Name: _____ Reg #: _____ Section: _____

National University of Computer and Emerging Sciences, Lahore Campus

Course:	**Natural Language Processing**	**Course Code:**	**CS 535**
Program:	**MS(Computer Science)**	**Semester:**	**Spring 2021**
Duration:	**3 hour**	**Total Marks:**	**70**
Paper Date:	**12-July-21**	**Weight**	**70%**
Section:	**CS**	**Page(s):**	**6**
Exam:	**Final Exam**		

Instruction/Notes: Attempt the examination on the question paper and write concise answers. You can use extra sheet for rough work. Do not attach extra sheets used for rough with the question paper. Don't fill the table titled Questions/Marks.

Questions	1	2-3	4-5	6-7	8-10	11	12/13	Total
Marks	/8	/8	/15	/16	/6	/9	/8	

Q1) Answer the following multiple choice questions. Suppose you have the following training data for Näıve Bayes. Encircle correct option. [8 Marks]

I liked the movie [LABEL=+]
I hated the movie because it was an action movie [LABEL=-]
Really cool movie [LABEL=+]

A) What is the unsmoothed maximum likelihood estimate of P(+)for this data?
i. 1/3 ii. 1/2 iii. 2/3 iv. 1

(B) What is the unsmoothed maximum likelihood estimate of P(movie|+)for this data?
i. 2/17 ii. 1/5 iii. 2/7 iv. 1/2

(C) Suppose we are given an unseen input sentence "the movie". What is the joint probability P(−,the movie)?
i. 2/300 ii. 4/98 iii. 1/12 iv. 1/3

(D) What prediction will the model make on sentence "the movie"?
i. Positive ii. Negative

Q2) Calculate the TFIDF for the terms listed below for documents 1 to 3. There are 1000 documents in a collection. The number of times each of these terms occur in documents 1 to 3 as well as the number of documents in the collections are listed below. Use this information to fill in the TFIDF scores for Doc 3 in the table below. [5 Marks]

Number of Documents Containing Terms:
_ Exam: 30
_ Fruit: 10
_ Apple: 80

	Raw Term Counts		
	Doc 1	Doc 2	Doc 3
Exam	4	54	3
Fruit	7	5	30
apple	25	34	9

Fill in the table below and show all working.

	Tf.IDF for terms in Doc 3
exam	
Fruit	
apple	

Q3) You are an English teacher and you ask your class to write a play in the style of Shakespeare. You want to score their plays using a trigram language model you computed from a corpus of all Shakespeare plays but you find that the data is too sparse and most of your students' sentences receive a score of zero. How would you use a back-off model to alleviate this problem? [3 Marks]

Q4) Following table gives co-occurrence counts based on syntactic dependencies of words. Write down context vector of the word "duty" using PPMI (Positive Pointwise Mutual Information) of words. (You can assume following table contains all words that can appear as object of a given a word. E.g. total count of words that appear as object of "assert" is 10. Sum of row counts represent total count of the word in collection. E.g. duty appears 22 times in collection. Total words in collection = N = 100) **[5 Marks]**

$$\text{PMI}(word_1, word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)}$$

	Object of assert	Object of assign	Object of avoid	Object of become	Modified by collective	Modified by assumed
duty	3	4	5	3	5	2
responsibility	2	2	7	4	2	7
taxes	0	0	3	0	0	1
danger	0	0	6	0	1	0
control	5	0	0	1	0	0

Q5) You are given the following training corpus: [1 + 1 + 2 + 3 + 3 = 10 Marks]

<s> I like oranges </s>
<s> oranges like I </s>
<s> We like cherries </s>
<s> I do not like cherries and oranges </s>

a) Calculate the probability of following test sentence. Include </s> in your counts just like any other token. λ_1 = trigram weight, λ_2 = bigram weight, λ_3 = unigram weight, $\lambda_1 = 0.4$, $\lambda_2 = 0.3$, $\lambda_3 = 0.3$

<s> I like bikes </s>

i. Unigram Model

ii. Bigram Model

iii. Trigram Model

iv. Trigram language model with linear interpolation.

