INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Stamp / Signature of the Invigilator

EXAMINATION (End Semester)											SEMESTER (Autumn)		
Roll Number										Section		Name	
Subject Number	C	S	6	0	0	5	7	Subject Name			*Natural Language Processing*		
Department / Center of the Student											Additional sheets		

Important Instructions and Guidelines for Students

1. You must occupy your seat as per the Examination Schedule/Sitting Plan.
2. Do not keep mobile phones or any similar electronic gadgets with you even in the switched off mode.
3. Loose papers, class notes, books or any such materials must not be in your possession, even if they are irrelevant to the subject you are taking examination.
4. Data book, codes, graph papers, relevant standard tables/charts or any other materials are allowed only when instructed by the paper-setter.
5. Use of instrument box, pencil box and non-programmable calculator is allowed during the examination. However, exchange of these items or any other papers (including question papers) is not permitted.
6. Write on both sides of the answer script and do not tear off any page. **Use last page(s) of the answer script for rough work.** Report to the invigilator if the answer script has torn or distorted page(s).
7. It is your responsibility to ensure that you have signed the Attendance Sheet. Keep your Admit Card/Identity Card on the desk for checking by the invigilator.
8. You may leave the examination hall for wash room or for drinking water for a very short period. Record your absence from the Examination Hall in the register provided. Smoking and the consumption of any kind of beverages are strictly prohibited inside the Examination Hall.
9. Do not leave the Examination Hall without submitting your answer script to the invigilator. **In any case, you are not allowed to take away the answer script with you.** After the completion of the examination, do not leave the seat until the invigilators collect all the answer scripts.
10. During the examination, either inside or outside the Examination Hall, gathering information from any kind of sources or exchanging information with others or any such attempt will be treated as **'unfair means'**. Do not adopt unfair means and do not indulge in unseemly behavior.

Violation of any of the above instructions may lead to severe punishment.

Signature of the Student

To be filled in by the examiner

Question Number	1	2	3	4	5	6	7	8	9	10	Total
Marks Obtained											

Marks obtained (in words)	Signature of the Examiner	Signature of the Scrutineer

Indian Institute of Technology Kharagpur
Midterm Autumn 2018

Speech and Natural Language Processing CS60057

Autumn 2018
Date: 22/11/2018

Endterm
Duration: 3 hours

Full Marks: 84

---

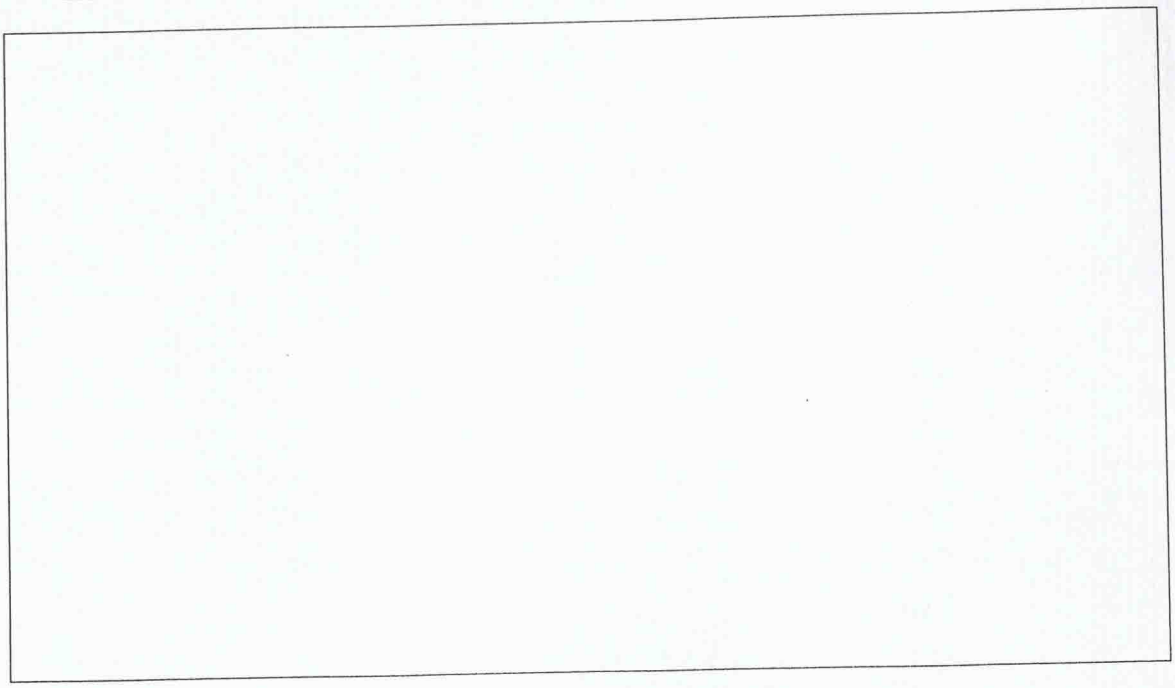
1. (a) (4 points) Although pre-trained word vectors work very well in many practical downstream tasks, in some settings it's best to continue to learn (i.e. 'retrain') the word vectors as parameters of our neural network.
   i. State why retraining may improve the results in certain situations.
   ii. Describe a situation where retraining the word vectors may hurt the model.

