Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: ML-Engineer.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Vector Search

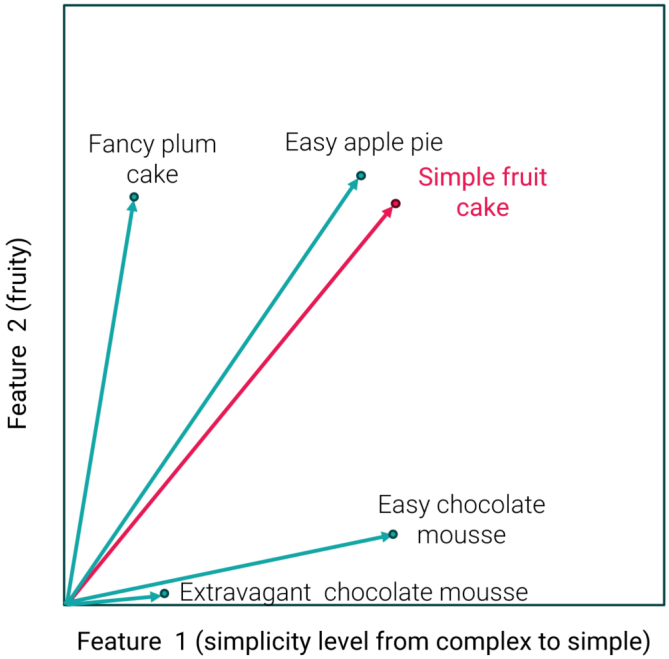
Vector search, which is also known as semantic search, is a technology that improves search accuracy by understanding the meaning (semantics) of the data and relations between its parts. Unlike traditional search, vector search efficiently handles synonyms, typos, ambiguous language, and broad or fuzzy queries. This is because it focuses on meaning, not just keywords.

Imagine that you are searching for a dessert to cook during the weekend. In a traditional search engine, the "simple fruit cake" query will reveal only websites that include these keywords. However, a vector search engine is able to provide results like "apple pie in 20 minutes" or "easy summer desserts", which capture the essence of the query and align with your desire for a straightforward dessert option, providing more valuable results to you.

What is a vector embedding?

A vector or vector embedding is a numerical representation of any kind of unstructured data (e.g. texts, images, videos, audio). It captures its meaning while being easy and efficient to compute with. Think of it like this: imagine you have a collection of cake recipes. You can convert each recipe into a vector embedding, which is like a unique numerical code that represents the recipe's characteristics (ingredients, cooking methods, flavors, etc.).

Once all the recipes are encoded into embeddings, we can perform a similarity search. This means we can compare the vectors to see how similar the recipes are. For example, the vector for an easy apple pie recipe would be close to the vector for a simple fruit cake recipe because they share similar characteristics (e.g. simplicity, fruitiness). On the other hand, the vector for an extravagant chocolate mousse cake would be farther away because it involves different ingredients and methods.



How to compare vectors?

i. Cosine similarity:

$$\boxed{cos(\Theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|}}$$

ii. Eucledian distance:

$$\boxed{d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|}$$

What is a vector database?

A vector database is a specialized database designed to store, manage, and search vectors efficiently. This efficiency is crucial for handling large datasets and performing fast vector similarity searches. Also, with a vector database, the knowledge of AI models can be improved, adapted, and updated. Therefore, today, most AI apps use a vector database.
