Semantic Search

Traditional search

StockCode	Description	Quantity	InvoiceDate	UnitPrice
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39

Traditional search

StockCode	Description	Quantity	InvoiceDate	UnitPrice
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39

SELECT * FROM Products WHERE UnitPrice > 3

Traditional search

StockCode	Description	Quantity	InvoiceDate	UnitPrice
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39

SELECT * FROM Products WHERE Quantity=6

Traditional search

StockCode	Description	Quantity	InvoiceDate	UnitPrice
85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39
84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75
84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39

```
SELECT * FROM Products WHERE Description LIKE '%HEART%';
```

Limitations

- We can only perform numeric comparisons and string search
- E.g. what if we wanted to query for all alcohol drinks

```
SELECT * FROM Products WHERE Description LIKE '%WHISKY%'
                          OR Description LIKE '%WHISKEY%'
                          OR Description LIKE '%GIN%'
                          OR Description LIKE '%RUM%'
                          OR Description LIKE '%BEER%'
                          OR Description LIKE '%WINE%'
                          OR Description LIKE '%CHAMPAGNE%'
                                    …
```

Semantic search

- What if there was a way to make a **semantic** query instead of a keyword-based one?

  e.g. “select all alcoholic beverages”

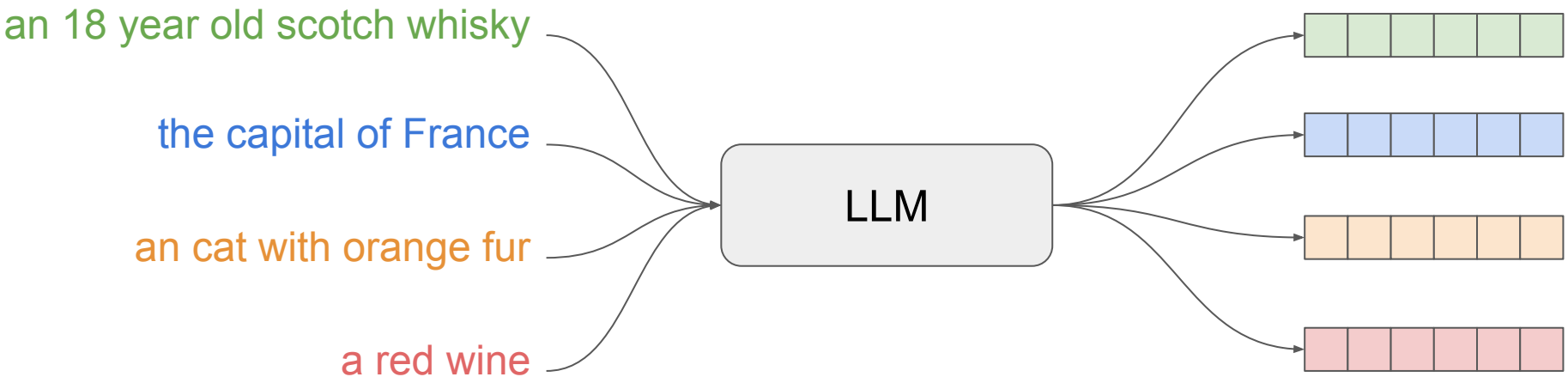
- How could we make this work?

  1. A way to extract the meaning from a sentence
  2. A way to query the existing data with the previously extracted semantics
  3. A database that could store the existing data efficiently

