Vectors in SQL Server

Article • 09/11/2024

Applies to: ✓ Azure SQL Database ✓ Azure SQL Managed Instance

Vectors are ordered arrays of numbers (typically floats) that can represent information about some data. For example, an image can be represented as a vector of pixel values, or a string of text can be represented as a vector or ASCII values. The process to turn data into a vector is called vectorization.

Embeddings

Embeddings are vectors that represent important features of data. Embeddings are often learned by using a deep learning model, and machine learning and AI models utilize them as features. Embeddings can also capture semantic similarity between similar concepts. For example, in generating an embedding for the words `person` and `human`, we would expect their embeddings (vector representation) to be similar in value since the words are also semantically similar.

Azure OpenAI features models to create embeddings from text data. The service breaks text out into tokens and generates embeddings using models pretrained by OpenAI. To learn more, see Creating embeddings with Azure OpenAI.

Once embeddings are generated, they can be stored into a SQL Server database. This allows you to store the embeddings alongside the data they represent, and to perform vector search queries to find similar data points.

Vector search

Vector search refers to the process of finding all vectors in a dataset that are similar to a specific query vector. Therefore, a query vector for the word `human` searches the entire dataset for similar vectors, and thus similar words: in this example it should find the word `person` as a close match. This closeness, or distance, is measured using a distance metric such as cosine distance. The closer vectors are, the more similar they are.

SQL Server provides built-in support for vectors via the `VECTOR` data type. Along with the data type there are dedicated functions to operate on vectors. For example it is possible to

find the distance between two vectors using the `VECTOR_DISTANCE` function. The function returns a scalar value with the distance between two vectors based on the distance metric you specify.

Since vectors are typically managed as arrays of floats, creating a vector can be done simply casting a JSON array to a `VECTOR` data type. For example, the following code creates a vector from a JSON array:

```SQL
SELECT CAST('[1.0, 2.0, 3.0]' AS VECTOR(3)) AS vector;
```

Same goes for converting a vector into a JSON array:

```SQL
DECLARE @v VECTOR(3)
SELECT CAST(@v AS NVARCHAR(MAX)) AS vector;
-- or
SELECT CAST(@v AS JSON) AS vector;
```

More details on how to use vectors in SQL Server can be found in the following articles:

- Vector Data Types
- Vector Functions
- Azure SQL DB Vector Search Samples

Related content

- Intelligent applications with Azure SQL Database

Feedback

Was this page helpful?   👍 Yes   👎 No

Provide product feedback | Get help at Microsoft Q&A