National University of Computer and Emerging Sciences

Islamabad Campus

AI-4001 Fundamentals of NLP

Final Exam

Total Time (Hrs): 3hrs
Total Marks: 100
Total Questions: 7

Date: May 23rd 2024

Course Instructor(s)

Mr. Saad Salman

Roll No _____ Section _____ Student Signature _____

Attempt all the questions.

In case you don't know the answer, please write question no and in the answer write "I don't know" to get 20% marks. In case you tried to attempt the question even by writing a word you will be ineligible for the marks.

[15 marks]

Q1: TF-IDF

Term frequency - Inverse document frequency ($tf—idf$), is a numerical statistic that is intended to reflect how important a word is to a document in a corpus. Assuming that $tf(t,d) = log(1 + f_{t,d})$ where $f_{t,d}$ is the raw count of a term t in a document d and $idf(t, D) = log\frac{N}{n_t}$ where N is the total number of documents in the corpus D and n_t is the number of documents containing the term t, for the subsequent questions consider the following documents:

```
corpus = [
        'data science is one of the most important fields of science',
        'this is one of the best data science courses',
        'data scientists analyze data'
        ]
```

The words in the corpus:

{'important', 'scientists', 'best', 'courses', 'this', 'analyze', 'of', 'most', 'the', 'is', 'science', 'fields', 'one', 'data'}

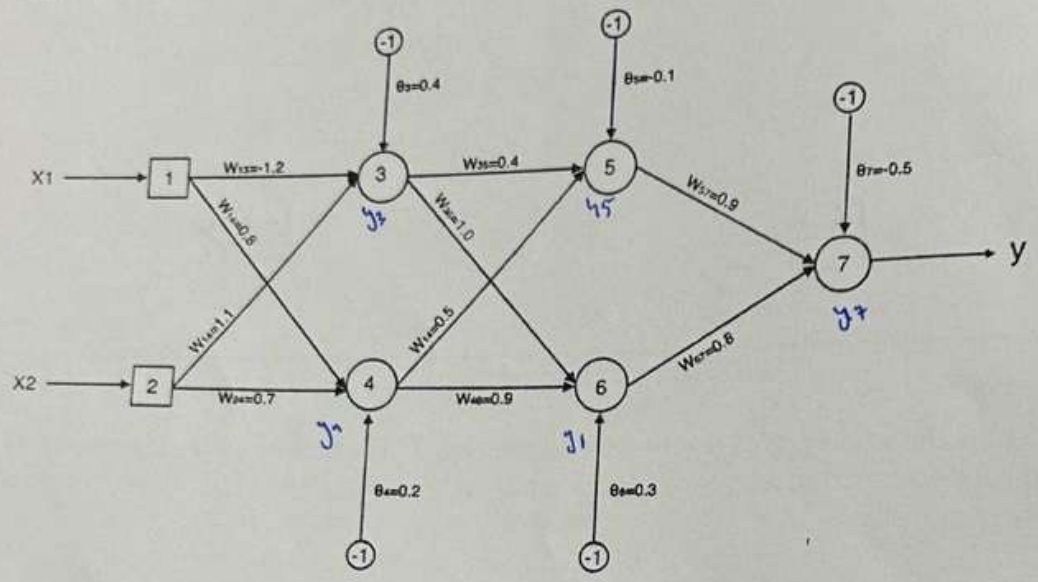
a) Compute the tf for the terms in T for each document.
b) Compute idf for the terms in T for the corpus.
c) Compute tf-idf(t,D,d) for the terms in T for each document in the corpus.

[25 marks]

Q2: Perceptron

Consider the neural network below. Perform one forward pass and fix the error and update the weight for the given inputs and outputs. You need to be accurate in your calculations. Please perform calculations up to 4 decimal places. This shows the complete working of the algorithm; fill in the table on the right, no marks will be awarded without that. Direct answers will not get any marks.

