Quiz 9 CS 725 2023

1. Consider the following word embedding matrix for the vocabulary of words = {"man","woman", "cat", "dog", "mouse"}.The (transpose of) embedding matrix is given as

$$E^T = \left\| \begin{array}{c|cccccc} \text{man} & 0.3 & 0.5 & 0.4 & -0.3 & -0.4 & 0.5 \\ \text{woman} & 0.4 & 0.3 & 0.5 & -0.4 & -0.5 & -0.3 \\ \text{cat} & 0 & 0 & 0 & 0.8 & -0.6 & 0 \\ \text{dog} & 0 & 0 & 0 & 0.8 & 0 & 0.6 \\ \text{mouse} & -0.5 & 0 & 0 & 0.5 & -0.5 & 0.5 \end{array} \right\| \quad (1)$$

   We define the similarity between two word embeddings as the dot product of the normalised word embedding vectors ($\frac{x_1 \cdot x_2}{\|x_1\|\|x_2\|}$). Which of the following pair have a similarity greater than 0.5?

   - man, dog
   - man, woman
   - cat, dog
   - cat, mouse
   - woman, mouse
   - woman, cat

   Answer:B,C,D. Observe that norms of all word embeddings is 1. Compute dot products and check for value greater than 0.5. For eg (Man,Woman) $= 0.3*0.4+0.5*0.3+0.4*0.5+(-0.3)*(-0.4)+(-0.4)*(-0.5)+0.5*(-0.3) = 0.64$.

2. You are given word embeddings of words "king" $= x_1$, "man" $= x_2$, "woman" $= x_3$, "male" $= x_4$, "female" $= x_5$. Which of the following would be nearly equal to the word embedding of the word "queen"?

   - $x_1 - x_2 + x_3$
   - $x_1 + x_2 - x_3$
   - $x_1 + x_2 + x_3$
   - $x_1 - x_4 + x_5$
   - $x_1 + x_4 - x_5$
   - $x_1 - x_4 - x_5$

   Answer: A,D. "queen" word embedding would be nearly equal to "king" - "man"/"male" + "woman"/"female"

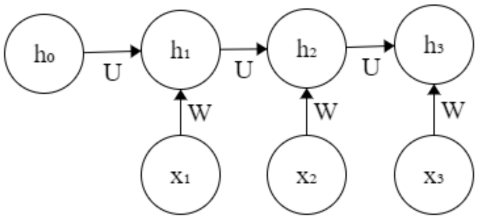
3. What are the drawbacks of using embedding matrix directly as one-hot encoding vectors of the vocabulary of training dataset?

   - One-hot encoding doesn't capture any semantic information about words. All words are treated as unrelated and orthogonal to each other.
   - This leads to high-dimensional vectors, especially for large vocabularies.
   - Most of the elements in a one-hot vector are zero, which can result in inefficiency during computation. It's a waste of memory to store all those zeros.
   - None of these

   Answer: ABC. All of these options are valid drawbacks.

4. Consider a simple (vanilla) RNN cell of the form $h_t = tanh(Wh_{t-1} + Ux_t + b)$ and $o_t = c + Vh_t$ which is passed through the input sequence x_1, x_2, x_3, x_4. Given the dimensions of $x_t \in \mathcal{R}^3$, $h_t \in \mathcal{R}^5$ and $o_t \in \mathcal{R}^7$, find the number of parameters in the RNN cell?

   Answer: 87. Calculate Dimensions of U,W,b,V,c. $5*3(U) + 5*5(W) + 5(b) + 7*5(V) + 7(c) = 87$



5. Part(a) Consider a simplified one dimensional ReLU-RNN cell, $h_t = ReLU(U \cdot h_{t-1} + W \cdot x_t)$ (all are scalars). Calculate $h_1 + h_2 + h_3$, where $U = 1, W = 2, h_0 = -3, x_1 = 1, x_2 = 2, x_3 = 0$.

   Answer: h1=0,h2=4,h3=4.Simple value insertions of h_0, h_1, h_2 recursively into ReLU formula given

6. Part (b) Calculate $\partial h_1 / \partial U$ for the forward pass of ReLU RNN in part (a). Use $\frac{\partial ReLU(x)}{\partial(x)}|_{x=0} = 0$

   Answer:0. Derivative by chain rule is $\frac{\partial ReLU(U \cdot h_0 + W \cdot x_1)}{\partial(U \cdot h_0 + W \cdot x_1)} * \frac{\partial(U \cdot h_0 + W \cdot x_1)}{\partial U}$. Since $U \cdot h_0 + W \cdot x_1 = -1$, hence ReLU's derivative is 0. For ReLU's derivative, see link