NATIONAL UNIVERSITY OF MODERN LANGUAGES

ISLAMABAD

Natural Language Processing (Lab)

Assignment No: 03

Submitted to
Dr. Qurat-ul-Ain Safder

Submitted By
Junaid Asif
(BSAI-144)

Submission Date: May 25, 2025

1. Introduction

Word Embeddings are dense vector representations of words that capture their semantic meaning. Unlike one-hot encoding, which creates sparse and high-dimensional vectors, word embeddings place semantically similar words closer together in a vector space.

2. Implementation

We used the Gensim library to implement Word2Vec embedding on a sample text. The process includes:

- Tokenizing the text into words.
- Training a Word2Vec model with defined vector size and context window.
- Viewing the embeddings and similarity between word pairs.

3. Code Explanation

- **Word2Vec:** Trains the model using skip-gram or CBOW.
- **vector_size:** Dimensionality of the word vectors.
- **window:** Context window around the target word.
- **similarity():** Measures cosine similarity between word vectors.

4. Output Examples

- Similarity between "word" and "vectors": *e.g., 0.45*
- **Word embedding for "word":**
  [0.0132 0.0273 -0.0213 ...] *(100-dimension vector)*

5. Conclusion

Word embeddings are powerful tools in NLP that enhance machine understanding of language. They are foundational to modern NLP systems such as chatbots, translators, and search engines.