Inputs: X1 = 1, X2 = 0,
Output: Y7 = 1
Alpha = 0.1



w13 = -1.2, w14 = 0.8, w23 = 1.1, w24 = 0.7, w35 = 0.4, w36 = 1.0, w45 = 0.5, w46 = 0.9, w57 = 0.9, w67 = 0.8
θ3 = 0.4, θ4 = 0.2. θ5 = -0.1, θ6 = 0.3, θ7 = -0.5

Formula for calculating the error gradient on output layer

$$\delta_k(p) = y_k(p).[1 - y_k(p)].e_k(p)$$

Formula for calculating the error gradient for neurons in input layer

$$\delta_j(p) = y_j(p).[1 - y_j(p)].\sum_{k=1}^{l} \delta_k(p).w_{jk}(p)$$

Sigmoid is used as activation function in each neuron

[Working on solution sheet.]

y_3	0.4680
y_4	0.6456
y_5	0.6201
y_6	0.6104
y_7	0.8243
e_7	0.1756

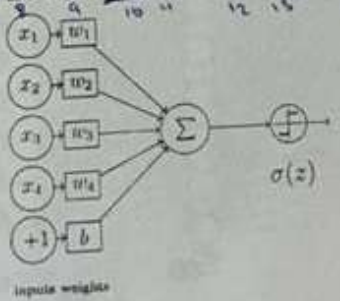
δ_7	0.0254
Δw_{57}	0.0846
Δw_{67}	0.0975
$\Delta\theta_7$	0.0021
δ_5	0.0054
δ_6	0.0483
Δw_{35}	0.02335
Δw_{45}	0.0485
Δw_{36}	
Δw_{46}	
$\Delta\theta_5$	
$\Delta\theta_6$	
δ_3	0.1665
δ_4	0.0105
Δw_{13}	
Δw_{23}	
Δw_{14}	
Δw_{24}	
$\Delta\theta 3$	
$\Delta\theta_4$	

[15 marks]

Q3: Logistic Regression

Consider the following sentence used for for training the following Logistic Regressor to detect sarcasm.

s1 = agar petrol mehnga ho jae, to app ne lena nhi hai. app ne ghabrana nhi hai



(NO s2 is given)

Given that s1 is labeled positive and and s2 is labeled as negative, and features x1 (count +ive lexicon), x2 (count -ive lexicon), x3 (negation words) and x4 (number of words) and the initial weight and bias vector: W, b = [0, 0, 0, 0, 0]. Determine the weights and bias vector after training and backpropagating the logistic regressor on the above sentence, given that the learning rate $\alpha = 0.1$

[10 marks]

Q4: *Text classification*

You are developing an emotion detection classifier that classifies sentence affect as angry(-), calm(=) or cool(+). Consider the following training corpus for the Multinomial Naive Bayes classifier with the given labels.

Training	
Sentence	Label
annoyed by his rage	-
very very annoyed at it	-
what does she do in her rage	-
he is cool	+
cool and cool is refreshing	+
calm and calm it is	=

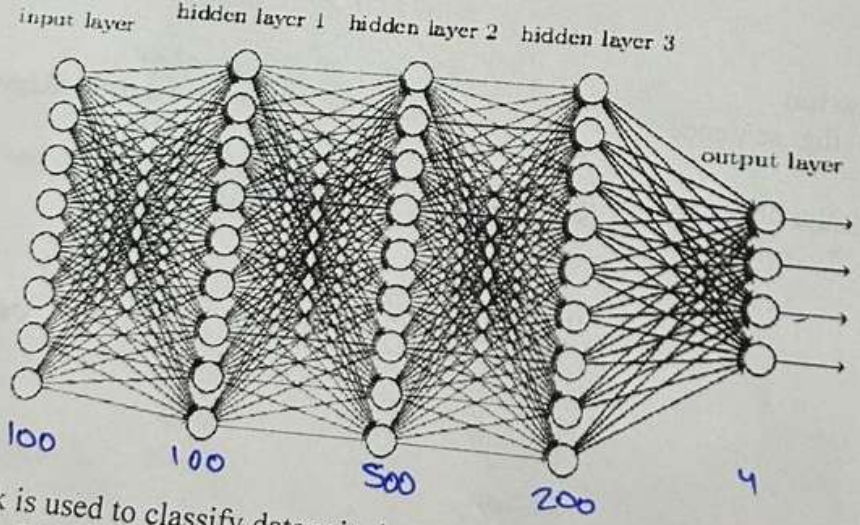
Considering that your classifier disregards stop words={what, very, and, he, his, her, she, by, is, at, it, in, do, does}, compute priors and likelihood probabilities for the given classes.

