ROLL NO._____NAME _____

CS 5316 – Natural Language Processing
Quiz 7 Solution
(Time limit: 12 minutes)

1. ((4 points) Consider a trained RNN for predicting whether a given word is a named entity (1) or not (0). Each word is represented by a vector of length 3 and there are 2 recurrent units with tanh activation functions in the recurrent layer and 1 unit with sigmoid activation function in the output layer. Determine the label for a word with representation [1, 0, 1] assuming that all parameters (weights, biases) and preceding time activations are all 1's.

$a_n = \tanh(W_{ax}x + W_{aa}a_o + b_a)$=
W_ax is a 2 by 3 matrix of 1's, W_a is a 2 by 2 matrix of 1's, and aa_o and b_a are 2 by 1 vectors of 1's.
a_n = tanh([2 2]^T + [2 2]^T + [1 1]^T)
a_n = tanh([5 5]^T)
a_n = [0.999 0.999]^T

$y = \sigma(W_{ya}a_n + b_y)$

W_ya is a 1 by 2 matrix of 1's, b_y is a 1 by 1 vector of 1's.
Y = sigmoid(1.998 + 1) = 0.953

Since y > 0.5 this word is a named entity.

2. (3 points) Differentiate between a GRU and a LSTM unit. Which one has greater representation capability. Be precise.

Each typical GRU has two gates: update and relevance. A typical LSTM unit has 3 gates: update, forget, and output. An LSTM based network is more expressive than an equal sized GRU based network.

3. (3 points)Compute the BLEU score for the generated sentence 'the lahore city garden' when given reference sentences: 'the garden city of lahore' and ' Lahore the city of garden'. Use modified precision (uni-, bi-, and tri-gram) and brevity penalty.

1-grams: the, Lahore, city, garden
Pre = 4/4 = 1
2-grams: the Lahore, Lahore city, city garden
Pre =0/3 = 0

3-grams: the Lahore city, Lahore city garden
Pre = 0/2 = 0

Avg. precision = 1/3

Brevity penalty = exp(1-R/C) = exp(1-5/4) = 0.778

Bue score = 1/3*0.778 = 0.259