Introduction to Large Language Models

Week-4 Assignment

Number of questions: 10   **Total mark: 10 X 1 = 10**

Question 1:

A one-hot vector representation captures semantic similarity between related words like "king" and "queen".

a) True
 b) False

Answer: b
Solution*:* One-hot vectors are orthogonal; no similarity is encoded.

Question 2:

Which method is used to reduce the dimensionality of a term-context matrix in count-based word representations?

a) Principal Component Analysis
 b) Matrix Inversion
 c) Singular Value Decomposition (SVD)
 d) Latent Dirichlet Allocation

Answer: c
Solution*:* SVD is used to obtain low-dimensional representations in latent semantic analysis.

Question 3:

 Which property makes tf-idf a better representation than raw term frequency?

a) It is non-linear
b) It accounts for the informativeness of words
c) It penalizes longer documents
d) It uses hierarchical clustering

Answer: c
Solution*:* IDF downweights common terms like "the" and emphasizes rare but important ones.

Question 4:

What is the purpose of using negative sampling in Word2Vec training?

a) To reduce dimensionality of word vectors
 b) To ensure gradient convergence
 c) To balance class distribution in classification
 d) To simplify softmax computation

Answer: d
Solution*:* Negative sampling avoids computing softmax over the entire vocabulary.

Question 5:

 In skip-gram Word2Vec, the model:

a) Predicts a word given its context
 b) Predicts the next sentence
 c) Predicts surrounding context words given a target word
 d) Learns n-gram frequencies

Answer: c
Solution*:* Skip-gram learns by predicting surrounding words given a center word.

Question 6:

Why does SVD-based word embedding struggle with adding new words to the vocabulary?

a) It uses online learning
 b) It lacks semantic interpretability
 c) It assumes word order
 d) It is computationally expensive to retrain

Answer: d
Solution*:* New words require recomputing the entire decomposition.

Question 7:

Which of the following best describes the term "distributional hypothesis" in NLP?

a) Words with high frequency have greater meaning
 b) Words are defined by their part-of-speech tags
 c) A word's meaning is characterized by the words around it
 d) Words should be normalized before vectorization

Answer: c

Question 8:

In Word2Vec, similarity between word vectors is computed using Euclidean distance.

a) True
 b) False

Answer: b
Solution*:* Similarity is computed using dot product or cosine similarity.

Question 9:

Which method solves the problem of OOV (Out-Of-Vocabulary) words better?

a) One-hot encoding
 b) CBOW
 c) Skip-gram with subsampling
 d) FastText embedding

Answer: d
Solution*:* FastText builds embeddings using character n-grams and handles unseen words.

Question 10:

If the word "economy" occurs 4 times in a corpus, and "growth" appears in a window of 5 words around it 3 times, what is the entry for (economy, growth) in a term-context matrix?

a) 1
 b) 2
 c) 3
 d) 4

Answer: c
Solution*:* It counts co-occurrences in the window — here, 3 times.