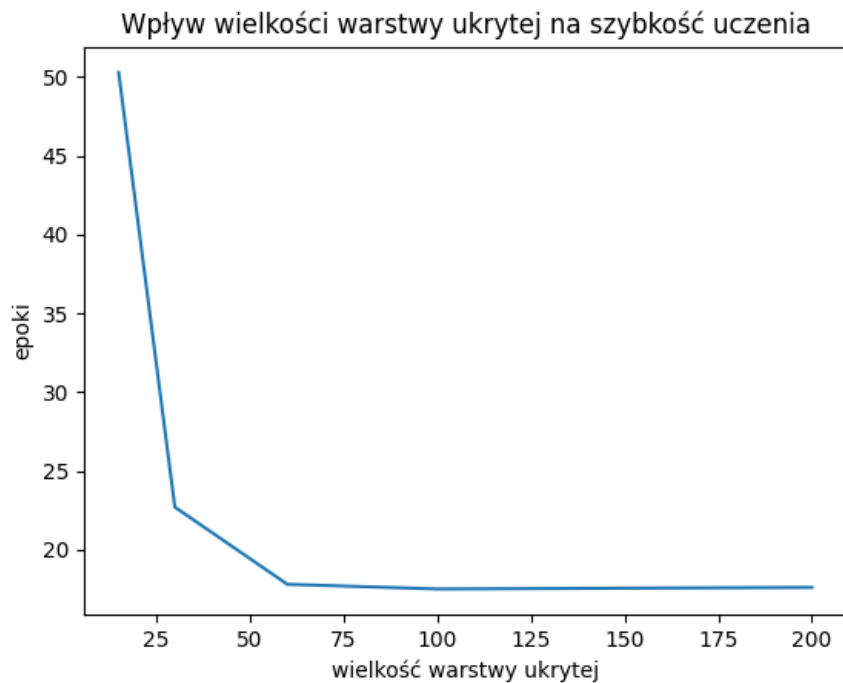
1 Badanie 1

Tabela 1: Parametry eksperymentu 1

θ	batch	funkcja	σ wag	σ biasów	próg przerywania uczenia	warstwy
0.1	100	sigmoid	0.1	0.1	95% dokładności	(784, badane, 10)

Tabela 2: Wpływ wielkości warstwy ukrytej na szybkość uczenia

wielkość warstwy ukrytej	15	30	60	100	200
epoki	50.3	22.7	17.8	17.5	17.6



Wnioski: Optymalna wielkość warstwy ukrytej dla zbioru danych MNIST wynosi ok. 100. Wszystkie wielkości warstwy ukrytej powyżej 60 dają bardzo podobne rezultaty.

