MIDTERM EXAM

Course: Natural Language Processing

Chapter 8: Recurrent Neural Networks (RNN) and LSTM

Time allowed: 60 minutes

Question	Points	Score
1	0	
2	0	
3	0	
4	0	
5	0	
6	0	
7	0	
8	0	
9	0	
10	0	
11	0	
12	0	
13	0	
14	0	
15	0	
16	0	
17	0	
18	0	
19	0	
20	0	
Total:	0	

Part I: Multiple Choice (8 questions, 4.0 points)

1. What key feature of a Recurrent Neural Network (RNN) makes it suitable for processing sequential data?

   A. It uses convolutional layers.

Name:..... Class:.....

B. Weight sharing across time steps and having a hidden state.

C. Each input is processed independently of the others.

D. There are no loops in its architecture.

2. In a simple RNN, what is the hidden state h_t at time step t calculated based on?

A. Only the current input x_t.

B. Only the previous hidden state h_{t-1}.

C. The current input x_t and the previous hidden state h_{t-1}.

D. Only the network's output at time step $t-1$.

3. What is the primary cause of the vanishing gradient problem in RNNs?

A. Using too many layers in the network.

B. The repeated multiplication of small values during backpropagation through time.

C. A learning rate that is too large.

D. The dataset size being too small.

4. In an LSTM, which gate is responsible for deciding what information to discard from the cell state?

A. Input Gate.

B. Forget Gate.

C. Output Gate.

D. Reset Gate.

5. Compared to an LSTM, what is a characteristic of a GRU?

A. It is more complex, with four gates.

B. It has separate cell state and hidden state vectors.

C. It is simpler, combining the forget and input gates into a single update gate.

D. It cannot solve the vanishing gradient problem.

6. Which gate in an LSTM uses a 'tanh' activation function to create new candidate values?

A. The Forget Gate.

B. Part of the Input Gate (to create $\tilde{C}_t$).

C. The Output Gate.

D. All three gates use 'tanh'.

7. What is the main purpose of using a Bidirectional RNN (Bi-RNN)?

A. To reduce training time.

B. To process two input sequences at once.

Name:..... Class:.....

C. To allow the network to use both past and future context at each time step.

D. To simplify the network architecture.

8. Which of the following tasks is a form of sequence labeling?

A. Machine Translation.

B. Text Classification.

C. Part-of-Speech Tagging.

D. Text Summarization.

Part II: Short Answer (7 questions, 10.0 points)

9. Write the formula for calculating the hidden state h_t in a simple RNN and explain the meaning of each term.

10. Explain why repeated multiplication during Backpropagation Through Time (BPTT) leads to vanishing or exploding gradients.

11. Write the full set of equations for an LSTM cell, including all gates and update equations.

12. Explain the role of the cell state (C_t) in an LSTM and how its additive update mechanism mitigates vanishing gradients.

13. Compare and contrast the architectures of a GRU and an LSTM, focusing on gates and the handling of the cell state.

14. Draw a block diagram of a Bidirectional RNN and explain how the final output at each time step is computed.

15. Describe how an RNN can be used as a language model.

Part III: Coding Practice (5 questions, 6.0 points)

16. Given $W_{xh} = \begin{pmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{pmatrix}$, $W_{hh} = \begin{pmatrix} 0.5 & 0.6 \\ 0.7 & 0.8 \end{pmatrix}$, input $x_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $h_{t-1} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$, compute the new hidden state h_t for a simple RNN with tanh activation and no bias.

17. Using NumPy, implement:

- a sigmoid function
- an LSTM forget gate function $lstm\_forget\_gate(xt, ht\_prev, Wf, bf)$.

18. Explain the steps to preprocess a raw text corpus for training an RNN-based language model.

Name:..... Class:.....

19. You are building a POS tagger using a Bi-LSTM. a) Why is a Bi-LSTM better than a unidirectional LSTM? b) What are the input and output tensor shapes?

20. In the Keras code below, identify the recurrent layer, number of hidden units, and activation function:

```
model = Sequential([
    Embedding(vocab_size, embedding_dim, input_length=max_length),
    SimpleRNN(32, activation='tanh'),
    Dense(1, activation='sigmoid')
])
```

Name:..... Class:.....

18. Explain the steps to preprocess a raw text corpus for training an RNN-based language model.

Tokenize → Create Vocabulary → Create (or download) embedding
→ Embed the tokens.

19. You are building a POS tagger using a Bi-LSTM. a) Why is a Bi-LSTM better than a unidirectional LSTM? b) What are the input and output tensor shapes?

a) By adapting information in the past and future, Bi-LSTM can learn richer context for more accurate predictions.

b) input : (N, T, embedding dim)

output : (N, T, num_classes)

N : Batch size

T : Sequence length.

20. In the Keras code below, identify the recurrent layer, number of hidden units, and activation function:

```
1 model = Sequential([
2     Embedding(vocab_size, embedding_dim, input_length=max_length),
3     SimpleRNN(32, activation='tanh'),
4     Dense(1, activation='sigmoid')
5 ])
```

Recurrent layer.

Activation: tanh (in recurrent layer)

sigmoid (in FNN)

num hidden units: 32