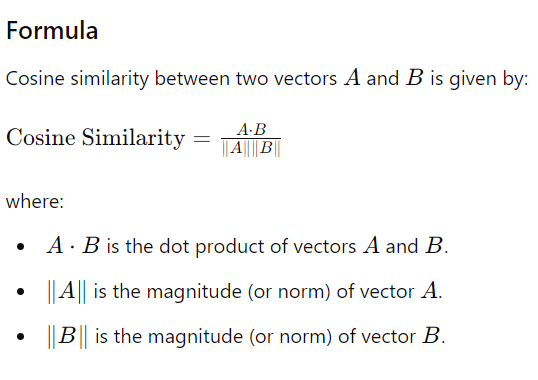
**Assignment No. 1**

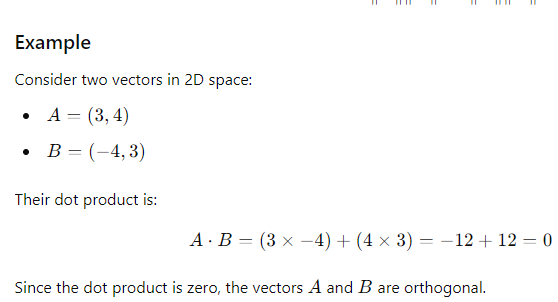
**Cosine Similarity**

* Cosine similarity is a metric used to measure how similar two vectors are in an inner product space.
* It’s commonly used in various applications, such as text analysis, to determine how similar two documents or pieces of text are.
* Here’s a basic rundown of how cosine similarity works:



* **Properties**
* **Range**: The cosine similarity value ranges from -1 to 1.
  + **1** indicates that the vectors are identical in direction.
  + **0** indicates orthogonality or no similarity.
  + **-1** indicates that the vectors are diametrically opposed (opposite directions).
* **Normalization**: Cosine similarity normalizes for the magnitude of the vectors, focusing only on the direction.
* **Applications**
* **Text Similarity**: Used in natural language processing to compare the similarity of documents or sentences.
* **Recommendation Systems**: Helps to find similar items or users based on their preferences.
* **Clustering**: In clustering algorithms like k-means, cosine similarity can be used to measure the similarity between data points.

**Orthogonal Vectors**

* Orthogonal vectors are vectors that are perpendicular to each other in a given space.
* In mathematical terms, two vectors A and B are orthogonal if their dot product is zero:
* A⋅B=0
* **Applications**
* **Computer Graphics**: Orthogonality is important for transformations and rotations.
* **Signal Processing**: Orthogonal signals can be analyzed separately without interference.
* **Data Science and Machine Learning**: Orthogonal vectors simplify various algorithms and models, such as Principal Component Analysis (PCA), where orthogonal components (principal components) capture the maximum variance.

**Application Of cosine Similarity: SVD & PCA**

**Singular Value Decomposition (SVD)**

**Definition**: SVD is a factorization method that decomposes a matrix MMM into three matrices: M = UΣVT where:

* U contains the left singular vectors,
* Σ is a diagonal matrix with singular values,
* VT contains the right singular vectors.

**Applications**:

* **Dimensionality Reduction**: Reduces the number of features while retaining the most significant ones.
* **Latent Semantic Analysis (LSA)**: In text analysis, SVD helps identify patterns and relationships in text data.
* **Noise Reduction**: SVD can be used to filter out noise from data by keeping only the most significant singular values.

**Relation to Cosine Similarity**:

* In the context of text analysis, after applying SVD, documents and terms can be represented in a reduced-dimensional space. Cosine similarity can then be used to measure the similarity between these lower-dimensional representations.

**Principal Component Analysis (PCA)**

**Definition**: PCA is a technique to reduce the dimensionality of data while preserving as much variance as possible. It transforms the original data into a new set of orthogonal axes (principal components), ordered by the amount of variance they capture.

**Applications**:

* **Dimensionality Reduction**: Simplifies data for analysis, visualization, and modeling.
* **Noise Reduction**: By focusing on the most significant principal components, noise and less important variations can be reduced.

**Relation to Cosine Similarity**:

* **Feature Space Transformation**: PCA transforms the feature space into a new basis where the axes (principal components) are orthogonal. After applying PCA, cosine similarity can be used in the transformed space to measure similarities.

**How They Interconnect**

1. **Text Analysis**:
   * **SVD** can be used to decompose term-document matrices, revealing latent structures and reducing dimensionality.
   * **PCA** can also be applied to these matrices (often after SVD) to reduce dimensionality further.
   * **Cosine Similarity** is then used in this reduced-dimensional space to measure the similarity between documents or terms.
2. **Data Reduction**:
   * Both SVD and PCA are methods for reducing dimensionality. After reducing dimensions using either method, cosine similarity can be applied to the lower-dimensional representations to find similarities.
3. **Recommendation Systems**:
   * **SVD** is often used to decompose user-item interaction matrices, revealing latent factors.
   * **PCA** can be used in conjunction to further reduce dimensions.
   * **Cosine Similarity** can then measure how similar users or items are in the reduced-dimensional space.

**Overview between SVD and PCA**

* **Cosine Similarity** is used to measure similarity between vectors and is often applied to the results of dimensionality reduction techniques like **SVD** and **PCA**.
* **SVD** and **PCA** both reduce dimensionality but do so in different ways and contexts.
* The use of cosine similarity in the context of these methods helps quantify the similarity of items, users, or documents in a reduced feature space.