

## SIECI NEURONOWE – ćwiczenie 7

Ćwiczenie 7 dotyczyć będzie sieci rekurencyjnych. Sieci rekurencyjne przydatne być mogą dla danych sekwencyjnych, gdzie porządek danych w czasie jest ważny. Warstwa rekurencyjna przyjmuje na wyjściu sekwencję wektorów i zwraca również sekwencję wektorów (przy czym często interesuje nas działanie na ostatnim wektorze tej sekwencji i np. torch domyślnie zwraca zarówno pełną sekwencję wyjść, jak i ostatni stan warstwy). Odnosnie działania, istotne jest że:

- Warstwa stosuje takie samo przekształcenie dla każdego kroku w sekwencji
- w kroku  $n$ , wejściem jest krok  $n$  wejściowej sekwencji, oraz wyjście z kroku  $n-1$
- konkretne typy warstw rekurencyjnych (RNN/GRU/LSTM) różnią się dodatkowym przekształceniem nakładanym na wyjście z poprzedniego kroku

Oprócz samych warstw rekurencyjnych, dwie istotne nowości:

- warstwa Embedding, której celem jest skonwertowanie wartości ze skończonego zbioru na wektory. Dla problemów językowych, zbiorem może być np. zbiór wszystkich słów języka. Warstwa taka to po prostu macierz (liczba słów) na (wymiar wektora), z której tak samo jak z każdej innej macierzy w obrębie struktury sieci neuronowej jesteśmy w stanie wyliczyć gradient. A to znaczy, że możemy ją optymalizować, co kończy się znajdowaniem dla każdego słowa wektora, który w jakiś sposób oddaje „znaczenie” (a w zasadzie: powiązania między podobnymi znaczeniami, zakodowane w podobieństwie/odległości wektorów).
- Padowanie sekwencji do jednakowej długości. Teksty w zbiorze IMDB są różnej długości, więc jeżeli chcemy uczyć w trybie batchowania, należy dopełnić wektory do równej długości zerami.

Należy skorzystać ze zbioru recenzji IMDB, dostępnego w torchtext oraz kerasie:

<https://keras.io/api/datasets/imdb/>

Do przebadania tym razem jest:

- Typ wybranej warstwy rekurencyjnej (RNN vs. LSTM)
- Wymiar warstwy rekurencyjnej
- Wpływ przycinania sekwencji do niepełnej długości (należy wyznaczyć długość maksymalną przy padowaniu batchy, i porównać wynik uzyskany w ten sposób z wynikami uzyskanymi przycinając wszystkie teksty do ustalonej z góry wartości, np. 10, 20 słów)

**Dla poprawienia wydajności, można pracować na podzbiore mniejszych rozmiarów niż pełny train.**

Ćwiczenie oceniane jest w skali 0-10 pkt, na jego wykonanie jest czas do końca semestru.