

Word2Vec [Reading Material]

What is Word2Vec?

Word2Vec (Word to Vector) is a technology used to represent the relationships between different words in the form of a graph. This technology is widely used in machine learning for embedding and text analysis. It was introduced in 2013 by researchers at Google.

Unlike earlier representations (like one-hot encoding), Word2Vec produces:

- Dense vectors (e.g., 100-300 dimensions)
- Semantic meaning captured in geometry
- Linear relationships (e.g., king - man + woman ≈ queen)

Why Do We Need Word2Vec?

We need Word2Vec because there are multiple problems with One-Hot Encoding.

For eg - if vocabulary size = 10,000:

- Each word = 10,000-dimensional sparse vector
- No semantic similarity captured
- "cat" and "dog" equally distant from "table" [which is not logical]

So Word2Vec solves this by:

- Learning vectors from context
- Encoding semantic similarity
- Reducing dimensionality

To learn more about Word2Vec, go through the following links:

1. [Link1](#)
2. [Link2](#)
3. [Link3](#)