COMP4446 / 5046 Sample Questions

1. WorNet and vectors both capture relationships between words.

    (a) (2 marks) What are the ways they represent relationships?

    > **Solution:** WordNet: A database of labeled relationships between words.
    > Vectors: Location in the space, e.g., proximity can indicate similarity.

    (b) (1 mark) What are the benefits of WordNet's approach?

    > **Solution:** Explicitly records the relationships in a human-interpretable way.

    (c) (1 mark) What are the benefits of using vectors?

    > **Solution:** Two options: (a) low human effort as they are determined automatically, (b) can represent soft relationships or varying degrees of a relationship,

2. (1 mark) What does it mean if two words have low distributional similarity?

    > **Solution:** The words are used in different contexts.

3. (1 mark) How could a Bag of Words model be adapted to account for word senses?

> **Solution:** Identify the word sense of each word and modify what goes into the bag to include it. An example is not needed, but you could say "bat (the animal)" is inserted rather than "bat".

4. (1 mark) Why do we usually care more about True Positives than True Negatives?

> **Solution:** In many NLP tasks we are looking for something, e.g., named entities. The true negatives are the words we aren't interested in (the non-entities), so they are less important to us.

5. (1 mark) When we talked about interpretability of hidden states in RNNs, we said things like "This position turns on inside quotes." What does 'turns on' mean?

> **Solution:** The value of that position in the hidden state (or cell state) is not zero / not very close to zero.

6. What do each of the following labels mean in NER?
   (a) (1 mark) B-PERSON

   > **Solution:** This token is the beginning of a person entity.

(b) (1 mark) I-LOCATION

> **Solution:** This token is inside of, but not the beginning of, a location entity.

(c) (1 mark) O

> **Solution:** This token is not part of an entity.

7. (1 mark) What do random sampling, top-k sampling, and top-p sampling all have in common?

> **Solution:** They all get a random sample from some set of options (the options vary across the methods).

8. (1 mark) In beam search, when does an item get removed from the beam?

> **Solution:** If an item is at the bottom of the beam (ie., has the lowest score of the current options) and a new option has a higher score, then the item is removed.

9. (1 mark) In the sentence below, which nouns would be put in the same clusters by a perfect coreference resolution system?

"I came to the room with the chocolate. I saw the chocolate. I ate the chocolate."

> **Solution:** [I, I, I], [the room], and [the chocolate, the chocolate, the chocolate]

10. (1 mark) Does the Viterbi algorithm only work with RNNs? Why / Why not?

> **Solution:** No, it can work with a wide variety of models. It just needs a model that gives a score to transitions and a score to emissions.

11. (1 mark) Consider the decoder in an encoder-decoder.

   (a) (1 mark) Where does the first input come from?

   > **Solution:** It is a fixed value that we provide.

   (b) (1 mark) Where do all the other inputs come from?

   > **Solution:** Each input token is the output from the previous step of the decoder.

12. (1 mark) We saw a range of different equations for different types of attention. What are those equations all calculating?

   > **Solution:** The scores used in the weighted average.

13. (1 mark) What is the value of having multiple heads in attention?

   > **Solution:** Each head can capture a different aspect of the input, e.g., focusing on syntactic relationships vs. semantic relationships.

14. (1 mark) What does a residual connection do with an input vector?

> **Solution:** Passes the vector over the computation (e.g., a feedforward layer) and adds it in afterwards.

15. One way to train a language model is with token edit detection. Another model (the editor) edits some of the input tokens and the model you are training (the predictor) has to predict "edited" or "same" for every token in the input.

    (a) (1 mark) What would go wrong if most of the tokens were edited?

    > **Solution:** The model might struggle to learn because there is no context to inform the decision of what has been edited.

    (b) (1 mark) What would go wrong if only one of the tokens was edited?

    > **Solution:** The model might learn very slowly as there is a very limited training signal.

16. (1 mark) How does an n-gram language model use data to inform its probabilities?

> **Solution:** N-grams in the data are counted and the counts are used to determine probabilities by dividing one count by another.

17. (1 mark) Given an example of how you can use the probabilities from a language model to classify a news article into one of these four topics: politics, entertainment, sport, tech.

> **Solution:** Provide the article as input, followed by "the topic of this article is:" and then look at the distribution of probabilities for the next token, comparing the four options listed and taking the one with highest probability.

18. (1 mark) When we use RAG, we calculate the similarity between things. What are we comparing and why?

> **Solution:** The query is compared with each piece of text in the database. This identifies text in the database that is related to the query.

19. (1 mark) When doing annotation, one option is to edit the outputs of a model.

(a) (1 mark) Where does that model come from?

Solution: The model is from prior work, trained on other data.

(b) (1 mark) What problem(s) could that cause?

Solution: The model may not be accurate on this data or have certain patterns of errors.

20. Consider the table of annotation counts for two people below.

		Annotator 1	
		Label A	Label B
Annotator 2	Label A	5	5
	Label B	5	5

(a) (1 mark) What will Cohen's Kappa be in this case?

> **Solution:** 0

(b) (1 mark) What would happen to Kappa if all of the numbers were doubled?

> **Solution:** The value would stay the same.

21. (2 marks) RLHF involves a model with a purpose we had not discussed earlier in the unit. What is the model and what is its purpose?

> **Solution:** A reward model. Its purpose is to give scores to different output options that are consistent with how a human would judge them.