Report

Perplexities

Pride and Prejudice

Model Name	Number of Ngrams	Average Train Perplexity	Average Test Perplexity
FFNN	N=3	39.3425	504.4886
	N=5	25.5988	900.2425
RNN	For any N	49.4136	134.8829
LSTM	For any N	53.9538	127.7136

Ulysses - James Joyce.txt

Model Name	Number of Ngrams	Average Train Perplexity	Average Test Perplexity
FFNN	N=3	171.1284	992.4623
	N=5	151.4799	1335.3407
RNN	For any N	56.4602	374.4795
LSTM	For any N	61.3177	339.9315

Comparison(Assignment1 & Assignment 2)

Comparison with Assignment 1 Models

- **N-gram models (Laplace, Good-Turing, Interpolation) have higher perplexity** due to their limited context.
- **Neural models (FFNN, RNN, LSTM) have lower perplexity** because they capture better word relationships.

Analysis and Observation:

1. FFNN Model: Overfitting Issue

- **Train perplexity decreases with larger n-grams (N=5 has lower perplexity than N=3)**, meaning that the model fits the training data better when given more context.
- **Test perplexity increases significantly when N=5**, suggesting **severe overfitting**—the model memorizes training sequences but generalizes poorly to unseen data.
- **N=3 performs better than N=5 on the test set**, reinforcing that a balance between context size and generalization is crucial.

2. RNN vs. LSTM: Better Generalization

- **LSTM has the lowest test perplexity (127.7136), outperforming both RNN and FFNN**. This indicates that the LSTM captures long-term dependencies effectively, making it more robust for generalizing to new sentences.
- **RNN performs slightly worse than LSTM (134.8829 vs. 127.7136), but still significantly better than FFNN**. However, RNNs struggle with longer dependencies due to vanishing gradients, explaining why LSTM has an edge.
- **Both RNN and LSTM have higher train perplexity compared to FFNN**, which suggests they **avoid overfitting** and learn meaningful generalizable patterns instead of memorizing.

Summary of Performance for Longer Sentences:

Model	**Performance for longer sentence**	**Why?**
LSTM	Best	Captures long-term dependencies and avoids vanishing gradients, making it ideal for longer sentences.
RNN	Moderate	Struggles with long sentences due to vanishing gradients, but performs decently on shorter sequences.
FFNN	Worst	Limited to fixed-size context windows, failing to capture long-term dependencies in longer sentences.

Model Ranking:

Rank	Model
1	LSTM
2	RNN
3	FFNN (N=3)
4	FFNN (N=5)