1 Understanding Outer Products

1. Consider two vectors: $\mathbf{a} = [1, 2, 3]$ and $\mathbf{b} = [1, 3, 1]$. Calculate their outer product $\mathbf{a} \otimes \mathbf{b}$.

2. Without exact calculations, sketch the resulting matrix from the outer product $\mathbf{u} \otimes \mathbf{v}$ of the following pairs of vectors.

   (a) $\mathbf{u} = [5, 4, 3, 2, 1]$ and $\mathbf{v} = [5, 4, 3, 2, 1]$

   (b) $\mathbf{u} = [1, 2, 4, 0, 0]$ and $\mathbf{v} = [3, 1, 2, 0, 0]$

   (c) $\mathbf{u} = [1, 1, 1, -1, -1]$ and $\mathbf{v} = [1, 1, 1, -1, -1]$

   Use shading to represent the relative values in the matrix (darker shades for larger values).

(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					

3. Look at the matrices you've sketched. If these matrices represented networks, what kind of network structures might each of them represent?

2 Decomposing Matrices

4. Consider the following matrix representing a small network:

$$\begin{bmatrix} 3 & 2 & 1 & 0 & 1 \\ 2 & 3 & 0 & 1 & 0 \\ 1 & 0 & 3 & 2 & 1 \\ 0 & 1 & 2 & 4 & 2 \\ 1 & 0 & 1 & 2 & 3 \end{bmatrix}$$

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.

5. Now consider this 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.

6. For the matrix in question 7, 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?

3 Neural Embeddings

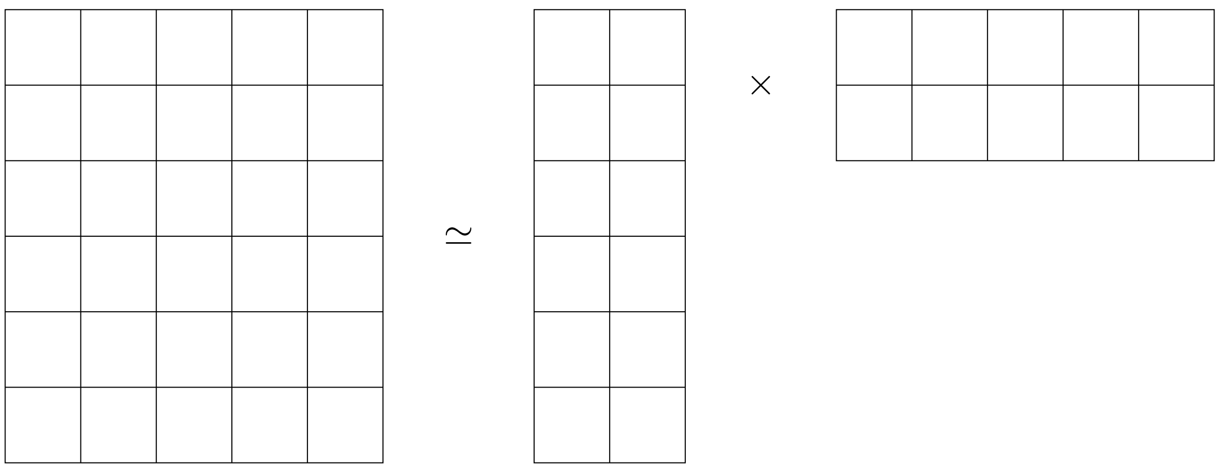
7. Fill in the following matrix with the number of times each pair of words co-occurs in the sentences below.

   1. He likes hot coffee in morning
   2. She loves hot tea in night
   3. He hates cold juice in morning
   4. They dislike cold rain in night
   5. People in Paris are warm
   6. Paris has cold rain in winter
   7. My relatives in London gave me a warm welcome
   8. London has many rain in winter

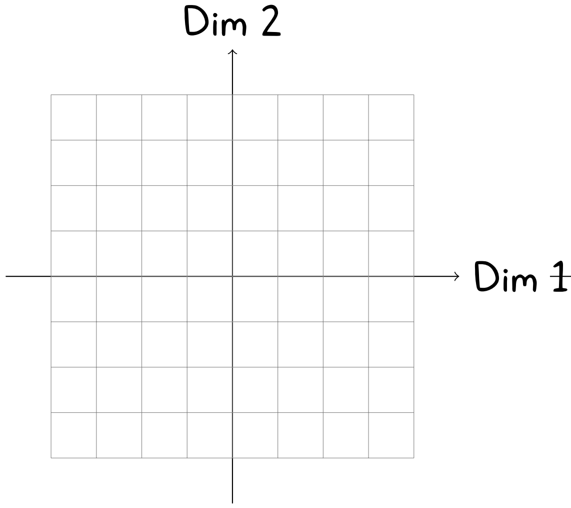
	hot	cold	warm	rain	in
likes					
loves					
hates					
dislike					
Paris					
London					

8. Fill in these matrices with shaded cells to represent the approximate decomposition of the co-occurrence matrix (darker shades for larger values).

