

Mikalai Stelmakh, 316951

Ishchuk Roman, 317362

WSI Lab 6 – Sieci Neuronowe

Zadanie

Zaimplementowanie sieci neuronowej do klasyfikacji ręcznie pisanych cyfr.

Uruchomienie programu

```
python3 nn.py [-h] -n N [--learn_rate LEARN_RATE] [--epochs EPOCHS]
```

Żeby dowiedzieć się więcej na temat argumentów należy użyć flagi "-h".

Rozwiązanie

Wagi

Model inicjalizowany jest losowymi wagami z przedziału $[-0.5; 0.5]$. Dodatkowo, do każdej warstwy oprócz wejściowej dodawany jest bias inicjalizowany zerami.

Funkcja aktywacji

Do aktywacji wartości wyjść warstw użyto funkcji sigmoidalnej.

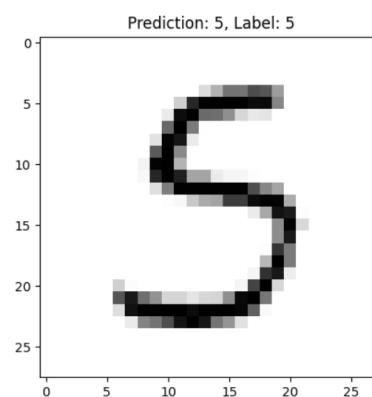
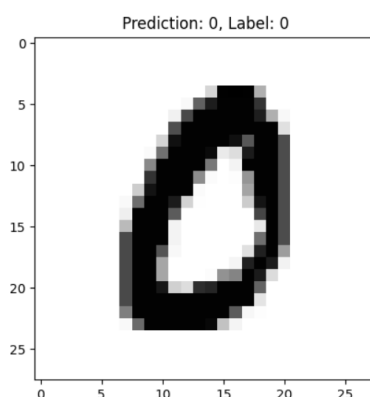
Funkcja aktywacji warstwy wyjściowej

Do aktywacji wartości wyjść warstwy wyjściowej użyto funkcji softmax.

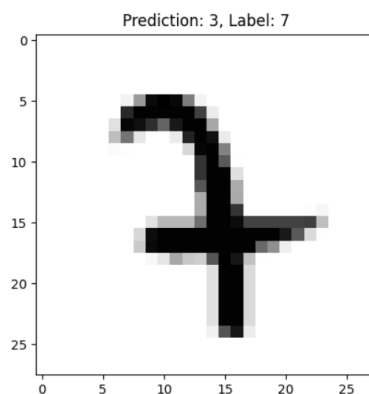
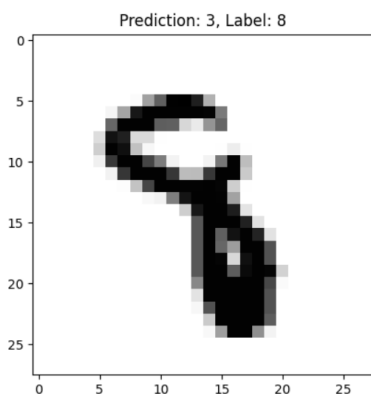
Wyniki

Dla ustawień opisanych w następnym punkcie otrzymano sieć o precyzji około 95%.

Kilka przykładów ze zbioru, które zostały sklasyfikowane poprawnie:



oraz które zostały sklasyfikowane niepoprawnie:



Badanie wpływu parametrów

Domyślne ustawienia:

- Liczba warstw ukrytych: 2
- Liczba neuronów warstwy ukrytej: 40
- Funkcja aktywacji: sigmoidalna
- Funkcja aktywacji warstwy wyjściowej: softmax
- Liczba epok: 3
- Współczynnik uczenia: 0.01

Wpływ danych parametrów na wartość precyzji klasyfikacji:

Liczba warstw ukrytych

Liczba warstw	min	średnia	max	odchylenie
1	94.5	94.75	94.9	0.13
2	94.7	94.83	95	0.12
3	94.7	94.85	95	0.12

Zwiększenie liczby warstw ukrytych powoduje większą precyzję. Natomiast zbyt duża liczba warstw spowoduje przeuczenie modelu, co oznacza, że będzie dobrze pasował do danych uczących, jednak będzie dawał gorsze wyniki, gdy zastosuje się do danych, z którymi nie zetknął się podczas uczenia.