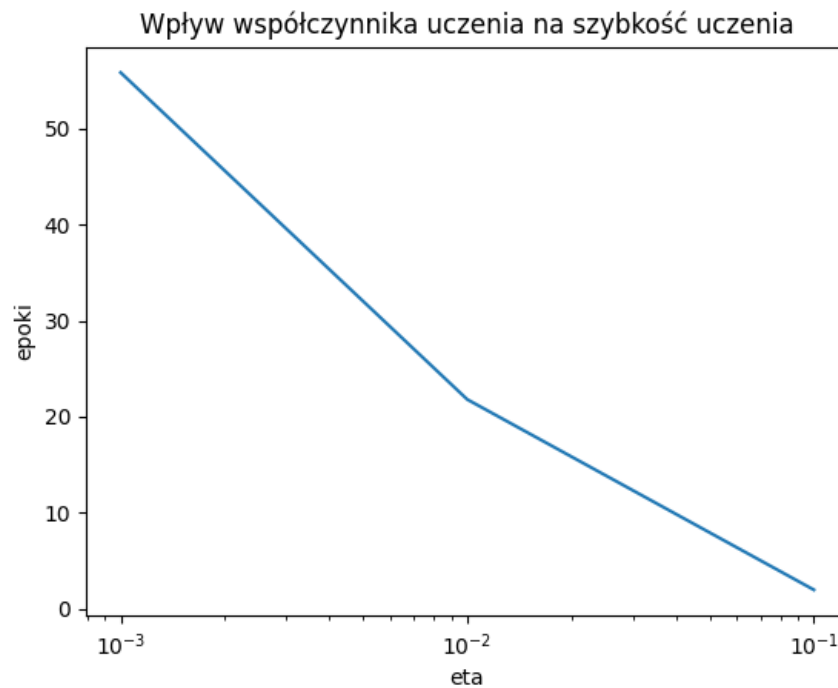
2 Badanie 2

Tabela 3: Parametry eksperymentu 2

θ	batch	funkcja	σ wag	σ biasów	próg przerywania uczenia	warstwy
Badane	100	sigmoid	0.1	0.1	95% dokładności	(784, 100, 10)

Tabela 4: Wpływ współczynnika uczenia na szybkość uczenia

θ	0.001	0.01	0.1
epoki	55.8	21.8	2.0



Wnioski: Dla mniejszych wartości współczynnika uczenia potrzebna jest większa ilość epok, aby osiągnąć dokładność 95%. Optymalną wartością współczynnika uczenia jest około 0.1

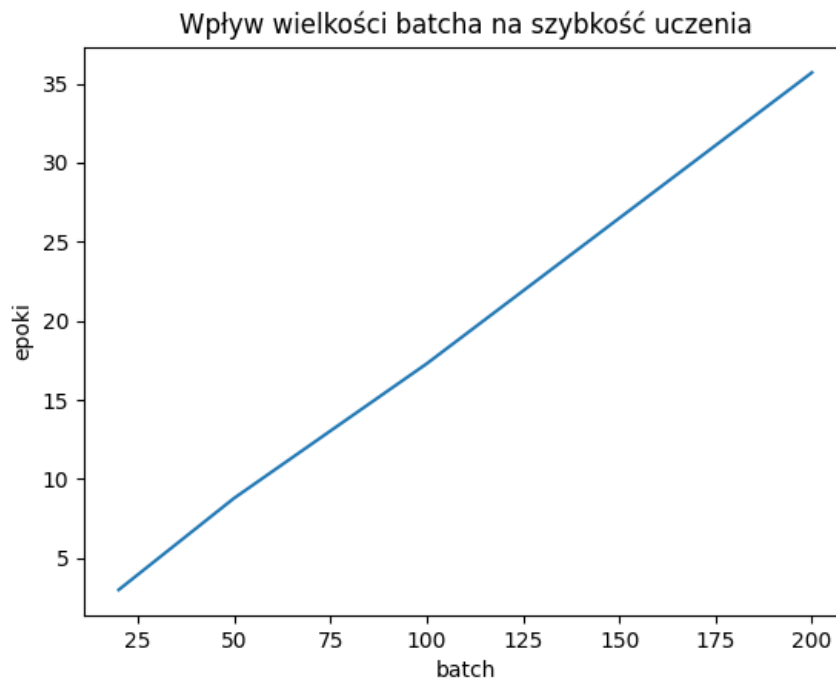
3 Badanie 3

Tabela 5: Parametry eksperymentu 3

θ	batch	funkcja	σ wag	σ biasów	próg przerwania uczenia	warstwy
0.1	Badane	sigmoid	0.1	0.1	95% dokładności	(784, 100, 10)

Tabela 6: Wpływ wielkości paczki na szybkość uczenia

batch	20	50	100	200
epoki	3.0	8.8	17.3	35.7



Wnioski: Dla mniejszego rozmiaru paczki sieć uczy się szybciej, wraz ze zwiększeniem rozmiaru paczki liniowo rośnie też ilość epok potrzebnych do osiągnięcia progu 95%, warto jednak zaznaczyć, że im mniejszy rozmiar paczki epoka trwa znacznie dłużej.

