
Sieć wielowarstwowa uczona metodą propagacji wstecznej

Autor: Rafał Behrendt

246643@student.pwr.edu.pl

Prowadząca: dr hab. inż. Urszula Markowska-Kaczmar

11 listopada 2021

Abstract

Celem ćwiczenia jest zapoznanie się z siecią wielowarstwową, uczeniem sieci za pomocą algorytmu propagacji wstecznej w wersji klasycznej (minimalizacja błędu średniokwadratowego oraz wpływem parametrów odgrywających istotną rolę w uczeniu sieci z propagacją wsteczną. Eksperymenty przeprowadzono na zbiorze MNIST zawierającym 60000 danych treningowych oraz 10000 danych testowych.

Kod źródłowy znajduje się w poniższym linku:

<https://github.com/aspirow/neuralNetworks/tree/list2>

1 Eksperymenty

Za domyślne hiperparametry przyjęto:

- Wielkość sieci: [20]
- Współczynnik uczenia: 0.3
- Funkcja aktywacji: tangens hiperboliczny
- Inicjalizacja wag za pomocą rozkładu normalnego o parametrach:
mi: 0
sigma: 0.5

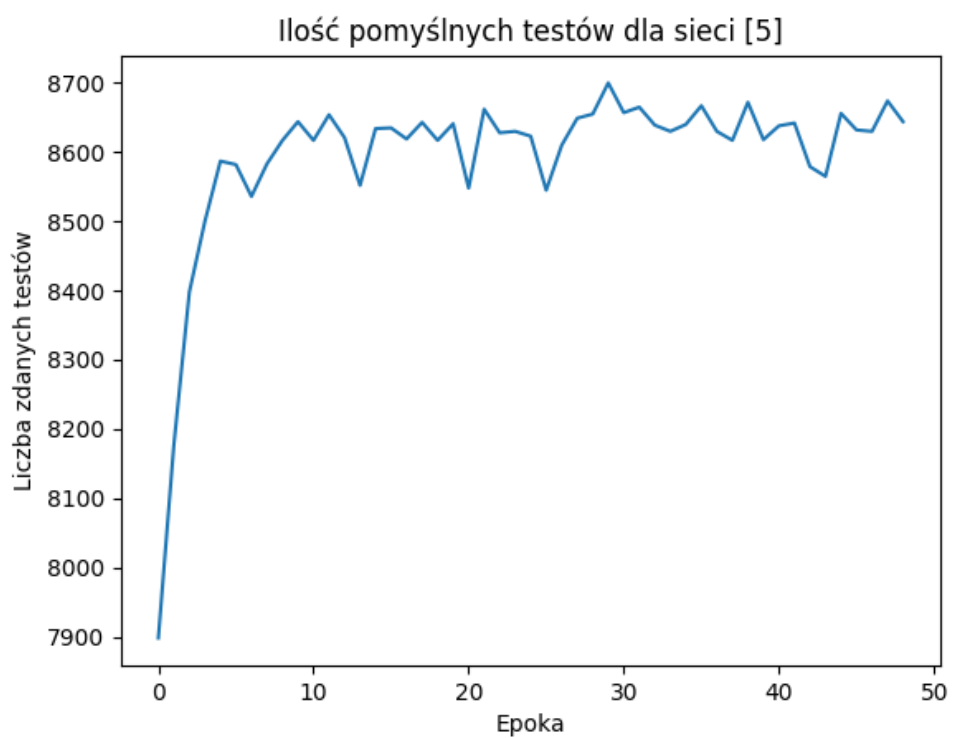
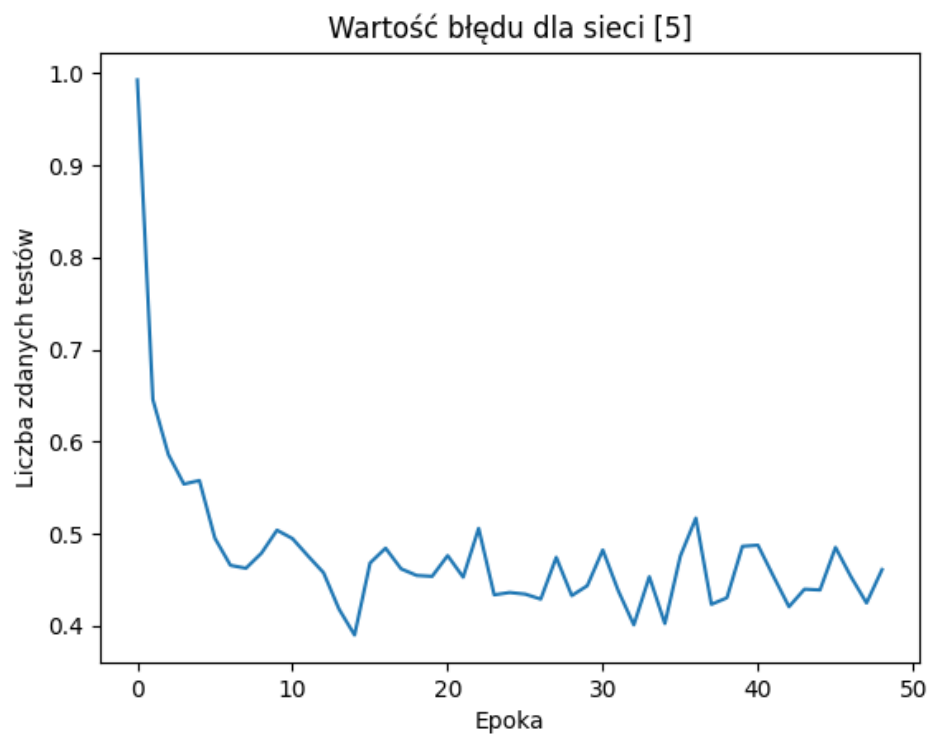
Każdy test przeszedł przez 50 epok zanim został zakończony.

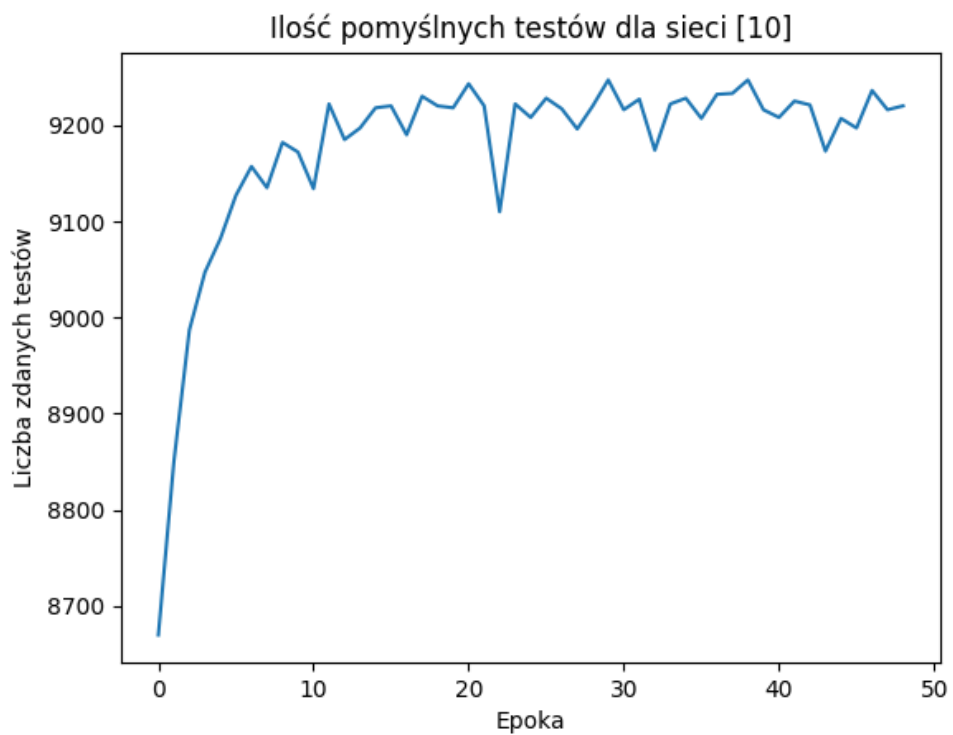
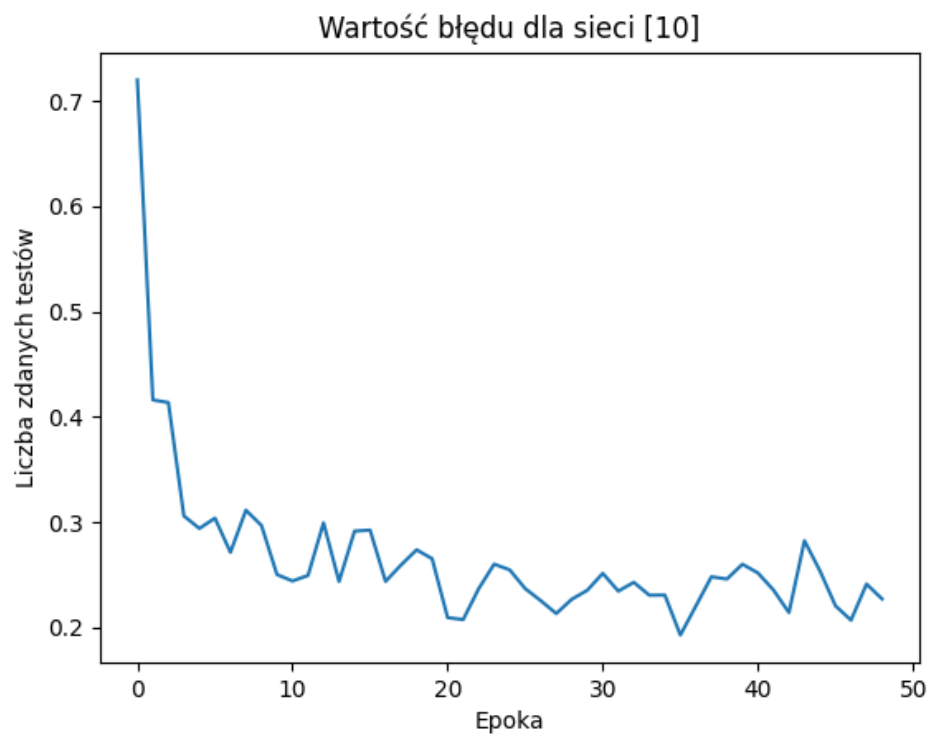
1.1 Test wielkości sieci

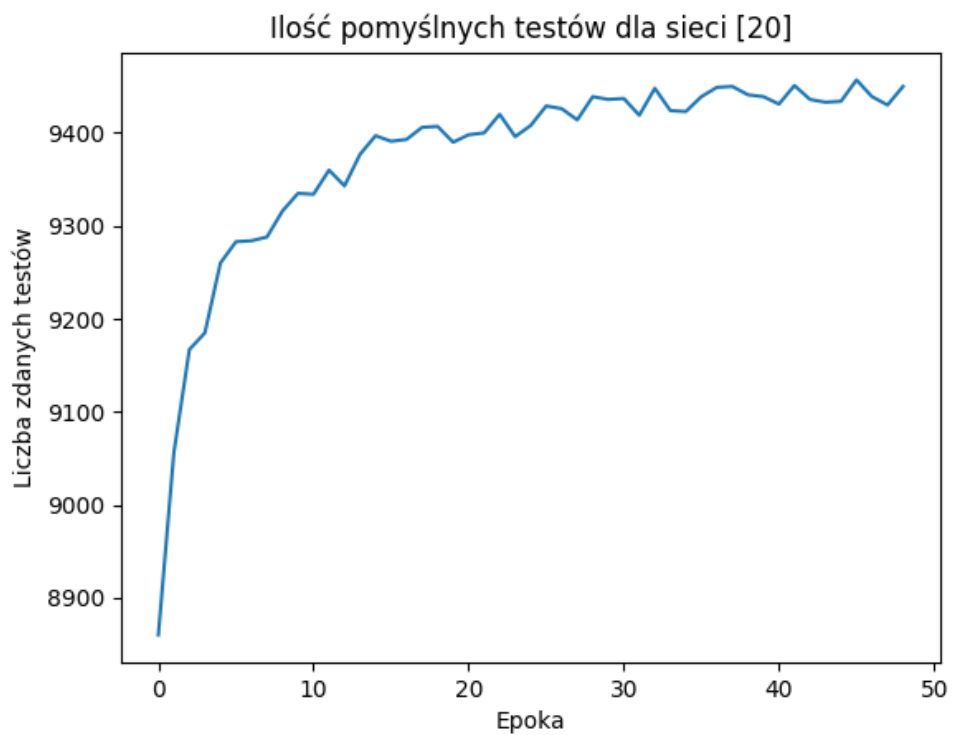
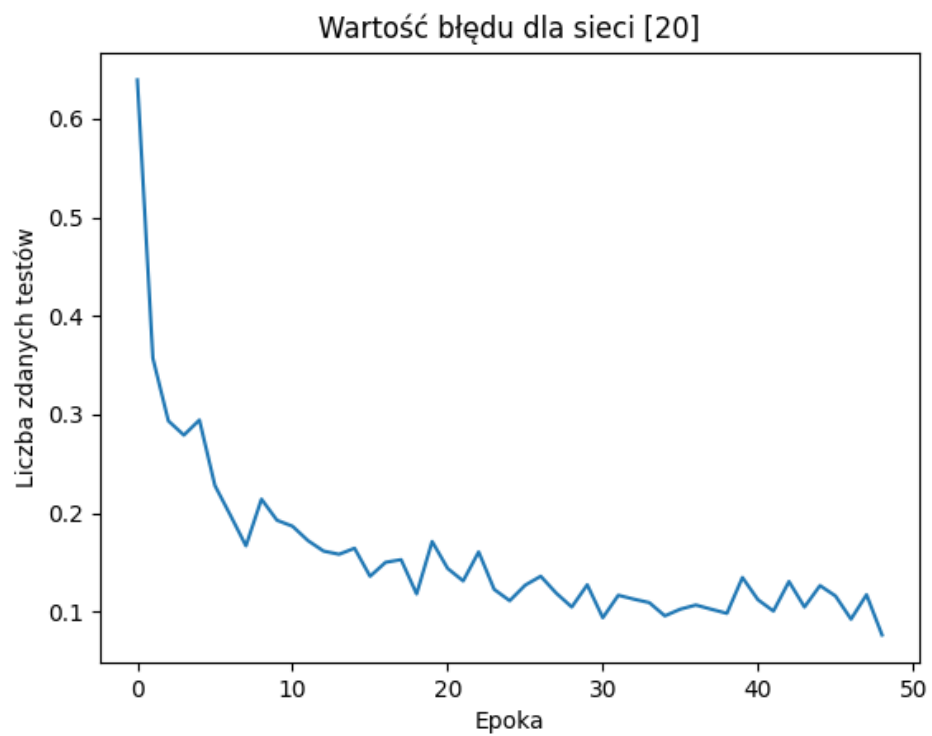
Zbadano jak wielkości sieci - liczba warstw wraz z liczbą neuronów w warstwach wpływa na skuteczność i szybkość uczenia. Zapis [5, 7] oznacza sieć neuronową z dwoma warstwami ukrytymi, z czego pierwsza zawiera 5 neuronów, a druga 7. Ilość neuronów na wyjściu wynosi zawsze 10. W tabelce zawarty jest uśredniony błąd oraz uśredniona liczba prawidłowo rozpoznanych liczb ze zbioru testowego. Niżej przedstawione są wyniki na wykresach.