Test	Label
in her rage she is calm and cool	?

Q5: Feed Forward Networks

Consider the Feed-Forward Deep Network architecture below:

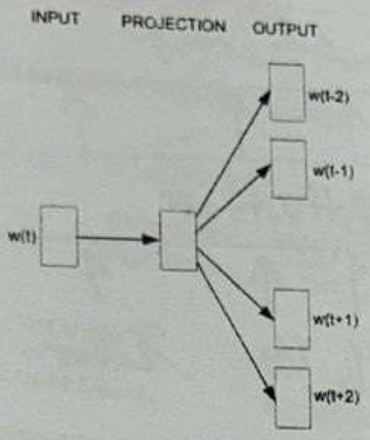
[10 marks]



) If the network is used to classify datapoits into four classes and there are 100 input features to the network and the hidden layers 1, 2, 3 have 100, 500, 200 neurons respectively, how many weights are there in total, including both weights and biases, in the entire neural network? How many neurons are there in total, in the neural network above?

Q6: Vector Semantics [10 marks]

Consider the test corpus ["the text has positive expressions like good"] for training the skip-gram model below for generating word embeddings, where the vocabulary size is 7.



Skip-gram

This Question is completly attempted on Solution sheet along with labelled diagram

a) What is the format of the words in the output layer? Show an example on the diagram above.

b) Label the diagram above to show the predicted words if the input word is "positive".

c) What is the significance of the size of the hidden layer in the above network?

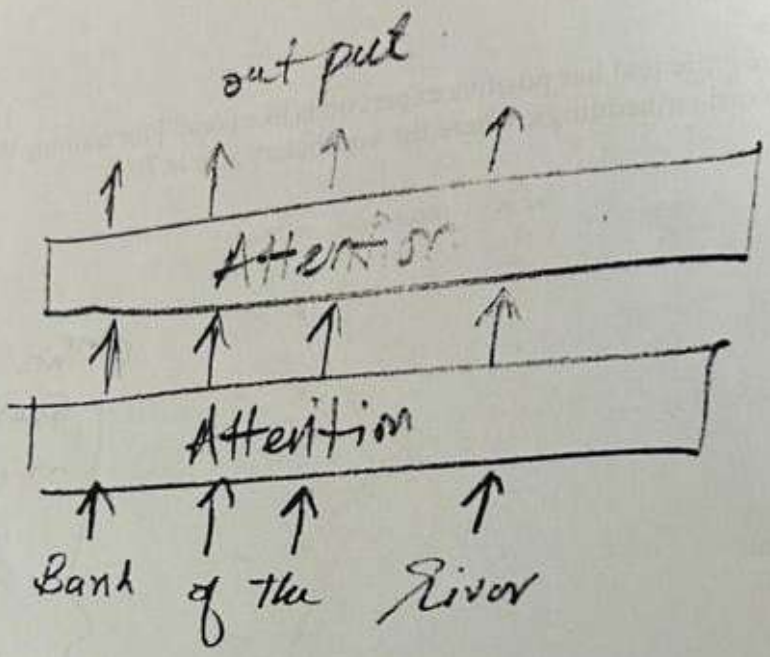
d) Consider the size of the hidden layer to be 3. Now draw the weights that would represent the embeddings of the word "positive" if the position of this word in the vocabulary is 4.

Q7: Attention [15 marks]

Consider the words and their vectors

Bank = [0.2 0.1 0.7 0]
Of = [0.1 0.8 0.1 0.3]
The = [0 0.2 0.1 0.3]
River =[0.8 0.11 0.3 0.6]

National University of Computer and Emerging Scien
Islamabad Campus



Pass the vectors through the above architecture and fin the output vectors.