4 Badanie 4

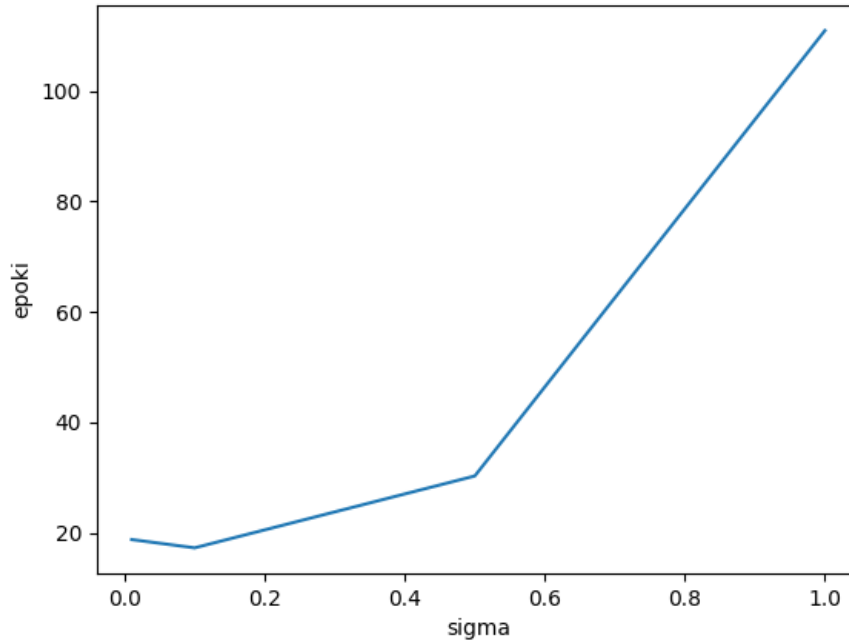
Tabela 7: Parametry eksperymentu 4

θ	batch	funkcja	σ wag	σ biasów	próg przerywania uczenia	warstwy
0.1	100	sigmoid	Badane	Badane	95% dokładności	(784, 100, 10)

Tabela 8: Wpływ odchylenia standardowego rozkładu mormalnego macierzy W oraz biasów na szybkość uczenia

σ dla W i biasów	0.01	0.1	0.5	1
epoki	18.8	17.3	30.3	110.9

Wpływ odchylenia standardowego rozkładu normalnego macierzy W oraz biasu na



Wnioski: Optymalną wartością odchylenia standardowego jest ok. 0.1, dalej im większe odchylenie standardowe liczba epok potrzebnych do osiągnięcia progu 95% dokładności rośnie eksponencjalnie.

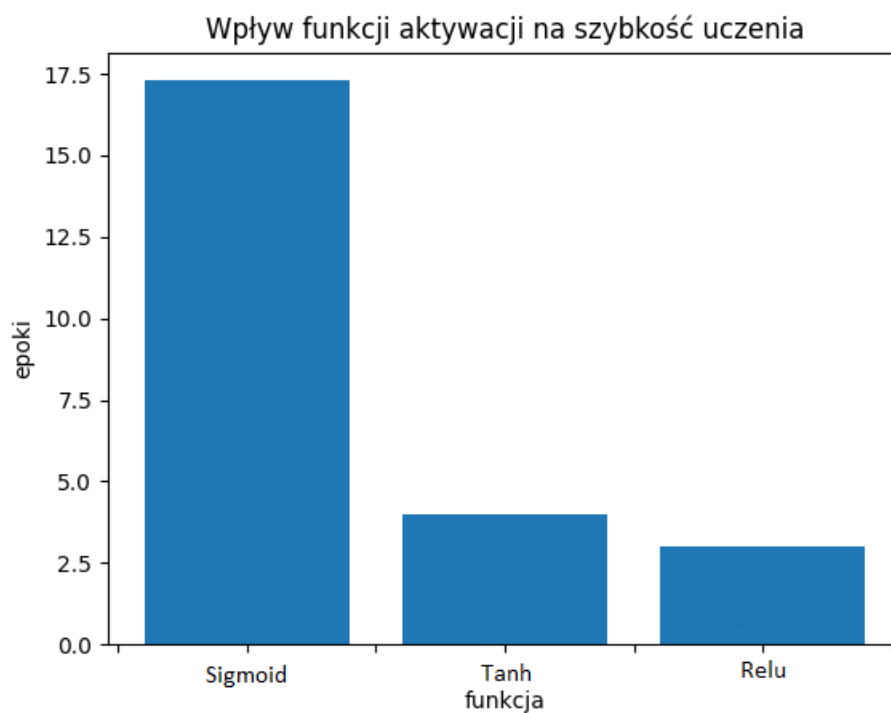
5 Badanie 5

Tabela 9: Parametry eksperymentu 5

θ	batch	funkcja	σ wag	σ biasów	próg przerwania uczenia	warstwy
0.1	100	Badane	0.1	0.1	95% dokładności	(784, 100, 10)

