

# HW2

## Q1

Difference:

Averaged Embeddings: Averaging pre-trained word embeddings to capture semantic information and context.

BoW: Using word frequencies to present text.

Averaged Embedding

Advantage: Captures semantic information

Disadvantage: Ignores word order, considers less meaningful words like "The"

BoW

Advantage: Easy to understand

Disadvantage: Ignore word order, inefficient

## Q3

### 3.1

Usefulness:

Self-attention considers the relationship between words, so considers the whole context instead of one word.

Drawbacks of not using Self-attention:

In reality, the same word can represent different meanings in different articles. If we do not use Self-attention, then we only consider word by word, so we will lose the influence between each word. And we will lose information from the relationship of words.

Similarity:

They both consider the relationship between words.

Difference:

Multi-head attention uses multiple self-attention in parallel, it can be more generalized.

### 3.2

Query(Q), Key(K), Value(V)