Chapter 11: Words, and the Company They Keep

Please ...

1. Read "Chapter 11: Words, and the Company They Keep" of Melanie Mitchell's "Artificial Intelligence: A Guide for Thinking Humans" book.
2. With respect to the questions presented for the "Chapter 11: Words, and the Company They Keep" reading, construct a document containing question/answer pairs, one pair of each question, where the answers, with perhaps just a small number of exceptions, are more or less lifted from Melanie Mitchell's text. Save your document as a **pdf** file.
3. Post your question/answer document to your web worksite.
4. Do your best to internalize your question/answer pairs in some sort of semantic sense, so that the answers are likely to come back to you when prompted by the questions.
5. Come to class for the discussion of "Chapter 11: Words, and the Company They Keep," when the time rolls around, prepared to participate in the discussion.
6. Please do all of this within one week of the "distribution" of this assignment.

Chapter 11: Words, and the Company They Keep - Questions

1. What is the phrase used to describe the field of AI that pertains to "getting computers to deal with human language."

2. List some topics that are included within the realm of NLP research.

3. TRUE/FALSE - Understanding nagural language is among AI's most difficult challenges, largely because language is inherently ambiguous, is deeply dependent on context, and assumes a great deal of background knowledge common to the communicating parties.

4. TRUE/FALSE - As with other areas of AI, the first several decades of NLP research focused on symbolic rule based approahces, which did not work very well since capturaing the subtleties of language by applying a set of explicit rules is far more challenging that was initially believed.

5. TRUE/FALSE - In the 1990's, rule-based NLP approaches were overshadowed by more successful statistical approaches, in which massive data sets were employed to train machine-learning algorithms, which today are informed by deep learning.

6. The task of trascribing spoken language into text in real time was deep learning's first major success in NLP. What is the phrase that is used to refer to this task?

7. TRUE/FALSE - Current speech recognition systems are so successful because they *understand the meaning* of the speach they are transcribing.

8. What famous rule of thumb that appies to any complex engineering project does MM reference in her discussion of automatic speech recognition?

9. What advance in NLP does MM believe will be required in order to deal with the outstanding issues of noise, unfamiliar accents, unknown words, and the fact that ambiguity and context sensitivity of language, that collectively impinge upon interpreting speech?

10. What does the word "sentiment" mean?

11. TRUE/FALSE - An AI system that could accurately classify a sentence (or longer passage) as to its *sentiment*–positive, negative, or some other dimension of opinion–would be solid gold to companies that want to analyze customers' comments about their products, find new potential customers, automate product recpmmendations ("people who liked X also like Y"), or selectively target their online advertisements.

12. TRUE/FALSE - Some early NLP systems looked for the presence of individual words or short sequences of words as indications of the sentiment of a text. What did the researchers who built these systems learn from their experiments? They learned that looking at single worlds or short sequences in isolation is generally not sufficeint to glean the overall sentiment; rather, it is necessary to capture the semantics of the words in the context of the whole sentence.

13. What kind of neural network tends to be used to process sentences (or melodies, or stock market trends), in such a way that a sequence of inputs of arbitrary lengths are mapped to a fixed size structure in order to capture, in some sense, a comprehensive representation of the inputs.

14. In abstract terms, explain the characteristic feature or recurrent neural network architecture.

15. In abstract terms, describe the process which governs recurent neural network execution for "encoding" networks.

16. TRUE/FALSE - Given a set of sentences that humans have labeled as "positive" or "negative", in sentiment, the encoder network can be trained from these examples via back-propogation.

17. Neural networks require their inputs to be *numbers*. So when processing a sequence of words, a mechanism is needed to map words to numbers. Name, and describe, the simplest such

mechanism that is presented in the text.

18. The NLP research community has proposed several methods for encoding words in a way that would capture certain semantic relationships. All of these methods are based on the same idea, which was expressed beautifully by the linguist John Firth in 1957: "You shall know a word by the company it keeps." What does this sentence mean with respect to the meaning of a word?

19. TRUE/FALSE - In linguistics, the idea captured by John Firth's poetic phrase is more formally known as **distributional semantics**, which expresses the following theory: "The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear."

20. Related to the theory of distributional semantics is the concept of a **semantic space**. What is a "semantic space"?

21. TRUE/FALSE - Once all the words in the vocabulary are properly placed in semantic space, the *meaning* of a word can be represented by its location in space–that is, by the coordinates defining its word vector. Moreover, it turns out that using word vectors as numerical inputs to represent words, as opposed to the simple one-hot scheme, greatly improves the performance of neural networks in NLP tasks.

22. In 2013 researchers at Google propose an algorithm, which they called "Word2Vec" for properly placing words into a semantic space in order to best capture the many dimensions of each word's meaning. In just a few sentences, describe the essence of the Word2Vec method that these researchers used to create the semantic space which they have since shared with the world.

23. What analogy did MM present as a segue to her discussion of word vectors giving rise to the notion of thought vectors, as an enchancement of word vector technology for capturing semantics?

24. TRUE/FALSE - Word vectors capture the biases inherent in the language data that produce them, and biases in word vectors might very well seep through neural networks to produce unexpected, hard-to-predict biases in widely used NLP applications.