Tabela 10: Wpływ funkcji aktywacji na szybkość uczenia

funkcja	sigmoid	tanh	relu
epoki	17.3	4.0	3.0



Wnioski: Najszybciej sieć uczy się przy użyciu funkcji aktywacji tanh oraz relu, można zaobserwować również znaczącą różnicę w czasie uczenia dla funkcji sigmoid gdzie epok potrzebnych do osiągnięcia progu 95% było zdecydowanie więcej.

6 Badanie 6

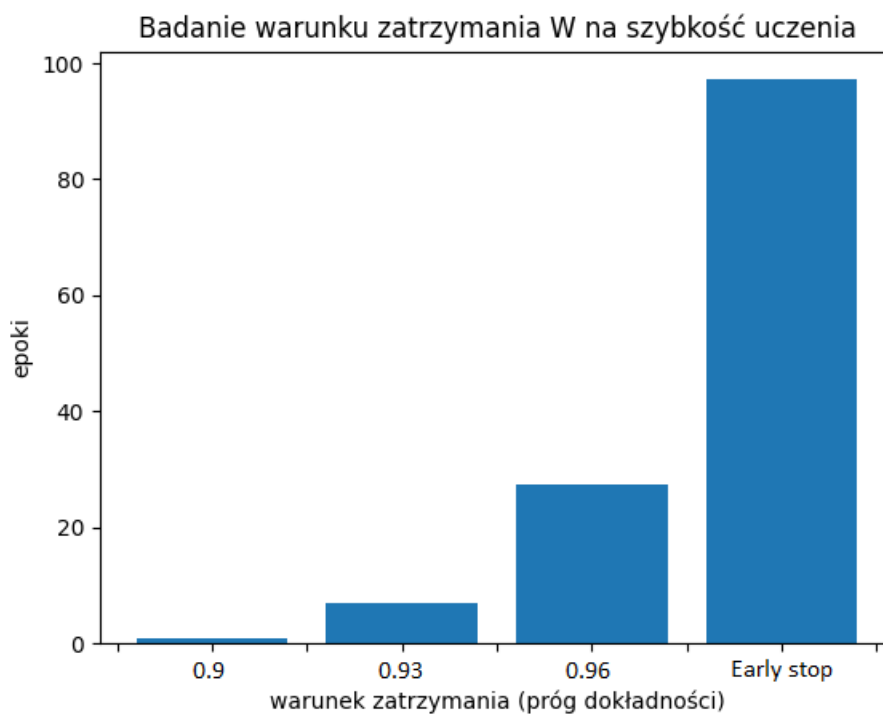
W celu sprawdzenia jak szybko sieć osiąga dany stopień dokładności i kiedy uczenie zostaje wstrzymane (early stopping) wykonałem również badanie dla zmiennego progu przerywania uczenia.

Tabela 11: Parametry eksperymentu 6

θ	batch	funkcja	σ wag	σ biasów	próg przerywania uczenia	warstwy
0.1	100	sigmoid	0.1	0.1	badane	(784, 100, 10)

Tabela 12: Wpływ funkcji aktywacji na szybkość uczenia

próg	0.9	0.93	0.96	early stop
epoki	1.0	7.0	27.3	97.2



Wnioski: Sieć, aby osiągnąć większy próg dokładności potrzebuje więcej epok, sieć automatycznie zatrzymywała uczenie po około 100 epokach (warto zaznaczyć, że dokładność wynosiła wtedy średnio ok 97.5%)