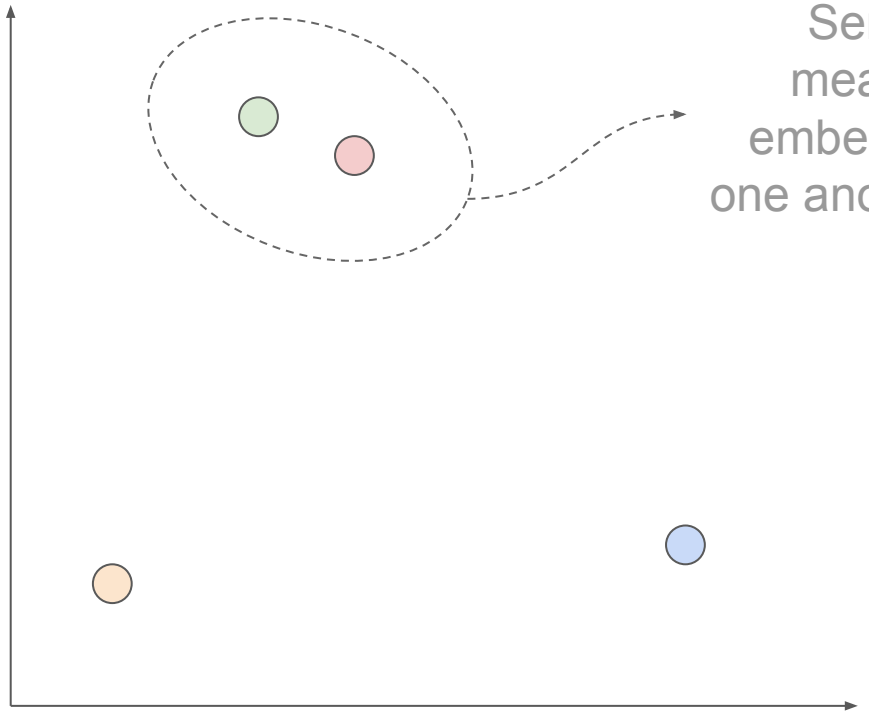
Component 1: Embeddings

- Nowadays we have LLMs that are able to extract the meaning from a word or even a sentence
- These output of an embedding model is a vector (or as we call them **embeddings**)
- The idea is that sentences with a similar meaning will produce embeddings that are closer in the vector space

Component 1: Embeddings



Component 1: Embeddings



Sentences with similar meanings will have their embeddings placed close to one another in the vector space

Component 2: Distance metric

- We need to use a distance metric to compare the embeddings
- Remember: closer embeddings → semantically related sentences
- The most common metric used is **cosine similarity**

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

How would this work in our example?

1. Extract the embeddings from all the descriptions

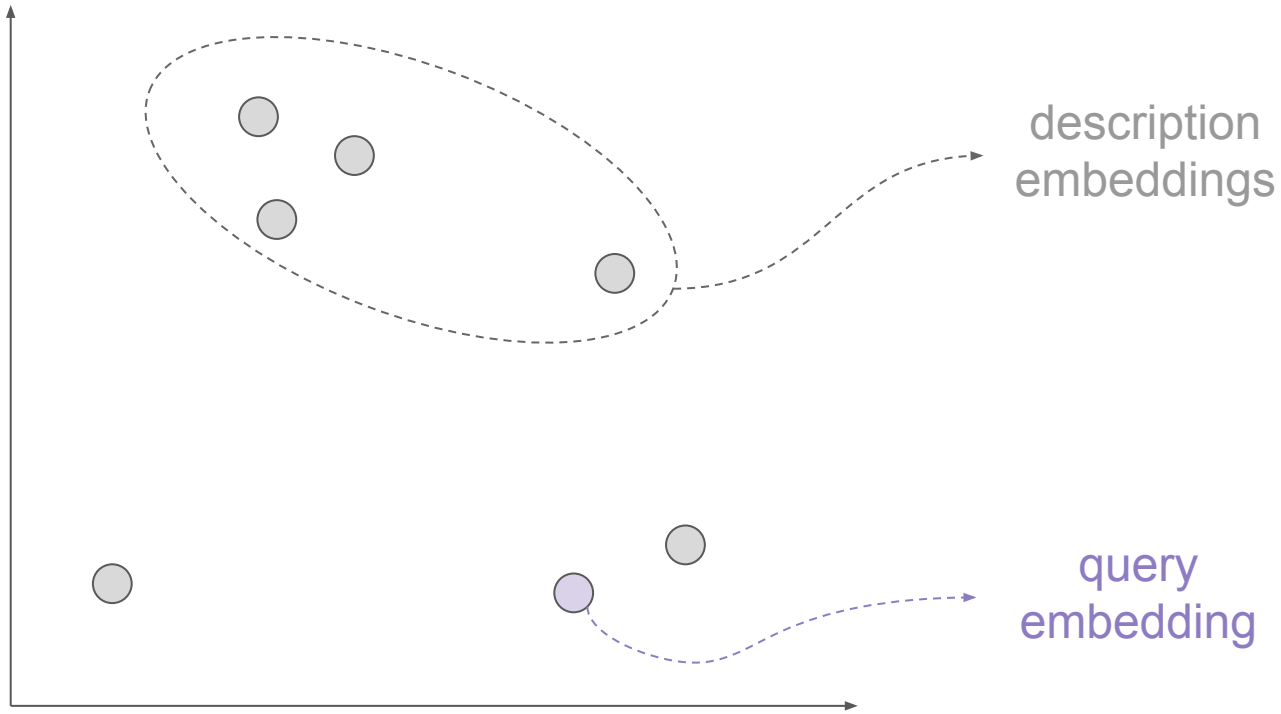
Description	Embedding
WHITE HANGING HEART T-LIGHT HOLDER	[-0.93, 1.15, -0.03, -1.92, 0.23, …]
CREAM CUPID HEARTS COAT HANGER	[-0.22, -1.02, 1.04, 0.43, -0.12, …]
KNITTED UNION FLAG HOT WATER BOTTLE	[0.18, 0.15, 0.65, 0.51, -0.44, …]
ARDBEG UIGEADAIL MALT WHISKY 700ML ISLEY	[1.13, 0.55, -0.49, -0.16, 1.42, …]
RED WOOLLY HOTTIE WHITE HEART.	[-0.82, -0.77, -0.13, -0.61, 0.55, …]

