Projekt NAI: Rozpoznawanie liter

uczenie sieci neuronowej

Autor: Paweł Wiszniewski, s13626

Spis treści

[1. Opis zadania 3](#_Toc506491602)

[2. Opis zbioru danych 3](#_Toc506491603)

[2.1. Źródło danych 3](#_Toc506491604)

[2.2. Przygotowanie danych 3](#_Toc506491605)

[3. Opis struktury sieci neuronowej 4](#_Toc506491606)

[4. Wyniki eksperymentalne 4](#_Toc506491607)

[4.1. Sieć z 16 neuronami w warstwie ukrytej: 5](#_Toc506491608)

[4.2. Sieć z 26 neuronami w warstwie ukrytej: 6](#_Toc506491609)

[4.3. Sieć z 52 neuronami w warstwie ukrytej: 7](#_Toc506491610)

[4.4. Sieć z 78 neuronami w warstwie ukrytej: 8](#_Toc506491611)

[4.5. Sieć z 104 neuronami w warstwie ukrytej: 9](#_Toc506491612)

[4.6. Wnioski: 10](#_Toc506491613)

# Opis zadania

Celem zadania jest stworzenie dwuwarstwowej sieci neuronowej i nauczenie jej rozpoznawania 26 liter angielskiego alfabetu, znajdujących się oryginalnie na czarno-białych obrazkach. W ramach kolejnych eksperymentów staramy się dobrać jak najlepsze parametry sieci neuronowej do realizacji tego zadania i pokazujemy przebieg uczenia w czasie.

# Opis zbioru danych

## Źródło danych

Dane używa do rozpoznawania liter pochodzą z Machine Learning Repository UCI i są dostępne pod linkiem:

<https://archive.ics.uci.edu/ml/datasets/letter+recognition>

Do utworzenia zestawu danych zostały użyte angielskie litery zapisane za pomocą jednego z 20 różnych rodzajów fontów, które zostały poddano losowym zniekształceniom. W ten sposób uzyskano 20 tysięcy różnych obrazków. Każdy z nich został następnie opisany za pomocą 16 atrybutów numerycznych przeskalowanych w liczby całkowite z zakresu 0 do 15.

Szczegółowy opis wszystkich atrybutów można znaleźć na stronie z danymi.

## Przygotowanie danych

Dane wejściowe zostały podzielone na 2 zbiory:

* Zbiór treningowy – 75% danych wejściowych – 15 000 rekordów
* Zbiór testowy – 25% danych wejściowych – 5 000 rekordów

Zbiór treningowy jest używany do nauczenia sieci rozpoznawania liter. Przed rozpoczęciem każdej nowej epoki uczenia zbiór treningowy jest mieszany.

Dane wejściowe zostały zestandaryzowane:

Ze względu na ilość danych wejściowych sieć jest uczona za pomocą algorytmu uczenia z pomocą zestawów danych (batch) – jednorazowo obliczamy wartości zmian dla zestawów paczek danych o ustalonej wielkości. Eksperymenty były przeprowadzane dla wartości batch ∈ { 10, 50, 100, 200 }.

# Opis struktury sieci neuronowej

Sieć dwuwarstwowa składa się z:

* warstwy wejściowej dla 16 atrybutów określających każdą literę
* warstwy ukrytej, której wielkość jest zmieniana w trakcie kolejnych eksperymentów i zawiera się w zbiorze {16, 26, 52, 78}
* warstwy wyjściowej – 26 neuronów. Każdy z neuronów będzie odpowiadał jednej literze angielskiego alfabetu. Neuron na warstwie wyjściowej o najwyższym poziomie aktywacji będzie wskazywał na rozpoznaną literę.

Wagi wszystkich połączeń między neuronami zostały zainicjalizowane losowo w przedziale [0, 1]. W trakcie kolejnych eksperymentów początkowe wagi sieci neuronowej są zapisywane tak aby można było porównywać między sobą wyniki kolejnych eksperymentów.

# Wyniki eksperymentalne

W trakcie badania wydajności sieci neuronowej zmieniane były trzy parametry:

