

---

# Sprawozdanie 3

---

**Konrad Gmyrek**  
247008

## **Abstract**

Poniżej przeprowadzono badania optymalizatorów oraz inicjalizatorów perceptronu wielowymiarowego, na szybkość uczenia. Głównym celem tych badań było sprawdzenie jak poszczególne inicjalizatory i optymalizatory wpływają na modele oraz wyrobienie sobie intuicji odnośnie ich działania.

Repozytorium kodu: <https://github.com/KonradGmy/Lab03>

Każde badanie zostało powtórzone 10 razy, a wyniki przedstawione w tabelach są uśrednione, maksymalna ilość epok wynosi 20, perceptron wielowarstwowy jest badany dla zbioru danych MNIST.

Tabela 1: Tabela parametrów stałych dla badań

$\theta$	batch	próg	warstwy
0.1	100	93%	(784, 100, 10)

# 1 Badanie optymalizatorów

Wpływ optymalizatorów na szybkość uczenia

Tabela 2: Hiperparametry optymalizatorów

Momentum $\gamma$	Momentum Nest. $\gamma$	Adagrad $\epsilon$	Adadelata $\epsilon \mid \gamma$	Adam $\epsilon \mid \beta_1 \mid \beta_2$
0.9	0.9	0.0001	0.0005 $\mid$ 0.9	0.1 $\mid$ 0.9 $\mid$ 0.9

Tabela 3: Parametry eksperymentu 1.1

funkcja	inicjalizator	optymalizator
sigmoid	standardowy	<b>badane</b>

Tabela 4: Wpływ optymalizatorów na szybkość uczenia dla funkcji sigmoidalnej

Standard	Momentum	Momentum Nest.	Adagrad	Adadelata	Adam
6.9	4.4	4.0	5.5	6.2	3.6

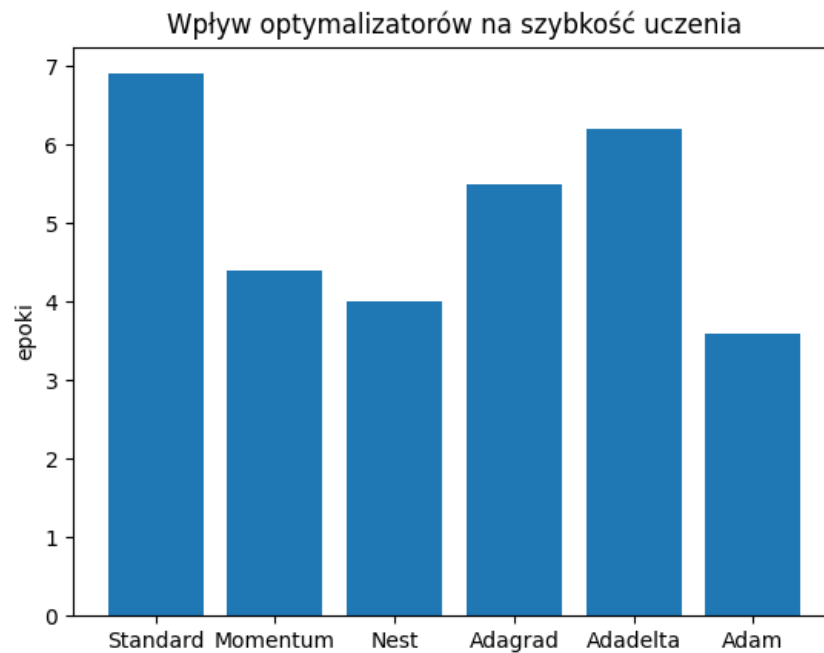
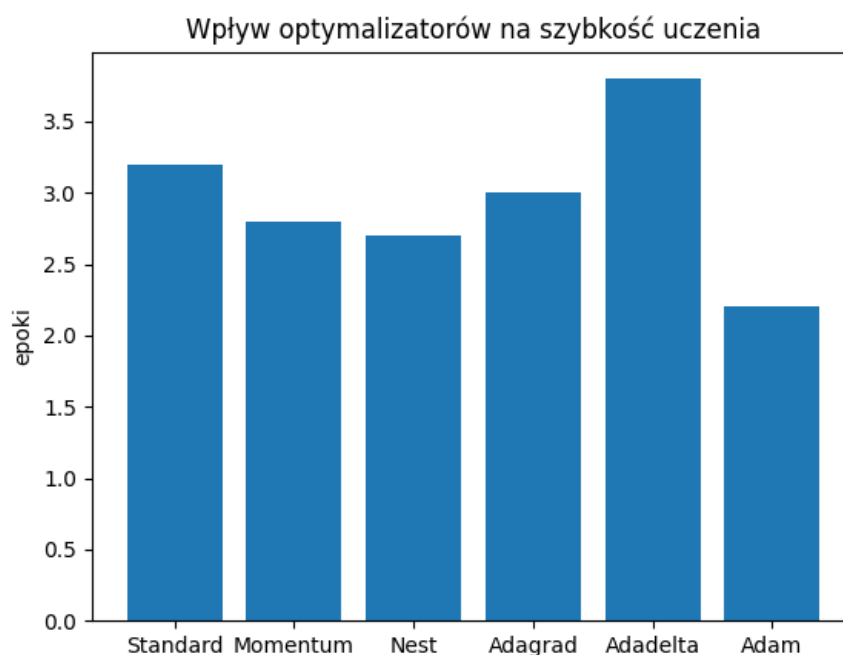


Tabela 5: Parametry eksperymentu 1.2

funkcja relu	inicjalizator standardowy	optymalizator <b>badane</b>
-----------------	------------------------------	--------------------------------

Tabela 6: Wpływ optymalizatorów na szybkość uczenia dla funkcji relu

Standard	Momentum	Momentum Nest.	Adagrad	Adadelta	Adam
3.2	2.8	2.7	3.0	3.8	2.2



Wnioski: Dla obu funkcji aktywacji można zaobserwować podobną tendencję, wszystkie optymalizatory zmniejszają ilość epok potrzebnych do osiągnięcia przez model wymaganego progu dokładności. Wyjątkiem jest tutaj adadelta, który dla funkcji relu daje gorsze wyniki. Najlepiej sprawdza się momentum Nesterova oraz adam. Wyniki dla funkcji relu są średnio lepsze niż w przypadku funkcji sigmoidalnej.