How would this work in our example?

2. When we want to perform a semantic query, extract the embedding from the query sentence

Query	Embedding
ALCOHOLIC BEVERAGE	[-0.11, -0.76, 1.12, 0.47, -0.71, …]

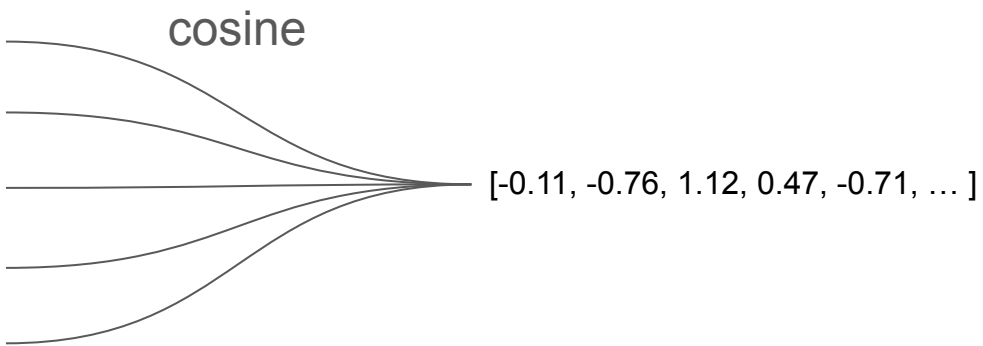
How would this work in our example?



How would this work in our example?

3. Compare the distance of the query's embedding to those of the descriptions

Description Embeddings
[-0.93, 1.15, -0.03, -1.92, 0.23, …]
[-0.22, -1.02, 1.04, 0.43, -0.12, …]
[0.18, 0.15, 0.65, 0.51, -0.44, …]
[1.13, 0.55, -0.49, -0.16, 1.42, …]
[-0.82, -0.77, -0.13, -0.61, 0.55, …]



How would this work in our example?

4. Return the N descriptions with the highest cosine similarity to the query or are above a confidence threshold

Description	Embedding	Similarity
WHITE HANGING HEART T-LIGHT HOLDER	[-0.93, 1.15, -0.03, -1.92, 0.23, …]	0.02
CREAM CUPID HEARTS COAT HANGER	[-0.22, -1.02, 1.04, 0.43, -0.12, …]	0.03
KNITTED UNION FLAG HOT WATER BOTTLE	[0.18, 0.15, 0.65, 0.51, -0.44, …]	0.01
ARDBEG UIGEADAIL MALT WHISKY 700ML ISLEY	[1.13, 0.55, -0.49, -0.16, 1.42, …]	0.78
RED WOOLLY HOTTIE WHITE HEART.	[-0.82, -0.77, -0.13, -0.61, 0.55, …]	0.05

Component 3: Vector database

- Specialized database for storing vectors
- Mainly used for similarity or semantic search, recommender systems, etc.
- Some DBs support storing metadata along with vectors and prefiltering according to regular queries
- Most use the Hierarchical Navigable Small World (HNSW) algorithm for indexing the vectors

Component 3: Vector database

Weaviate