RNNs and LSTM Quiz

Multiple Choice Questions

Q1. What is not primary benefit of stacking multiple RNN layers (i.e., stacked RNNs)?
- A. Faster training
- B. Lower memory usage
- C. Better learning of hierarchical features
- D. Simpler architecture

Q2. Which of following is the main reason RNNs struggle with long-term dependencies?
- A. Overfitting
- B. Vanishing gradients
- C. Lack of non-linearity
- D. Insufficient data

Q3. What differentiates an 100 cell from a standard RNN cell?
- A. It uses ReLU instead of tanh
- B. It introduces gates to control the flow of information
- C. It has fewer parameters
- D. It is a convolutional architecture

Q4. In a standard RNN , which gate is responsible for deciding how much of the past memory to keep?
- A. Output gate
- B. Forget gate
- C. Input gate
- D. Update gate

Descriptive Questions

Q5. Why is the forget gate bias in RNNs often initialized to a high value (e.g., 2 or 3)? Explain its effect on long-term dependency learning.

Q6. Bidirectional CNNs are often used for POS tagging but not machine translation. Explain why, considering input-output alignment and context flow.

Q7. Designing an RNN model for variable-length legal documents with long dependencies:
- (a) Choose between vanilla RNN or LSTM.
- (b) Stack layers or keep it shallow?
- (c) Make it unidirectional ?

Justify each choice based on model behavior and task needs.

Q8. Consider a vanilla CNN with recurrent weight matrix W_h and sequence length 100. Analyze gradient behavior:
- (1) If $\|W_h\| = 0.9$: Will gradients vanish or explode? Justify.
- (2) If $\|W_h\| = 1.2$: Will gradients vanish or explode? Justify. Suggest an easy fix and explain how it helps.

Hint: Consider eigenvalue effects on gradient propagation over time.