Wielkość	Błąd	Skuteczność
[5]	0.478	8594
[10]	0.27	9176
[20]	0.15	9374
[30]	0.11	9496
[5, 5]	0.5	8548
[5, 10, 20]	0.45	8659

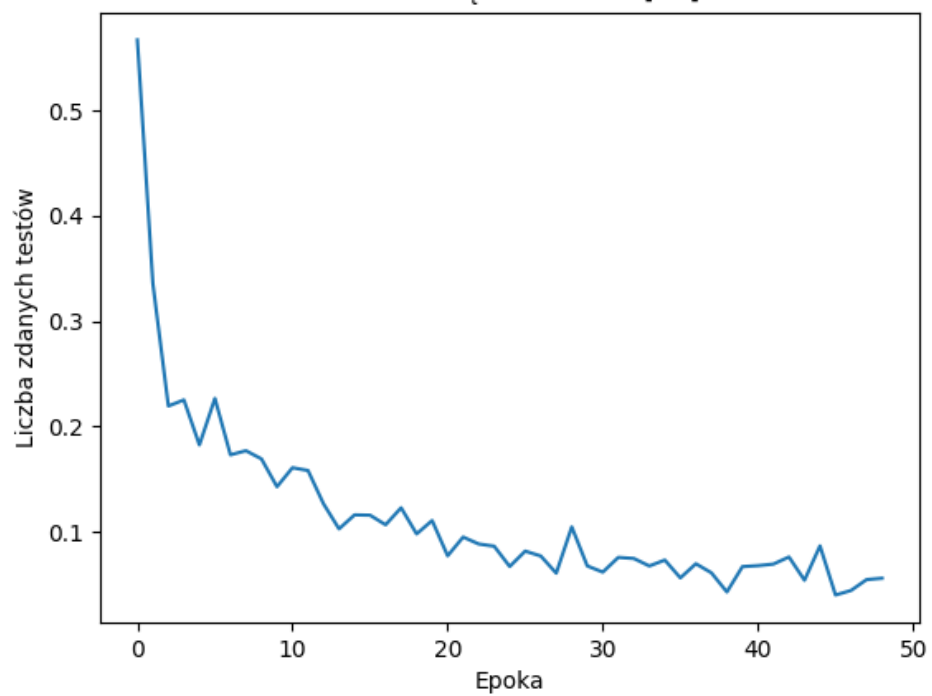
Tablica 1: Wielkość sieci



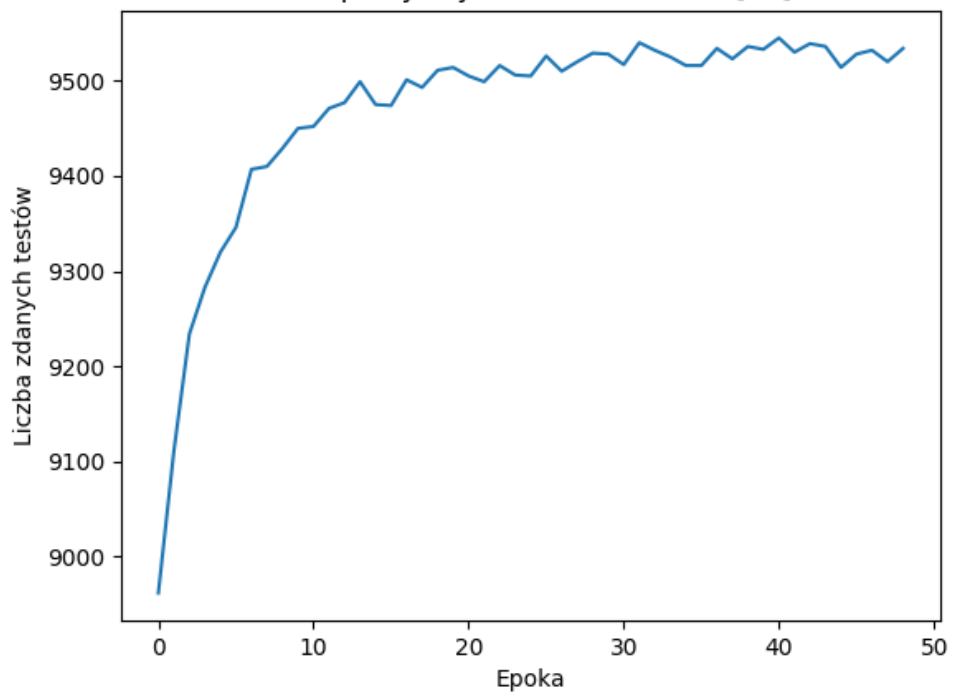


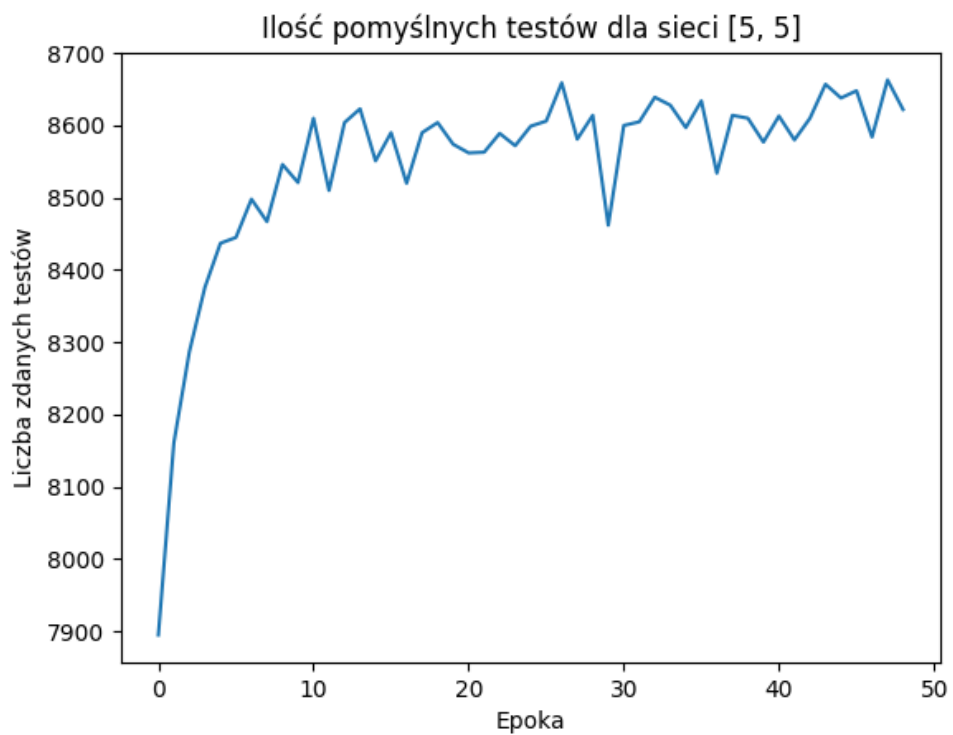
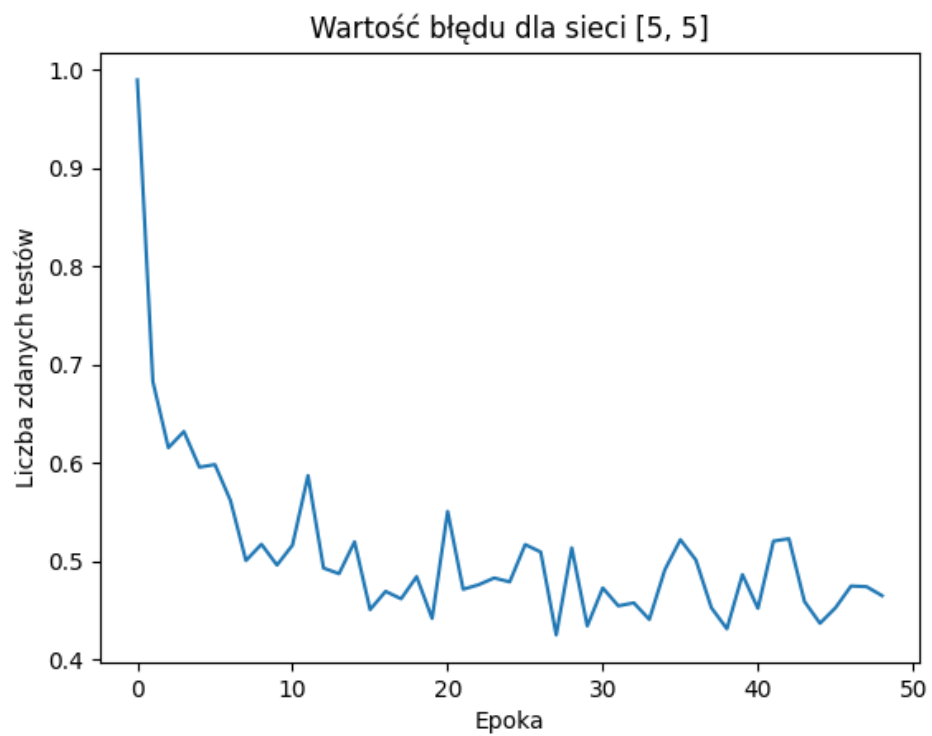


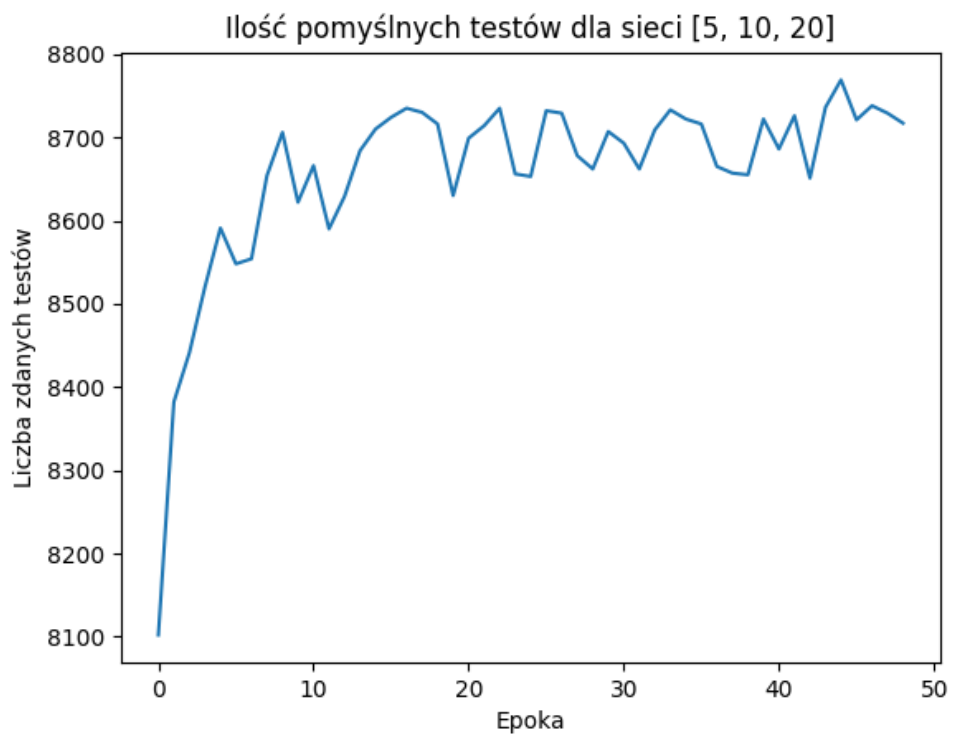
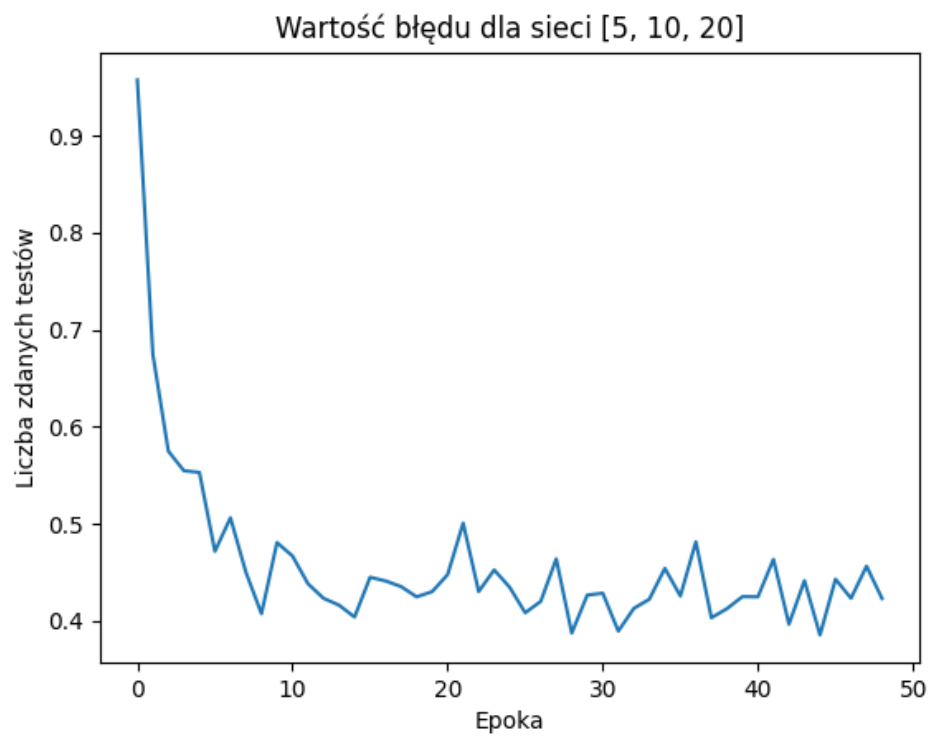
Wartość błędu dla sieci [30]



Ilość pomyślnych testów dla sieci [30]







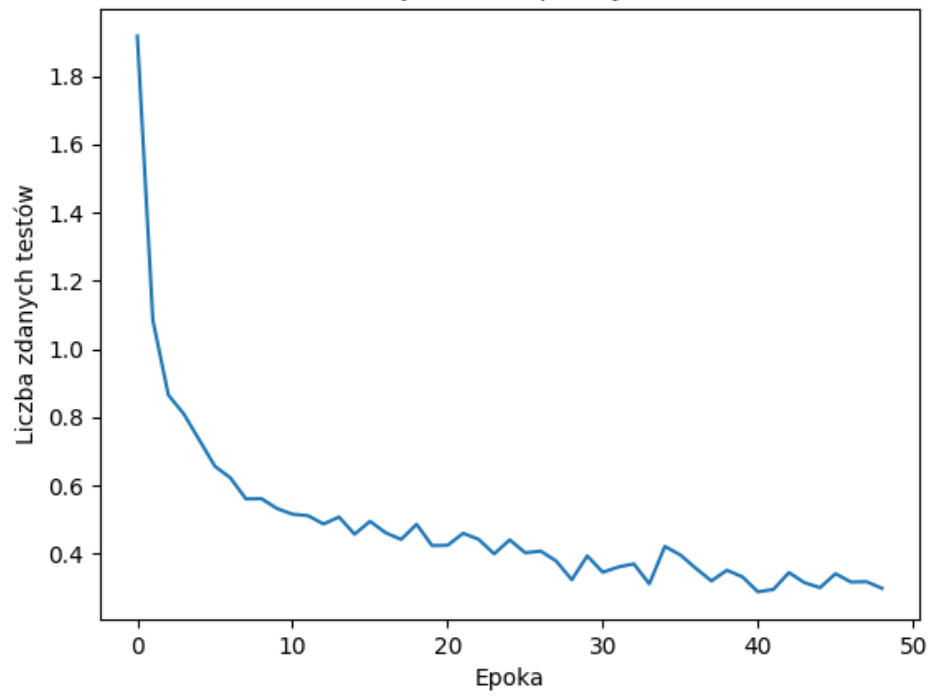
1.2 Test wielkości współczynnika uczenia

Zbadano jak współczynnik uczenia wpływa na spadek błędu i skuteczność sieci. W tabelce zawarty jest uśredniony błąd oraz uśredniona liczba prawidłowo rozpoznanych liczb ze zbioru testowego. Niżej przedstawione są wyniki na wykresach.

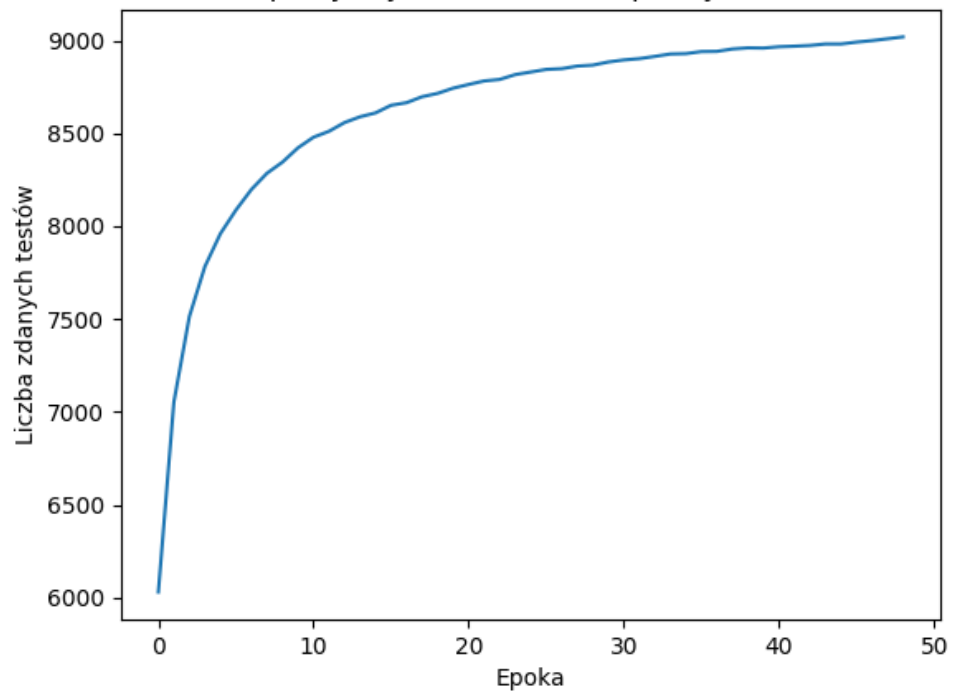
Współczynnik uczenia	Błąd	Skuteczność
0.01	0.478	8594
0.05	0.27	9176
0.1	0.2	9309
0.2	0.17	9320
0.3	0.15	9401
0.5	0.13	9441
0.75	0.15	9395
1	0.16	9353

Tablica 2: Współczynnik uczenia

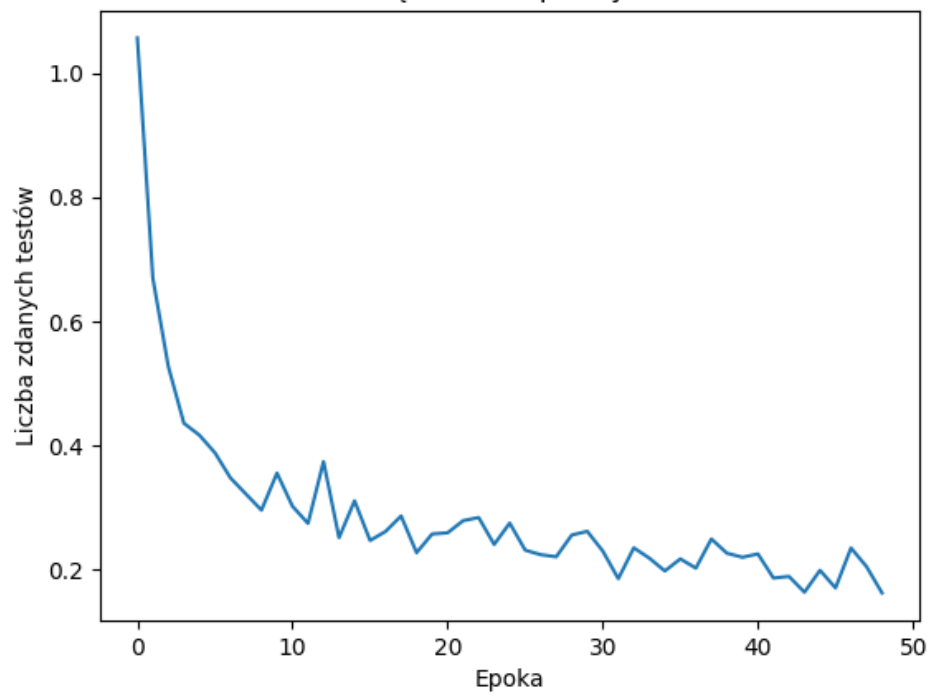
Wartość błędu dla współczynnika 0.01



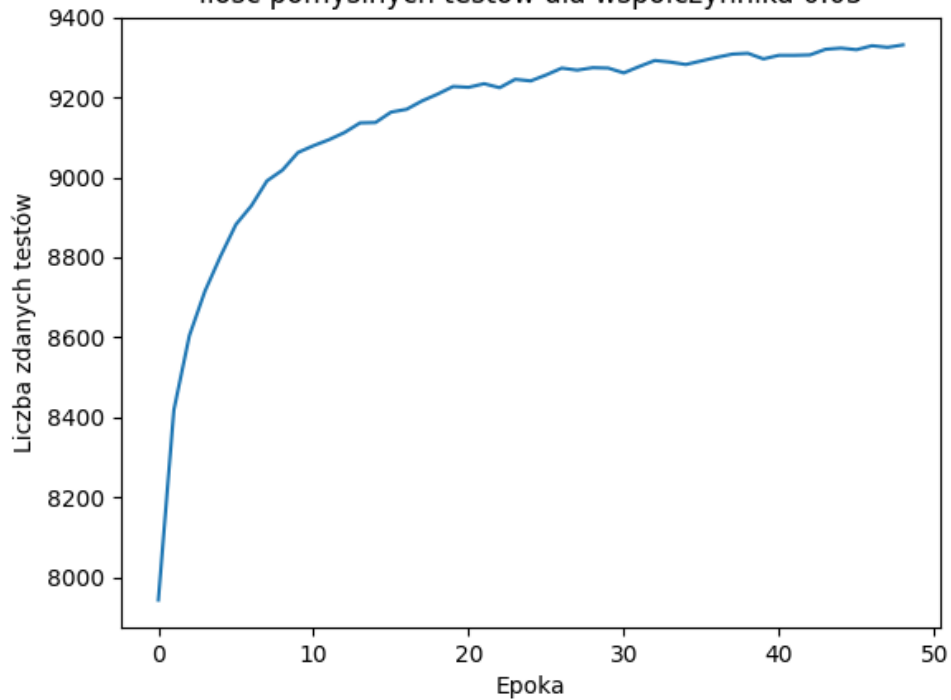
Ilość pomyślnych testów dla współczynnika 0.01

