Text embeddings similarity

We perform below steps to find similarity between two embeddings:

1. **Vector Representation**: Text is first converted into vectors using methods like Word2Vec, GloVe, or transformer models (e.g., BERT). Each vector represents the semantic meaning of the text.
2. **Dot Product**: The cosine similarity between two vectors is calculated using the dot product. For two vectors AAA and BBB:
   Dot Product=A⋅B=∑i=1nAi×Bi\text{Dot Product} = A \cdot B = \sum_{i=1}^{n} A_i \times B_iDot Product=A⋅B=i=1∑nAi×Bi
3. **Magnitude**: Each vector's magnitude (or length) is computed:
   Magnitude of A=∑i=1nAi2\text{Magnitude of A} = \sqrt{\sum_{i=1}^{n} A_i^2}Magnitude of A=i=1∑nAi2 Magnitude of B=∑i=1nBi2\text{Magnitude of B} = \sqrt{\sum_{i=1}^{n} B_i^2}Magnitude of B=i=1∑nBi2
4. **Cosine Similarity Formula**: The cosine similarity is then calculated as:
   Cosine Similarity=A⋅B // A // × // B // \text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \times \|B\|}Cosine Similarity= // A // × // B // A⋅B
   where // A // \|A\| // A // and // B // \|B\| // B // are the magnitudes of vectors AAA and BBB, respectively.

Interpretation:

- The value of cosine similarity ranges from -1 to 1.
  - **1**: Indicates that the vectors are identical in direction (high similarity).
  - **0**: Indicates orthogonality (no similarity).
  - **-1**: Indicates opposite directions (completely dissimilar).