   (b) (4 points) Word2Vec represents a family of embedding algorithms that are commonly used in a variety of contexts. Suppose in a recommender system for online shopping, we have the transaction sequences for items purchased by each user.For example, a user u_i purchases items $p_{i1}, p_{i2}, \ldots p_{iN_i}$ The total set of items is given by $x_1, x_2, \ldots, x_n$. Explain how you would use ideas similar to Word2Vec to recommend similar items to users who have shown interest in any one of the items.

2. (a) (2 points) State two disadvantages of a ngram based language model.

   (b) (2 points) State two advantages of a RNN based language model compared with a ngram language model.

   (c) (1 point) State one advantage of a LSTM language model compared to a RNN language model.



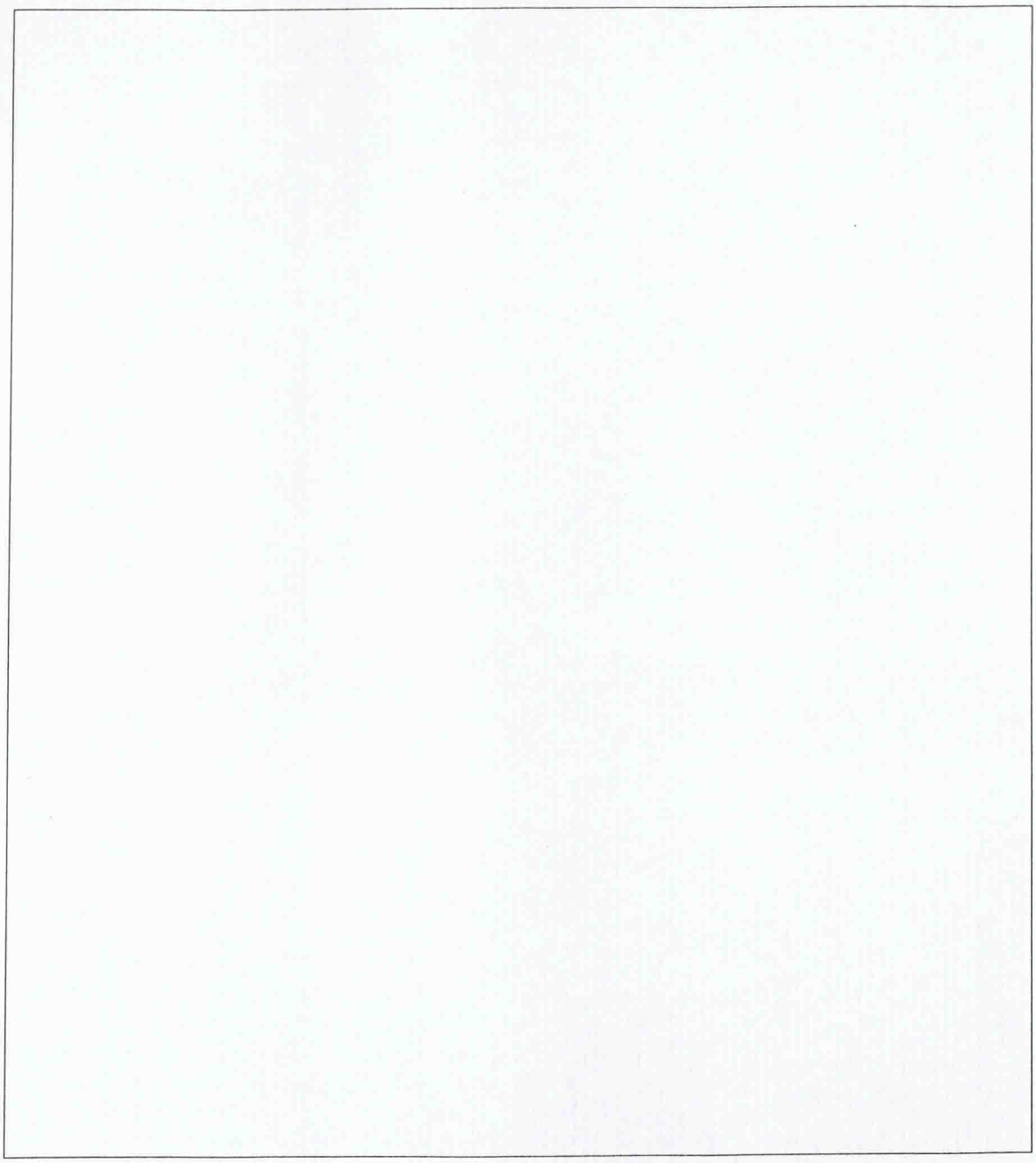
3. (a) (4 points) Assuming the grammar below, show the parse tree for the sentence

