# Sample Final Exam

What will be tested:

- Basics (what, how)
- Understanding of the topics (why)
- What was covered:
    - After Midterm - around 80% weight
   - Before midterm – around 20% weight

What will NOT be tested:

- Python and NLP libraries

1. [3 points] The following diagram shows a recurrent neural network unrolled over time which is being trained for part-of-speech tagging. When the network generates the incorrect output "VBD", the error will be propagated from which nodes to which nodes? Answer using the labels shown on the nodes.
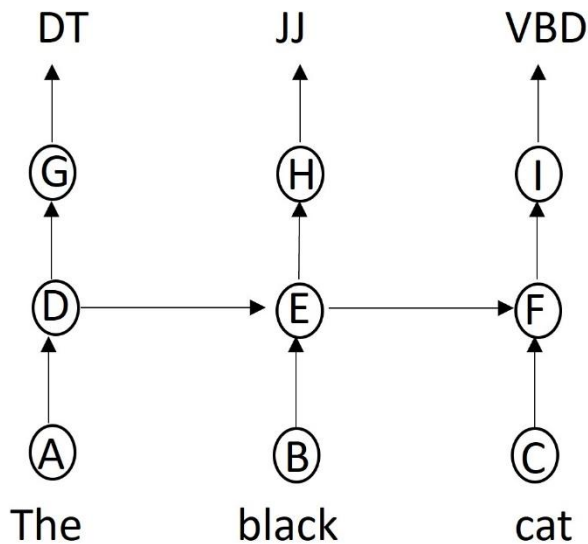
2. [4 points] Following is a diagram of the transformer architecture. Name its components labeled from A to G. Which of these components use(s) encoder-decoder attention and which use(s) self-attention?
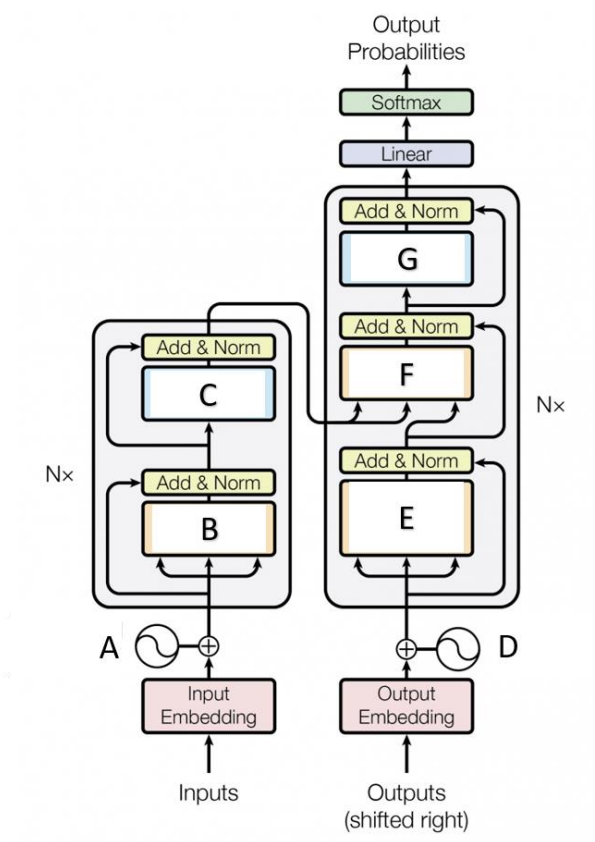
3. [2 points] Given a sentence with a word which could have multiple senses. Why is its embedding obtained through the output of BERT superior to its embedding obtained using earlier methods (such as skip-gram)? Answer in one or two sentences.

4. [4 points] Given the following machine translated sentence and the three reference translations, compute the BLEU score (use only unigrams and bigrams). Leave the answer as a mathematical expression, no need to do calculations.

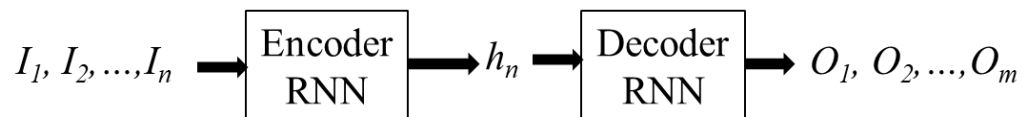Machine: It will be a good weather day with a few chance of rain.

Reference 1: The weather will be pleasant with a chance of rain.

Reference 2: The weather will be good with a possibility of rain.

Reference 3: Good weather with a probability of precipitation.

5. [3 points] A system needs to be built to automatically do the following: given a tweet and its reply, determine whether the reply is in agreement or in disagreement with the tweet. It is decided to employ BERT architecture for this task. Describe each in one or two sentences: (i) What should be the appropriate input given to BERT? (ii) What output from BERT will be relevant for the task?

6. [2 points] A robot has been designed that has pre-built functionalities to do cleaning inside a house. Currently it takes commands only in its own machine language. To make it more user-friendly, the capability to take commands in natural language needs to be added to the robot. Adding this capability should be treated as an instance of which NLP task? What type of training data will you provide if machine learning is to be used to built a system for this task? Answer each question in one or two sentences.

7. [3 points] Mention a major drawback of the following architecture where I and O are input and output sequences respectively, and h_n is the context vector. Name a method that can counter that drawback (details of the method are not needed).

$$I_1, I_2, ..., I_n \longrightarrow \boxed{\begin{array}{c} \text{Encoder} \\ \text{RNN} \end{array}} \longrightarrow h_n \longrightarrow \boxed{\begin{array}{c} \text{Decoder} \\ \text{RNN} \end{array}} \longrightarrow O_1, O_2, ..., O_m$$

8. [2 points] What component of transformer is missing in BERT? Why it does not need it?

9. [2 points] Why does the transformer architecture lose the order of words in an input sequence? What is done to remedy it?