Please answer in Chinese or English.

**Vector Semantic and Embedding**

**（一）Suppose we have the following document collection, where each document contains the following words:**

**- Document 1: cat, like, sleep**

**- Document 2: dog, like, play**

**- Document 3: cat, dislike, play**

**- Document 4: dog, like, sleep, play**

**Please calculate the TF-IDF vectors for the words "like" and "sleep" and compute the cosine similarity between them.**

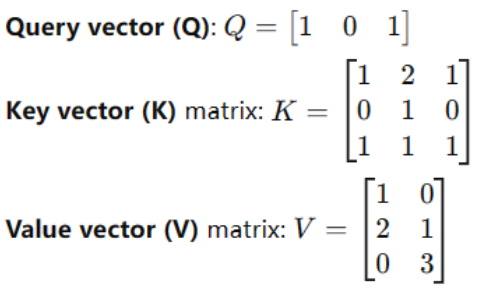
**（二）In the Word2Vec model, what are the purposes of subsampling and negative sampling? What is the difference between the two?**

**（三）In word embedding models, why is cosine similarity commonly used to measure similarity between word vectors? Please explain the advantages of cosine similarity and discuss its differences from Euclidean distance.**

**Neural Language Model**

**（一）Explain the gate structures in LSTM and the purpose of each gate.**

**（二）Here is a calculation question about the attention mechanism. In a simple attention mechanism, given the following matrices:**

****

**Please calculate the attention weights of the query vector Q for each key vector and compute the final output vector based on these weights. The similarity can be calculated by dot product.**

**（三）Explain the main mechanisms and corresponding loss functions in BERT pre-training, and describe the principles and functions of each.**

**Part of Speech Labeling**

**（一）How do common ambiguity issues impact tagging accuracy in part-of-speech (POS) tagging? Provide examples of how the word "back" changes its POS in different contexts, and discuss how POS tagging algorithms (such as the Hidden Markov Model) handle such ambiguities, focusing on how context modeling is applied.**

**（二）In the part-of-speech tagging problem, how is the Forward Algorithm used to calculate the probability of an observation sequence? Please explain the specific steps and advantages of the Forward Algorithm in the context of the Hidden Markov Model structure.**