

Assignment 3

ECSE344L (NLP)

Word Vector/Embedding Generation

In the last Lab we have created word embedding using SVD. In this lab we will use Iterative method to create word embedding.

Utilize your IITB English/Hindi corpus.

Question 1. Create One hot encoding for every word in your corpus.

Question 2- Use Continuous bag of word with negative sampling to create word embedding $N=50, 100$. Where N is embedding vector dimension.

Question 3. Use skip gram model with negative sampling to create word embedding $N=50, 100$

Question 4- Use Continuous bag of word without negative sampling to create word embedding $N=50, 100$ Where N is embedding vector dimension.

Question 5. Use skip gram model without negative sampling to create word embedding $N=50, 100$