Liczba neuronów warstwy ukrytej

Liczba neuronów	min	średnia	max	odchylenie
10	90.5	91	91.8	0.46
40	94.7	94.9	95.1	0.14
160	95.9	96	96.08	0.05

Liczba neuronów warstw ukrytych wpływa na wartość precyzji podobno do liczby warstw ukrytych. Mała liczba neuronów spowoduje niedopasowanie modelu, natomiast zbyt duża liczba neuronów spowoduje przeuczenie modelu. Dostaliśmy najlepsze wyniki dla liczby neuronów w przedziale pomiędzy rozmiarem warstwy wejściowej a wyjściowej.

Liczba epok

Liczba epok	min	średnia	max	odchylenie
1	85.26	85.67	86.1	0.35
3	95.6	95.75	95.9	0.08
5	97.5	97.6	97.74	0.08

Tak samo jak dla liczby neuronów oraz liczby warstw ukrytych nie można od razu powiedzieć jaka liczba epok jest optymalna. Wartość ta zależy od kilku parametrów, takich jak złożoność problemu lub rozmiar zbioru danych trenujących.

Współczynnik uczenia

Współczynnik uczenia	min	średnia	max	odchylenie
0.01	94.6	94.9	95.1	0.15
0.1	92.8	93	93.3	0.13
0.3	25.6	38.8	51.5	8.18

Współczynnik uczenia jest parametrem określającym podczas propagacji wstecznej wielkość kroku w każdej iteracji, zbliżając się do minimum funkcji straty. Zbyt mały krok będzie wymagał większej ilości epok do znalezienia optymalnych wartości wag. Tym niemniej do otrzymania wysokiej precyzji potrzebujemy stosunkowo małego

współczynnika uczenia dlatego, że dla dłuższego kroku algorytm “przeskakuje” właściwe wartości i nigdy nie trafia w minimum funkcji straty.

Porównanie wyników uzyskanych podczas implementacji

1	
min	94.58
avg	94.779
max	95.06
odchylenie	0.128
2	
min	94.65
avg	94.926
max	95.13
odchylenie	0.157
3	
min	94.75
avg	94.861
max	94.98
odchylenie	0.081

W przypadku liczby ukrytych warstw: zwiększenie liczby ukrytych warstw prowadzi do większej dokładności: 1 jest zła, 2 - gorsza, 3 lepsza niż 1

10	
min	90.64
avg	90.899
max	91.15
odchylenie	0.178
40	
min	94.79
avg	94.946
max	95.1
odchylenie	0.123
160	
min	95.81
avg	95.941
max	96.02
odchylenie	0.071

Dla liczby neuronów w warstwach ukrytych: 10 jest złe, 40 jest lepsze, 160 ma jeszcze lepszy wpływ na odchylenia.

1	
min	85
avg	85.713
max	86.17
odchylenie	0.427
3	
min	95.46
avg	95.693
max	95.92
odchylenie	0.125
5	
min	97.44
avg	97.585
max	97.75
odchylenie	0.088

Dla epoki: 1 epoka za duże odchylenie, 2 epoka - odchylenie minimalne, 5 epoka - odbywa się przekwalifikowanie

0.01	
min	94.71
avg	94.919
max	95.06
odchylenie	0.097
0.1	
min	92.79
avg	93.115
max	93.39
odchylenie	0.222
0.3	
min	35.57
avg	42.095
max	53.13
odchylenie	4.969

W przypadku szybkości uczenia się: aby uzyskać wysoką dokładność, potrzebujemy niskiego współczynnika uczenia się: 0,01 - wynik doskonały, 0,1 - wynik nieco gorszy, 0,3 - okropny wynik

Podział pracy

Mikalai Stelmakh: zaimplementowanie sieci neuronowej

Roman Ishchuk: zaimplementowanie testów do raportu, badanie wpływu parametrów