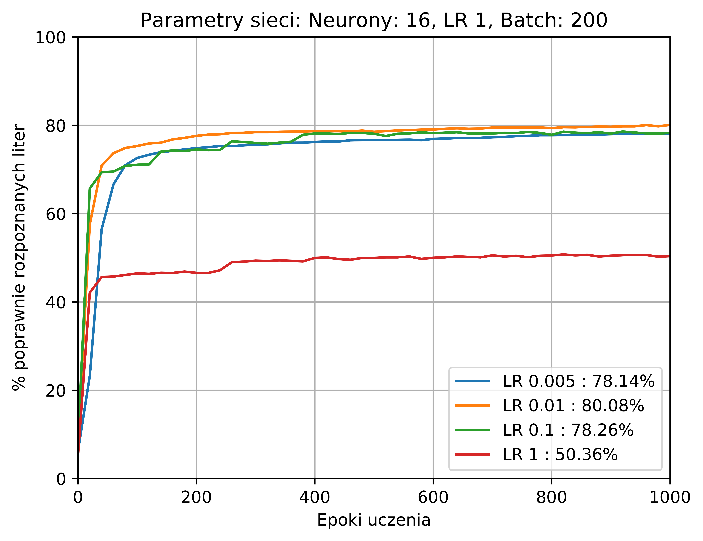
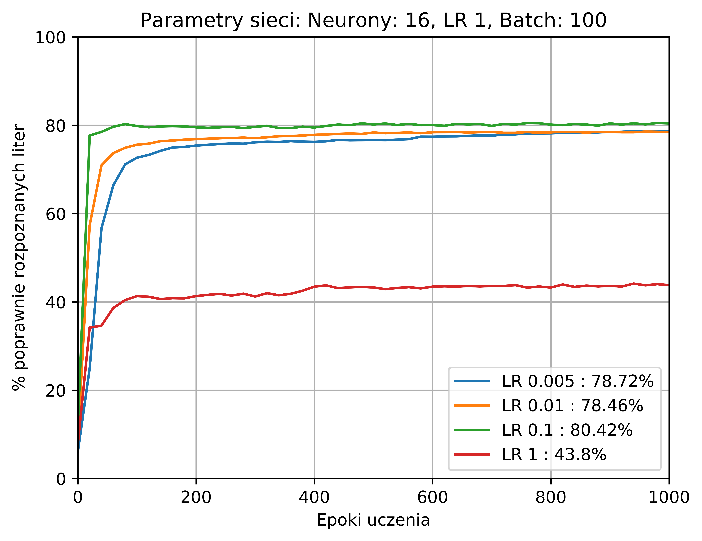
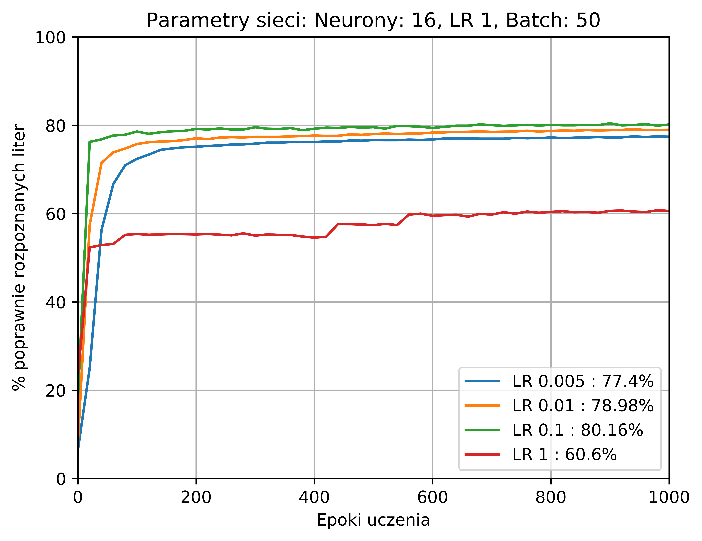
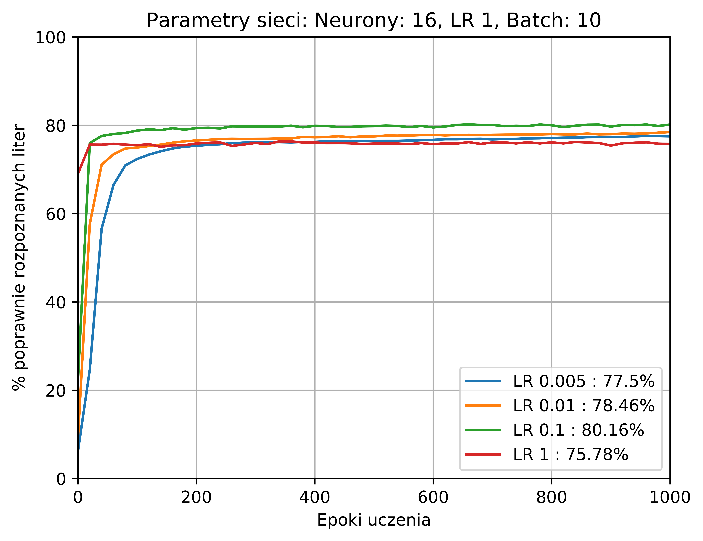
* Szybkość uczenia się (learning rate) ∈ { 0.005, 0.01, 0.1, 1 }
* Ilość neuronów w warstwie ukrytej ∈ { 16, 26, 52, 78, 104 }
* Wielkość zestawów danych (batch) używanych do uczenia sieci ∈ { 10, 50, 100, 200 }

W każdym eksperymencie co 20 epok uczenia sieci neuronowej sprawdzana była precyzja sieci w rozpoznawaniu danych z zestawu testowego. Na koniec eksperymentu generowany był wykres pokazujący precyzję sieci w kolejnych epokach w zależności od użytych parametrów oraz ostatecznie osiągnięte wyniki precyzji rozpoznawania danych z zestawu testowego.

