WIDS-2025
Project Report week2

Boda Prabanjan Jadav
24B1001
Mentor : Aditya Sanapala

Contents

1 Overview

In Week 2, I studied the fundamentals of deep learning approaches for Natural Language Processing (NLP). This week focused on understanding how machines learn semantic meaning from text rather than relying only on word frequencies. I explored word embeddings, recurrent neural networks, attention mechanisms, and transformer models, and implemented models to compare classical and deep learning methods for text classification.

2 Word Embeddings: Word2Vec and GloVe

I learned that word embeddings represent words as dense vectors that capture semantic and contextual relationships. Unlike one-hot encoding, embeddings allow similar words to be close to each other in vector space.

I studied Word2Vec, focusing on the Continuous Bag of Words (CBOW) and Skip-Gram architectures. I learned how these models use local context windows to learn word representations and how training over multiple epochs refines the quality of embeddings. I also understood the effect of hyperparameters such as vector dimension, window size, and minimum word frequency.

I then studied GloVe embeddings, which use global word co-occurrence statistics to learn word vectors. I learned how GloVe combines the advantages of global matrix factorization and local context-based learning, resulting in meaningful semantic relationships between words.

3 Recurrent Neural Networks: RNN, GRU, and LSTM

I explored recurrent neural networks and learned why basic RNNs struggle with long sequences due to vanishing and exploding gradients. This led me to study gated architectures.

I learned how Long Short-Term Memory (LSTM) networks use input, forget, and output gates to capture long-term dependencies in text. I also studied Gated Recurrent Units (GRU), which provide a simpler gated structure while retaining strong performance. I understood how these models are suitable for NLP tasks where word order and context are important.

4 Attention Mechanism

I studied the attention mechanism and learned how it allows models to focus on relevant parts of an input sequence. I understood that attention assigns weights to different words based on their importance to the task, solving the information bottleneck problem present in traditional sequence-to-sequence models. This concept helped me understand why attention significantly improves performance in NLP tasks.

5 Introduction to Transformers

I learned the basic architecture of transformer models and how they rely entirely on self-attention instead of recurrence. I studied how positional encodings provide order information and how transformers enable parallel computation. This section helped me understand why transformers outperform recurrent models on many NLP tasks and form the foundation of modern language models.

6 Embedding Visualization

I learned how dimensionality reduction techniques such as Principal Component Analysis (PCA) and t-SNE are used to visualize high-dimensional word embeddings. These visualizations helped me confirm that semantically similar words cluster together, indicating effective learning of word representations.

7 Mini-Project: TF-IDF vs LSTM

As part of the mini-project, I compared a TF-IDF based Logistic Regression model with an LSTM-based classifier for sentiment analysis. I observed that the TF-IDF model provided a strong and fast baseline, while the LSTM model was able to capture contextual information more effectively but required more training time and tuning. This comparison highlighted the trade-off between model simplicity and expressive power.

8 Conclusion

Week 2 significantly improved my understanding of how deep learning models process and understand natural language. I developed a clear understanding of word embeddings, sequence models, attention mechanisms, and transformers. This week bridged the gap between classical NLP techniques and modern deep learning approaches and strengthened my foundation for advanced NLP models.