BRAC UNIVERSITY
Department of Computer Science and Engineering

Examination: Final | Semester: Summer 2025
Duration: 90 minutes | Full Marks: 45

CSE 440: Natural Language Processing II

Figures in the right margin indicate marks.

Answer all 3

1. A. You are asked to perform named entity recognition on the following sentence: ***Abid has reached BRACU***. Your possible tags are (in this exact order): LOC, ORG, GPE, PER, O. Given the LSTM equations and the weight matrices on the other side, what is the most probable tag for ***BRACU***? [12]

   B. What is the difference between hidden state and cell state in an LSTM? [3]

2. A. Write two disadvantages of Hidden Markov Models. How can you solve them? [5]

   B. Do you think you will need a Bidirectional HMM with Viterbi Algorithm for better performance? Why? Why not? Explain your answer. [5]

   C. What properties do we generally look for while picking an activation function? Compare tanh and sigmoid as activation functions. [5]

3. A. Discuss how BERT is pre-trained. What are masked language modeling and next sentence prediction? Give examples for both of these. [5]

   B. If you have two languages: one has 3 words (a, b and c) and the other has 4 (w, x, y and z), how many word level translation probabilities will you need to calculate? What will be the initial values of those probabilities? How many alignments can there be in the sentence-pair a b c → x y z? What will be those alignments' probabilities? [10]

Equations for the LSTM:

$f_t = \sigma(W_f[h_{t-1},x_t] + b_f)$
$i_t = \sigma(Wi[h_{t-1},x_t] + b_i)$
$o_t = \sigma(Wo[h_{t-1},x_t] + b_o)$
$\acute{C}_t = \tanh(W_C[h_{t-1},x_t] + b_C)$
$C_t = f_t \ⓧ C_{t-1} + i_t \ⓧ \acute{C}_t$
$h_t = o_t \ⓧ \tanh(C_t)$
$y_t = \text{softmax}(h_t V)$

yt will indicate BRACU's tag.

Now, the weights and the variables:

W_f = [0 1, 2 1, 2 1, 0 0, 1 0]
b_f = [0 0]
W_i = [0 -1, -1 1, 0 -1, 0 0, -2 1]
b_i = [-1 -1]
W_o = [1 2, 2 0, 0 1, 1 0, 1 1]
b_o = [0 1]
W_C = [1 1, 2 0, 1 0, 0 0, 1 1]
b_C = [-3 -1]
V = [1 0 1 0, 1 1 0 0]
h_{t-1} = [1 1]
C_{t-1} = [1 1]
X_t = BRACU = [0 1 1]

- ⓧ in the equation means elementwise multiplication and + means elementwise addition
- Use upto 2 decimal points in your calculations