Najlepszy wynik dla sieci podczas jej trenowania uzyskano dla następujących parametrów:

LR: 0.01, Batch: 50, neurony w warstwie ukrytej: 78, liczba epok: 2000. **Precyzja 91.1 %**

## Sieć z 16 neuronami w warstwie ukrytej:



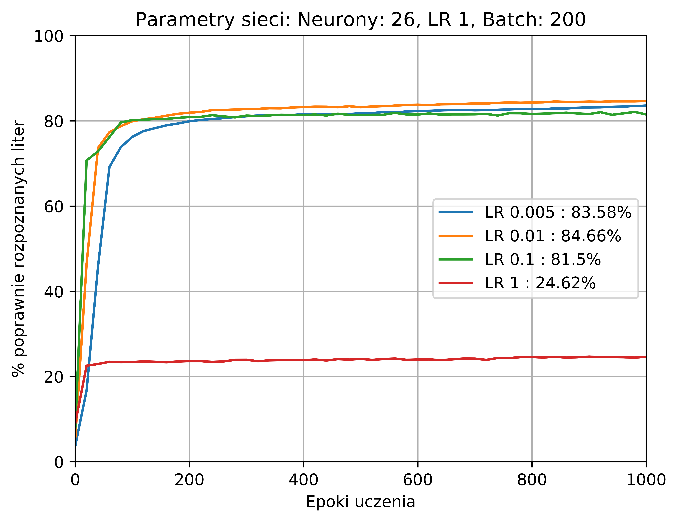
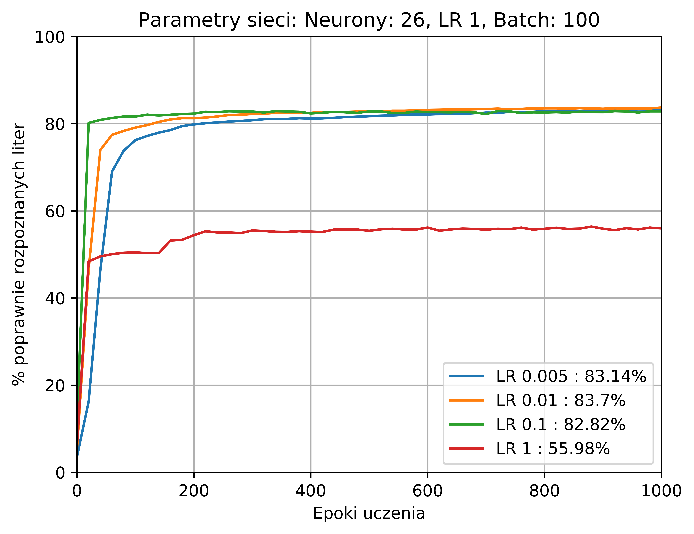
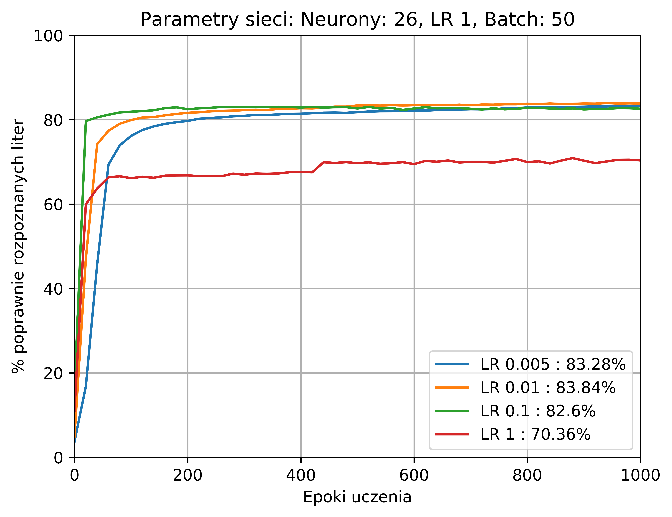
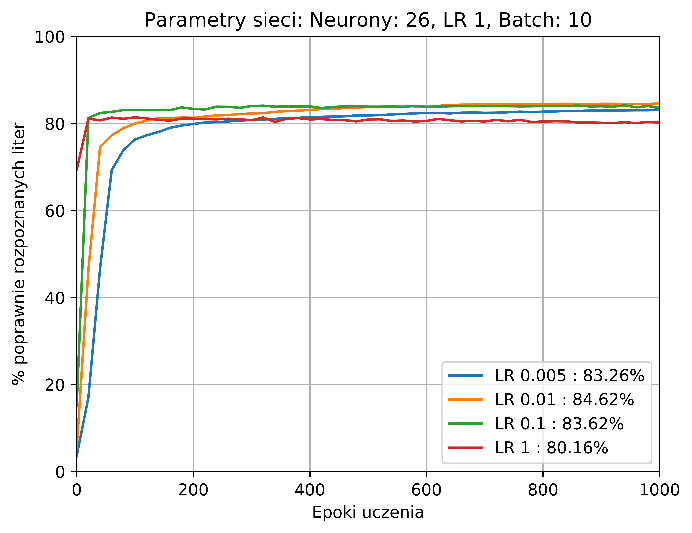
Najlepsze wyniki po 1000 epok uczenia się uzyskano dla sieci z parametrami LR: 0.1 i Batch: 100:

80.42%

Dla LR = 1 sieć szybko osiągała swoją maksymalną precyzję, ale nie jest w stanie dobrać optymalnych parametrów, co szczególnie widać przy dużych zestawach danych uczących. Zapewne przy tak dużej wartości LR próby zejścia zgodnie z gradientem funkcji błędu powodują przestrzelenie i oscylację. Jeżeli pojawiają się skoki w precyzji to są dosyć gwałtowne.

Mała wartość LR pozwala być może w przyszłości osiągnąć większą precyzję sieci, ale w zakresie pierwszych 1000 epok uczenia zbyt wolno dochodziły do maksymalnej wartości precyzji. Niestety zwiększa też szansę na to, że sieć utknie w maksimum lokalnym.

## Sieć z 26 neuronami w warstwie ukrytej:



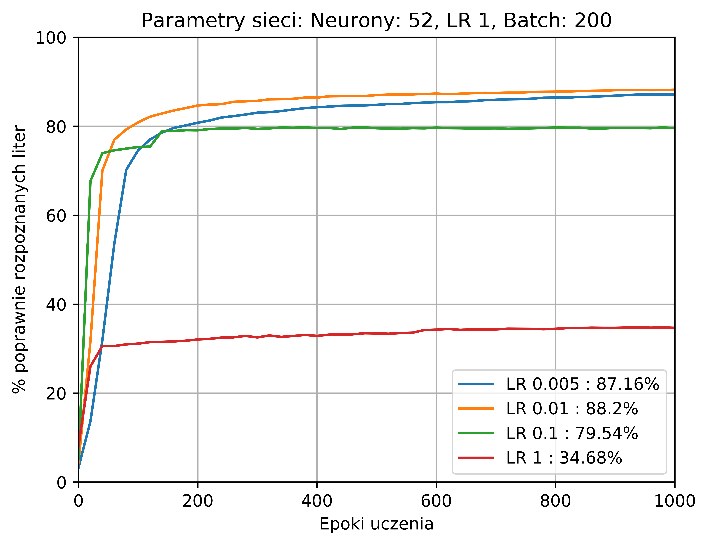
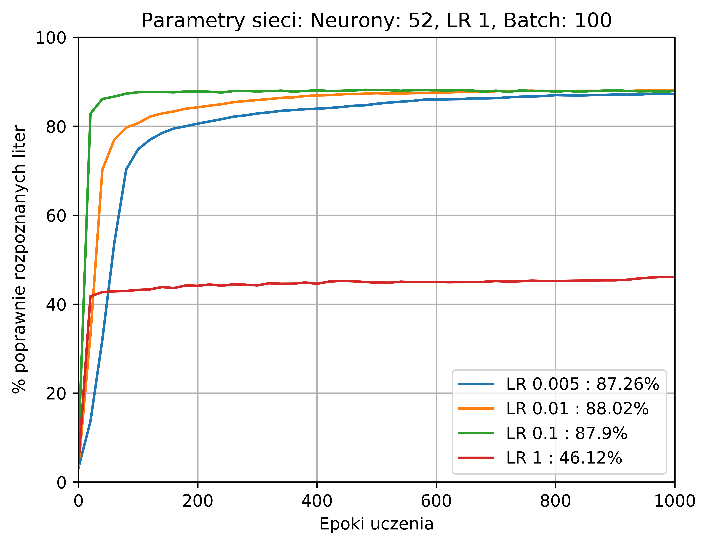
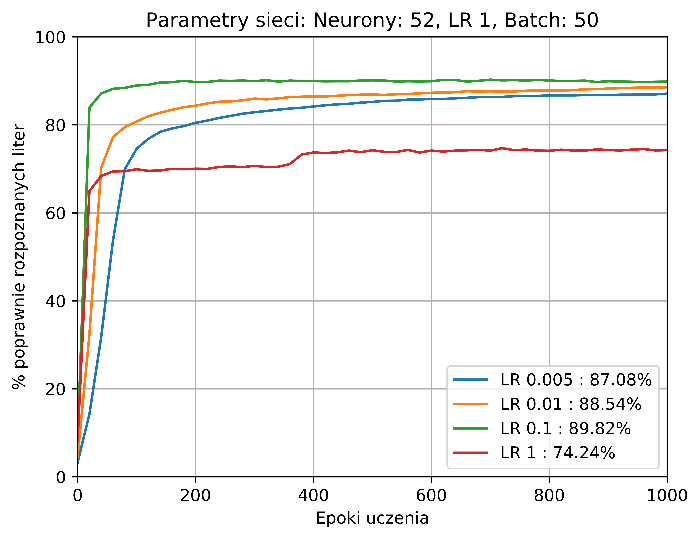
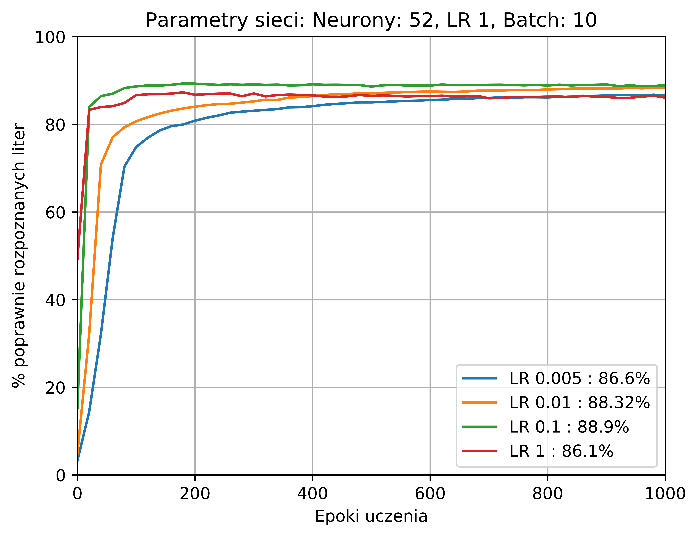
Po tysiącu epok uczenia najwyższą precyzję osiągają sieci z LR = 0.01:

