

Sprawozdanie

Ćwiczenie 2

Sieć wielowarstwowa uczona metodą propagacji wstecznej

Piotr Swirkaitis

246753

Wprowadzenie

W ramach ćwiczenia opracowano model wielowarstwowej sieci neuronowej, wykorzystującej technikę SGD, przy uczeniu metodą propagacji wstecznej. Zaimplementowano również funkcjonalność podziału zbioru treningowego na tak zwane paczki (batch), gdzie aktualizacja wag, wykonuje się po przejściu wszystkich elementów z jednej paczki.

Badania

Jako podstawowe parametry wprowadzane do sieci podczas badań przyjęto:

- Rozmiar paczki (batcha): 32 obrazki
- Współczynnik uczenia: 0.5
- Wagi: rozkład normalny
- Domyślne funkcja aktywacji w warstwie: sigmoid
- Liczba warstw ukrytych: 2
- Liczba neuronów w kolejnych warstwach ukrytych:
 - 128
 - 64

Badanie wpływu liczby neuronów w warstwie na uczenie sieci neuronowej

Warstwa ukryta	Wartość trafności predykcji w procentach		
	1 epoka	5 epoka	10 epoka
256 - 128	70.92	89.06	91.23
128 - 64	75.83	88.52	91.161
64 - 32	75.59	87.8	90.457
128 – 64 - 32	75.142	88.247	90.895

Wnioski:

Zauważono nieznaczne wahania wyników, podczas uczenia sieci neuronowej przy różnych rozmiarach warstw. Badaniu poddano 3 modele z dwoma warstwami ukrytymi oraz jeden z 3 warstwami. Nie zauważono znaczących odchyleń w wynikach predykcji, natomiast można zaznaczyć, że najlepsze osiągnął model dwuwarstwowy z warstwami po kolejno 256 i 128 neuronów. Przy użyciu modelu 3 warstwowego nie zaobserwowano znacznych odchyleń od wyników pozyskanych w innych doświadczeniach.

Badanie wpływu rozmiaru paczki (batch) na uczenie sieci neuronowej

Rozmiar paczki	Wartość trafności predykcji w procentach		
	1 epoka	5 epoka	10 epoka
16	82.228	90.27	92.23
32	75.83	88.52	91.161
64	68.247	86.190	88.4
128	57.361	75.83	85.76

Wnioski:

Z przeprowadzonych badań wynika, że sieć uczy się lepiej przy mniejszym rozmiarze paczki. Może to wynikać z częstotliwości aktualizacji wag, która następuje częściej przy mniejszym rozmiarze paczki. Najlepsze wyniki osiągnięto przy rozmiarze paczki 16, przy czym przy 2 razy większej paczce wynik osiągnięty przez sieć po 10 epokach jest nieznacznie mniejszy. Najgorszy wynik osiągnięto przy rozmiarze paczki równym 128 obrazków, gdzie wynik testu na zbiorze testowym wyniósł 85.76%.

Badanie wpływu współczynnika uczenia na uczenie sieci neuronowej

Wartość współczynnika uczenia	Wartość trafności predykcji w procentach		
	1 epoka	5 epoka	10 epoka
0.5	75.83	88.52	91.161
0.1	39.14	77.98	83.94
0.05	34.19	67.79	77.47
0.01	21.67	40.104	51.285

Wnioski:

Zauważono, że wartość współczynnika uczenia ma znaczący wpływ na szybkość uczenia się sieci neuronowej. Najlepsze wartości zaobserwowano przy współczynniku uczenia równym 0.5, natomiast najgorsze przy współczynniku równym 0.01. Współczynnik uczenia ma znaczący wpływ na proces uczenia sieci oraz jego szybkość, ponieważ odpowiada bowiem za tempo aktualizacji nowych wag.

Badanie wpływu inicjalizacji wag na uczenie sieci neuronowej

Badając wpływ inicjalizacji wag zastosowano różne wartości odchylenia standardowego podczas generowania wag z rozkładu normalnego.

Odchylenie standardowe	Wartość trafności predykcji w procentach		
	1 epoka	5 epoka	10 epoka
0.5	86.54	92.95	94.30
1	75.83	88.52	91.161
2	51.87	72.6	76.524
5	13.2	39.09	48.74

Wnioski:

Podsumowując przeprowadzone badania można stwierdzić że odchylenie standardowe rozkładu normalnego inicjalizującego wagi, ma wpływ na szybkość oraz jakość uczenia modelu. Najlepsze wyniki zaobserwowano dla wartości odchylenia standardowego równego 0.5, natomiast wraz z zwiększaniem się wartości odchylenia standardowego, wartość trafności predykcji modelu malała. Odchylenie standardowe podczas generowania wag ma duży wpływ na rozrzut z jakiego wyznaczone będą wartości.

Badanie wpływu funkcji aktywacji w warstwie na uczenie sieci neuronowej

Funkcja aktywacji	Wartość trafności predykcji w procentach		
	1 epoka	5 epoka	10 epoka
ReLU	81.258	90.27	90.23
Sigmoid	75.83	87.62	91.155
Tanh	80.567	89.972	90.47

Wnioski:

Z przeprowadzonych badań można wysnuć wnioski, że właściwie dobrana do problemu funkcja aktywacji warstwy może być znacząca dla szybkości uczenia modelu. Przeprowadzone badania na różnych funkcjach aktywacji oscylują wokół podobnych wartości, więc nie można jednoznacznie stwierdzić przewagi jednej funkcji nad innymi, co może wynikać z niskiej liczby epok, natomiast można zaznaczyć, że właściwie dobrana do zadanego problemu funkcja może przyspieszyć proces uczenia się sieci neuronowej.

Wnioski

Analizując wyniki z przeprowadzonych badań można stwierdzić, że szybkość uczenia wielowarstwowej sieci neuronowej zależy od zadanych parametrów uczenia. Zauważono znaczący wpływ wielkości paczki (batch), gdzie najlepsze wyniki osiągnięto przy rozmiarze paczki równym 32. Dużą rolę w sukcesie uczenia sieci odgrywa również parametr współczynnika uczenia, który odpowiada za tempo aktualizacji wag. Najlepsze wyniki sieć osiągnęła przy parametrze współczynnika uczenia równym 0,5. Duży wpływ wywierała na sieć inicjalizacja początkowych wag. Stwierdzono, że przy małym odchyleniu standardowym rozkładu normalnego, z którego losowano wagi sieć osiąga dużo lepsze wyniki niż w sytuacji gdzie wagi są losowane z dużego rozrzutu wartości. Badając wpływ liczby neuronów w warstwach, nie zauważono znaczących odchyłeń od wyników, natomiast stwierdzono, że czas wykonywania operacji znacznie się zwiększa przy większej liczbie warstw oraz większej liczbie neuronów w tych warstwach. Analizując wyniki badania wpływu funkcji aktywacji na proces uczenia sieci, nie można jednoznacznie wybrać najlepszej funkcji dla zadanego problemu, co może być spowodowane zbyt małą liczbą epok podczas treningu. Można jednak stwierdzić, że dobranie funkcji aktywacji odpowiednio dla zadanego problemu, ma wpływ na proces uczenia.