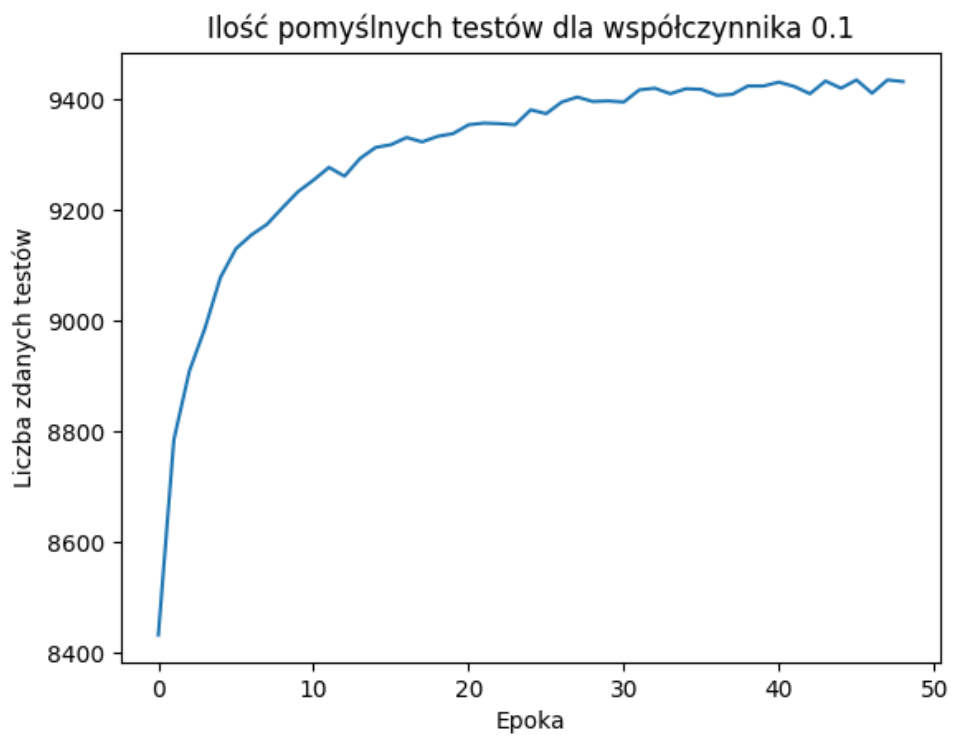


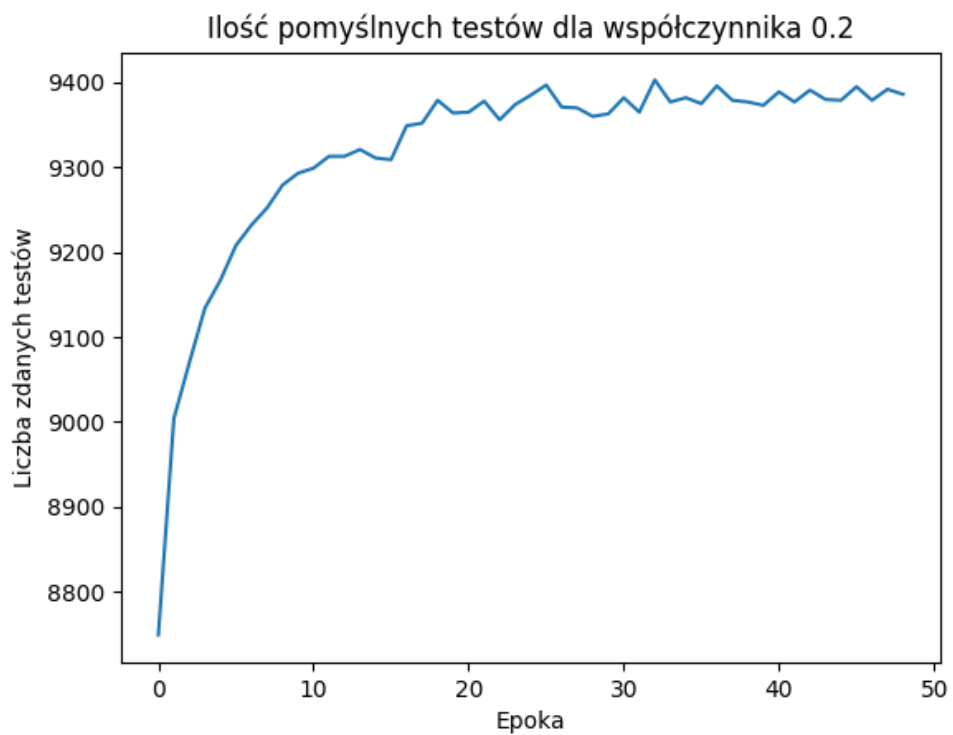
Wartość błędu dla współczynnika 0.05

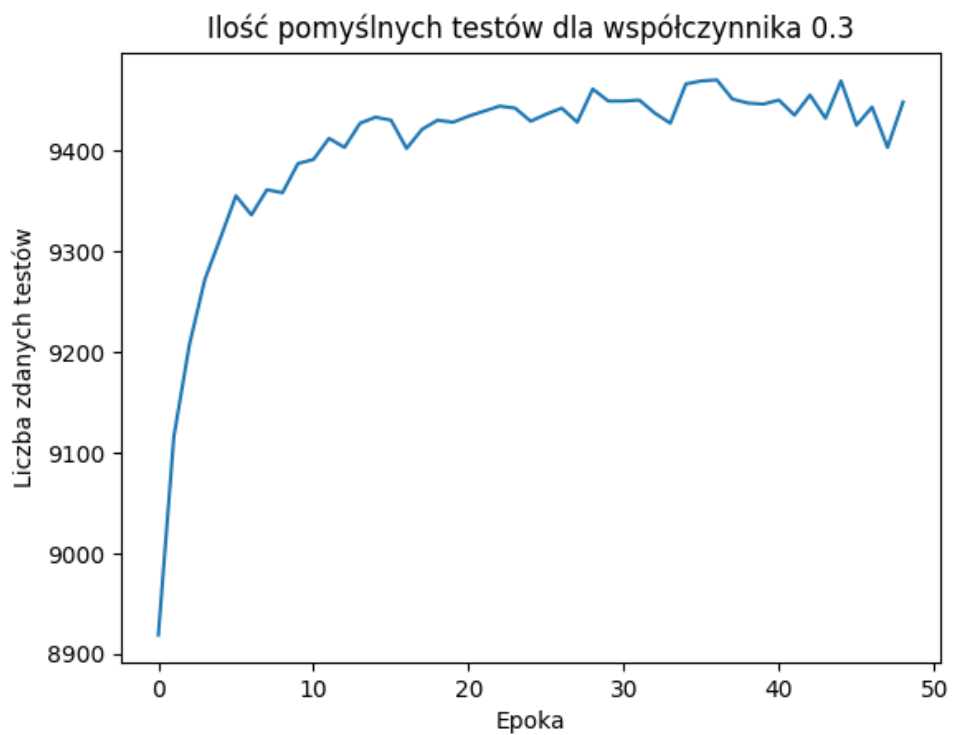


Ilość pomyślnych testów dla współczynnika 0.05

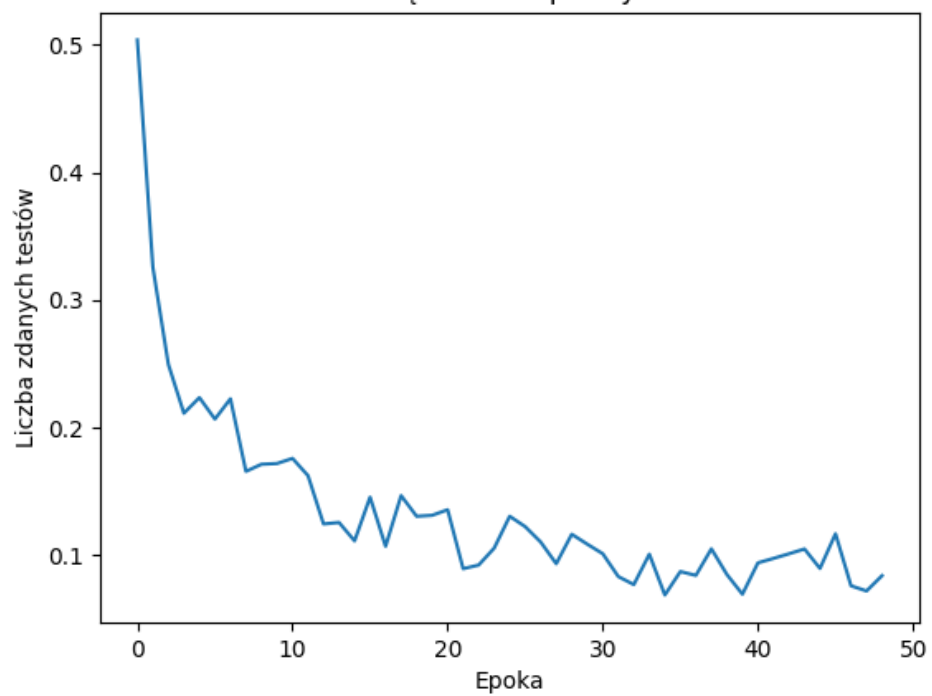




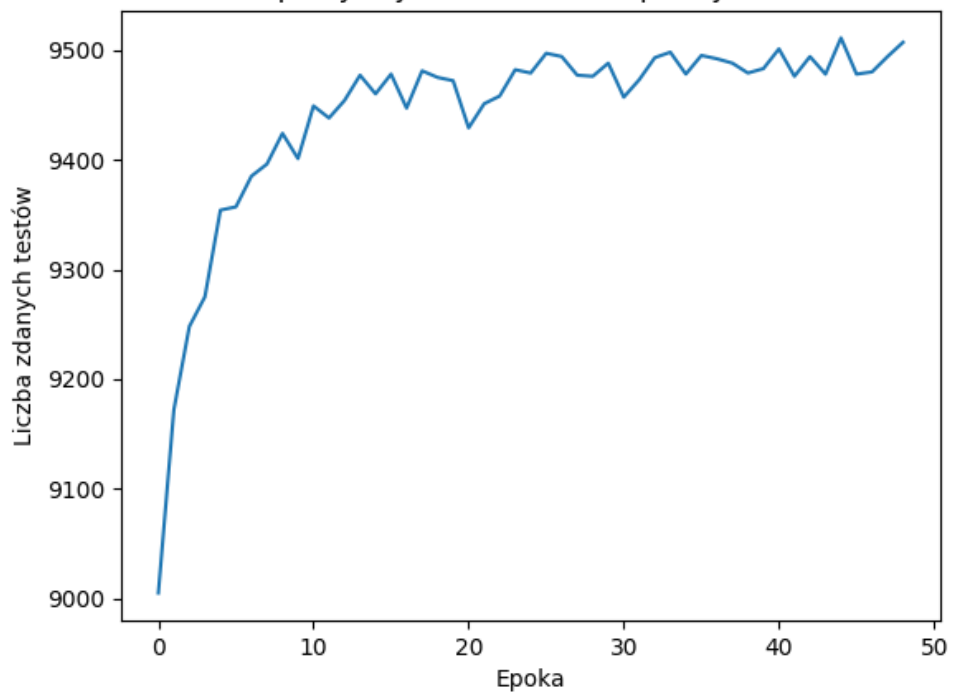


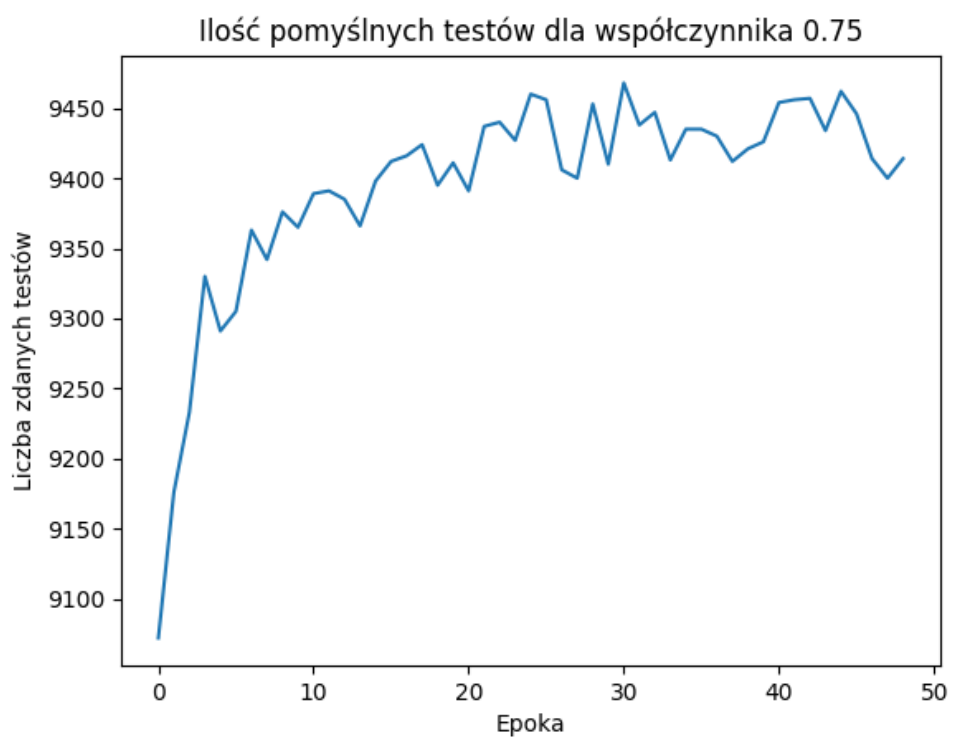
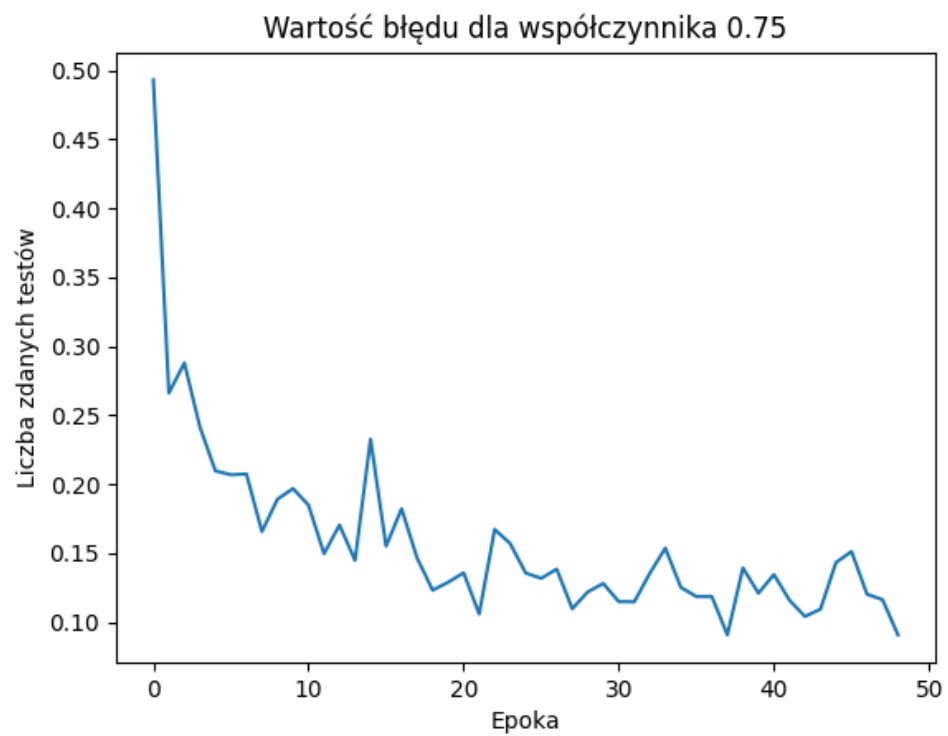


