

# Understanding Embeddings

## What are Embeddings?

**Vector Representation** of objects (words, images, users) in continuous space where:

- Similar items are closer together
- Dissimilar items are farther apart
- Reduces complex data to fixed-length vectors

## Types of Embeddings

### Static:

- Word2Vec
- GloVe
- FastText

### Contextual:

- BERT
- GPT
- ELMo

## Real-World Applications

- **NLP:** Translation, sentiment analysis
- **RecSys:** User-item matching
- **Vision:** Image recognition
- **BioMed:** Molecular analysis

## Key Properties

- **Dimensionality:** 50-300 features
- **Dense:** Continuous values
- **Contextual:** Context-aware meaning

## Vector Structure

Example 3D vector:

$[0.2, -0.7, 0.5]$

- Fixed length
- Numerical features
- Comparable via similarity metrics

## Visualization & Evaluation

### Visualization:

- PCA
- t-SNE
- UMAP

### Evaluation:

- Intrinsic (vector properties)
- Extrinsic (task performance)

## Embedding Space Visualization

