NLP Final Exam 2023

Question 1: Multiple Choice Questions (Only ONE correct choice, and no negatives):

1. The following statement is true about Skip-Gram:
   a)   Predicts target word from context
   b)   Predicts context from given target word
   c)   Has one hidden layer that is fully-connected
   d)   None of the above

2. Calculate cross-entropy loss of a single output of a model that predicts one of 4 classes uniformly randomly
   a)   -0.25 log(0.25)
   b)   -4 log (1/4)
   c)   -log (1/4)
   d)   -4 log (4)

3. It is better to use **Macro** average when performance in all classes is equally important.
   a)   True
   b)   False

4. The following statement is NOT an advantage of dense vectors to sparse vectors:
   a)   Easier to use as a feature in other Machine Learning models
   b)   Handles unknown words unlike long sparse vectors
   c)   More general representation of rare words
   d)   Dense vectors are shorter than sparse vectors

5. The unknown words problem is solved by:
   a)   Laplace smoothing
   b)   interpolation
   c)   backoff
   d)   None of the above

6. The CKY algorithm can parse a sentence under any context-free grammar.
   a)   True
   b)   False

7. Encoding bigrams using a one-hot encoding, gives a vector representation of size ....
   a) $|V|$
   b) $|V^2|$
   c) $|2V|$
   d) $|V|^2$

8. (0|1)*000 is a regular expression that matches …..
   a)   All even binary strings
   b)   All binary strings that are divisible by 4
   c)   All binary strings that are divisible by 8
   d)   All binary strings ending with 00

9. Truncated backpropagation through time that can be used with RNNs ….
   a)   Limits forward but not backward
   b)   Limits backward but not forward
   c)   Limits both forward and backward
   d)   Limits neither forward nor backward

10. The Markov chain representing a bi-gram language model is computed by:

   a) $P(w_1) \prod_{i=2}^{n} p(w_i | w_{i-1}, w_{i-2})$

   b) $P(w_1) P(w_2 | w_1) \prod_{i=2}^{n-1} p(w_{i+1} | w_i, w_{i-1})$

   c) $P(w_1) P(w_2 | w_1) \prod_{i=1}^{n} p(w_i | w_{i-1}, w_{i-2})$

   d) $P(w_2 | w_1) \prod_{i=2}^{n} p(w_i | w_{i-1}, w_{i-2})$

11. The Gazetteer that can be used in NER is …..
   a)   List of people names
   b)   List of animals names
   c)   List of places names
   d)   None of the above

12. Attention mechanism allows the encoder to get information from all the hidden states of the decoder.
   a)   True
   b)   False

13. Mean Reciprocal Rank (MRR) of a factoid Question Answering model decreases indicating an increase in performance
   a)   True
   b)   False

14. Teacher Forcing method that omits redundancy by using a single set of embeddings at the input of the softmax layers.
   a)   True
   b)   False

15. What is the CORRECT statement about Logistic Regression
   a)   Better to use when the features are strongly correlated
   b)   Trains faster than Bayes's classifier
   c)   Considered a generative model
   d)   All of the above

Question 2:

A] CKY Problem

Non terminal:	Terminal:
S -> NP VP	NP -> b
VP$_1$-> V NP	V -> b
VP$_2$-> NP V	N -> a
VP -> VP VP	
NP -> N NP	

Required word: 'babbbb', is it generatable using above grammar rules? (Yes/No)

b	a	b	b	b	b
NP,V	phi	VP	S	VP	S
	N	NP	VP	S	VP
		NP,V	VP	S	VP
			NP,V	VP	S
				NP,V	VP
					NP,V

B] Viterbi Problem

Given the tags (a1, a2, a3) and their initial probabilities (π), transition probabilities matrix (A), emission probabilities matrix (B), and given the sentence "b3b1b3b3b2".

π:

a1	a2	a3
0.6	0.2	0.2

A (Transition Probabilities Matrix):

	a1	a2	a3
a1	0.8	0.1	0.1
a2	0.2	0.7	0.1
a3	0.1	0.3	0.6

B (Emission Probabilities Matrix):

	b1	b2	b3
a1	0.7	0	0.3
a2	0.1	0.9	0
a3	0	0.2	0.8

Required to fill the Viterbi table and get the best tags for the given sentence.

C] Give examples of Intrinsic and Extrinsic dense vectors performance evaluation.

D] Describe **very** briefly how Fasttext handles Unknown words.

Question 3:

A]

An RNN with $U_{1\times1} = V_{1\times1} = W_{1\times1} = 1$, and linear activation & output functions $g(x) = x, f(h_i) = h_i$. What does it learn after training?

B] RNN Bottleneck Problem

You use an RNN to do sentiment analysis on hotel reviews. Each review receives a score of 0 (very negative) to 5 (very positive). The output depends on the last hidden layer only. The RNN was given this input: "The vacation was ruined, the food was bad, the rooms were bad. But surprisingly the staff seemed very happy". The model outputs "very positive". Why did this misclassification occur?

