

WSI - ćwiczenie nr 5

Macias Marcin (DOKT) <marcin.macias.dokt@pw.edu.pl>

Fri 4/21/2023 2:30 PM

To: Basan Anton (STUD) <anton.basan.stud@pw.edu.pl>; Bieńkowski Maciej 2 (STUD) <maciej.bienkowski2.stud@pw.edu.pl>; Gieleta Konrad (STUD) <konrad.gieleta.stud@pw.edu.pl>; Jasiński Bartosz 2 (STUD) <bartosz.jasinski2.stud@pw.edu.pl>; Kiernożek Jakub (STUD) <jakub.kiernożek.stud@pw.edu.pl>; Kisz Monika (STUD) <monika.kisz.stud@pw.edu.pl>; Kowalski Jakub 9 (STUD) <jakub.kowalski9.stud@pw.edu.pl>; Król Mateusz 3 (STUD) <mateusz.krol3.stud@pw.edu.pl>; Kwiatkowski Jan (STUD) <jan.kwiatkowski.stud@pw.edu.pl>; Kwieciński Bartłomiej (STUD) <bartlomiej.kwiecinski.stud@pw.edu.pl>; Marton Paweł (STUD) <pawel.marton.stud@pw.edu.pl>; Ruciński Dawid (STUD) <dawid.rucinski.stud@pw.edu.pl>; Styczeń Wojciech (STUD) <wojciech.styczen.stud@pw.edu.pl>; Targosiński Wiktor (STUD) <>wiktor.targosinski.stud@pw.edu.pl>; Turycz Mateusz (STUD) <mateusz.turycz.stud@pw.edu.pl>

 1 attachments (15 MB)

WSI_5_Sieci_neuronowe.pptx;

Witam Państwa,

tematem piątych ćwiczeń są sieci neuronowe. Zadanie to będą wykonywali Państwo w parach, które zadeklarowane zostały w przesłanym przeze mnie wcześniej Excelu:

 **WSI23L - sieci neuronowe.xlsx**

Osobom, które wciąż nie mają pary sugerowałbym sformować zespoły z innymi osobami, które znajdują się w tej samej sytuacji. Z tego co widzę to tylko 3 osoby nie mają grupy, więc prosi się o zespół 3 osobowy.

W ramach piątych ćwiczeń będą Państwo musieli zaproponować architekturę, zaimplementować, wytrenować i przeprowadzić walidację sieci neuronowej do klasyfikacji ręcznie pisanych cyfr. Zbiór danych do użycia: MNIST - - <http://yann.lecun.com/exdb/mnist/>. (zbiór ten dostępny jest też w ramach sklearn, co może być dla Państwa prostsze w pracy <https://scikit-learn.org/0.19/datasets/mldata.html>)

Należałoby uzyskane rezultaty porównać dla różnej liczby neuronów ukrytych i różnej liczby warstw (sprawdzić tylko kilka przypadków z uwagi na czas nauki sieci). Warto sprawdzić też wpływ współczynnika uczenia na pracę modelu. W raporcie warto opisać wygląd przykładowej sieci i przeprowadzić wstępną analizę zbioru danych. Pokazać jakie przykłady ze zbioru MNIST zostały sklasyfikowane poprawnie a jakie nie (dosłownie kilka obrazów np. że cyfry 1 i 7 są błędne albo że 2 jest wykrywane poprawnie).

