# Cosine Similarity

Implement normalized cosine similarity to evaluate the embedding model.

Chapter Goals:

- Learn about cosine similarity and how it's used to compare embedding vectors
- Create a function that computes cosine similarities for a given word

## A. Vector comparison

In mathematics, the standard way for comparing vector similarity is through *cosine similarity*. Since word embeddings are just vectors of real numbers, we can use also cosine similarity to compare embeddings for different words.

For two vectors, $\mathbf{u}$ and $\mathbf{v}$, the equation for cosine similarity is

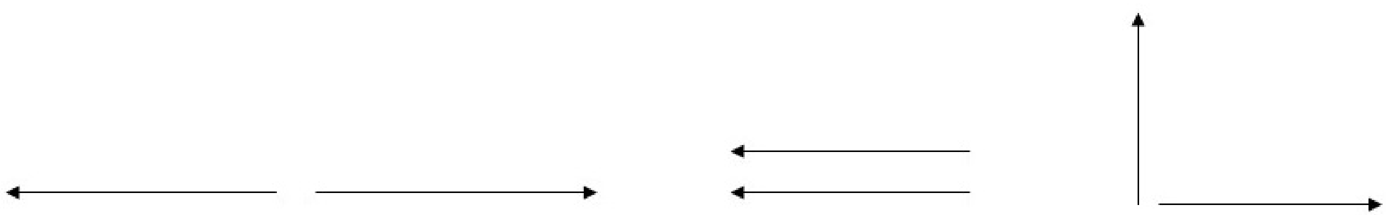
$$\text{cos sim} = \frac{\mathbf{u}}{||\mathbf{u}||_2} \cdot \frac{\mathbf{v}}{||\mathbf{v}||_2}$$

where $||v||_2$ represents the L2-norm of vector $v$, and $\cdot$ represents the dot product operation.

We refer to the quantity $\frac{v}{||v||_2}$ as the L2-normalization of vector $v$.

## B. Correlation

The cosine similarity measures the *correlation* between two vectors, i.e. how closely related the two vectors are. The range of values for cosine similarity is [-1, 1]. A value of 1 means the vectors are perfectly identical, a value of -1 means the vectors are complete opposites, and a value of 0 means the vectors are *orthogonal* (i.e. completely uncorrelated).



Note that the cosine similarity values are based on a spectrum, so we can measure correlation based on the cosine similarity's proximity to 1, 0, or -1. For example, we would expect the word embeddings for "orange" and "juice" to have a cosine similarity close to 1, since they are often used in the same context in conjunction with one another. On the other hand, we would expect "good" and "bad" to have a negative cosine similarity, since they are antonyms. And in most text corpuses, "chocolate" and "fence" would have a cosine similarity near 0, since they tend to be unrelated.

For example, imagine two vectors $v_1$ and $v_2$.

$$v_1 = \begin{bmatrix} 2 \\ 0 \\ 6 \end{bmatrix}, v_2 = \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}$$

The L2 norm of $v_1$ is $\sqrt{2^2 + 0^2 + 6^2} = 6.324$

The L2 norm of $v_2$ is $\sqrt{4^2 + 2^2 + 5^2} = 6.708$

$$\frac{v_1}{6.324} = \begin{bmatrix} 0.316 \\ 0 \\ 0.949 \end{bmatrix}$$

$$\frac{v_2}{6.708} = \begin{bmatrix} 0.596 \\ 0.298 \\ 0.745 \end{bmatrix}$$

$$\begin{bmatrix} 0.316 \\ 0 \\ 0.949 \end{bmatrix} \cdot \begin{bmatrix} 0.596 \\ 0.298 \\ 0.745 \end{bmatrix} = 0.895$$

This number is very close to one, which means that $v_1$ and $v_2$ are very similar vectors