   The young girl went to the movie.

```
S-> NP VP
VP -> VP PP
VP -> verb NP
VP -> verb
NP -> DET NOM
NOM -> ADJ NOM
NOM -> NOUN
PP -> PREP NP
DET -> the
ADJ -> young
NOUN -> girl
VERB -> went
PREP -> to
NOUN -> movie
```

   (b) (4 points) Show how you would have to modify the grammar above to handle the following sentence.

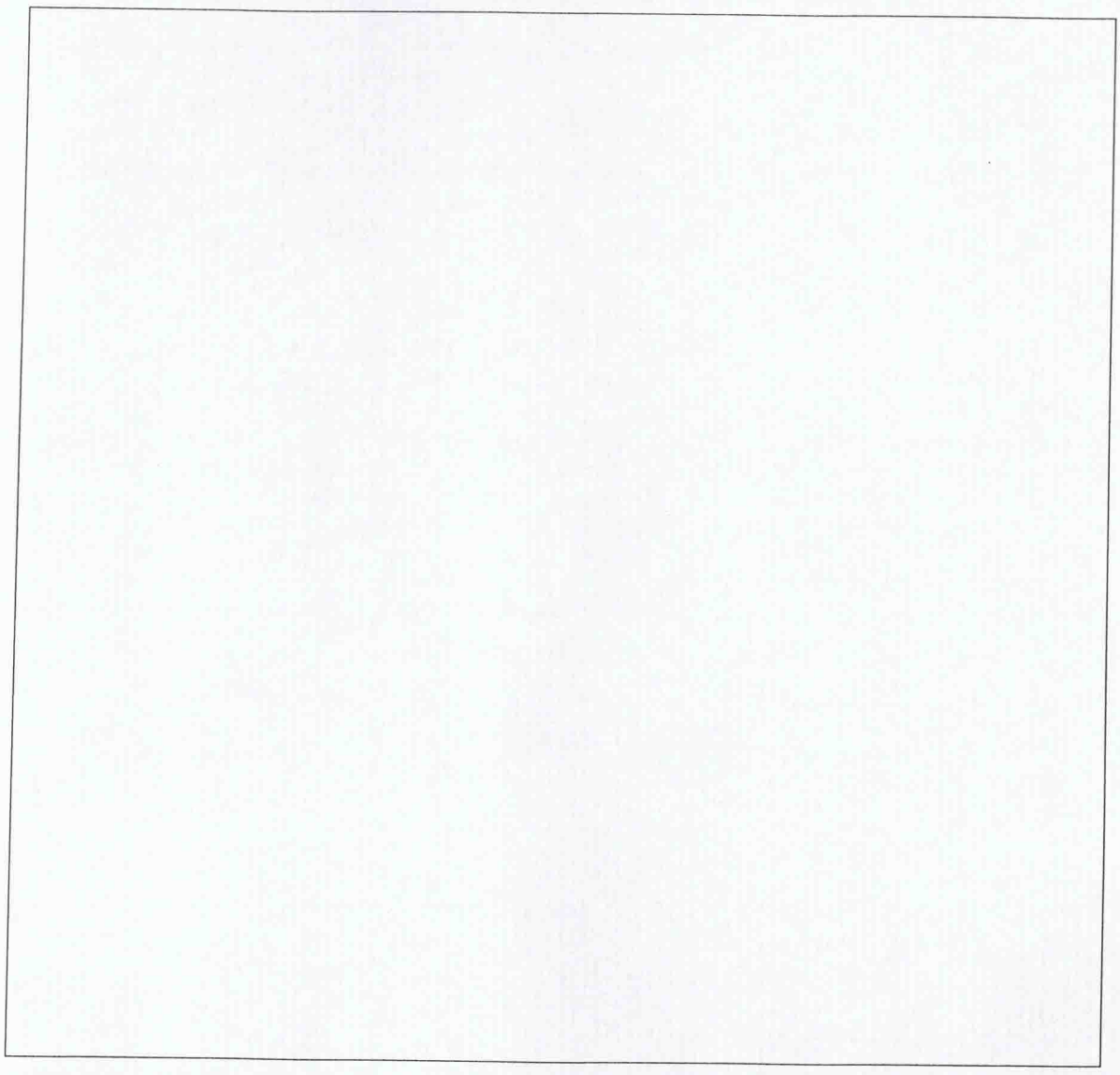
   The girl in the red dress went inside the park.