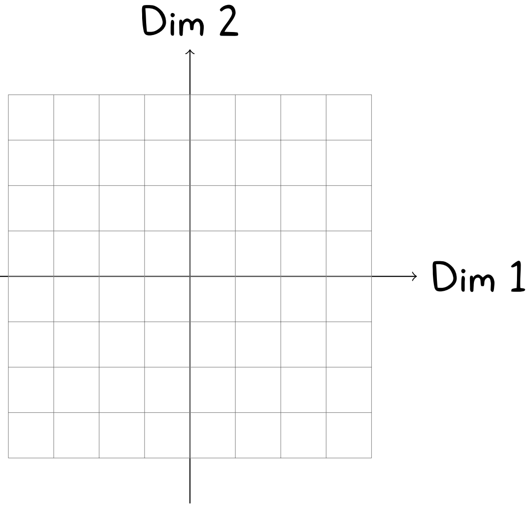
Co-occurrence matrix Query matrix Key matrix



9. In the grids below, use the first matrix to represent the words in 2D space (left), and the second matrix to represent the context of the words in 2D space (right).



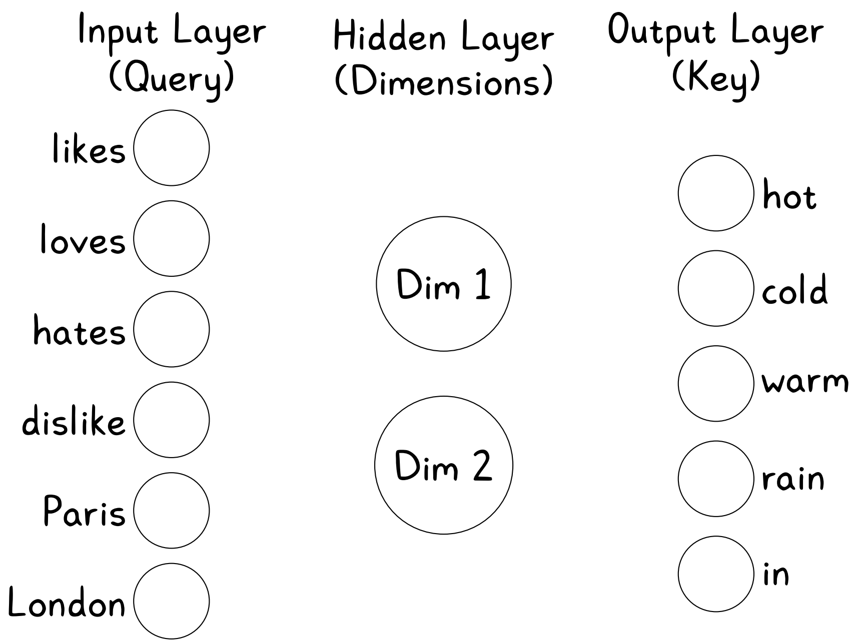
Query matrix



Key matrix

10. On your 2D plot for the first matrix, mark where you would expect these new words to appear: "enjoyed", "Berlin", "dislikes". Explain your reasoning.

11. Sketch the matrix decomposition as a neural network. Represent the absolute value of the matrix entries by the width of the lines connecting the words and dimensions.

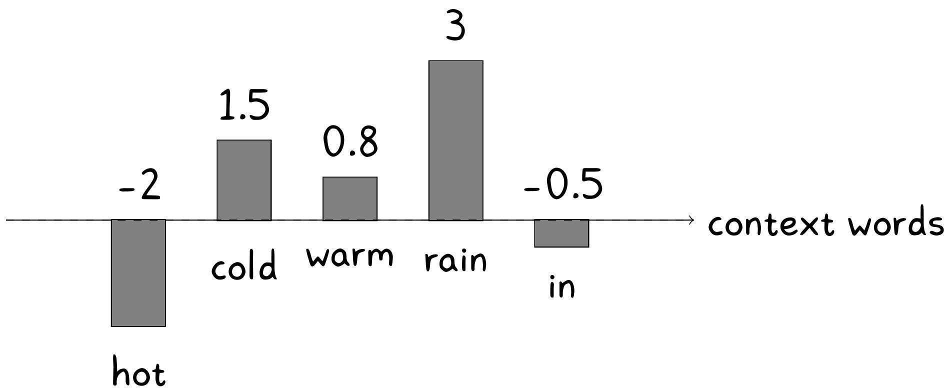


Part 4: Softmax

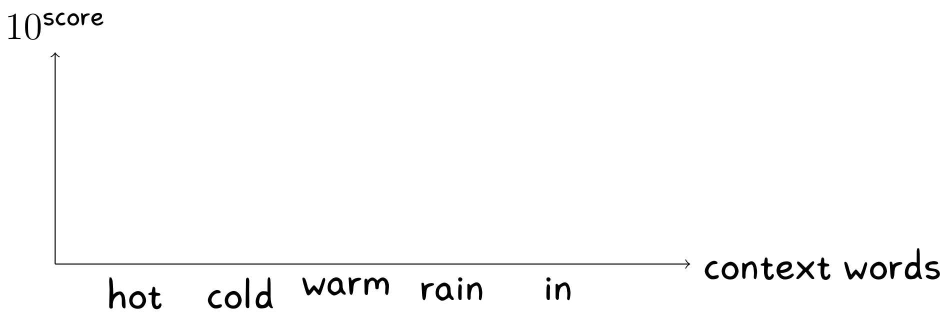
Consider the word "Paris". After taking the matrix decomposition and then reconstructing the co-occurrence matrix, we get the following values for its co-occurrence with other words:

	hot	cold	warm	rain	in
Paris	-2	1.5	0.8	3	-0.5

The following is the bar chart of these values:



12. Now, imagine these values are exaggerated by taking 10 to the power of each value. Sketch how the relative heights would change without exact calculations. Put the your estimate of the values on the y-axis for each word.



13. Based on your estimate, normalize these values so that they sum to 1.