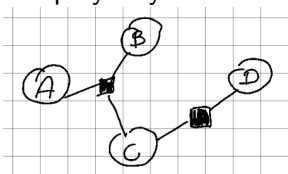


1. Zakładając poniższy korpus z dwoma typami bytów nazwanych: osoba PER i lokalizacja LOC (używamy 5 znaczników: B-LOC, I-LOC, B-PER, I-PER, O):

- Harvey Ford – występujący 9 razy jako PER i raz jako LOC
- Harvey Park – występujący 9 razy jako LOC i raz jako PER
- Myrtle Ford – występujący 9 razy jako PER i raz jako LOC
- Myrtle Park – występujący 9 razy jako LOC i raz jako PER

Oblicz prawdopodobieństwo otagowania „Harvey Ford” jako osobę przez model MEMM, zakładając że użyty klasyfikator bazowy idealnie zamodelował prawdopodobieństwa warunkowe.

2. Dla podanego grafu rozkładu na czynniki podaj postać modelu rozkładu prawdopodobieństwa. Skonstruuj odpowiadający mu nieskierowany graf Markowa i zapisz inny rozkład na czynniki, który byłby z nim kompatybilny.



3. Rozpisz wzór na model liniowy CRF obliczający prawdopodobieństwo sekwencji tagów 'N V N' dla zdania 'Ala ma kota'.
4. Zapisz wzór na algorytm Viterbiego dla modelu CRF
5. Czy do implementacji warstwy CRF w sieci neuronowej należy wykorzystać algorytm propagacji zaufania (w przód i w tył) czy można wykorzystać standardowy autograd? Czy implementując sieć neuronową należy zaimplementować zarówno wektory w przód jak i w tył?
6. Na wykładzie zaprezentowano warstwy CRF z funkcjami kompatybilności inspirowanymi modelem HMM. Narysuj graf rozkładu na czynniki takiego modelu. Zaproponuj własne rozszerzenie (inne od tych podawanych na wykładzie) warstwy CRF o wykorzystywanie dodatkowych informacji. Określ jakie zalety i wady ma zaproponowany model względem warstwy CRF prezentowanej na wykładzie.

7. Oblicz sekwencję wyników zwracanych przez rekurencyjną sieć neuronową dla zdania „Ala ma kota”. Sieć ma dwa neurony rekurencyjne, które mają zerowy wyraz wolny i tożsamościową funkcję aktywacji. Macierz wag tego neuronu to

$$W = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ -1 & 0 \\ -1 & 1 \\ -1 & -1 \\ 0 & 2 \end{bmatrix}$$

Wyjście sieci jest obliczane jednym neuronem sigmoidalnym, również z zerowym wyrazem wolnym. Wektor wag tego neuronu to

$$U = [-1, 1]^T$$

. Słowa wejściowe są kodowane kodowaniem „1 z n” (kolejność: Ala, kota, ma, STOP), stan początkowy sieci to

$$h_0 = [0, 0]^T$$

8. Wyznacz wzór na gradient rekurencyjnej sieci neuronowej po błędzie, który popełnia ona w drugiej chwili czasowej. Gradient wyznacz jedynie dla macierzy wag mnożonej przez stan ukryty (tj. pomiń macierz/część macierzy wag mnożonej przez wejście).
9. Opisz jak wygląda etap predykcji przy wykorzystaniu modelu predykcji sekwencji opartego o rekurencyjną sieć neuronową (bez warstwy CRF). Czy aby otrzymać optymalne dekodowanie możemy wykorzystać programowanie dynamiczne?
10. Rekurencyjna sieć neuronowa modeluje rozkład $P(y_i | x_1^i)$. W jaki sposób należy ją zmodyfikować aby
- modelowała rozkład $P(y_i | x_1^n)$
 - modelowała rozkład $P(y_i | y_{i-1}, x_1^i)$
11. Sieć rekurencyjną można wykorzystać do zadania modelowania języka. Opisz trening takiego modelu i generację tekstu przy jego użyciu. Czy są jakieś warianty sieci rekurencyjnych, których nie można wykorzystać do modelowania języka?
12. Aby ograniczyć problem eksplodującego gradientu stosujemy technikę ograniczania długości gradientu (zmniejszamy go jeśli jest za długi). Dlaczego problemu zanikającego gradientu nie rozwiązujemy poprzez wydłużanie gradientu jeśli jest za krótki?

13. Jednym z najważniejszych problemów w przetwarzaniu sekwencji jest tzw. problem modelowania długich zależności. Jaki związek ma problem zanikającego/eksplodującego gradientu z tym problemem, zakładając modelowanie sekwencji przy użyciu rekurencyjnej sieci neuronowej?
14. Mając nauczony model rekurencyjny np. do modelowania języka, w jaki sposób można empirycznie sprawdzić jak długie zależności są modelowane?
15. Macierz wag W_{hh} w sieci rekurencyjnej może być inicjalizowana losowo, tak jak w przypadku klasycznej sieci neuronowej. Czy może być ona zainicjalizowana samymi zerami? Czy istnieje jakaś inna inicjalizacja tej macierzy stałymi wartościami, która miałaby praktyczny sens i interpretację?
16. Zapisz wzór na neuron LSTM i wyjaśnij intuicje za nim stojące. Jaka jest interpretacja poszczególnych sygnałów? Skąd wzięła się nazwa „long short term memory”?
17. W jaki sposób neuron LSTM rozwiązuje problem zanikającego gradientu?
18. Zapisz wzór na neuron GRU oraz podaj interpretacje poszczególnych neuronów/sygnałów.