4. Consider the problem of text classification where a document needs to be classified into one of K classes.
   (a) (6 points) Design a simple CNN for the text classification task. Draw the corresponding network and describe the input, the output, the layers and their functions.
   (b) (2 points) What is one advantage of using a CNN compared to using a RNN for text classification?

5. (a) (4 points) How is noisy channel model applied to machine translation? What are the two components of the noisy channel model?

   (b) (5 points) Describe the steps in the generative process of IBM model 1 that can be used to generate a string f in the target language from an English string e.

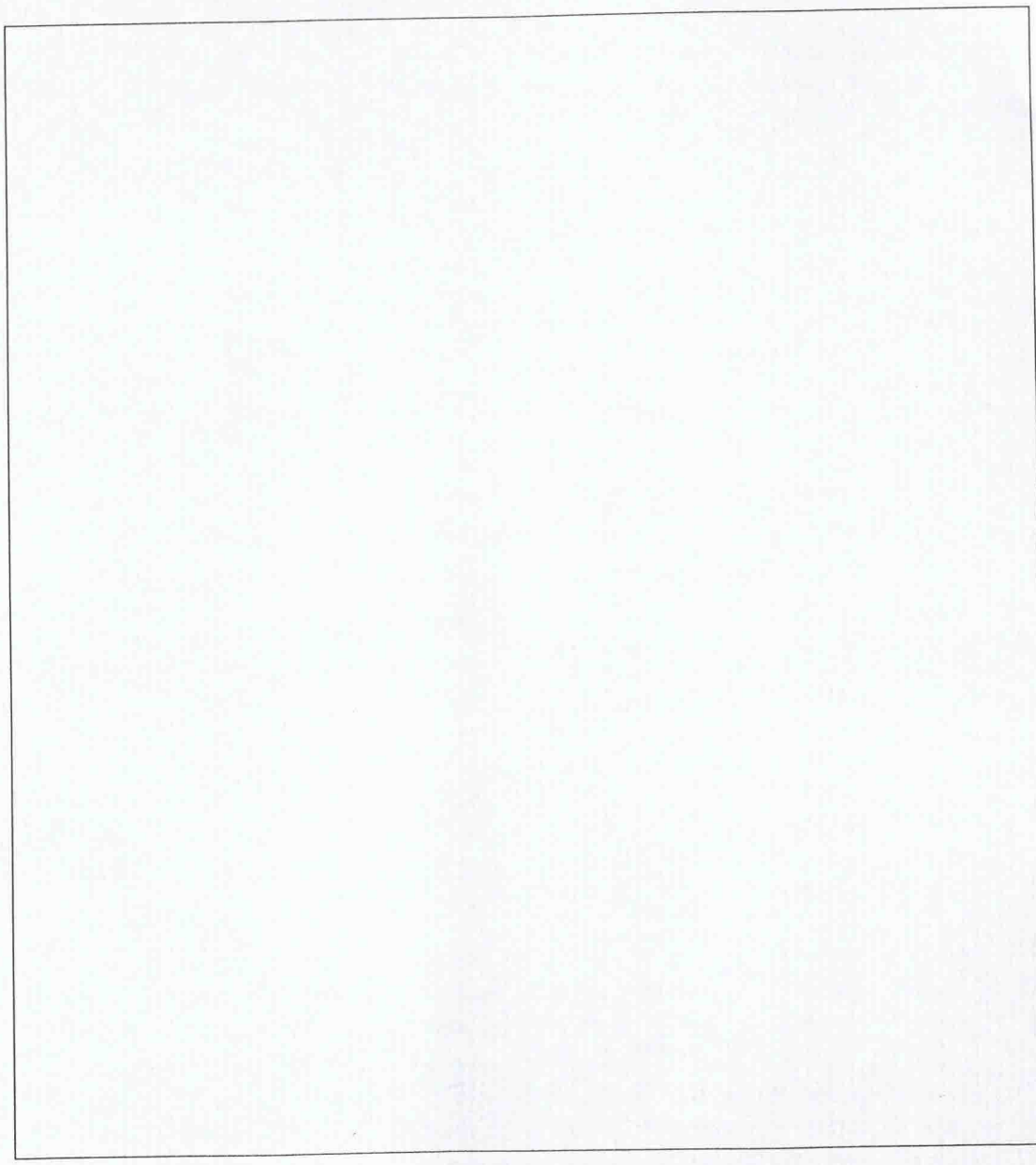


6. (a) (10 points) Assume a basic encoder-decoder based neural machine translation system without attention. Assume a single layer RNN based system.

