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."

That field is "natural language processing," (178).

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

Some topics included within the realm of NLP research are: "speech recognition, web search, automated question answering, and machine translation," (178).

3. TRUE/FALSE - Understanding natural 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 approaches, which did not work very well since capturing the subtleties of language by applying a set of explicit rules is far more challenging than 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 transcribing 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?

The transcription task is known as "automated speech recognition," (180).

7. TRUE/FALSE - Current speech recognition systems are so successful because they *understand the meaning* of the speech they are transcribing.
8. What famous rule of thumb that applies to any complex engineering project does MM reference in her discussion of automatic speech recognition?

Mitchell writes, "There's a famous rule of thumb in any complex engineering project: the first 90 percent of the project takes 10 percent of the time and the last 10 percent takes 90 percent of the time," (181).

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?

Mitchell believes that machines will need to be able to understand the semantics of language in order to deal with the problems listed above (182).

10. What does the word "sentiment" mean?

A sentiment is the "degree" or value of an opinion–it can be positive, negative, or another value (182).

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 recommendations ("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 words or short sequences in isolation is generally not sufficient 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.

"Recurrent neural networks" are often used to process sentences, (184).

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

The characteristic feature of recurrent neural networks is the fact that it considers previously processed words in its calculation of activation of the current input to mimic how humans read a sentence from left-to-right, performing sentiment analysis as they go along (184-185).

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

The positive sentiment is calculated for an input word at each time step, where a time step denotes one input in sequential order (based on position in the sentence). The activation (encoding) is calculated and used for calculating the activations of future time steps. This continues until the machine receives the END symbol, denoting that the end of the sentence has been reached. The final sentiment is calculated and encoding is completed (186).

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-propagation.
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.

"One-hot encoding" assigns numeric value to words given to a neural network. When a given word is input into the neural network, it is "switched on" or "hot" (value equals 1), while all other words are "off" (value equals 0) (187).

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?

This sentence illustrates the idea that words share semantic relationships with other words and often occur in similar situational contexts with one another (188).

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"?

A semantic space uses dimensions to model the semantic similarity between different words. Since words can have multiple meanings, i.e. charm meaning wit or jewelry, dimensions are used to model each of the different meanings. The proximity of the words on a plane likewise demonstrates the degree of similarity between their meanings (188-189).

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 proposed 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.

Google used their large collection of Google News documents to train Word2Vec. Using these documents, the neural network associates pairs of words together, discarding common words like "the," to create a basis for predicting words adjacent to an input word. Word2Vec uses one-hot encoding to store the words in its vocabulary. Activation corresponds to "...the network's confidence that the corresponding word in the vocabulary has occurred adjacent to the input word," (192). This training model is used to create the semantic space of Word2Vec (191-192).

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 enhancement of word vector technology for capturing semantics?

Mitchell writes, "...to a person with a hammer, everything looks like a nail; to an AI researcher with a neural network, everything looks like a vector," (195).

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.