

Multiple choice. Circle all that apply

1. Which of the following are true about a valid dependency tree?

- a. There must be a path from 0 to every node i in the tree
- b. Every node must have at least one parent
- c. Each arc is labeled with a word
- d. Nodes must not form a cycle.

2. In a fully-connected, feed-forward neural network

- a. Some neuron in layer i is connected to every neuron in layer $i+1$
- b. Each neuron in layer i is connected to some neuron in layer $i+1$
- c. Each neuron in layer i is connected to every neuron in layer $i+1$
- d. Some neuron in layer i is connected to some neuron in layer $i+1$

3. A single layer perceptron

- a. has many input neurons but only one output neuron
- b. has one hidden layer
- c. has no hidden layers
- d. can only learn functions where the input is linearly separable

4. True or False: The grammar shown below is ambiguous for the sentence “The man called the girl on a chair.”

S \rightarrow NP VP
NP \rightarrow Det NOM
NOM \rightarrow noun
NOM \rightarrow NOM PP
VP \rightarrow VP PP (typo)
VP \rightarrow verb
VP \rightarrow VP NP (this rule was missing from original question as posed)
PP \rightarrow prep NP

5. A good source of seed patterns to bootstrap information extraction for biographies is:

- a. Google search results
- b. Manual patterns
- c. Wikipedia entries for people
- d. WordNet

6. Which of the following word pairs is an example of a hypernym?

- a. *animal* is a hypernym of *dog*
- b. *black* is a hypernym of *white*
- c. *big* is a hypernym of *large*
- d. *bank* is a hypernym of *building*

Short Answer

1. Provide two sentences that could be generated by the following AMR
(r / read-01
arg0 (b1 /boy)
arg1 (b2/ book
:poss (a /amr-unknown))
2. Name two sources of disambiguation for POS tagging and describe how they are used in hidden markov modeling.
3. How many gates does an LSTM have? State what they are and what each gate does (one sentence/gate).
4. State two different kinds of information that would be useful for learning a dependency parser and explain why (1 sentence each).
5. What is the difference between distributional representations and distributed representations?
6. State two ways in which BERT is different from Word2Vec.
6. In the following text, circle the mentions. What does the term *wikification* mean and why is it useful to develop a program that does this?

Donald John Trump (born June 14, 1946) is the 45th and current President of the United States, in office since January 20, 2017. Before entering politics, he was a businessman

and television personality. Trump was born in the New York City borough of Queens. He earned an economics degree from the Wharton School of the University of Pennsylvania and followed his grandmother Elizabeth and father Fred in running the family real estate company.

7. A phrase-based MT system uses a language model, a phrase-based model and a distortion model. Give a one sentence definition of each in the context of MT, referring to how they are used.

8. What is the vanishing gradient problem and when does it arise in neural networks? What are two techniques that can be used to combat it? (Your answers can apply to any kind of network you want, e.g., a recurrent network.)

Problem solving

3. [20 points]. Summarization. Suppose you want to develop a machine learning approach to summarization that extracts phrases rather than full sentences and puts together the phrases to form a sentence for the summary.

You need to create training data that your learner can use. You have access to the Zipf-Davis corpus, which consists of articles associated with a human abstract which has quite a bit of rewriting of the input article.

1. Describe how you might use a parser and what code you would have to write to access phrases below the sentence level.
2. Describe how the following methods of determining when an extracted clause from the summary matches a clause from the article would result in different kinds of abstraction:
 - a. Using word overlap only
 - b. Using WordNet also
3. Many summarization systems use language models. Explain how they are used as part of a neural net summarization system. Provide a diagram of the summarization system architecture.

3. [14 points]. Information Extraction and bootstrapping. Suppose you are building an information extraction system to identify the city and state in which a person was born. You want to use bootstrapping to do this.

- a. [8 points] You know where Barack Obama was born (Honolulu, Hawaii) but you don't know where any other famous person was born. Describe how you could use this information, along with a combination of data from Google and Wikipedia, to find patterns that could be used in general to determine place of birth. Be specific. Show the algorithm and some examples of the algorithm in operation (you can make up data that you think would be available).

- b. [4 points] Why is it better to use both Google and Wikipedia rather than one corpus alone? What advantage does each corpus have?