84.66%

Tak jak poprzednio, sieci z wyższą wartością LR szybciej osiągają maksymalną precyzję, ale nie są zbyt dokładne, przez mają gorsze wyniki niż LR = 0.01. W szczególności dla LR = 1 sieć po osiągnieciu pewnej precyzji nie jest w stanie się już lepiej uczyć.

Zwiększenie ilości neuronów w warstwie ukrytej podniosło wyniki w porównaniu do identycznych parametrów z 16 neuronami ukrytymi.

## Sieć z 52 neuronami w warstwie ukrytej:



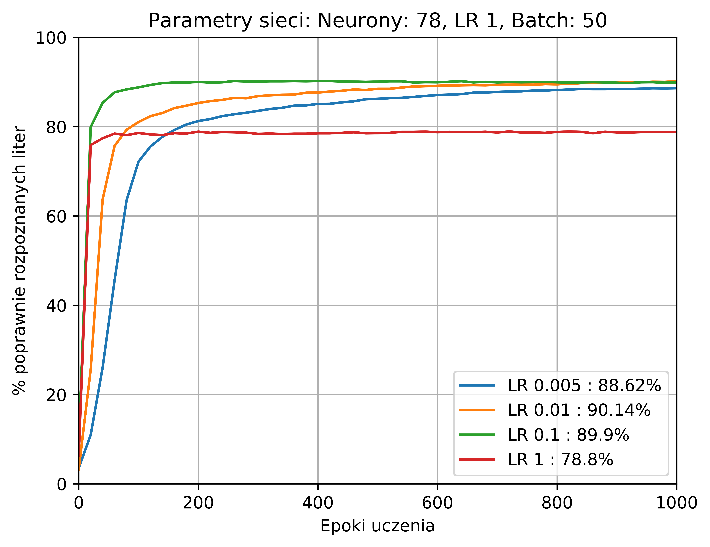
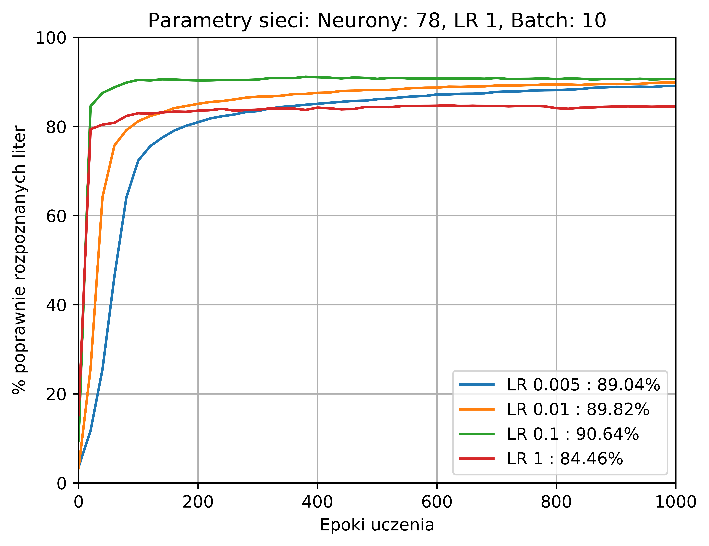
Po tysiącu epok uczenia najwyższą precyzję osiągają sieci z LR = 0.1:

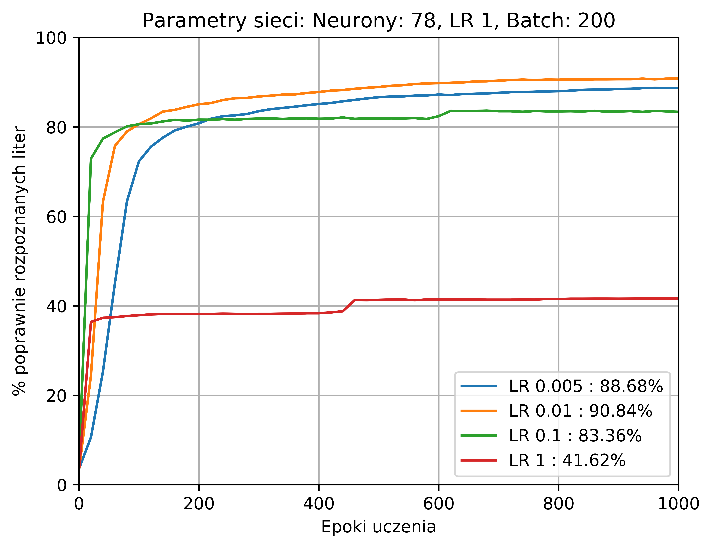
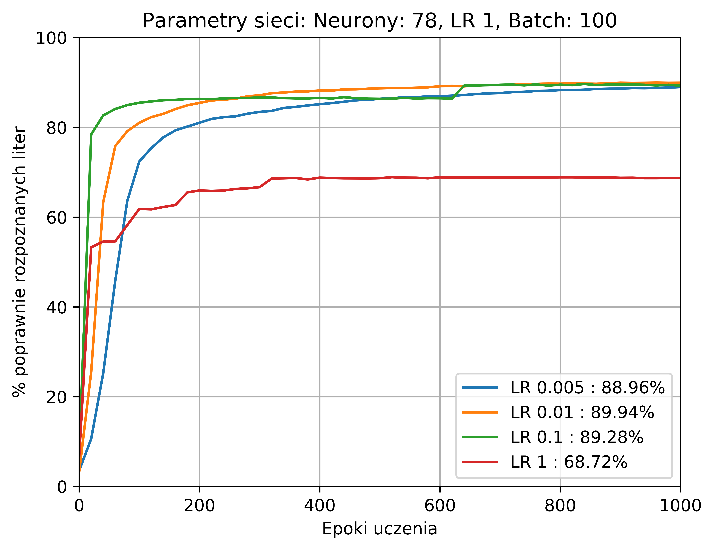
89.82%

ale sieci z LR = 0.01 mają wyniki bardzo zbliżone – najlepszy wynik to 88.54%. Tak niewielka różnica może wynikać z specyfiki losowego wygenerowania danych początkowych. Na wykresie możemy też zauważyć, że wprawdzie LR = 0.1 zwycięża, ale precyzja dla sieci LR = 0.01 ciągle rośnie i być może prześcignęłaby wyższe LR przy większej liczbie epok.

Zwiększenie ilości neuronów w warstwie ukrytej podniosło wyniki w porównaniu do identycznych parametrów z 26 neuronami ukrytymi.

