Sieci Neuronowe – laboratorium  
  
Ćwiczenie nr. 3  
  
Kajetan Pynka, 254495

Spis treści

[Wstęp teoretyczny 3](#_Toc119533434)

[Wpływ liczby neuronów 4](#_Toc119533435)

[Liczba neuronów - 10 4](#_Toc119533436)

[Liczba neuronów - 20 5](#_Toc119533437)

[Liczba neuronów - 30 6](#_Toc119533438)

[Liczba neuronów - 50 7](#_Toc119533439)

[Wnioski 8](#_Toc119533440)

[Wpływ współczynnika uczenia 9](#_Toc119533441)

[Współczynnik uczenia – 1e-4 9](#_Toc119533442)

[Współczynnik uczenia 5e-5 10](#_Toc119533443)

[Współczynnik uczenia 1e-5 11](#_Toc119533444)

[Współczynnik uczenia – 1e-6 12](#_Toc119533445)

[Wnioski 13](#_Toc119533446)

[Wpływ wielkości paczki (batcha) 14](#_Toc119533447)

[Wielkość paczki – 6 14](#_Toc119533448)

[Wielkość paczki – 60 15](#_Toc119533449)

[Wielkość paczki – 600 16](#_Toc119533450)

[Wielkość paczki – 6000 17](#_Toc119533451)

[Wnioski 18](#_Toc119533452)

[Wpływ inicjalizacji wag – odchylenie standardowe 19](#_Toc119533453)

[Odchylenie standardowe – 0.1 19](#_Toc119533454)

[Odchylenie standardowe – 0.01 20](#_Toc119533455)

[Odchylenie standardowe – 0.001 21](#_Toc119533456)

[Odchylenie standardowe – 0.0001 22](#_Toc119533457)

[Wnioski 23](#_Toc119533458)

[Wpływ funkcji aktywacji 24](#_Toc119533459)

[Funkcja aktywacji – sigmoid 24](#_Toc119533460)

[Funkcja aktywacji – tangens hiperboliczny 25](#_Toc119533461)

[Funkcja aktywacji – RELU 26](#_Toc119533462)

[Wnioski 27](#_Toc119533463)

# Wstęp teoretyczny

W ramach ćwiczenia zaimplementowałem sieć MLP. Wszelkie algorytmy zostały napisane od podstaw, natomiast do przeprowadzenia obliczeń macierzowych wykorzystałem bibliotekę ‘numpy’. W ramach badań analizowałem z osobna następujących pięć czynników: liczbę neuronów warstwy ukrytej, wartość współczynnika uczenia, wielkość paczki (batcha), początkowe wartości wag oraz rodzaj funkcji aktywacji warstwy ukrytej. Każdy eksperyment był oceniany ze względu na liczbę epok potrzebnych do wyuczenia modelu (model uznawany był za wyuczony gdy różnica między wartością funkcji kosztu dla epoki poprzedniej i aktualnej wynosiła poniżej 30). Dodatkowo, dla każdej epoki zapisywany był aktualny błąd na zbiorze testowym (w skali [0;1] gdzie 0.1 oznacza, że model miał 10% błędnych predykcji = 90% poprawnych predykcji). Każdy eksperyment został powtórzony 10 razy (z takimi samymi wartościami hiperparametrów). Macierz pomyłek jest średnią z tych 10 uruchomień, dla każdego uruchomienia obliczana dla zbioru testowego PO zakończeniu nauki.

Legenda do wykresów:

* Kolor czerwony oznacza przebieg eksperymentu o najgorszym finalnym rezultacie (najwyższym błędzie)
* Kolor niebieski oznacza średnią ze wszystkich 10 powtórzeń eksperymentu
* Kolor zielony oznacza przebieg eksperymentu o najlepszym finalnym rezultacie (najniższym błędzie)

Dla każdego eksperymentu przyjęte zostały następujące wartości hiperparametrów (tam, gdzie dany parametr nie był oczywiście przedmiotem badania):

* Liczba neuronów warstwy ukrytej – 20
* Rozmiar zbioru treningowego – 60000
* Rozmiar zbioru testowego – 10000
* Rozmiar paczki (batcha) – 60
* Dopuszczalna wartość błędu między funkcją kosztu – 30
* Współczynnik uczenia – 1e-5 (0.00001)
* Early stopping – aktywny, przy czym jeśli przez 3 kolejne epoki i tak następuje regres to proces nauki był przerywany
* Tylko jedna warstwa ukryta
* Wagi inicjalizowane za pomocą rozkładu normalnego gdzie
  + Mediana / wartość oczekiwana – 0
  + Odchylenie standardowe – 0.001

# Porównanie sieci konwolucyjnej i MLP