Chapter 8 Practice Exercises: RNN, LSTM, Seq2Seq, Encoder–Decoder
Speech and Language Processing (3rd draft)

Instructions

The following exercises are designed to help practice the concepts of RNNs, LSTMs, Seq2Seq, and Encoder–Decoder models. You should implement them in `PyTorch`, using either toy datasets or small datasets from `torchtext` or HuggingFace.

Exercises

Exercise 21. Basic RNN Language Model
Implement a simple RNN using `torch.nn.RNN` to train on a small text corpus (e.g., *Alice in Wonderland*). Goal: predict the next character. Practice: build `char2idx`, `idx2char`, train for a few epochs, and generate sample text.

Exercise 22. Compare RNN and LSTM
On the same dataset from Exercise 1, replace RNN with LSTM. Compare the **perplexity** of the two models.
Practice: implement an `evaluate_perplexity(model, data)` function.

Exercise 23. LSTM from Scratch
Manually implement the forward pass of an LSTM cell (without using `torch.nn.LSTM`). Practice: define a class `MyLSTMCell`, and test it on a batch of inputs.

Exercise 24. Seq2Seq without Attention
Build a simple Seq2Seq model for a toy translation task (e.g., mapping

"two four" → "24"). Do not use attention. Practice: encoder with LSTM, decoder with LSTM, training with teacher forcing.

Exercise 25. Seq2Seq with Attention
Add attention (Luong or Bahdanau) to the model from Exercise 4. Compare BLEU score with the no-attention version. Practice: implement an `Attention` layer, and visualize attention heatmaps.

Exercise 26. Bidirectional RNN
Train a BiRNN for sentiment classification on the IMDB dataset (positive/negative). Practice: load data with `torchtext`, compare unidirectional vs bi-directional RNNs.

Exercise 27. Encoder–Decoder for Machine Translation
Train a Seq2Seq (LSTM encoder–decoder + attention) on a small dataset from HuggingFace (e.g., `opus_books` en–fr). Practice: train for 1–2 epochs, print sample translations.

Exercise 28. Padding and Masking
Write code to handle **padding and masking** during Seq2Seq training. Practice: compare loss when not masking vs masking (ignoring `<pad>` tokens).

Exercise 29. Teacher Forcing vs Free Running
Compare translation results when training with **teacher forcing ratio = 1.0** and **0.5**. Practice: log the loss curves and print sample outputs.

Exercise 30. Inference with Beam Search
Add **beam search decoding** to the Seq2Seq model (from Exercise 7). Compare BLEU scores between greedy search and beam search. Practice: implement `beam_search(decoder, encoder_output, beam_size=3)`.

Implementation Tips

- Structure your code modularly: `Encoder`, `Decoder`, `Seq2Seq`, `Attention`.

- Always log metrics (loss, perplexity, BLEU) to compare models.

- For datasets, use `torchtext`, HuggingFace `datasets`, or custom toy data.