Uwagi do pracy:

- wykonują Państwo to zadanie w parach dlatego w raporcie należy wykazać kto wykonał jakie zadania
- rozwiązanie przesyła do mnie tylko jedna osoba z zespołu
- podobnie jak w poprzednich ćwiczeniach nie chcę otrzymywać od Państwa zbioru danych
- na potrzebę propagacji wstecznej przyda się Państwu przypomnieć wiadomości z algorytmu spadku gradientu z pierwszych ćwiczeń
- sugerowałbym użyć bibliotekę numpy do reprezentacji danych (bardziej dla Państwa wygody, to nie jest wymóg jako taki)

Bonusowo mogą Państwo w ramach ciekawości porównać wyniki uzyskane w ramach implementacji z pytorch albo keras/tensorflow i opisać w raporcie (brak tego nie będzie negatywnie oceniany, natomiast może to pozytywnie wpłynąć na ocenę, jeśli będą mieli Państwo tylko drobne błędy w pracy. Podstawą oceny za to ćwiczenie jest bazowa implementacja sieci neuronowej)

Bardziej interesuje mnie poprawność implementacji Państwa sieci (radzę nie przekombinować) niż jej precyzja klasyfikacji. Osobiście nie lubię sieci neuronowych, gdyż można zrobić wszystko dobrze i uzyskać słabe wyniki z powodu słabej nauki modelu, więc mam świadomość, że Państwa rozwiązania nie muszą działać perfekcyjnie. Słaba sieć to rozumiem coś ok. 80%, dlatego przy rozwiązaniach poniżej 50% radzę sprawdzić swoją implementację.

Podpowiedź do zadania:

W skład sieci neuronowej wchodzi:

- warstwa wejściowa (w naszym przypadku 784 piksele)
- neurony warstwy ukrytej
- wagi i wartość progu między warstwami
- funkcja aktywacji (możemy założyć funkcję sigmoidalną)
- warstwa wyjściowa (w naszym przypadku docelowa klasa czyli 0-9 - warto użyć softmax)

Model inicjalizowany jest losowymi wagami i w trakcie nauki modelu wykorzystuje się propagację i wsteczną propagację gradientu by je odpowiednio ustawić. W skrócie obliczenia na macierzach i redukcja błędu predykcji. W pewnym momencie dostaniemy sensowne wagi, które będą oddawały naturę naszego problemu. Oczywiście jak będziemy dalej kontynuować ten proces to przeuczymy sieć.

Pierwotnie chciałem, żeby Państwo sprawdzili wpływ sortowania na działanie sieci neuronowej, ale uznałem, że nauka modelu może trochę zająć, więc napiszę od razu co by Państwo zaobserwowali. Sieć uczy się i ustawia wartości wag zależnie od otrzymanych danych. Jeżeli damy jej dane wybrane w sposób pseudolosowy to dostosuje się ona w miarę ogólnie. Natomiast jakbyśmy posortowali dane klasami to istnieje ryzyko, że sieć najpierw nauczy się pierwszej klasy, po czym jak dojdzie do obserwacji klasy kolejnej to uzna całą tę naukę za szum i nauczy się domyślnie wykrywać tylko elementy klasy drugiej. Oczywiście jest to bardzo uproszczony przykład, ale wynika to głównie z natury procesu nauki takiej sieci.

Jednym z problemów, który mogą Państwo zaobserwować w swoim rozwiązaniach może być problem zaniku gradientu. Dzieje się to, gdy wagi sieci wstępnie ustawimy na zero lub przyrost gradientu jest bardzo mały. Stąd też współcześnie zamiast funkcji sigmoidalnej używa się ReLU, tanH lub innych tego typu funkcji. W Państwa przypadku nie ma to aż tak dużego znaczenia, ale warto o tym pamiętać.

Otrzymali Państwo zadanie z przetwarzania obrazów, dlatego nie spodziewam się bardzo wysokich wyników, gdyż jest kilkanaście cyfr w zbiorze MNIST, które są bardziej bazgrołami niż cyframi i klasyczna sieć neuronowa sobie z nimi nie poradzi. Do tego typu zadań zwykle używa się sieci splotowych (ang. convolutional neural network), ale nie są one tematem WSI.

Formalnie zadanie z sieci neuronowych zaczynamy 28.04 i wypadają nam zajęcia z uwagi na majówkę i Juwenalia, więc oddawanie ćwiczeń będzie dopiero 19.05.

Pozdrawiam,

Marcin Macias