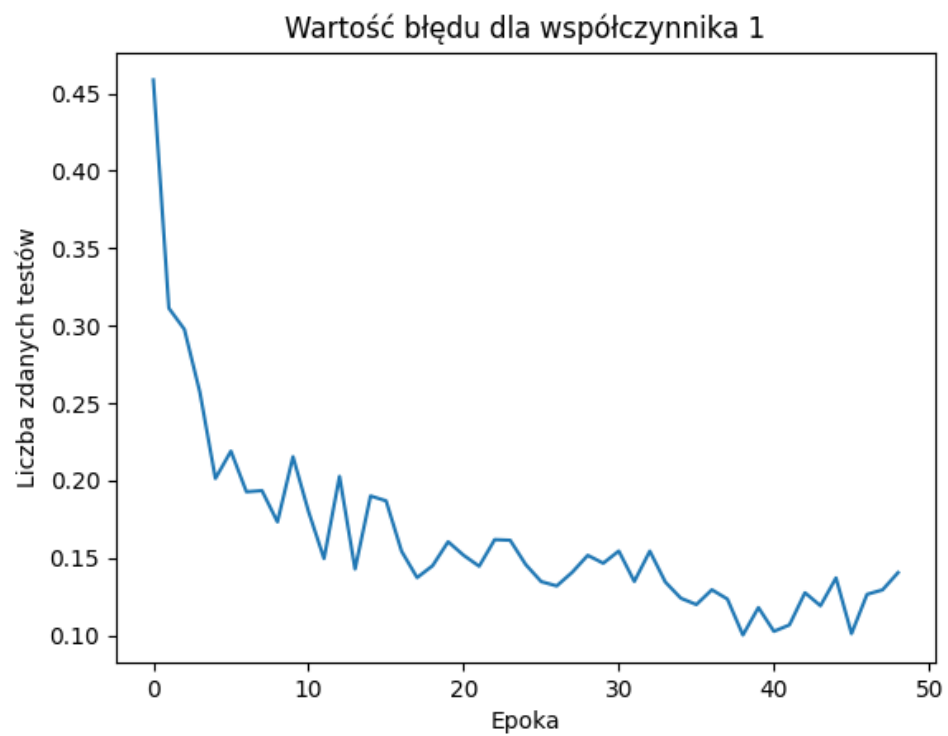
Wartość błędu dla współczynnika 0.5



Ilość pomyślnych testów dla współczynnika 0.5





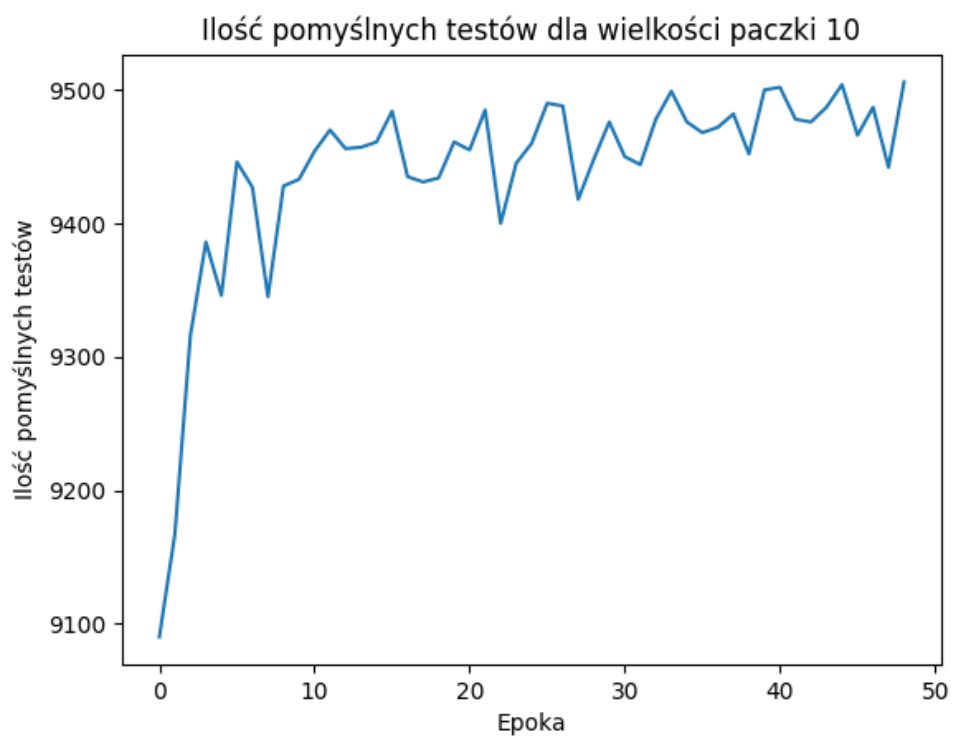


1.3 Test wielkości paczki

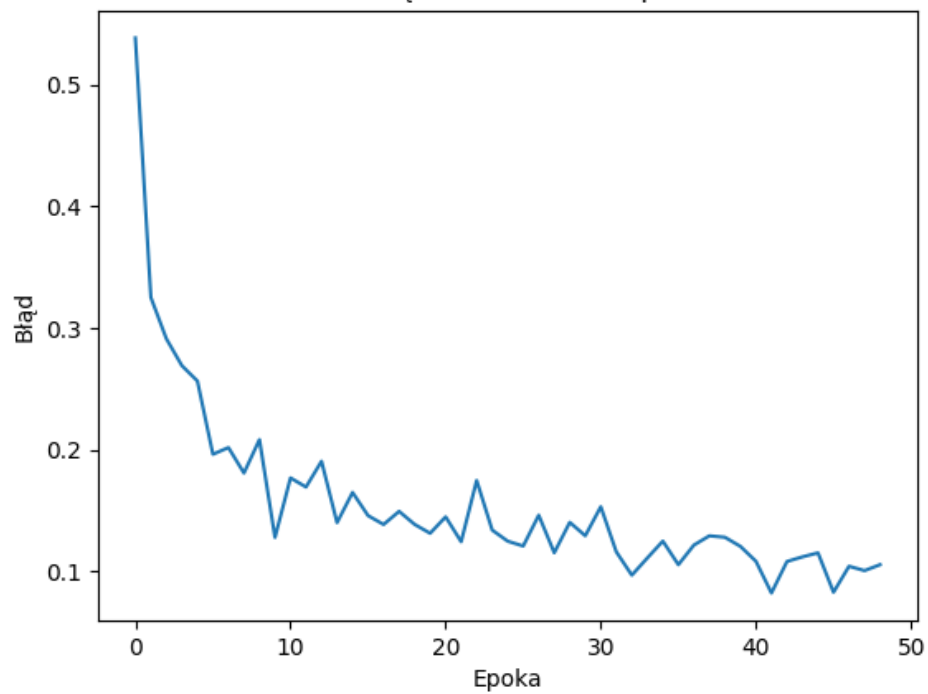
Zbadano jak wielkość paczki wpływa na spadek błędu i skuteczność sieci. W tabelce zawarty jest uśredniony błąd oraz uśredniona liczba prawidłowo rozpoznanych liczb ze zbioru testowego. Niżej przedstawione są wyniki na wykresach.

Wielkość paczki	Błąd	Skuteczność
10	0.15	9440
50	0.15	9400
100	0.19	9312
500	0.31	9053
1000	0.39	8893
10000	0.39	8884
60000	0.38	8902

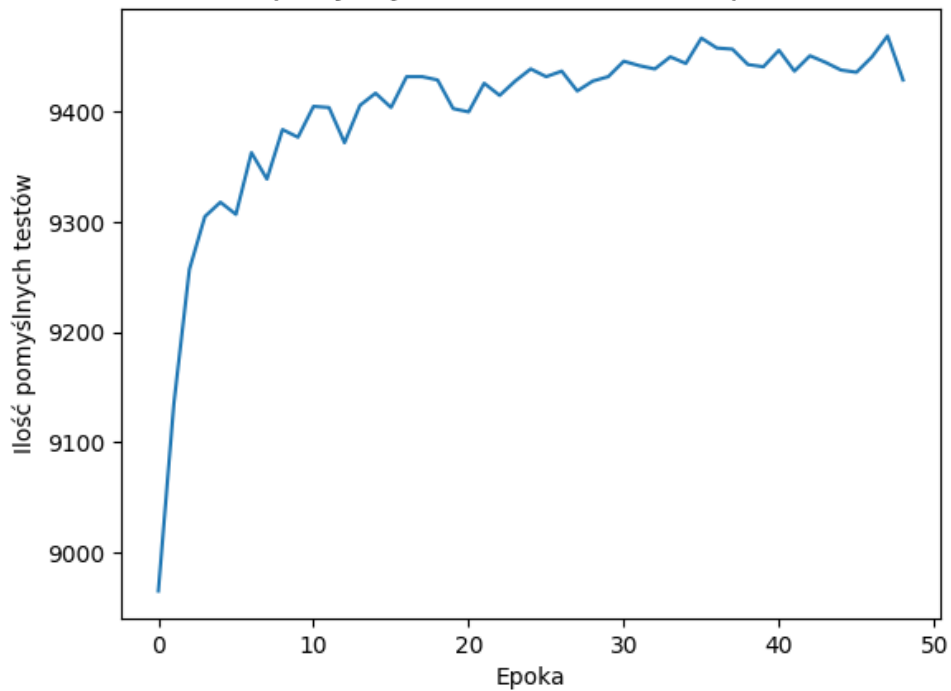
Tablica 3: Wielkość paczki

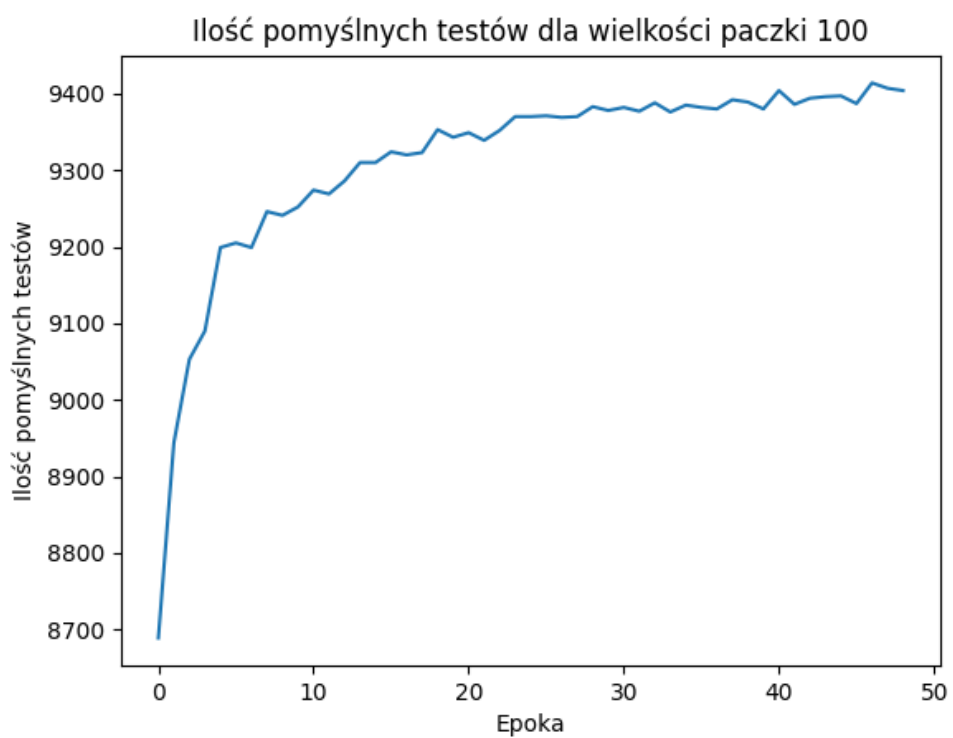
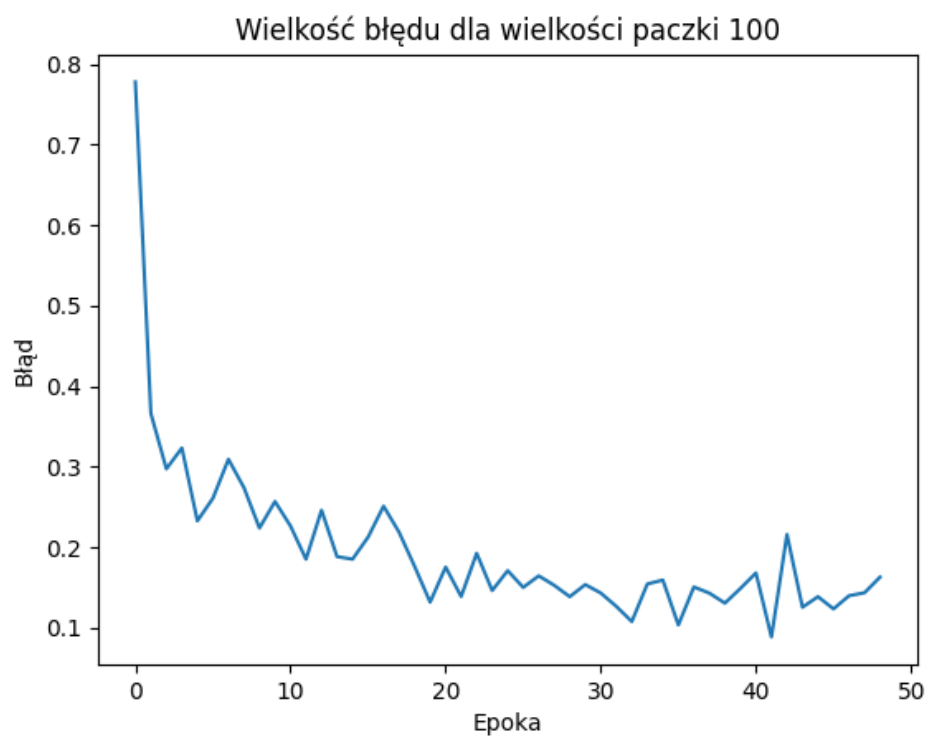


