Exercise 1: Outer Product

Step 1.1 Calculate the outer product $\mathbf{a} \otimes \mathbf{b}$ for the vectors $\mathbf{a} = [1, 2, 3]$ and $\mathbf{b} = [1, 3, 1]$.

> *Think:* How does each element in the first vector interact with each element in the second vector? Which elements of the resulting matrix become the largest?

Step 1.2 For each of the following vector pairs, visualize their outer product using shading patterns:

1. $\mathbf{u} = [5, 4, 3, 2, 1]$ and $\mathbf{v} = [5, 4, 3, 2, 1]$
2. $\mathbf{u} = [1, 2, 4, 0, 0]$ and $\mathbf{v} = [3, 1, 2, 0, 0]$
3. $\mathbf{u} = [1, 1, 1, -1, -1]$ and $\mathbf{v} = [1, 1, 1, -1, -1]$

(a)

	1	2	3	4	5
1					
2					
3					
4					
5					

(b)

	1	2	3	4	5
1					
2					
3					
4					
5					

(c)

	1	2	3	4	5
1					
2					
3					
4					
5					

Exercise 2: Matrix Decomposition

Step 2.1 Consider the following matrix representing a small network:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Try to "decompose" this matrix into the outer product of two vectors with minimal error. Sketch the vectors and use shading to represent their values.

Step 2.2 Now examine this more complex matrix:

$$\begin{bmatrix} 8 & 4 & 2 & 1 & 0 \\ 4 & 4 & 2 & 0 & 0 \\ 2 & 2 & 2 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Can you decompose this into a sum of two outer products? Sketch the vectors for each outer product and use shading.

Step 2.3 Building on your decomposition from Step 2: If you had to keep only one of the two outer products, which would you choose and why? What information about the network would be preserved?

Exercise 3: Word Embeddings and Document Analysis

Step 3.1 Examine these short news headlines:

1. "Tech company launches new smartphone"
2. "Tech news update"
3. "Major tech company announces news update about smartphone technology advances"
4. "Smartphone sales rise in tech sector"
5. "New restaurant opens downtown"
6. "Local restaurant offers new menu"

Count how many times each word appears in each headline and fill in the grid below:

Word appearances in each headline:

	D1	D2	D3	D4	D5	D6
tech						
smartphone						
restaurant						
new						

Think: What patterns do you notice about how words are distributed across different types of headlines?

Step 3.2 Compare these two headlines:

i. Short headline: "Tech news update"

ii. Longer headline: "Major tech company announces news update about smartphone technology advances"

Think: The word "tech" appears once in each headline. Why should these appearances be weighted differently? How does the length of each headline affect the significance of the word?

Step 3.3 Let's quantify word importance using term frequency (TF):

$$\text{TF} = \frac{\text{number of times word appears in document}}{\text{total number of words in document}}$$

Calculate the TF for "tech" in both headlines:

Headline	Word Count	Times Words	Total Frequency	Term (TF)
Short				
Long				

Think: How does TF help capture the relative importance of words in documents of different lengths?

Step 3.4 Now let's consider how specific or general each word is across all documents using the inverse document frequency (IDF):

$$\text{IDF} = \log\left(\frac{\text{total number of headlines}}{\text{number of headlines containing the word}}\right)$$

Word	Appears in how many docs?	IDF (Relative Specificity) (darker = more specific)
tech		☐
smartphone		☐
restaurant		☐
new		☐

Think: What does IDF tell us about which words are topic-specific versus general-purpose?

Step 3.5 Combine TF and IDF to create word embeddings:

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

TF-IDF for each word and document:

	D1	D2	D3	D4	D5	D6
tech						
smartphone						
restaurant						
new						

> *Think:* How does TF-IDF combine the benefits of both TF and IDF to create more meaningful word representations?

