Week 1 LLM Learning Plan: In-Depth Overview

Your Name

May 26–June 2, 2025

Introduction

This document provides an in-depth yet brief overview of Week 1 (May 26–June 2, 2025) LLM learning, covering tokenization, word embeddings, RNNs, transformers, and fine-tuning. It includes advanced resources for revision, ideal for a full-stack web engineer to revisit concepts and deepen NLP expertise.

Topic Overview

Tokenization

Overview: Tokenization splits text into tokens (words, subwords) for numerical processing in LLMs. NLTK and BERT's WordPiece handle different granularities.
Key Insights: Tokens convert to IDs; subword tokenization (e.g., "LLMs" to "ll" + "ms") manages rare words.
Week 1 Application: Tokenized "I love learning about LLMs" (NLTK) and "Transformers power modern LLMs" (BERT).
Advanced Resources:

- Hugging Face Tokenizer Guide: Details WordPiece, BPE.
- Subword Tokenization Paper: Explores efficiency in LLMs.

Word Embeddings

Overview: Embeddings (e.g., Word2Vec, GloVe) map tokens to vectors, capturing semantic context.
Key Insights: Skip-gram predicts context; GloVe uses co-occurrence matrices. Cosine similarity measures relatedness (e.g., "king" vs. "queen": 0.7839).
Week 1 Application: Computed similarity with `gensim`'s GloVe.
Advanced Resources:

- Word2Vec Paper: Deep dive into Skip-gram/CBOW.
- GloVe Project: Co-occurrence mechanics.

RNNs and LSTMs

Overview: RNNs process sequences sequentially; LSTMs mitigate vanishing gradients for longer contexts.
Key Insights: RNNs struggle with long dependencies; LSTMs use gates to retain memory.
Week 1 Application: Studied limitations vs. transformers' parallel attention.
Advanced Resources:

- LSTM Paper: Original architecture details.
- Colah's RNN Guide: Advanced RNN variants.

Transformers

Overview: Transformers use self-attention and FFNs for parallel, context-aware processing.
Key Insights: Attention weights token relationships; BERT's encoder-only design excels in classification.
Week 1 Application: Explored attention, tokenized with BERT, ran sentiment analysis (score: 0.99983).
Advanced Resources:

- Attention Is All You Need: Transformer paper.
- The Illustrated BERT: Visualizes BERT's flow.

Fine-Tuning

Overview: Fine-tuning adapts pre-trained models (e.g., DistilBERT) to specific tasks like sentiment classification.
Key Insights: Hyperparameters (learning rate, epochs) impact accuracy. Validation loss monitors overfitting.
Week 1 Application: Fine-tuned DistilBERT on IMDB (92.4% accuracy, later 87.7% after tweaks).
Advanced Resources:

- Hugging Face Fine-Tuning: Advanced techniques.
- BERT Paper: Fine-tuning strategies.