C]

A CBOW word2vec model learns the word vector matrix W_1, and a co-occurrence count based method also learns another word vector matrix W_2.

1. Are W_1, W_2 identical? (Yes/No)
2. If the CBOW model was retrained, would the new vector matrix be identical to the old one W_1? (Yes/No)
3. If the co-occurrence model was retrained, would the new vector matrix be identical to the old one W_2? (Yes/No)

D]

Given the following sentence "NLP exam is held at faculty of engineering". Apply **BIOES** tagging. Entities of interest are "COURSE", "ORGANIZATION", and "PERSON".

E]

Given a random model and a test set containing a sentence of 30 characters drawn randomly from the English alphabet. (Note that the English alphabet contains 26 characters) Compute the Perplexity of this test sentence.

Question 4:

A] Tweets Confusion Matrix Problem (Binary Classification)

	Golden "RUMOR"	Golden "NOT RUMOR"
System "RUMOR"	100	50
System "NOT RUMOR"	200	150

    i. Calculate Precision, Recall, F1

    ii. If it is very harmful to predict that tweets that are true "RUMOR" as "NOT RUMOR", is the above system the right choice? (Yes/No)

    iii. What is the reason for your answer to the question above?

B] Three Regex Problems. Write the output of the following code snippets:

    i.

```
import re

text = "NLP NLP exam --- exam"
regex = r"\b([a-zA-Z]+)\s+\1\b"
print([x.group() for x in re.finditer(regex, text)])
```
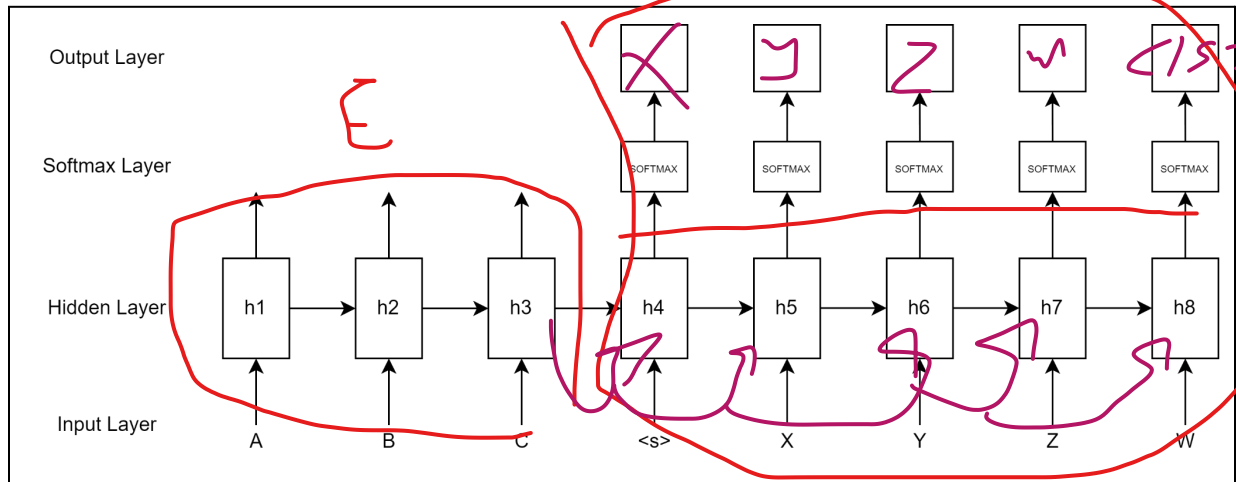
    ii.

```
import re

text = "I am $n$o$t getting %% in this exam"
text = re.sub(r"%%", "Excellent", text)
text = re.sub(r" \$\w\$\w\$\w", '', text)
print(text)
```

    iii.

```
text = "Natural lanuage processing (NLP)"
print(text.split())
```

C] Encoder-Decoder Model Question

Output Layer								
Softmax Layer			SOFTMAX	SOFTMAX	SOFTMAX	SOFTMAX	SOFTMAX	
Hidden Layer	h1	h2	h3	h4	h5	h6	h7	h8
Input Layer	A	B	C	<s>	X	Y	Z	W

i. Draw a rectangle around the encoder part.

ii. Draw a rectangle around the decoder part.

iii. What is the hidden state vector in the encoder that represents the context vector? (____)

iv. Complete the output layer of the decoder in the above diagram.

v. Draw a potential solution for the problem of the influence of the context vector being decreased as the decoder generates output.

vi. There is still a problem, the context vector is a bottleneck, which can be fixed using (____) attention

vii. Mention the **name** of one score that calculates the relevance of each encoder's hidden state to the decoder. (_____) dot Product