Wielkość błędu dla wielkości paczki 50

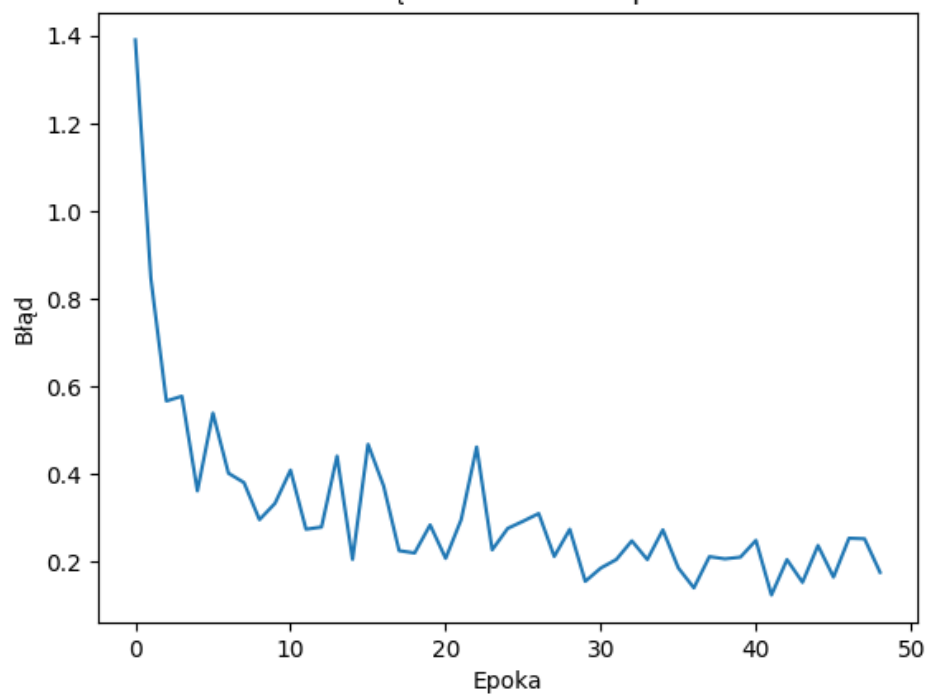


Ilość pomyślnych testów dla wielkości paczki 50

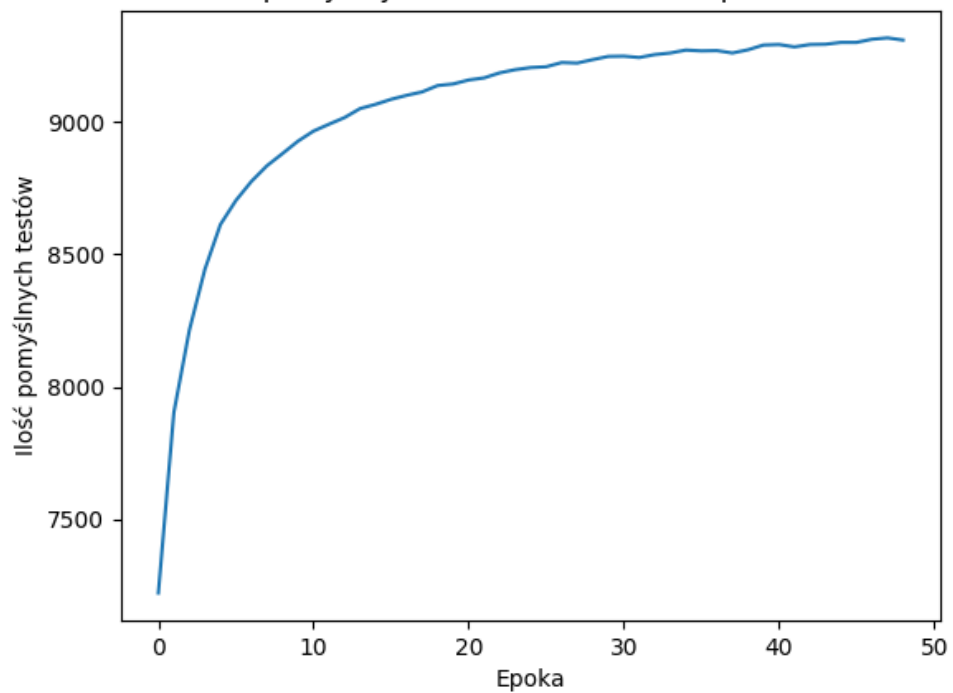


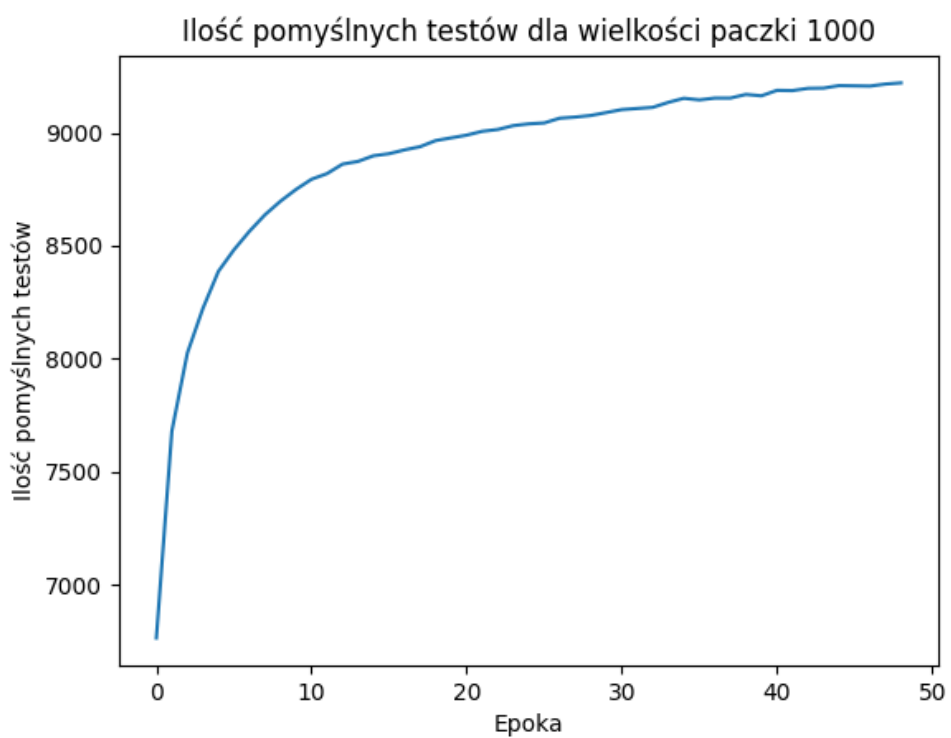


Wielkość błędu dla wielkości paczki 500

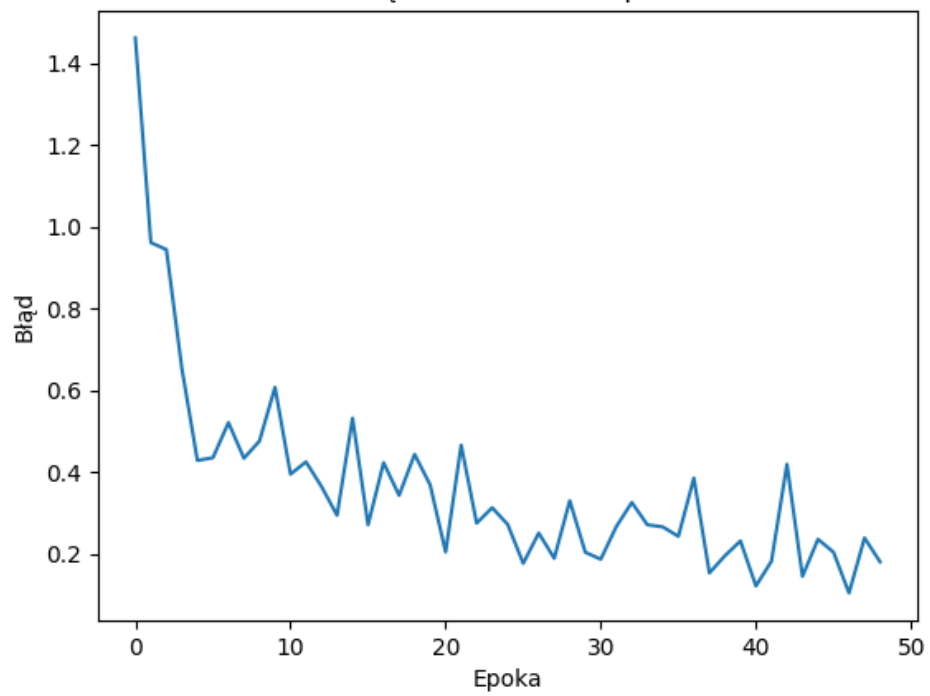


Ilość pomyślnych testów dla wielkości paczki 500

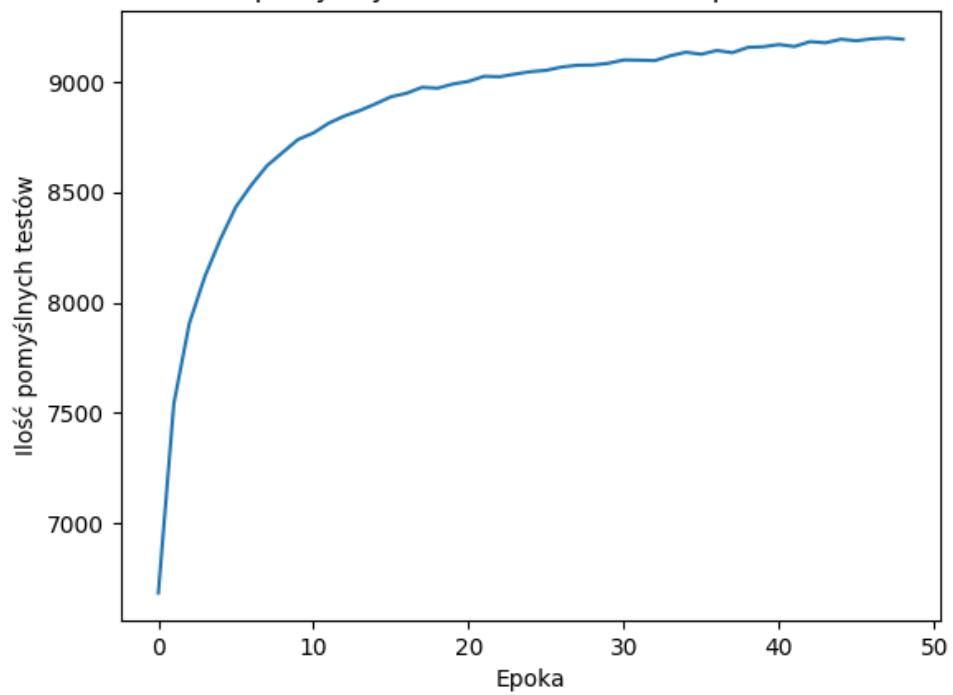




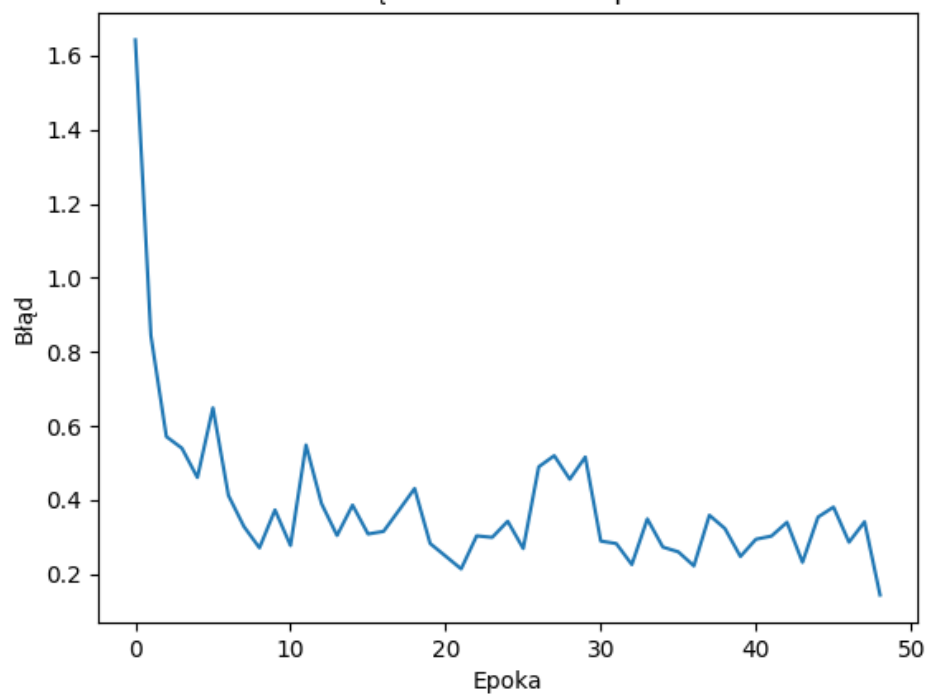
Wielkość błędu dla wielkości paczki 2000



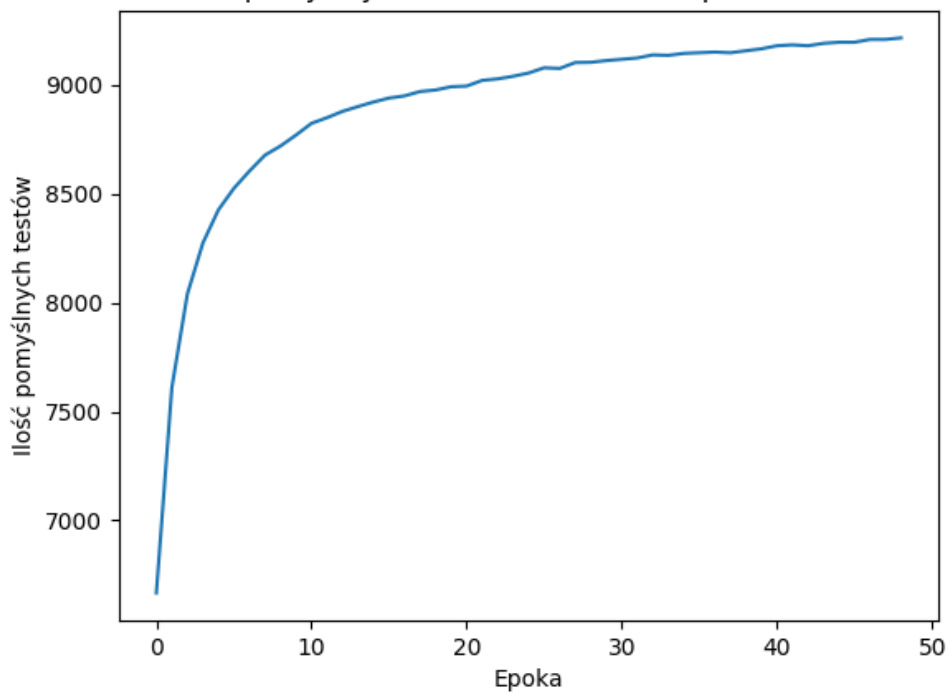
Ilość pomyślnych testów dla wielkości paczki 2000



Wielkość błędu dla wielkości paczki 60000



Ilość pomyślnych testów dla wielkości paczki 60000



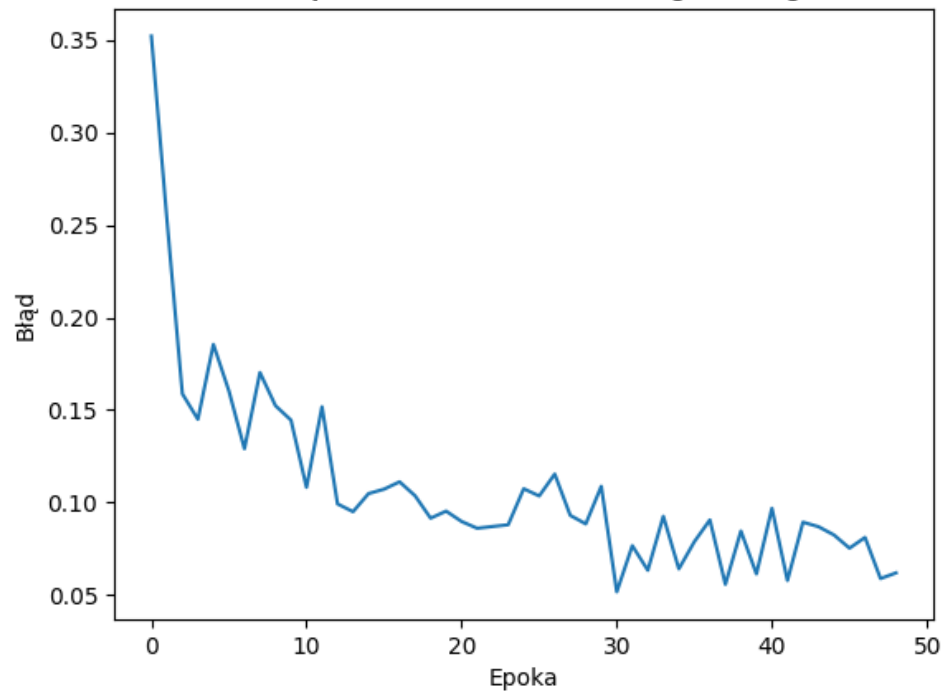
1.4 Test inicjalizacji wartości wag

Zbadano jak sposób inicjalizacji wag wpływa na spadek błędu i skuteczność sieci. Do inicjalizacji wag wykorzystano rozkład normalny. Mediana ustawiona jest zawsze na zero, a badany jest wpływ zmiany odchylenia standardowego. W tabelce zawarty jest uśredniony błąd oraz uśredniona liczba prawidłowo rozpoznanych liczb ze zbioru testowego. Niżej przedstawione są wyniki na wykresach.

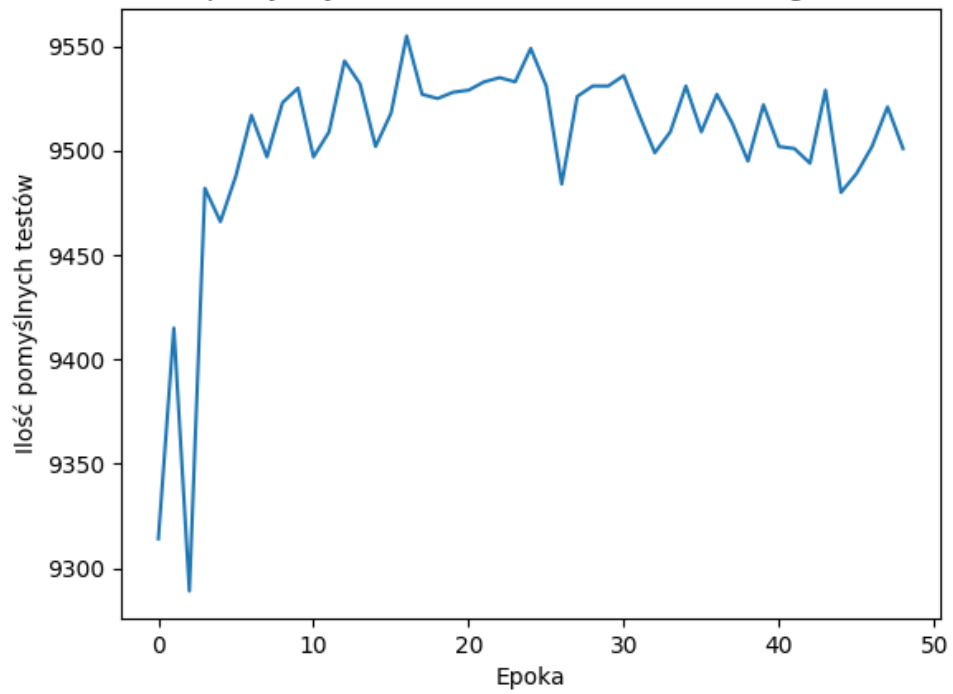
Sigma	Błąd	Skuteczność
0.1	0.1	9504
0.2	0.11	9466
0.3	0.13	9441
0.4	0.13	9411
0.5	0.17	9332
0.6	0.18	9312
0.7	0.2	9281
0.8	0.22	9271
0.9	0.24	9155
1	0.27	9146