## Sieć z 78 neuronami w warstwie ukrytej:

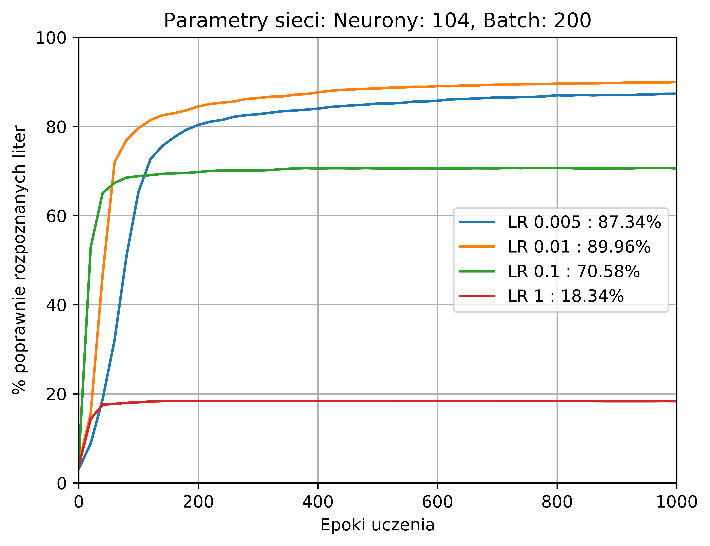
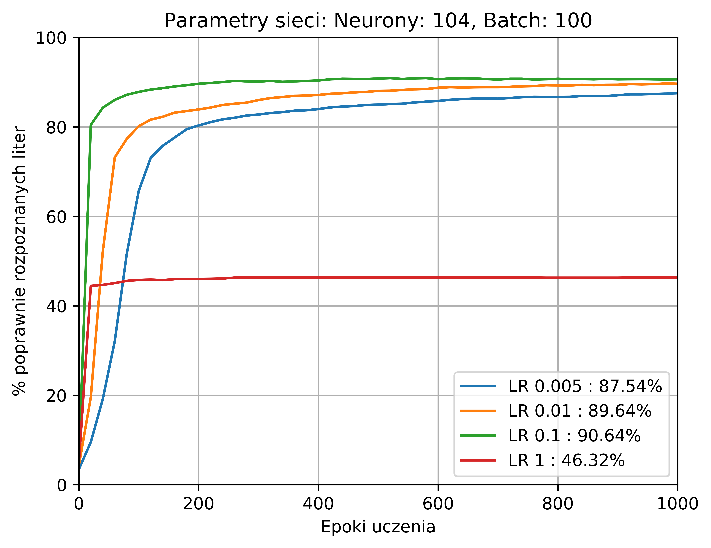
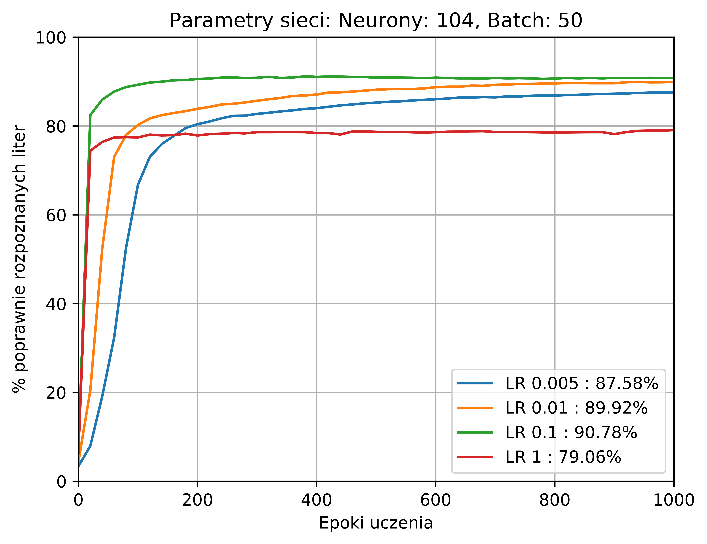
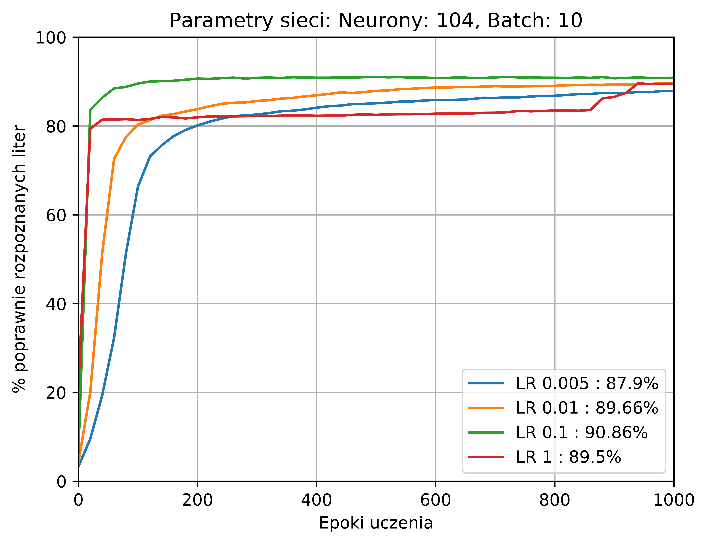




Dla 78 neuronów najlepszy wynik po 1000 epok to 90.84% dla LR = 0.01 i batch = 200. Prawie identyczny wynik 90.64% osiągnęliśmy przy LR = 0.1 i batch = 10. Warto jednak zwrócić uwagę, że dla w drugim przypadku wykres wydaje się już osiągnąć swoje maximum, zaś w przypadku LR = 0.01 widać jeszcze lekką tendencję wzrostową.

Wzrost liczby neuronów z 52 do 78 tylko nieznacznie poprawił precyzję, za to wydłużył czas nauki w każdej epoce.

## Sieć z 104 neuronami w warstwie ukrytej:



Dla 104 neuronów najlepsze wyniki ciągle oscylują w okolicy 90.80%. Wprawdzie najlepszy wynik tej sieci jest o 0.02% lepszy niż w przypadku 78 neuronów ale nie jest to znacząca zmiana przy kolejnym wzroście czasu uczenia w każdej epoce.

Przy tej ilości neuronów LR = 0.1 generuje najlepsze wyniki, ale być może jest to kwestia zbyt małej ilości epok uczenia.