## 2 Badanie inicjalizatorów

Wpływ inicjalizatorów na szybkość uczenia

Tabela 7: Parametry eksperymentu 5.1

funkcja	inicjalizator	optymalizator
sigmoid	<b>badane</b>	standardowy

Tabela 8: Wpływ inicjalizatorów na szybkość uczenia dla funkcji sigmoidalnej

Standard	Xavier	He
6.9	6.7	6.8

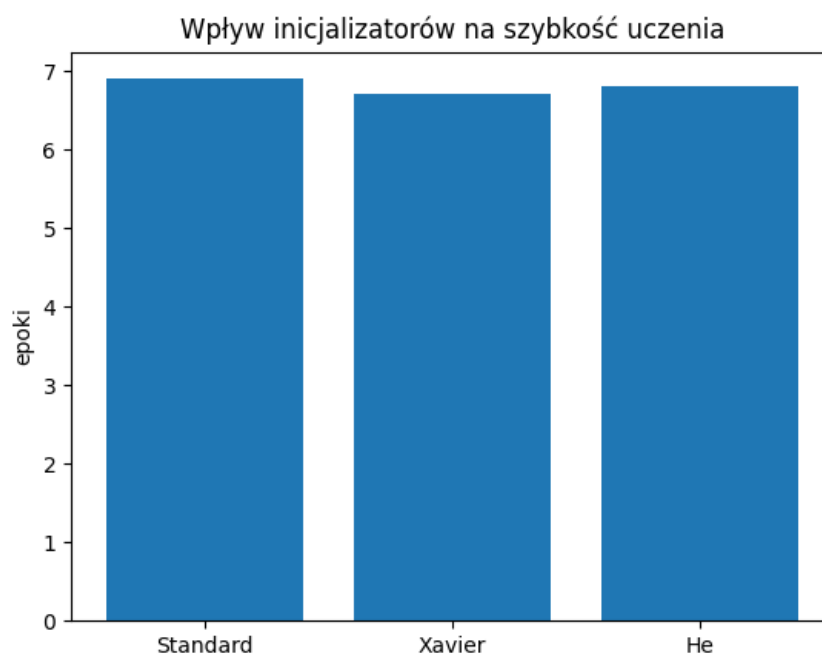
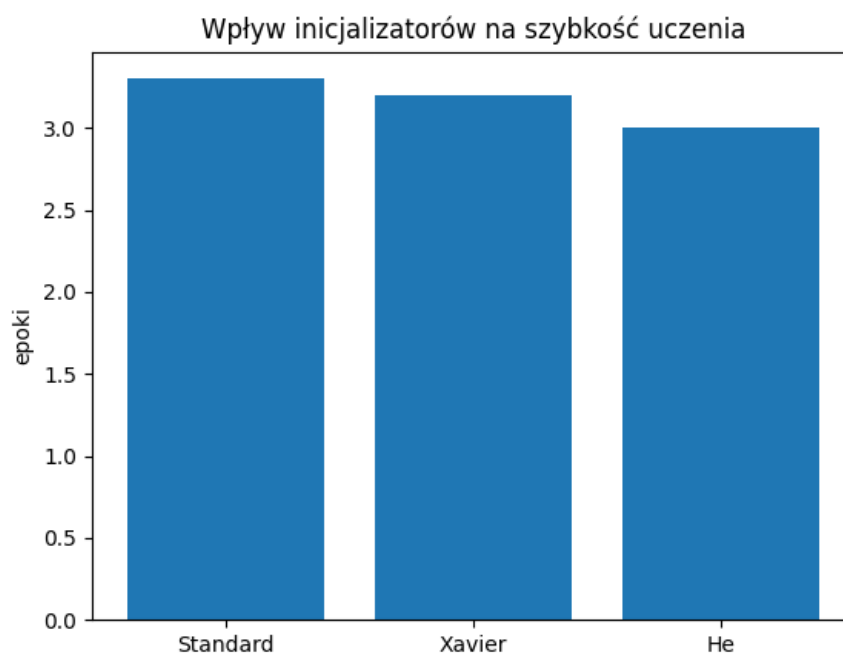


Tabela 9: Parametry eksperymentu 5.2

funkcja	inicjalizator	optymalizator
relu	<b>badane</b>	standardowy

Tabela 10: Wpływ inicjalizatorów na szybkość uczenia dla funkcji relu

Standard	Xavier	He
6.9	6.7	6.8
3.3	3.2	3.0



Wnioski: W porównaniu do inicjalizatora standardowego (odchylenie standardowe = 0.1), Xavier oraz He nieco przyspieszają proces uczenia jednak zmiana ta nie jest na tyle znacząca jak w przypadku optymalizatorów. Zastosowanie funkcji aktywacji relu ponownie przyspiesza proces uczenia w porównaniu z funkcją sigmoidalną.