   During training, the system takes as input pairs of sentences in two languages. Consider a sentence in source language $f_1 f_2 \ldots f_n$ and its translation $e_1 e_2 \ldots e_m$.

   Assume that the vocabulary size of the source language is V_f and that of the target language is V_e. Assume word vectors of size 300. State your assumptions about the other dimensions.

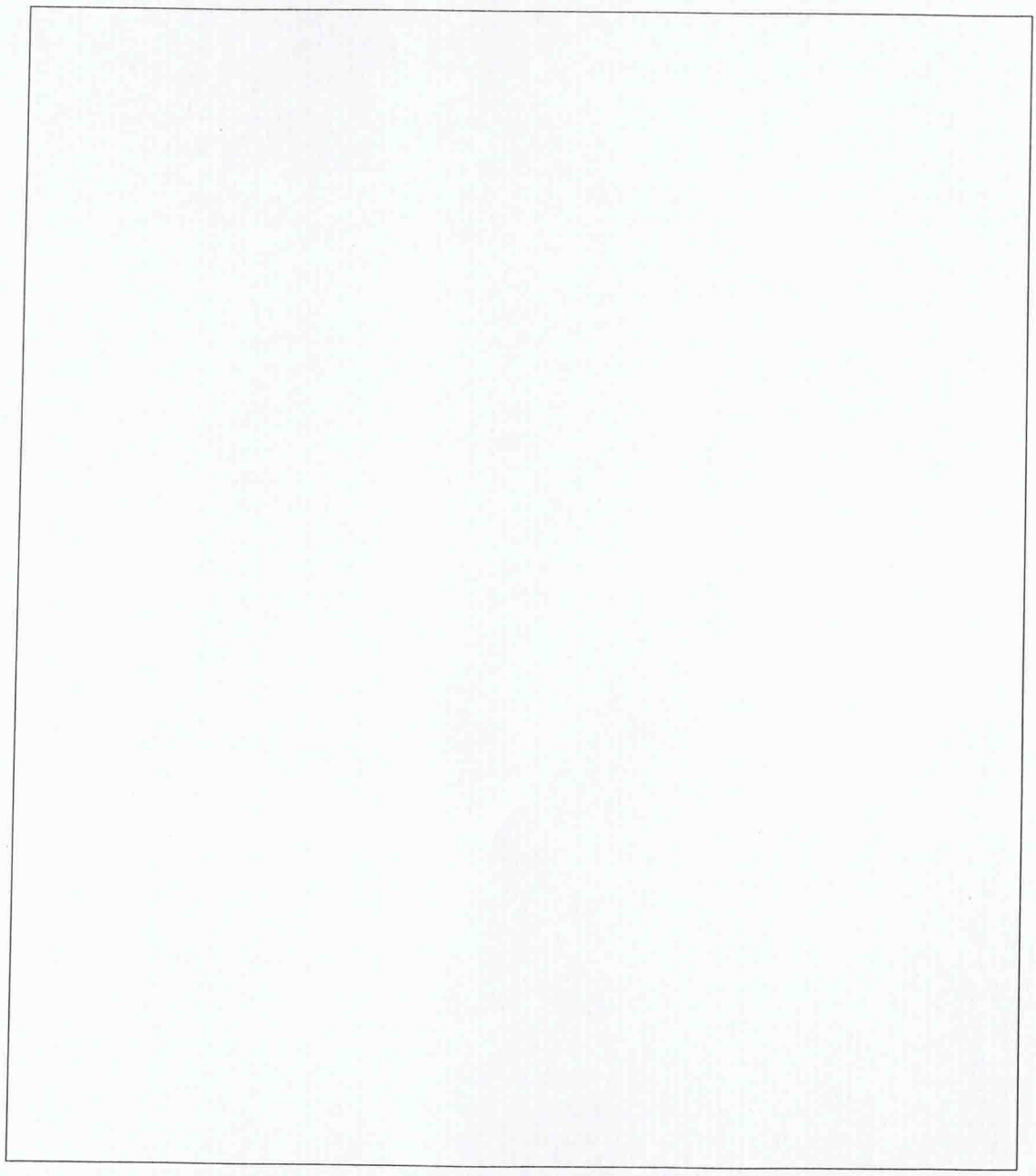
   Draw the encoder decoder cells and all other units for this example. Show all the input and output connections for each cell in the encoder and the decoder. Indicate the various parameters and their dimensions.



