Name:_____ **Roll Number:**_____

Date: 09-10-2024 **Quiz -2** **Time allowed: 20 mins**

NLP-A **Total Marks: 11**

Q1 a) How does a neural language model differ from a traditional n-gram language model? (2)

Sparsity problem and storage problem is reduced

b) In the Skip-Gram model, if a target word has 4 context words, how many different pairs of input-output examples will be generated from a single target word in the training set? and how? (2)

for a target word with 4 context words, the Skip-Gram model will generate **4 different input-output pairs**. Each pair consists of the target word as the input and one of the context words as the output.

c) What is the role of word embeddings in neural language models? (2)

- Dimensionality Reduction
- Capturing Semantic Relationships
- Handling Sparsity
- Facilitating Transfer Learning
- Improving Model Performance
- Contextual Representations

d) If two words have very similar context words across the corpus, how will their embeddings be affected in the Skip-Gram model? (2)

In the Skip-Gram model, if two words share very similar context words, their embeddings will be closely positioned in the vector space, reflecting similar meanings or usages. This leads to high cosine similarity between their embeddings, indicating frequent co-occurrence in similar contexts. Additionally, the embeddings will capture shared semantic features, enhancing their similarity.

Q2. Draw a CBOW architecture with vocabulary 10,000, embedding dimension to be 100. What will be the weight matrix for the context and target words. Mention the dimensions at each step of the network. (4)

- **Input Layer: 1x10,000 (representing the one-hot encoded vector of the current word)**
- **W1 (Input to Hidden): 10,000x100 (mapping words to their 100-dimensional embeddings)**
- **Hidden Layer (H): 1x100 (where N is the number of hidden units)**
- **W2 (Hidden to Output): 100 x 10,000**
- **Output Layer: 1 x 10,000 (predicting the probability distribution over the vocabulary)**

Name:_____ **Roll Number:**_____

1. What is the concept of static and customized embeddings in NLM?
   Static such as word to vec...which are not learned when used for any problem such as in NLM.... customized embeddings are contextualized on your corpus and backpropoagated during the training process.

2. Consider a text corpus with the following attributes: (3+2+2)
   - Total words: 100,000
   - Vocabulary size: 4,000
   - Embedding dimension for each word: 200
   - Context window: 5 words
   - Hidden layer: 50 units

   You are tasked with training a **Neural Language Model (NLM)** on this corpus.
   - Draw the architecture of this NLM, specifying the dimensionality of each and weight matrices.

   d= 200, N=50
   1xV Vxd 1xd dxN 1xN Nx V 1xV (change this acc to the context size)

   - How would the dimensions of the embedding layer differ if the vocabulary size were 10,000 instead of 4,000, and embedding dimension is changed to 100?
     - **d= 100, N=50 and V=10000**
     - 1xV Vxd 1xd dxN 1xN Nx V 1xV

   - Discuss how the *context window size* affects the input layer dimensions and the number of parameters (that is weights) in the model.

   In neural language models (NLMs), the context window size determines the number of words considered at a time to predict the next word. A larger context window allows the model to capture longer-range dependencies between words, while a smaller context window focuses on more local relationships. But it will be computationally expensive...so num of weights will increase in the input layer since more words used in this layer..and cost too.

Name:_____ **Roll Number:**_____

3. In skipgram model, how are the negative samples decided for words with a very low frequency?

Could pick w according to their unigram frequency P(w)

More common to use $p_\alpha(w)$

$$P_\alpha(w) = \frac{count(w)^\alpha}{\sum_w count(w)^\alpha}$$

α= ¾ works well because it gives rare noise words slightly higher probability

4. **Given a target word "dog" and its context words ["barks", "loud", "runs"], explain how the Skip-Gram model would update the embeddings for the word "dog" and its context words using negative sampling.**

In the Skip-Gram model with negative sampling, the target word "dog" and its context words ("barks", "loud", "runs") are updated by:

1. Computing the dot product between "dog" and each context word (positive samples).
2. Selecting random words as negative samples and computing their dot products with "dog".
3. Applying the sigmoid function to both positive and negative pairs.
4. Using gradient descent to update the embeddings, making "dog" closer to its true context words and farther from negative samples.
5. The result is that "dog" and its context words are closer in the vector space, improving the word embeddings.

5. Consider a text corpus with the following attributes: (5)

- **Total words: 500,000**
- **Vocabulary size: 5,000**
- **Embedding dimension for each word: 100**
- **Context window: 4 words**
- **Hidden layer: 50 units**

You are tasked with training a Neural Language Model (NLM) on this corpus.

- Draw the architecture of this NLM, specifying the dimensionality of each layer (input, hidden, output) and weight matrices.

Name:_____ **Roll Number:_____**

4 words will be in the input layer.... V=5000, d=100, hidden =50

1x4(d) 4(d)x50 1x50 50x5000 1x5000