Q6) Suppose we are training a LSTM language model for the sentence ”computers are able to see, hear, and learn”
One hot encoded vector of words is given as follows: [5 Marks]
computers = x_1: [1 0 0 0 0 0 0 0]
are = x_2: [0 1 0 0 0 0 0 0]
able = x_3: [0 0 1 0 0 0 0 0]
to = x_4 : [0 0 0 1 0 0 0 0]
see = x_5 : [0 0 0 0 1 0 0 0]
hear = x_6 : [0 0 0 0 0 1 0 0]
and = x_7: [0 0 0 0 0 0 1 0]
learn = x_8: [0 0 0 0 0 0 0 1]

Suppose the input at 7 different time stamps is as follws:
x_1 = computers, x_2 = are, x_3 = able, x_4 = to, x_5 = see, x_6 = hear, x_7 = and,

The predicted output distribution of words at different time stamps is as follows:
y_1 = [0 0.2 0.1 0.1 0.4 0.2 0 0]
y_2 = [0.1 0.2 0.3 0.3 0 0 0.1 0]
y_3 = [0 0.1 0 0.3 0.4 0.2 0 0]
y_4 = [0 0.1 0.1 0 0.6 0.2 0]
y_5 = [0 0 0.1 0 0 0.4 0.3 0.2]
y_6 = [0 0 0.1 0 0 0 0.4 0.5]
y_7 = [0 0 0.1 0 0.1 0 0.5 0.3]

Compute the cross entropy loss for this sentence.

Name: _____ Reg #: _____ Section: _____

Q7) (a) Suppose we have following language model: [4+4 = 8 Marks]

• input sequence of length 5 (lets say 5 words).
• Hidden layer units are 4.
Embedding vector size = 6
• V = vocabulary = 8

i. Draw RNN architecture diagram with dimensions of all layers and weight matrices

ii. Give the update equations for a simple RNN unit in terms of x, y, and h (input x, output y, and recurrent state h). Assume it uses tanh non-linearity.

Q7) (b) What is the role of gates in LSTM? How are gates implemented? [3 Marks]

Q8) Which of the following statements is INCORRECT? [2 Marks]

A. Recurrent neural networks can handle a sequence of arbitrary length,while feedforward neural networks can not.
B. Training recurrent neural networks is hard because of vanishing and ex-ploding gradient problems.
C. Gradient clipping is an effective way of solving vanishing gradient prob-lem.
D. Gated recurrent units (GRUs) have fewer parameters than LSTMs.

Q9) What is the probable approach when dealing with "Exploding Gradient" problem in RNNs? [2 Marks]
A) Use modified architectures like LSTM and GRUs
B) Gradient clipping
C) Dropout
D) None of these

Q10) If calculation of reset gate in GRU unit is close to 0, which of the following would occur? [2 Marks]
A) Previous hidden state would be ignored
B) Previous hidden state would be not be ignored

Q11) (a) What are problems of greedy decoding and how beam search resolves these problems? Describe in context of neural machine translation. [3 Marks]

Q11) (b) What are some advantages of neural machine translation as compared to statistical machine translation. [3 Marks]

Q11) (c) What is effect of changing beam size k on neural text generation? [3 Marks]

Q12 is only for MS students

Q12) (a) Describe some smoothing techniques used in neural language modeling? [4 Marks]

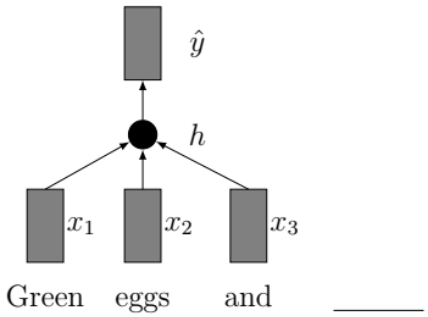
Q12) (b) What are advantages of using dense word vectors like word2vec as compared to sparse word vectors? [4 Marks]

Name: _____ Reg #: _____ Section: _____

Q13 is only for PhD students

Q13) (a) If we chose to update our word vectors when training the LSTM model on sentiment classification data, how would these word vectors differ from ones not updated during training? Explain with an example. Assume that the word vectors of the LSTM model were initialized using word2vec. [4 Marks]

Q13) (b) A feedforward neural network language model (NNLM) can be used as another architecture for training word vectors. This model tries to predict a word given the N words that precede it. To do so, we concatenate the word vectors of N previous words and send them through a single hidden layer of size H with a tanh nonlinearity and use a softmax layer to make a prediction of the current word. The size of the vocabulary is V. The model is trained using a cross entropy loss for the current word. Let the word vectors of the N previous words be x1,x2,...,xN, each a column vector of dimension D, and let y be the one-hot vector for the current word. [4 Marks]



$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}$$

$$\mathbf{h} = \tanh(W\mathbf{x} + \mathbf{b})$$

$$\hat{\mathbf{y}} = \text{softmax}(U\mathbf{h} + \mathbf{d})$$

$$J = \text{CE}(\mathbf{y}, \hat{\mathbf{y}})$$

$$CE = -\sum_i y_i \log(\hat{y}_i).$$

State two important differences between NNLM the WordToVec language model we learned in class. Explain how each might affect the word vectors learned.

Name: _____ Reg #: _____ Section: _____