Summary: Co-occurrence, Vector Representation, Matrix Construction

1. Co-occurrence

- Definition: Two words appear together within a context window (typically size k).
- Example: In 'I love music more than any other genre', 'love' and 'music' co-occur if k=2.

2. How to Say 'Co-occurrence' in English

- Co-occurrence relationship
- 'They co-occur' or 'They occur together'

3. Vector Representation

- Representing words as numerical vectors for computation.
- Methods: One-hot encoding, Count Vector / TF-IDF, Word Embeddings (Word2Vec, GloVe).

4. Vector Space

- A space where vectors (e.g., word vectors) live.
- Enables comparison by distance, angle, etc.

5. W/D vs W/W Matrices

Word-by-Document (W/D):

- Rows: Words, Columns: Documents
- Values: Word frequency in documents

Word-by-Word (W/W):

- Rows/Columns: Words
- Values: Frequency of two words co-occurring within window k

6. Matrix Construction

- W/D: Count word frequency per document.
- W/W: Slide a window over text and count all word pairs inside.

7. Beyond Word Pairs

- n-gram (e.g., 3-gram: 'I love NLP')
- Higher-dimensional tensors

- Word combinations within a window

Word-by-Word (W/W) Co-occurrence Matrix (window size = 1)

Word	I	love	NLP	deep	learning	loves	me
I	0	2	0	0	0	0	0
love	2	0	1	1	0	0	0
NLP	0	1	0	0	0	0	0
deep	0	1	0	0	2	0	0
learning	0	0	0	2	0	1	0
loves	0	0	0	0	1	0	1
me	0	0	0	0	0	1	0

Word-by-Document (W/D) Matrix

Word	Doc1	Doc2	Doc3
I	1	1	0
love	1	1	0
NLP	1	0	0
deep	0	1	1
learning	0	1	1
loves	0	0	1
me	0	0	1