Assignment 4
Sequence Models: RNN, LSTM, GRU

Part A — Foundations

22 questions

★	**Must-know** — asked very frequently in GenAI / ML interviews (6 questions). Prepare these first.
[CODE]	**Hands-on task** — write and run the code (2 tasks).

1. What is the difference between an RNN and an ANN? ★
2. How does backpropagation in an RNN differ from an ANN? What is Backpropagation Through Time (BPTT)?
3. How does an RNN handle variable-length inputs?
4. What are the three weight matrices used in an RNN? Explain parameter sharing.
5. What is a bidirectional RNN? When is it useful?
6. Why do RNNs suffer from vanishing gradients and how can it be solved? ★ · Asked: NLP interview rounds
7. Explain the LSTM cell: cell state, hidden state and the three gates (forget, input, output). ★ · Asked: NLP interview rounds
8. How is the cell state updated in an LSTM? Write the equations.
9. How does LSTM address vanishing gradients and long-term dependencies? ★
10. What is the difference between the hidden state and the cell state?
11. What is a GRU? Explain its gates (update, reset). How does it differ from LSTM? ★
12. How would you decide between LSTM and GRU?
13. Advantages and disadvantages of RNN, LSTM and GRU.
14. Why did Transformers replace RNNs/LSTMs for most NLP tasks? ★
15. What does return_sequences=True do in an LSTM layer and when do you need it?
16. How do you calculate the number of trainable parameters in RNN, LSTM and GRU layers?
17. What is padding? What is truncating? Explain pre vs post padding and the related hyperparameters (maxlen, padding, truncating).
18. What is the Embedding layer in Keras? How many parameters does it have?
19. Explain dropout and recurrent dropout.

20. Write code for padding sequences. [CODE]

21. Build networks with (a) SimpleRNN, (b) LSTM, (c) GRU and (d) Bidirectional layer for sentiment classification. [CODE]

22. How can you reduce RNN training problems without moving directly to LSTM/GRU? · Asked: NLP interview rounds