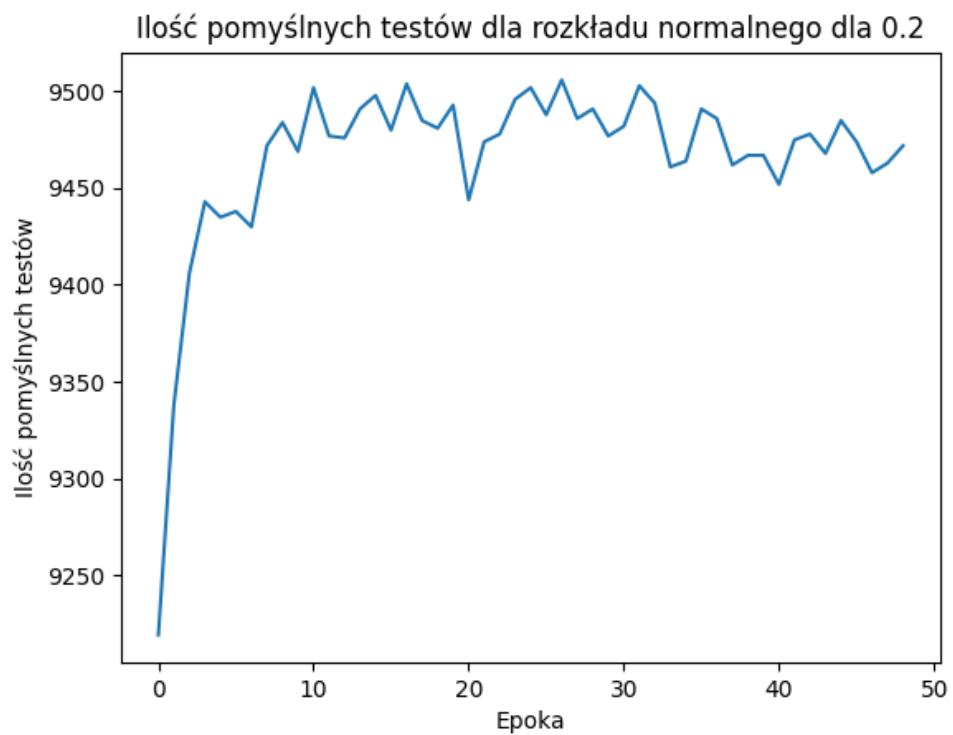
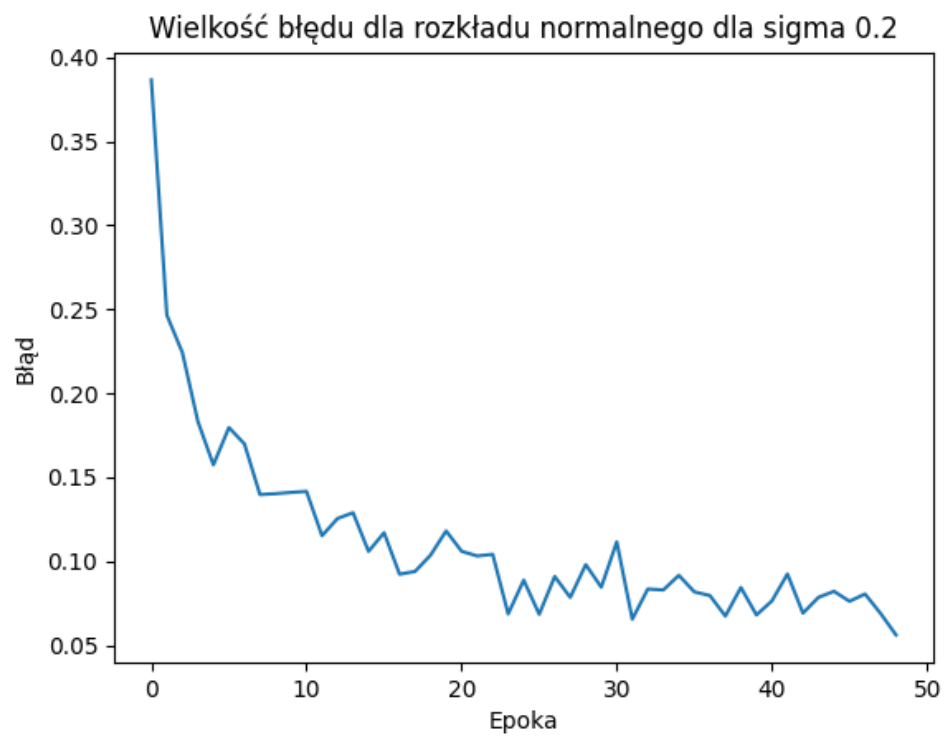
Tablica 4: Wielkość odchylenia standardowego

Wielkość błędu dla rozkładu normalnego dla sigma 0.1

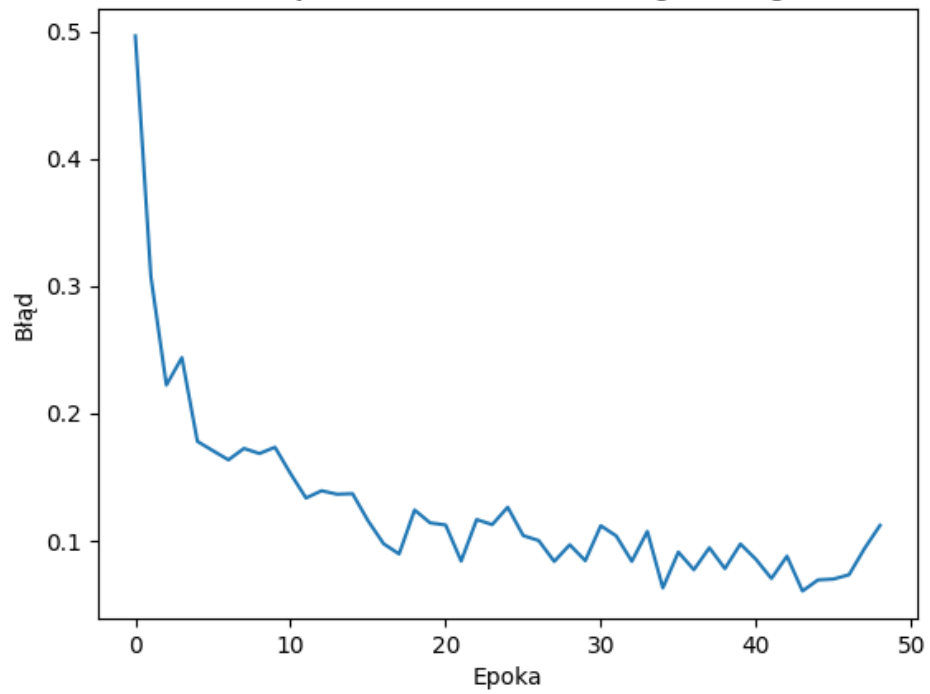


Ilość pomyślnych testów dla rozkładu normalnego dla 0.1

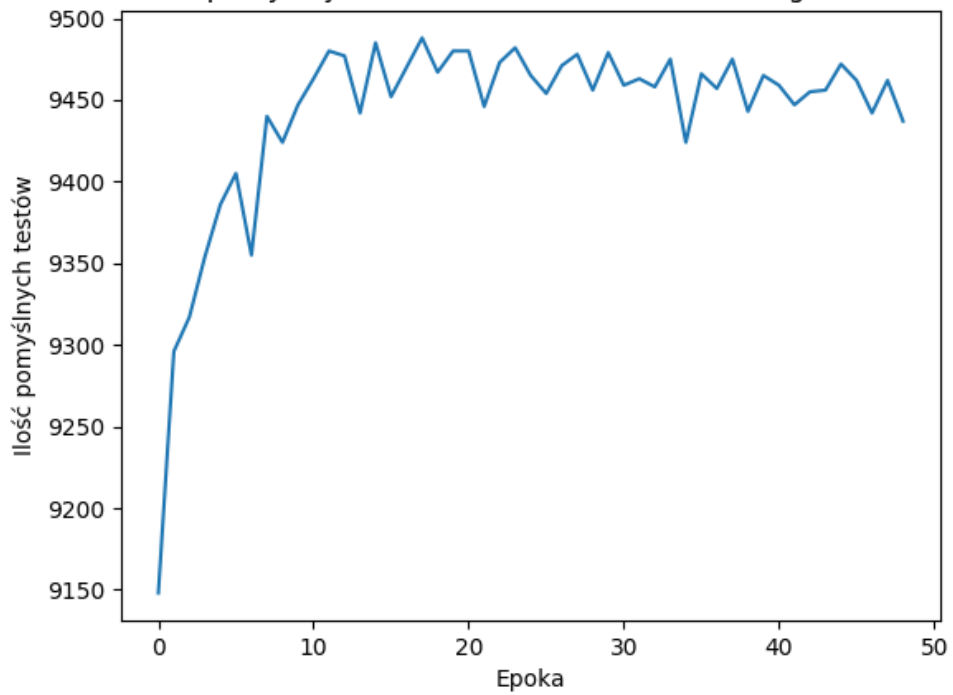




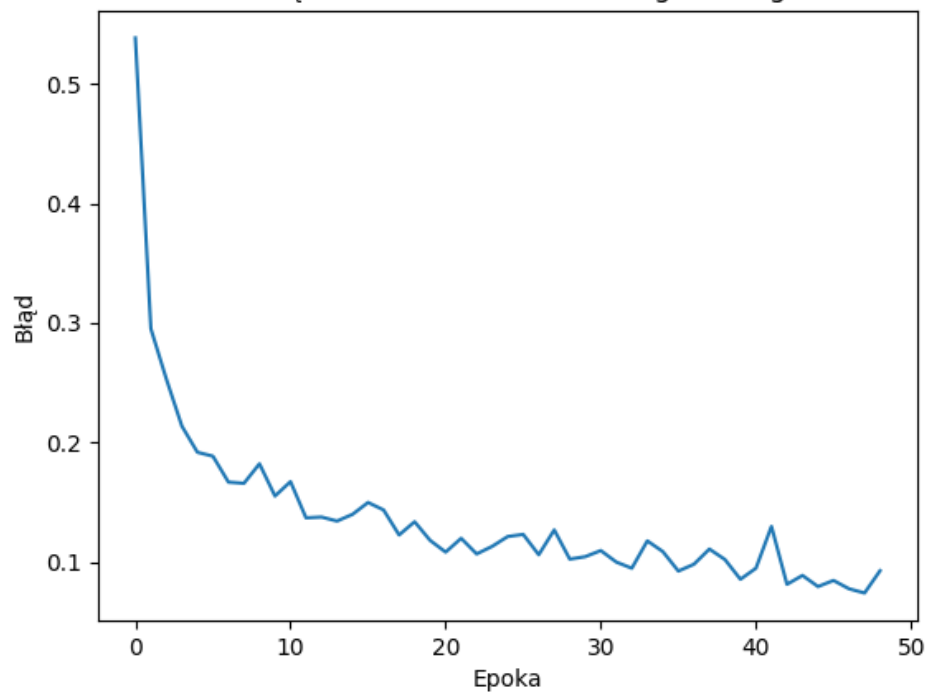
Wielkość błędu dla rozkładu normalnego dla sigma 0.3



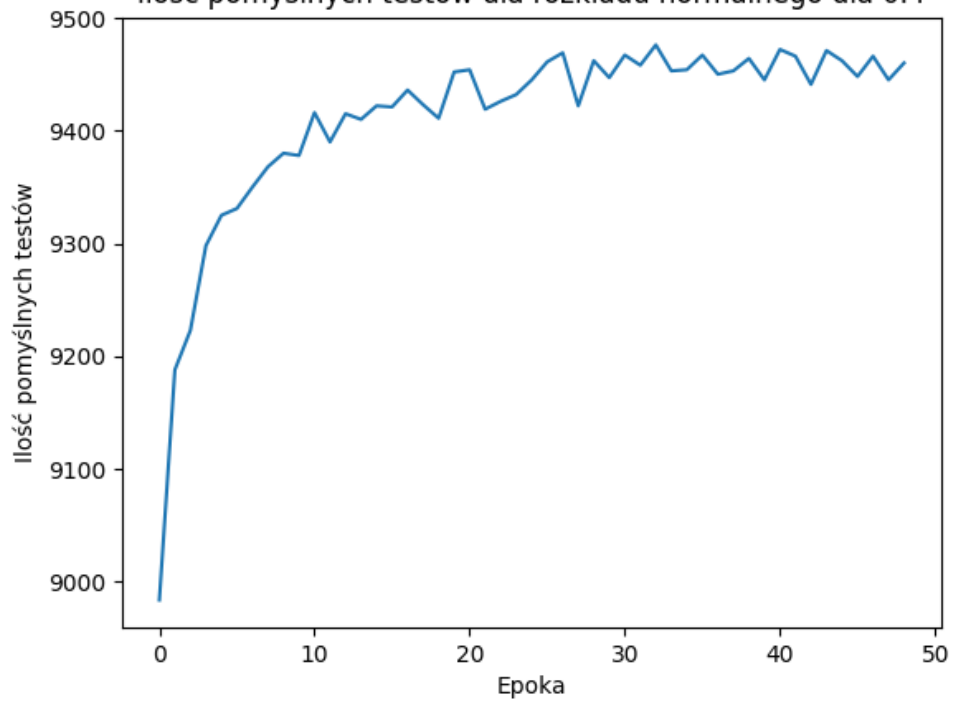
Ilość pomyślnych testów dla rozkładu normalnego dla 0.3



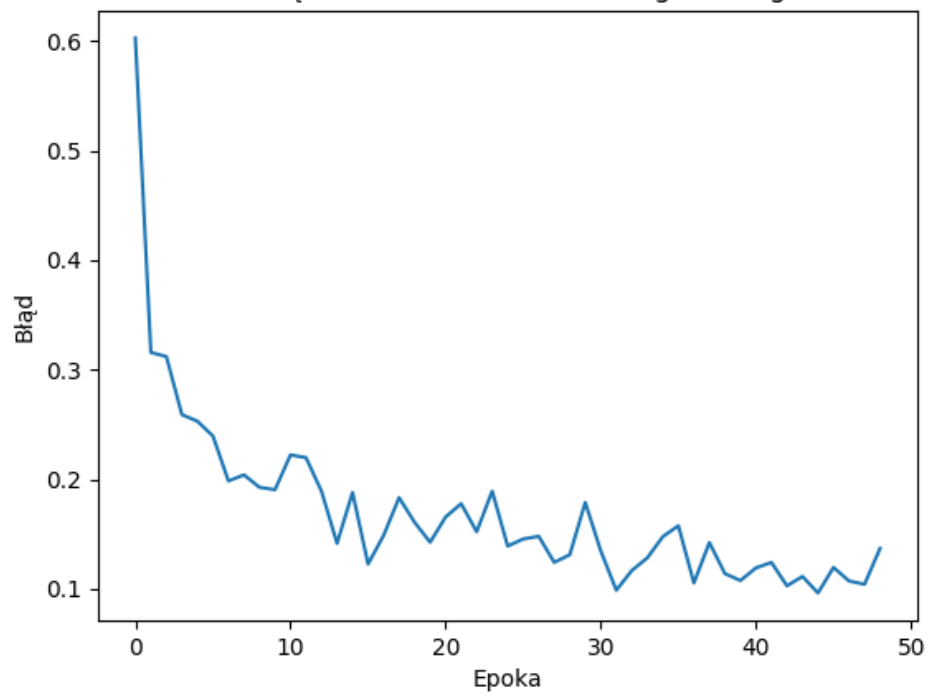
Wielkość błędu dla rozkładu normalnego dla sigma 0.4



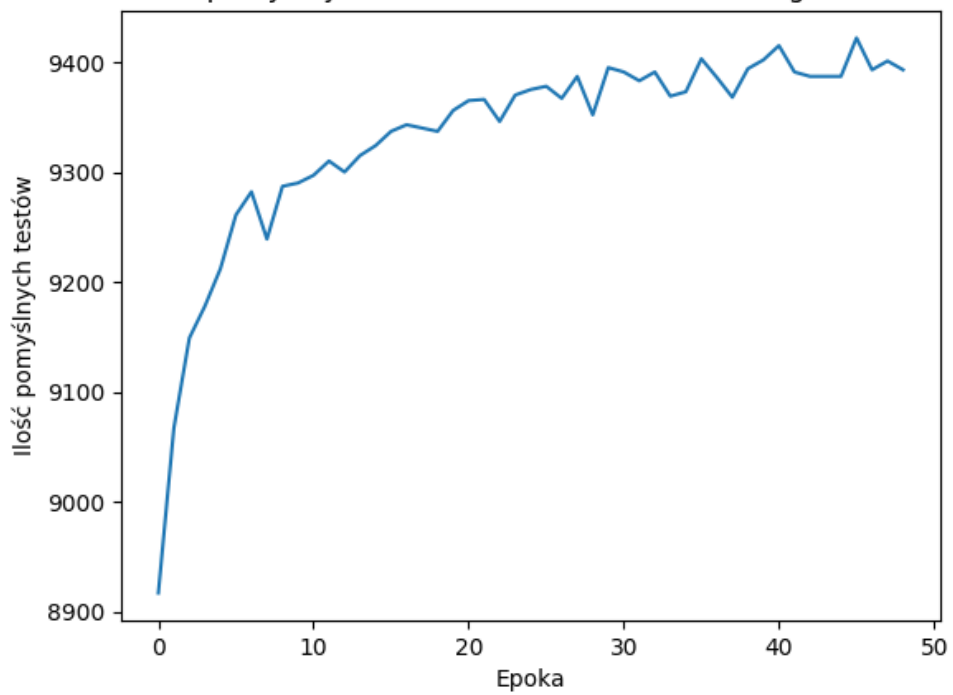
Ilość pomyślnych testów dla rozkładu normalnego dla 0.4



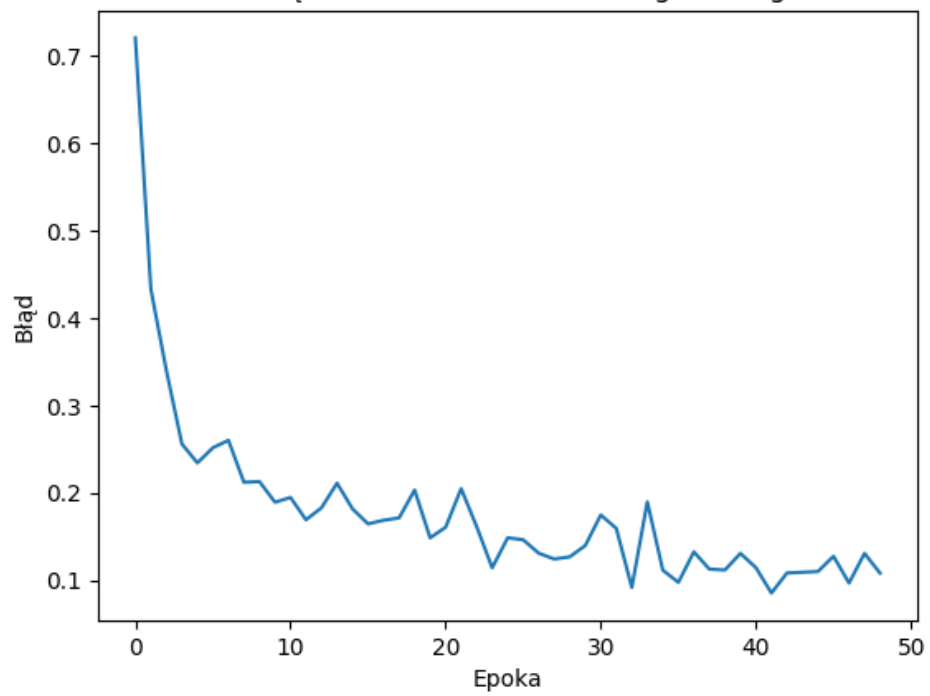
Wielkość błędu dla rozkładu normalnego dla sigma 0.5