(b) (10 points) Suppose at test time you feed a sentence $\phi_1\phi_2 \dots \phi_k$. Explain precisely how the translated sentence in the target language is recovered when you use the following decoders:

i. greedy decoder

ii. beam search based decoder



(c) (2 points) Why is the model not expected to work well when the source sentence is too long?

(d) (4 points) The encoder-decoder model may be enhanced with attention. For an attention computation method of your choice, show how exactly the alignment score and the attention is computed when decoding the first word.

(e) (2 points) Explain how the attention model helps to get better translation for longer sentences.

7. Consider the task of sentiment analysis. The input is a sentence and the output is a sentiment class between 1 (very negative) and 5 (very positive). Consider the following alternative models for this task:

   1. RNN model
   2. LSTM model
   3. CNN model

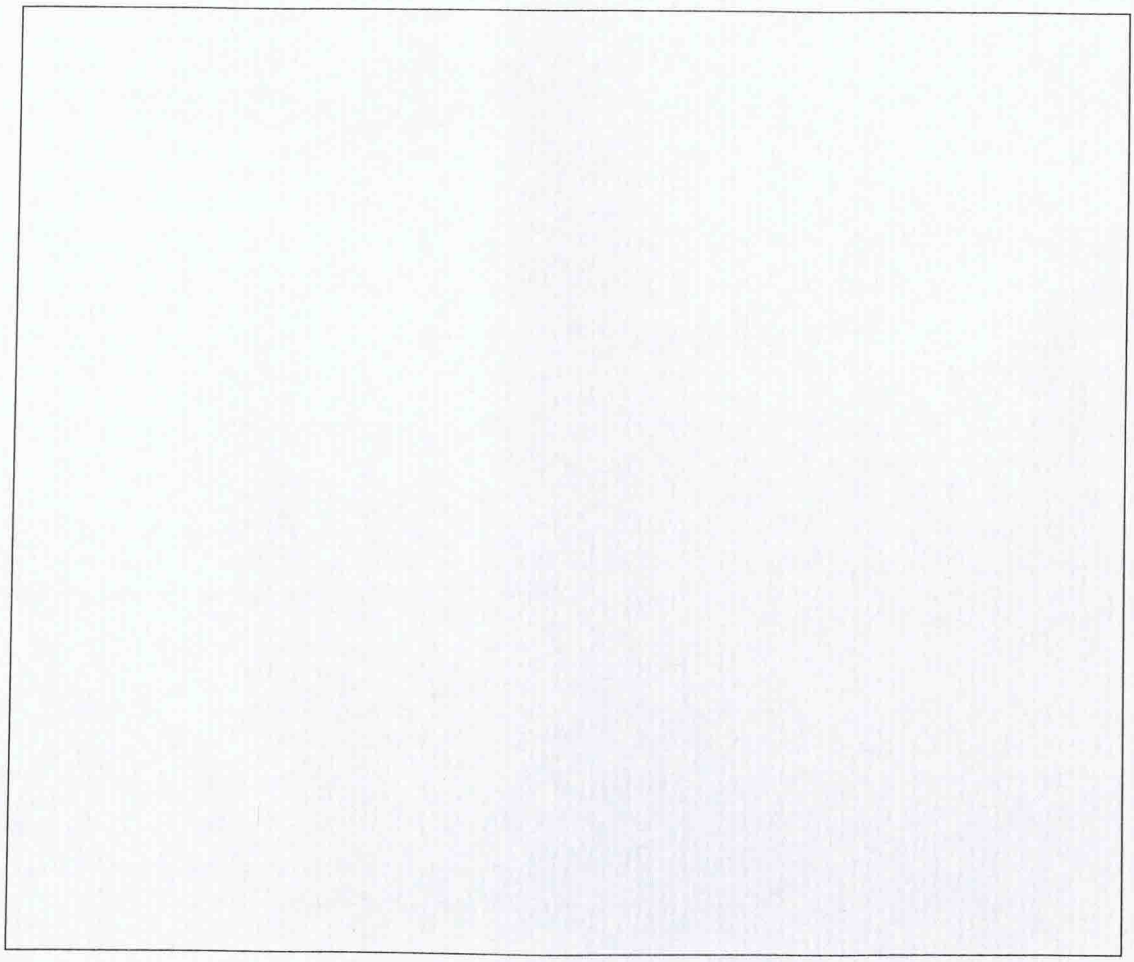
   For all the models below, assume that the vocabulary size is V and you use D-dimensional word embedding vectors.