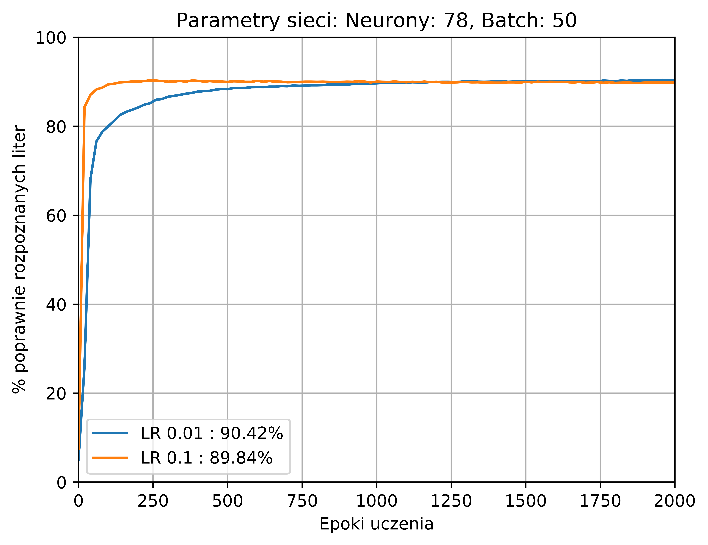
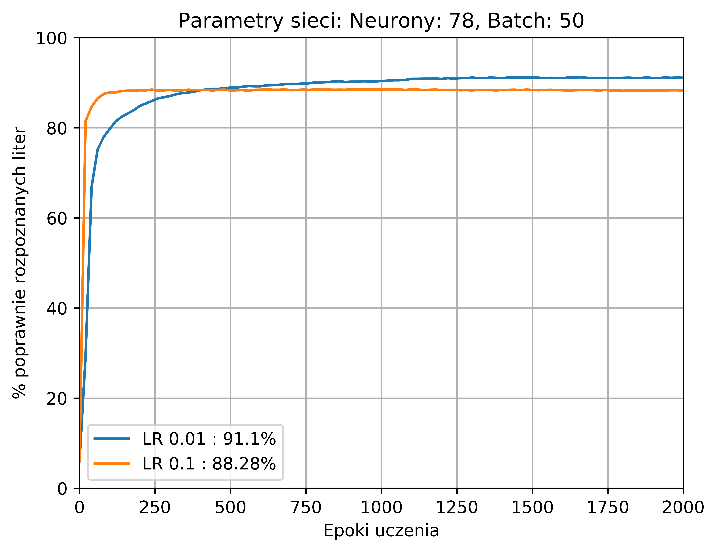
## Wnioski

Zbyt mała ilość neuronów utrudnia sieci osiągnięcie wysokiego poziomu precyzji, tym niemniej nie ma sensu zbyt agresywne powiększanie ukrytej warstwy sieci – przyrost precyzji jest w pewnym momencie niezauważalny. Najlepsze wyniki sieć osiągała w przedziale <52, 78> neuronów (pomiędzy 2x i 3x wielkość warstwy wyjściowej). Większa warstwa ukryta spowalnia również obliczenia – jest ich po prostu dużo więcej do wykonania.

Większe wartości LR spowalnia szybkość uczenia się ale zwiększa maksymalną precyzję sieci. Podobny efekt ma wielkość zestawów uczących (batch) – większe zestawy zwiększają szybkość uczenia się sieci ale zmniejszają maksymalny poziom precyzji. Niestety dodatkowo zmniejszenie wielkości zestawów znacząco wydłuża liczbę wszystkich obliczeń dla każdej epoki – komputer przeprowadzać duże ilości obliczeń dla mniejszych paczek danych i przełączać się między wywołaniami kodu uczącego – czego nie widać na wykresie. Przykładowo różnica czasu wyuczenia 1000 epok dla batch = 10 i batch = 50 była dwukrotna (wolniej dla batch = 10). Trzeba ostrożnie wyważyć między sobą te dwa parametry.

Wykres precyzja sieci neuronowej w zależności od ilości epok również jest dosyć specyficzny. Widać wyraźnie, że wartości powyżej 80% precyzji sieć osiąga już po mniej niż 200 epokach uczenia. Kolejne epoki wprawdzie zwiększały (dla większości parametrów) precyzję, ale przyrosty te są coraz wolniejsze z kolejnymi epokami. Znaczące zwiększenie epok nie wpłynie więc już zbyt poważnie na precyzję sieci. Dużo ważniejszy jest dobry dobór wielkości sieci i prędkości uczenia się.

Na koniec wykonałem jeszcze porównanie wyników sieci neuronowej dla 2000 epok uczenia.



Interesujące jest, że nawet przy tych samych parametrach sieci, wyniki mogą się różnić o niewielki procent. W powyższym przykładzie dla 78 neuronów, batch = 50 i 2000 epok uczenia, różnice wyników dla takiego samego LR są w zakresie <0.68%, 1.56%>. Jak widać maksymalna precyzja sieci może więc zależeć od wartości wylosowanych początkowo wag.