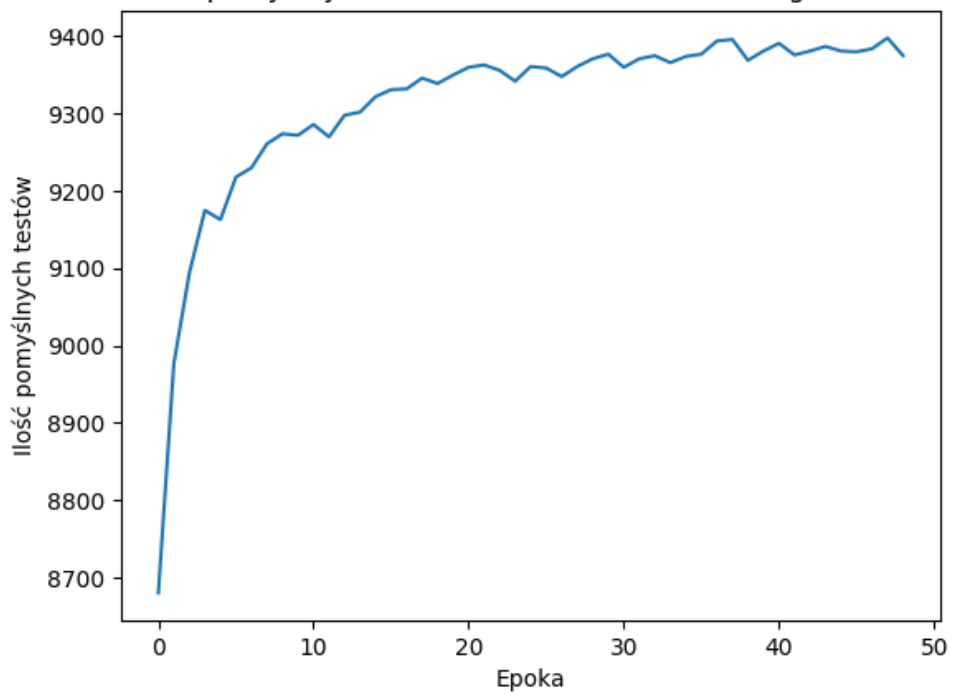
Ilość pomyślnych testów dla rozkładu normalnego dla 0.5



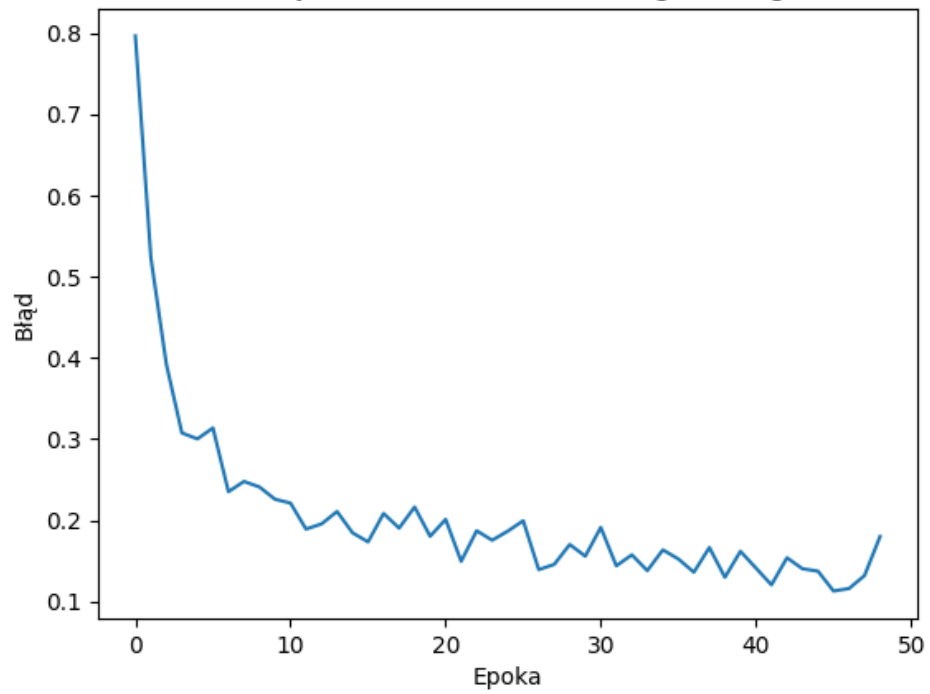
Wielkość błędu dla rozkładu normalnego dla sigma 0.6



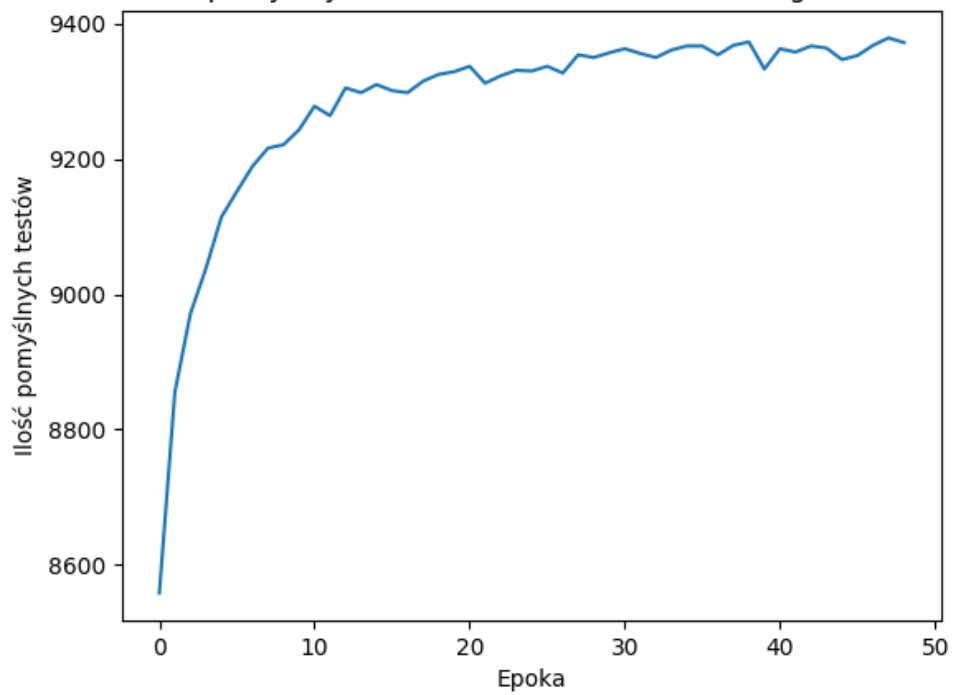
Ilość pomyślnych testów dla rozkładu normalnego dla 0.6



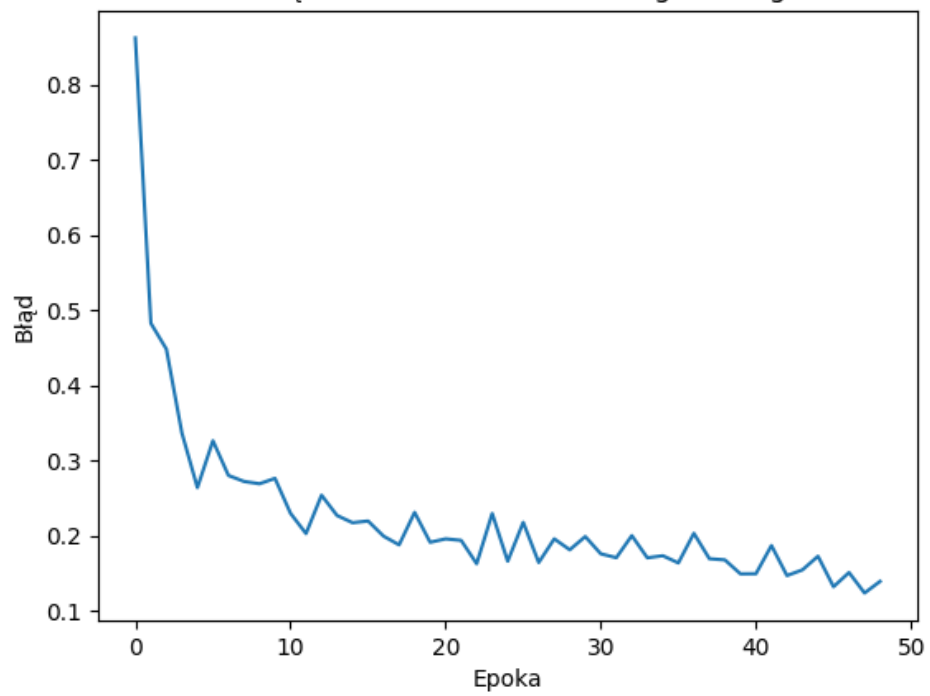
Wielkość błędu dla rozkładu normalnego dla sigma 0.7



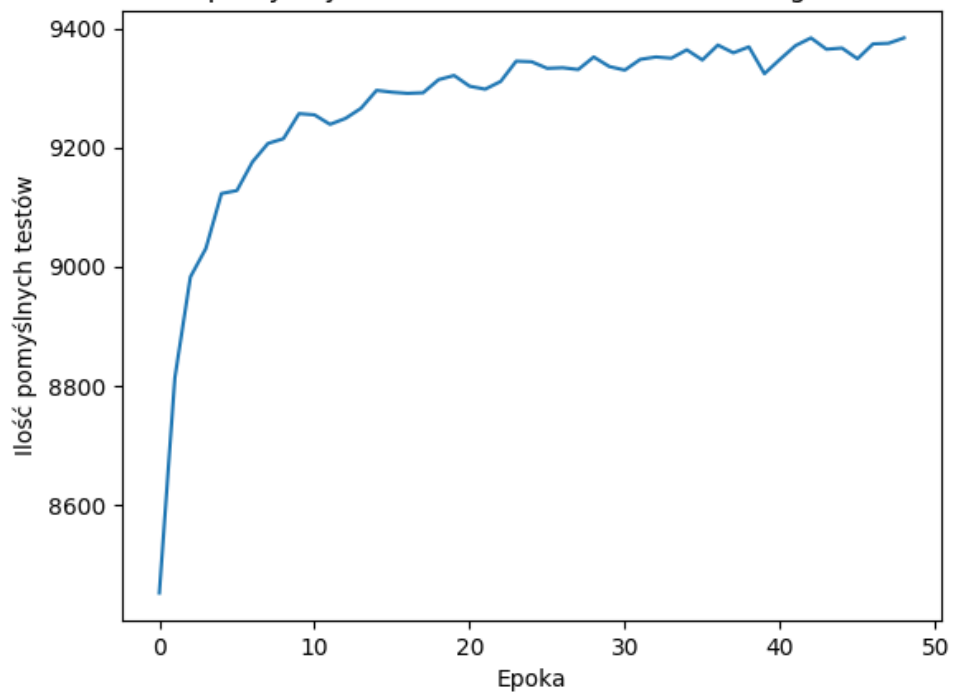
Ilość pomyślnych testów dla rozkładu normalnego dla 0.7



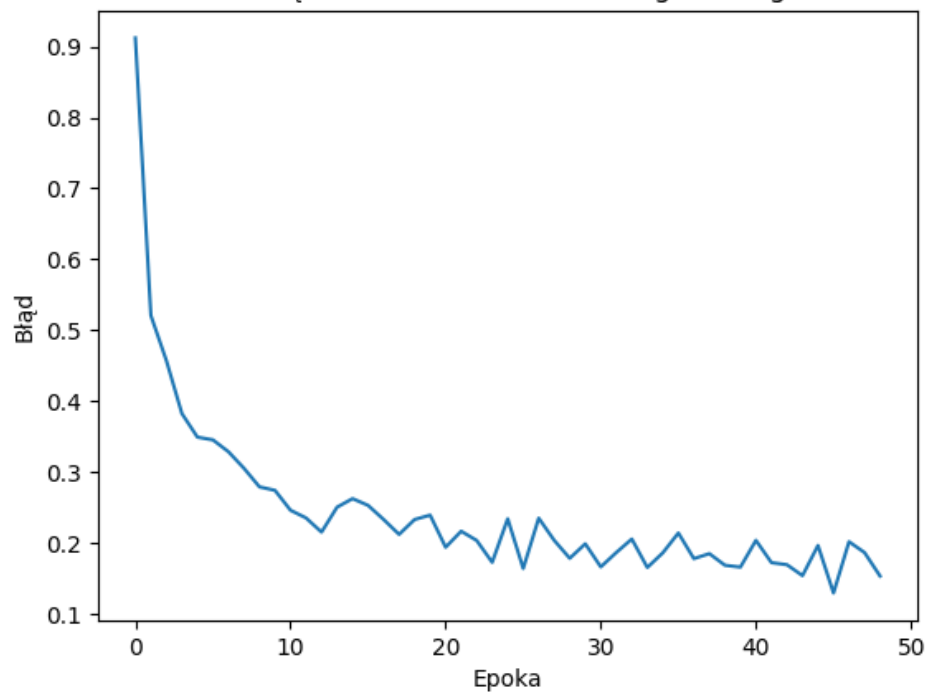
Wielkość błędu dla rozkładu normalnego dla sigma 0.8



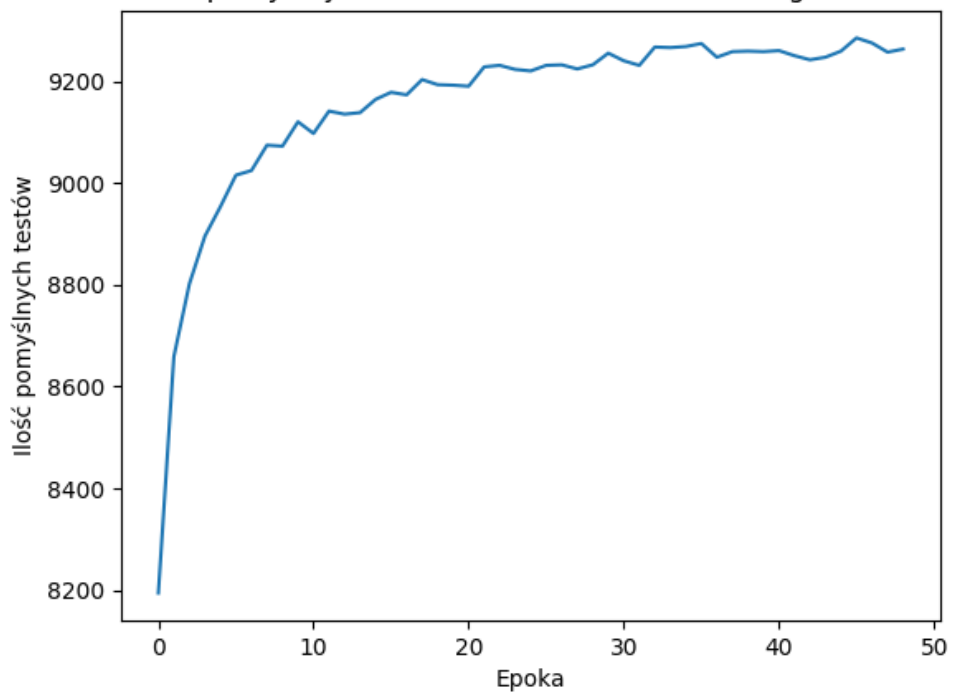
Ilość pomyślnych testów dla rozkładu normalnego dla 0.8



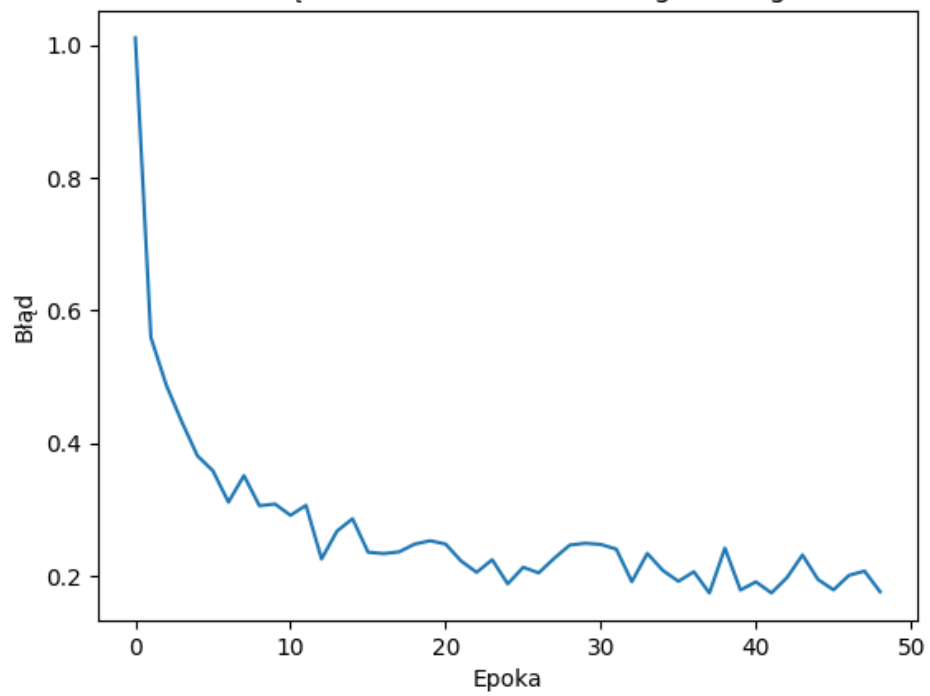
Wielkość błędu dla rozkładu normalnego dla sigma 0.9



