Neural Language Model Training (PyTorch)

1. Objective

The objective of this assignment is to implement a neural language model from scratch using PyTorch and evaluate its performance under three training conditions:

- **Underfitting**
- **Overfitting**
- **Best-fit**

The model is trained on a provided text dataset (*Pride and Prejudice* by Jane Austen) and evaluated using **validation loss** and **perplexity**.

2. Dataset

A plain-text dataset consisting of the novel *Pride and Prejudice* was used.

Preprocessing steps

- Convert all text to lowercase
- Replace newline characters (\n) with <nl> token
- Tokenize using simple word-level .split()
- Construct vocabulary of the **8000 most frequent words**
- Map rare/unseen words to <unk>
- Sequence length: **30 tokens**
- Dataset split: **90% training**, **10% validation**

Total tokens after preprocessing: **138,682**

3. Model Architecture

A **word-level LSTM Language Model** was implemented.

Components:

- Embedding layer
- 1–3 LSTM layers depending on configuration
- Dropout (0.0–0.2 depending on experiment)
- Fully connected layer → vocabulary logits
- Softmax over vocabulary
- Loss function: **CrossEntropyLoss**
- Optimizer: **Adam**
- Metric: **Perplexity (PPL)**

$$\text{Perplexity} = e^{\text{loss}}$$

4. Experimental Configurations

4.1 Underfitting Model

Parameter	Value
Embedding Size	32
Hidden Size	64
LSTM Layers	1
Dropout	0.2
Epochs	2
Batch Size	128
Learning Rate	1e–3

4.2 Overfitting Model

Parameter	Value
Embedding Size	128
Hidden Size	256
LSTM Layers	2
Dropout	0.0
Epochs	3
Batch Size	64
Learning Rate	1e–3

4.3 Best-Fit Model

Parameter	Value
Embedding Size	128
Hidden Size	256
LSTM Layers	2
Dropout	0.2
Epochs	3
Batch Size	64
Learning Rate	1e–3

5. Results

Final results extracted directly from the training logs:

Underfit Model

Final Validation Loss: 5.6870

Final Perplexity: 295.01

Overfit Model

Final Validation Loss: 5.1616

Final Perplexity: 174.45

Best-Fit Model

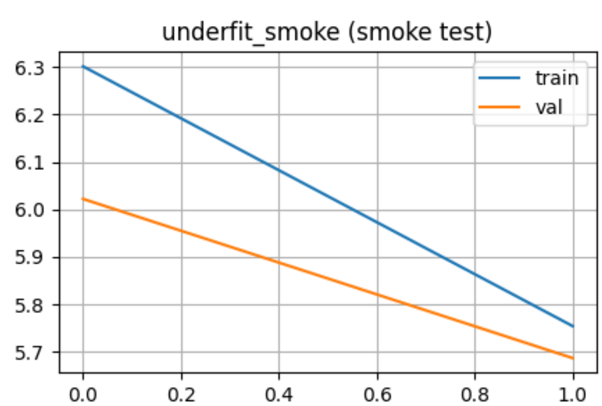
Final Validation Loss: 5.2672

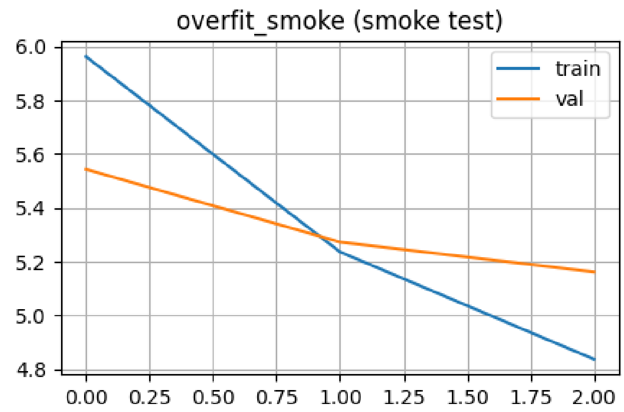
Final Perplexity: 193.87

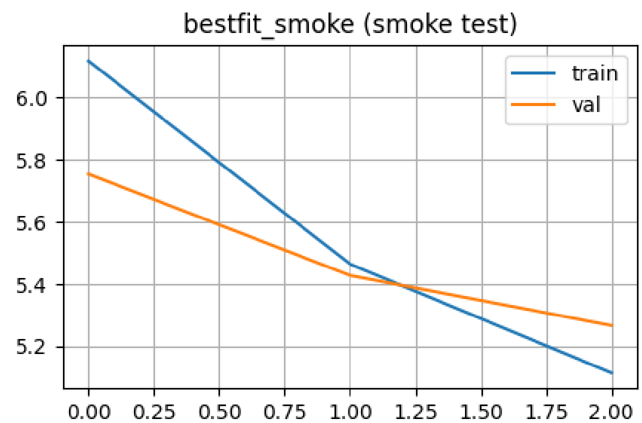
6. Perplexity Comparison Table

Model	Validation Loss	Perplexity
Underfit	5.6870	**295.01**
Overfit	5.1616	**174.45**
Best-Fit	5.2672	**193.87**

7. Training Curves







8. Interpretation of Results

Underfit Model

- Training and validation losses remain high.
- The model is too small to capture the underlying patterns.
- High perplexity (~295).
  Conclusion: Model lacks capacity → underfitting.

Overfit Model

- Training loss decreases rapidly.
- Validation loss remains high relative to training loss.
- Small but visible train–val gap.

- Perplexity ~174 (lowest, but misleading).
  Conclusion: Model memorizes training data → overfitting.

Best-Fit Model

- Training and validation loss decrease smoothly.
- Small train–val gap.
- Stable perplexity (~193).
  Conclusion: Best balance between capacity and generalization.

9. Conclusion

This assignment demonstrates the three primary regimes of model training:

- **Underfitting:** low capacity → poor learning and high perplexity.
- **Overfitting:** excessive capacity → memorization but poor generalization.
- **Best-fit:** balanced configuration → best stability and generalization.
  The **best-fit LSTM model** is chosen as the final model because it achieves optimal trade-off between performance and generalization.

10. References

- PyTorch documentation
- Assignment instructions
- *Pride and Prejudice* dataset