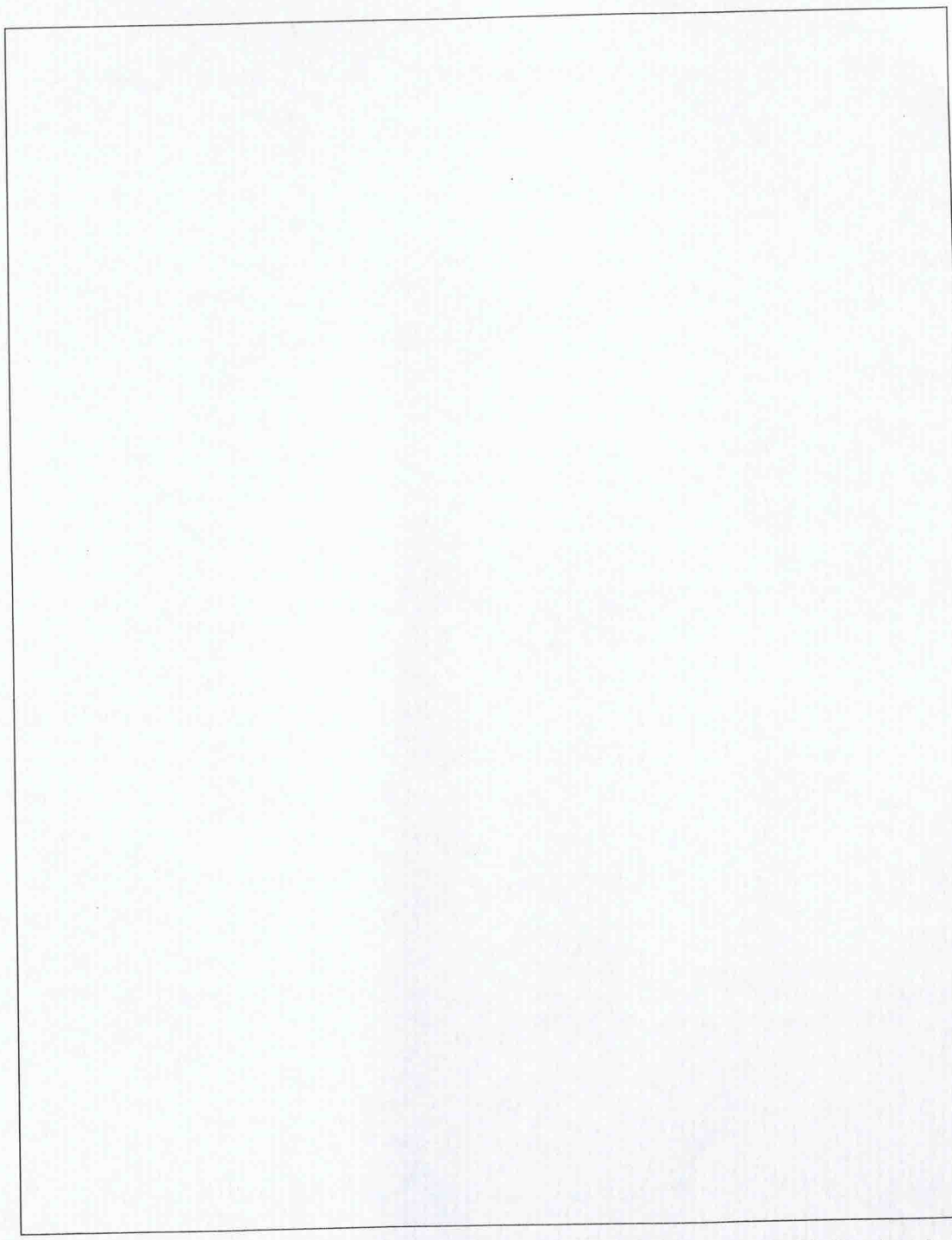
   (a) Consider a RNN implementation for sentiment analysis. Assume that there is a single hidden layer with H units.

   i. (3 points) Give the sketch of the RNN. Clearly indicate how the input is fed and where the outputs are.