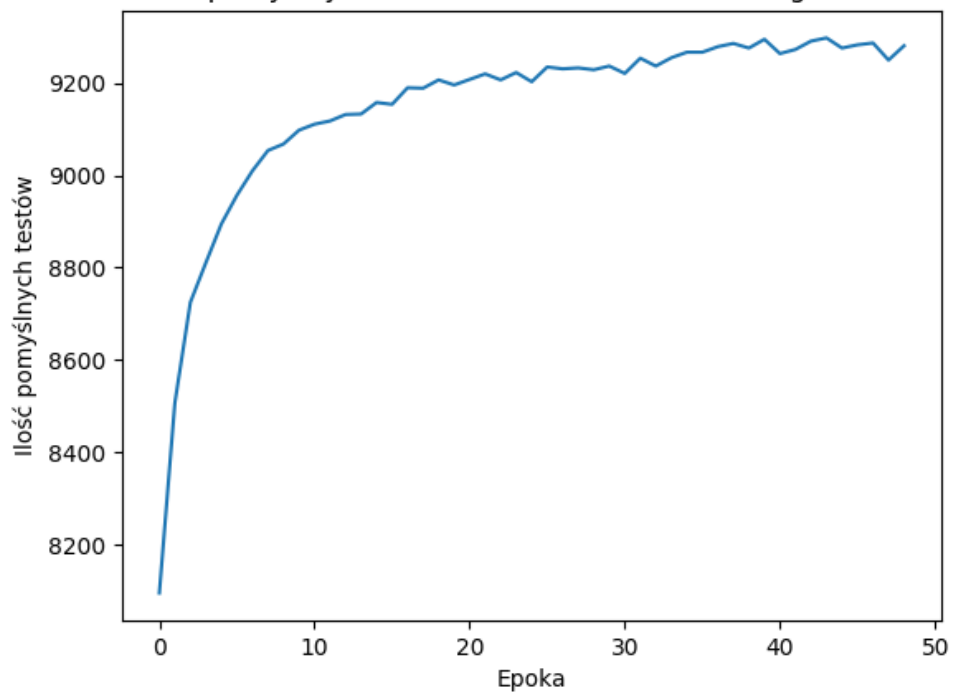
Ilość pomyślnych testów dla rozkładu normalnego dla 0.9



Wielkość błędu dla rozkładu normalnego dla sigma 1.0



Ilość pomyślnych testów dla rozkładu normalnego dla 1.0

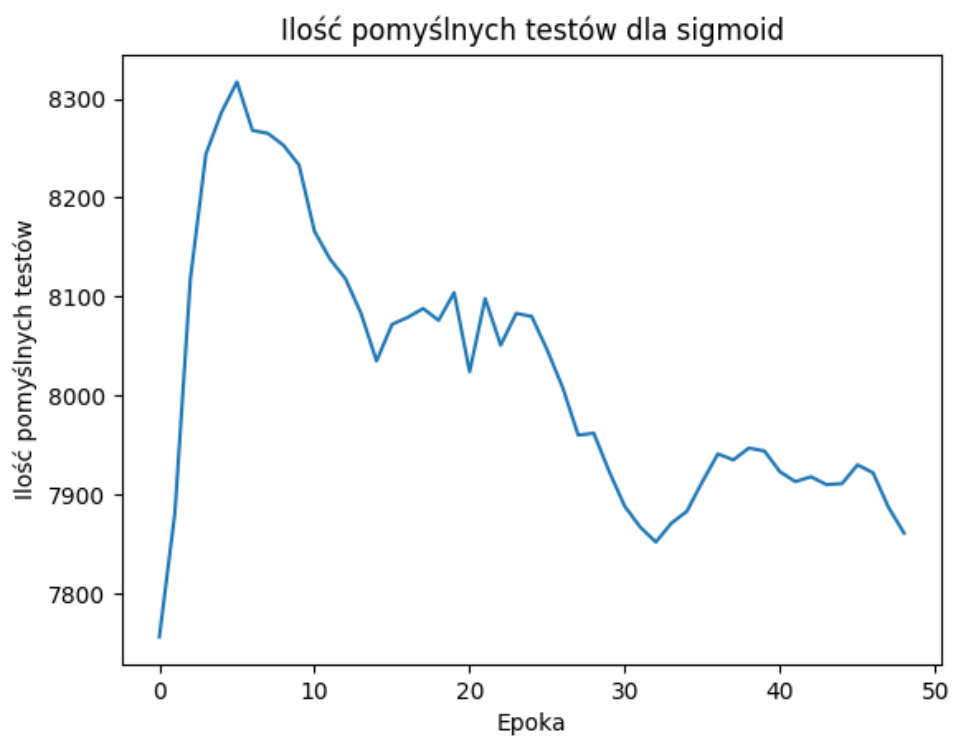
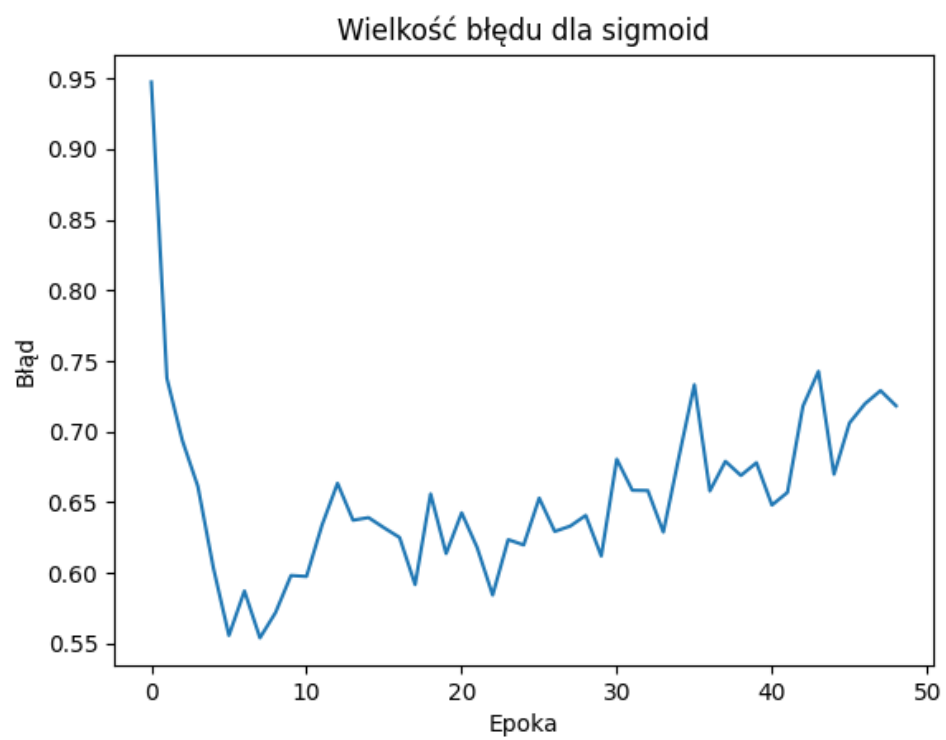


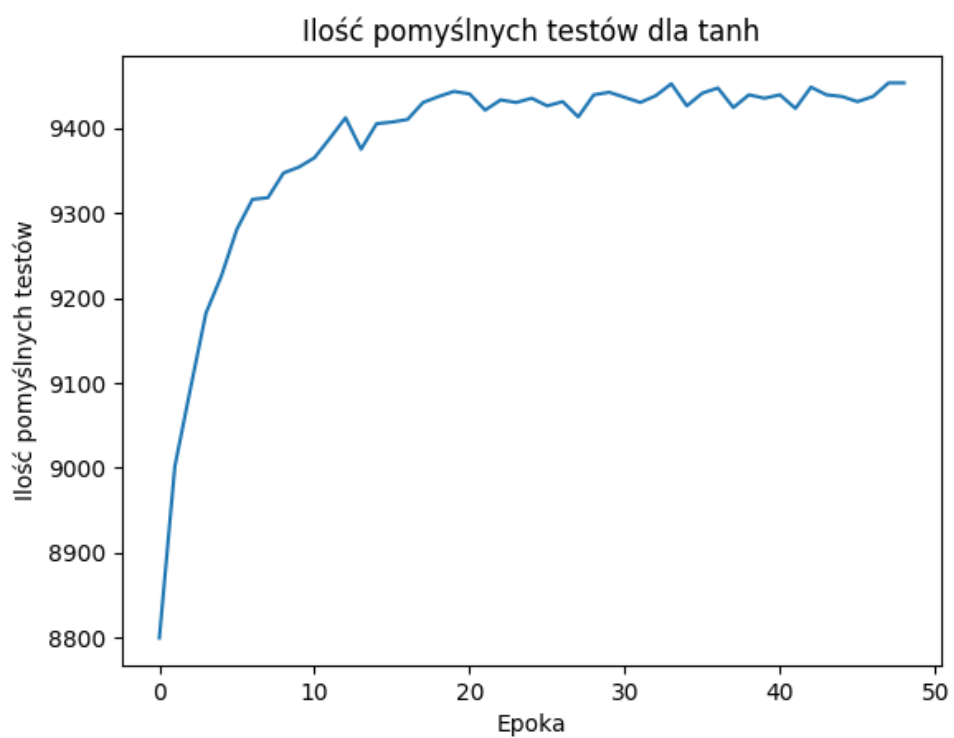
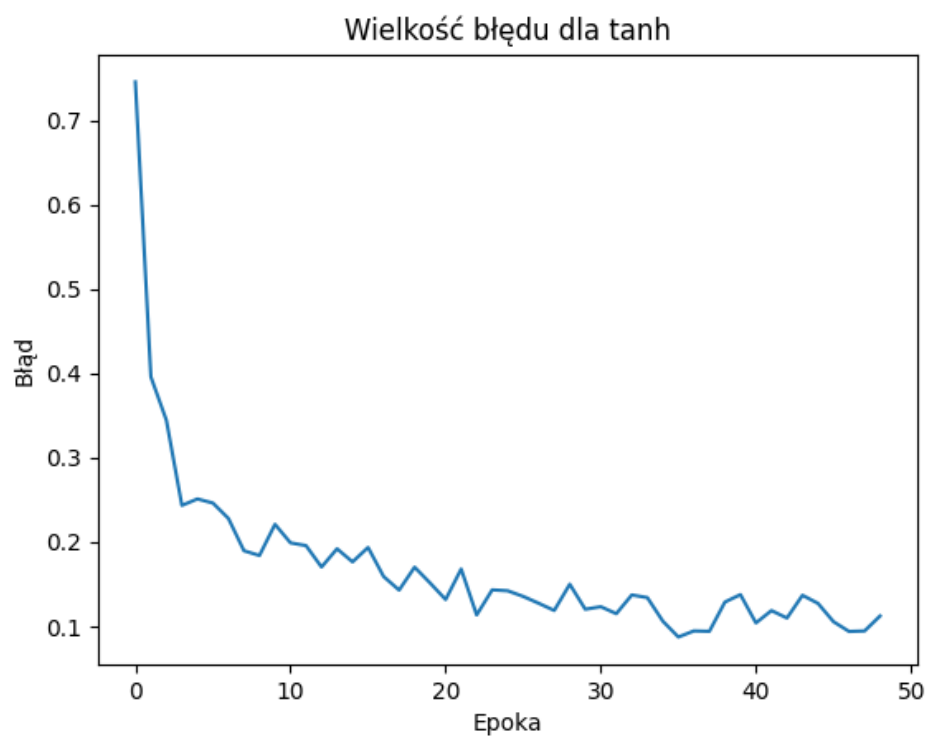
1.5 Test funkcji aktywacji

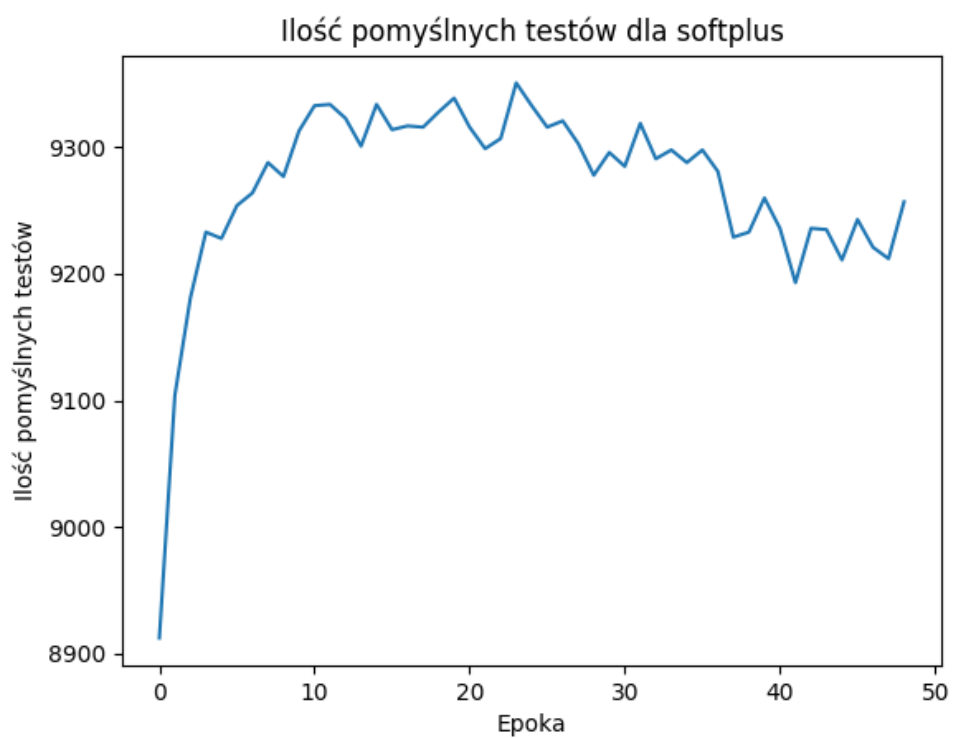
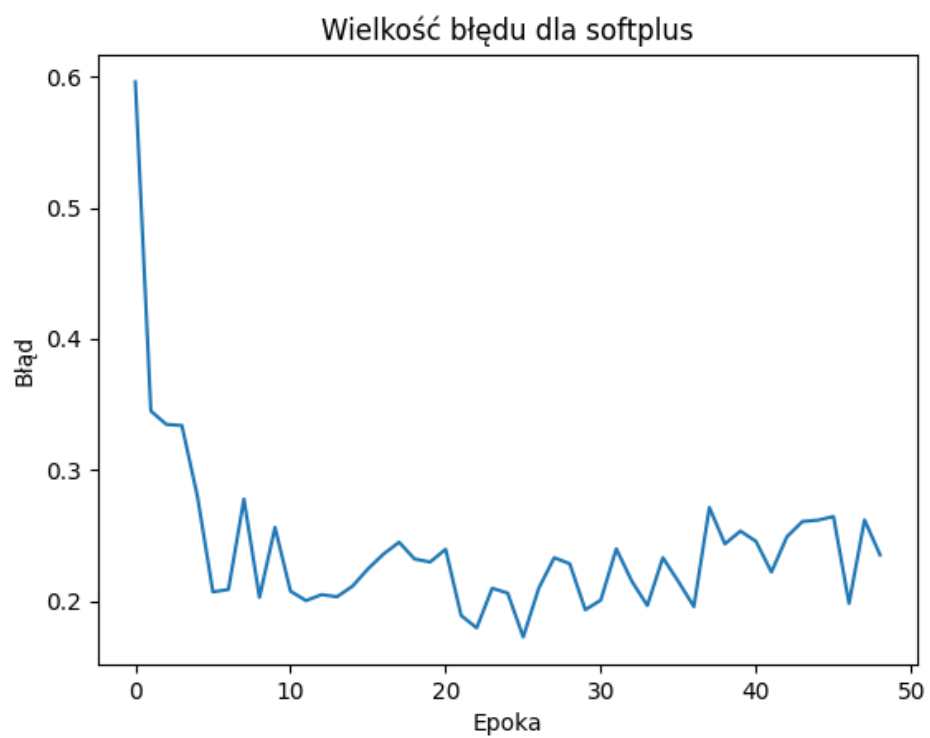
W sieci neuronowej przetestowano 3 różne funkcje aktywacji: sigmoidalną, tangensa hiperbolicznego oraz softplus. W tabelce zawarty jest uśredniony błąd oraz uśredniona liczba prawidłowo rozpoznanych liczb ze zbioru testowego. Niżej przedstawione są wyniki na wykresach.

Funkcja aktywacji	Błąd	Skuteczność
sigmoid	0.65	8021
tanh	0.17	9380
softplus	0.24	9270

Tablica 5: Funkcja aktywacji







2 Wnioski

2.1 Wielkość sieci

Ilość warstw ukrytych wydaje się nie wpływać znacząco na skuteczność sieci. Ilość neuronów na warstwę ma z kolei widoczny pozytywny wpływ. Duże ilości z zwiększają rozmiary matryc i ilości działań które należy wykonać, zabierają zatem więcej czasu.

2.2 Współczynnik uczenia

Najlepszy wynik uzyskany został dla pośredniej wartości 0.5. Zbyt niski wynik dawał wysoki błąd, a wraz z wzrost powyżej optymalnej wartości błąd również spadał. Niski współczynnik uczenia może wyuczyć sieć lepiej, jednak wymaga ona przejścia przez więcej epok. Zbyt duży współczynnik może spowodować, że sieć przeoczy optymalny punkt w którym najlepiej rozpoznaje wzorce.

2.3 Wielkość paczki

Testy wskazują że mniejsza wielkość paczki skutkuje poprawą wyników. Mała wielkość paczki ma za zadanie poprawić prędkość działania sieci, za cenę wprowadzenia szumu, czyli pewnych niedokładności. Można zauważyć że dla większych wartości paczki mimo że średni wynik jest gorszy niż te niższe, to dla każdej epoki jest on bardziej stabilny - wyniki są na jednym poziomie.

2.4 Sposób inicjalizacji wag

Im mniejsze odchylenie standardowe tym można zauważyć lepsze wyniki - chcemy zatem aby wartości były bliskie 0. Zbyt duże odchylenie standardowe skutkuje mniejszą dokładnością sieci.

2.5 Użyta funkcja aktywacji

Można zauważyć, że z zadaniem problemem najlepiej radzi sobie funkcja tangensa hiperbolicznego. Niewiele gorzej spisuje się funkcja softplus (użyta jako substytut funkcji ReLU). Dużą niestabilnością wykazuje się funkcja sigmoidalna.

Przedstawione wnioski odnoszą się do problemu rozwiązywanego na zajęciach, czyli rozpoznawania liczb ze zbioru MIST. Mogą one nie być uniwersalne i dla innych rodzajów problemów inne hiperparametry mogą okazać się skuteczniejsze.