Step 3.6 Decompose the TF-IDF matrix into meaningful components:

TF-IDF matrix

$\simeq$

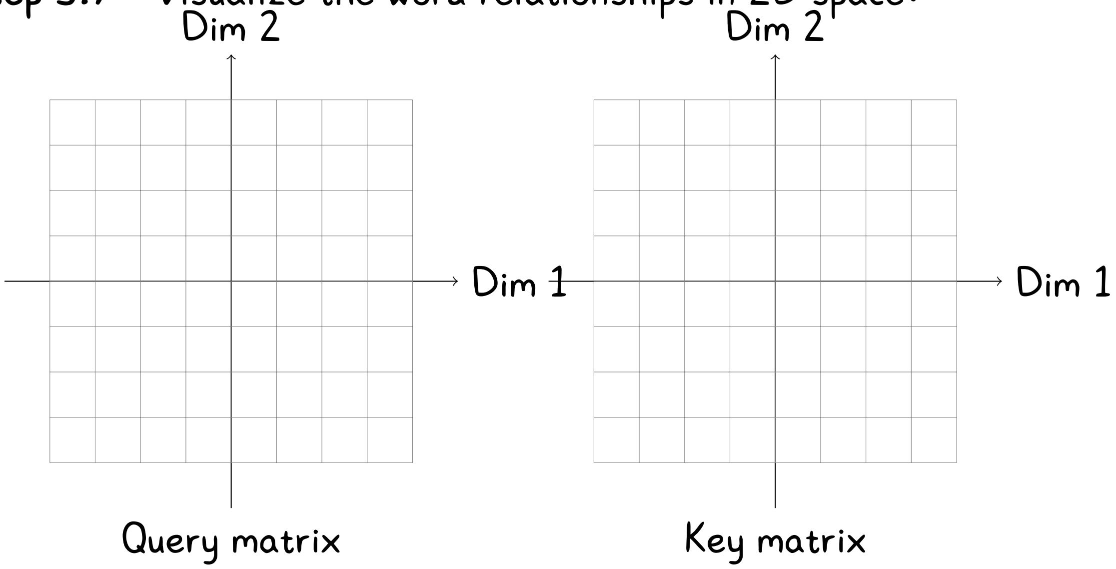
Query matrix

$\times$

Key matrix

> *Think:* What aspects of word/document meaning are captured by the query and key matrices?

Step 3.7 Visualize the word relationships in 2D space:



Place "apple", "banana", and "curry" on the first matrix.

> *Think:* What semantic relationships between words are revealed by their positions in the 2D space?

Exercise 4: Pointwise Mutual Information

Step 4.1 Consider again our news headlines, but now let's look at which words appear together in the same headline. Create a word-word matrix:

Word co-occurrences:

	tech	smartphone	restaurant	new
tech				
smartphone				
restaurant				
new				

Think: What is the relationship between the word-document matrix and the word-word matrix?

Step 4.2 Consider the words "tech" and "smartphone". They appear together in 3 headlines (Probability of 3/6 = 0.5). If words appeared randomly, how many times would they appear together?

Step 4.3 How about "tech" and "restaurant"?

Step 4.4 Calculate the expected co-occurrence and compare with observed:

	tech-smartphone	tech-restaurant
Observed co-occurrences		
Expected co-occurrences		
Ratio (Observed/Expected)		

Think: What does it mean when the ratio is greater than 1? Less than 1?

Step 4.5 The Pointwise Mutual Information (PMI) is defined as:

$$\text{PMI}(word1, word2) = \log\left(\frac{\text{P(word1, word2)}}{\text{P(word1) P(word2)}}\right)$$

where $P(word1, word2)$ is the probability of co-occurrence. It is the co-occurrence count divided by the sum of all co-occurrence counts. $P(word1)$ is the probability of word1 (e.g., 0.5 for "tech"), and $P(word2)$ is the probability of word2 (e.g., 0.5 for "smartphone"). Complete the PMI matrix for our words:

PMI values (darker = stronger positive association):

Think: How does PMI capture different information than raw co-occurrence counts?