   ii. (6 points) Specify how the RNN would be used at test time and specify

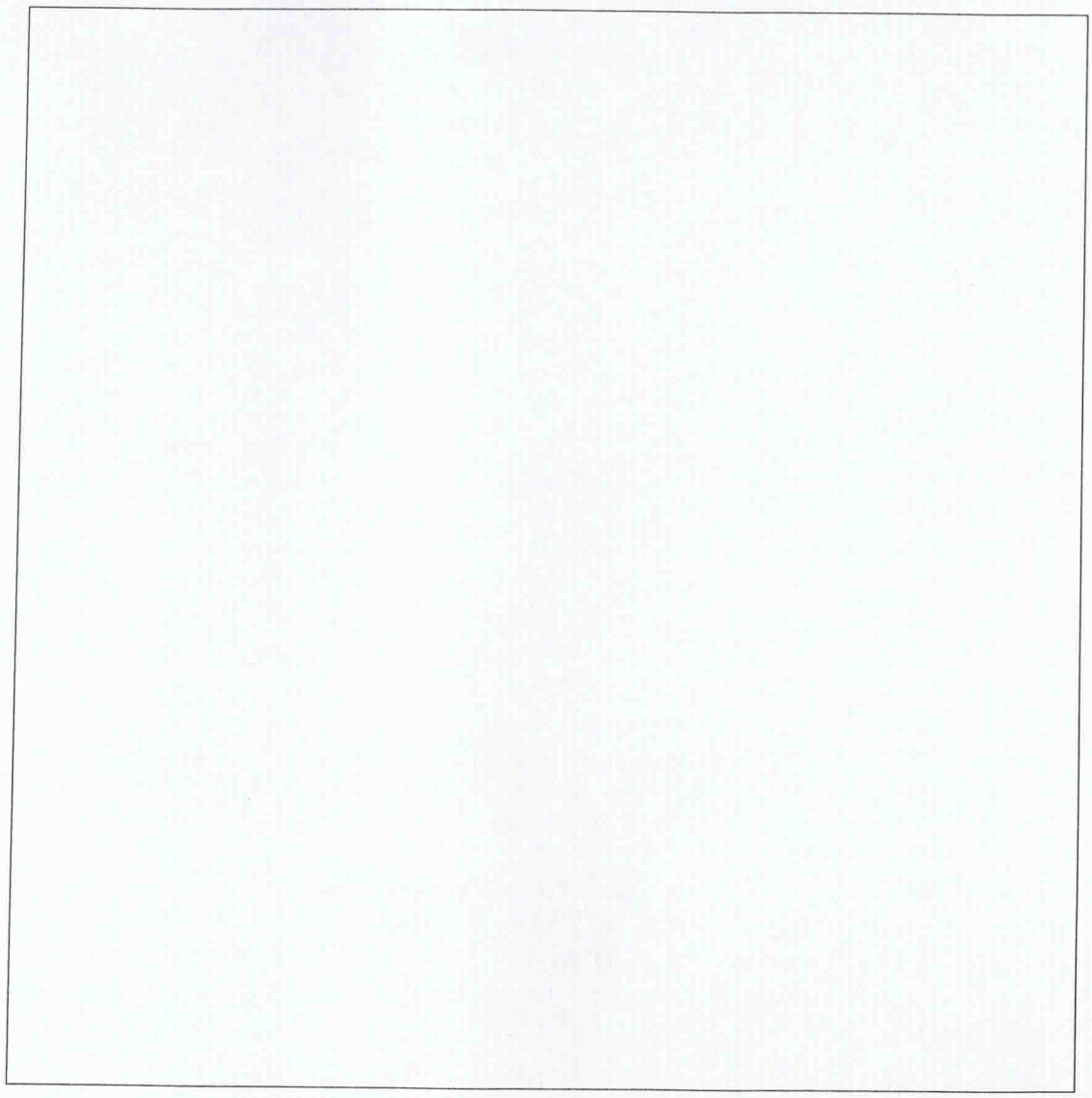
   1. how many outputs
   2. what is the form for each output
   3. which inputs are fed at each time step to produce each output
   4. What are the weights (parameters) associated with this network? Give the number of dimensions of each weight vector.





(b) (5 points) Consider a CNN implementation. Give the sketch of a CNN for the above problem. State any assumption about the number of parameters and the hyper-parameter values.

What are the weights (parameters) associated with this network? Give the number of dimensions of each weight vector.



(c) (4 points) i. Give an example of a sentence for which a RNN based model is more likely to give a wrong answer while a LSTM model may give a better answer.

ii. What benefit may a CNN model have over a RNN/LSTM based model for this task?

8. Consider mention ranking which is the task of assigning to each mention its highest scoring antecedent. Assume that the first stage of mention detection has been done. Model this as a Machine Learning problem. Suggest a ML model and specify the input and output of the model. Suggest a loss function or optimization objective for this task.