ECE365 Spring 2021 Quiz 2: 05/04/2021

Instructions: Show your work to receive full credit. Make sure your writing is legible.
Total points: 60

1. [**Language Modeling** (15 points)]

   A) Name two NLP applications where language modeling is useful (2.5 points). For each application mention how language modeling can help the application (2.5 points).

   B) In order to compute the probability of the sentence **`I see you running around`** using a **4-gram language model**, write P(`I see you running around`) in terms of the necessary probabilities. You can pad as many <start> and <end> tags as needed. (5 points)

   C) Suppose you are given a large corpus and you have very good smoothing techniques to train a good language model. Despite this, what could be one reason (related to the property of language) that makes an n-gram language model insufficient to model language in general? (5 points)

2. [**Sequence Labeling** (15 points)]

   A) Tag ambiguity (5 points)
   We are given the sentence: `The representative put chairs on that table.`
   We also know the following about the parts-of-speech tags of words:

   - `Representative` can be `NOUN` or `ADJECTIVE`
   - `The` is `DT`
   - `That` can be `DT, ADVERB` or `CONJUNCTION`
   - `On` is `PREPOSITION`
   - `Put` can be `NOUN` or `VERB`
   - `Chairs` can be `VERB, or NOUN`
   - `Table` is `NOUN`

   Question: How many tag sequences are possible?

   B) Consider the following HMM for tagging (all transition and emission probabilities needed to answer this question have been provided).

Emission Probabilities	**Transition Probabilities**
P(*she*|PRON) = 0.1	P(PRON|START) = 0.5
P(*can*|AUX) = 0.2	P(AUX|PRON) = 0.2
P(*can*|NOUN) = 0.001	P(NOUN|PRON) = 0.001
P(*run*|VERB) = 0.01	P(VERB|AUX) = 0.5
P(*run*|NOUN) = 0.001	P(NOUN|AUX) = 0.001
	P(VERB|NOUN) = 0.2
	P(NOUN|NOUN) = 0.1

Which of the following statements is/are true? Choose **only one** of a-e as your answer. (5 points)

1. Given that a word is a NOUN, the probability that the word is *can* is 0.001.
2. The probability that the word after an AUX is **not** a VERB is 0.5.
3. P(*she*/PRON *can*/AUX) > P(*she*/PRON *can*/NOUN).

a) Only 1 and 3
b) Only 1
c) Only 1 and 2
d) Only 2 and 3
e) 1, 2 and 3

C) Suppose we have the following POS-tagged training data:

- *viruses/NOUN are/VERB dangerous/ADJ*
- *flying/ADJ planes/NOUN are/VERB numerous/ADJ*
- *I/PRON saw/VERB Mary/PROPN draw/VERB planes/NOUN*
- *He/PRON planes/VERB shelves/NOUN*

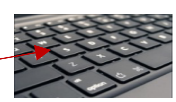
Calculate the likelihood probabilities (probability of a word given the POS tag) for the following words (no smoothing needed). (5 points)

a. P(planes|VERB) =
b. P(planes|NOUN) =

3. [**Lexical Semantics** (15 points)]

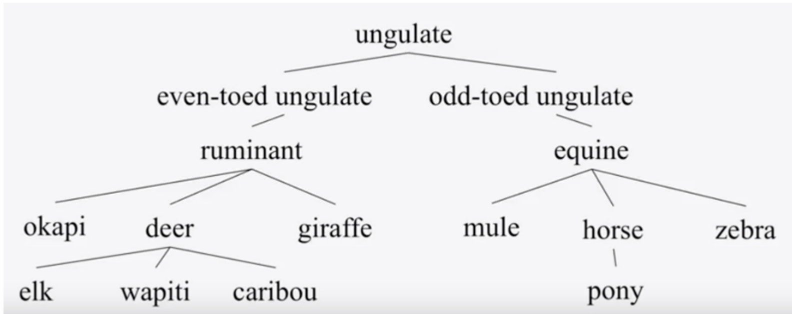
A) In the class we saw some semantic relations that have been defined between words. Name the relation between the following pairs of words by choosing among **synonym, antonym, hypernym, hyponym, meronym** and **holonym**. (10 points)

a. *cold* is a/an ___________ of *icy*
b. *friend* is a/an __________ of *enemy*
c. *head is* a/an ___________ of *body*
d. *Politician* is a/an ____________ of *Donald Trump*



e. *Keyboard* is a/an ____________ of *the 's' key*

B) Consider the following hierarchy of some animals from WordNet. Suppose words are compared by identifying the nodes where the different words appear and then counting the number of links that are needed to get from one to the other. Let us also assume that the smaller the number of links between words, the similar they are in meaning.



Using this notion of similarity (also called WordNet similarity) mention which of the word pairs mentioned below is most similar. (5 points)

a. deer-elk b. deer-horse c. deer-pony d. okapi-caribou

4. [**Word Embeddings** (15 points)]

A) One-hot word vectors: Consider the following 2-sentence corpus. 1) `have a good day.` and, 2) `have a great day.` Notice that the sentences are similar in meaning. Let the complete sorted vocabulary based on this corpus be V = {a, day, good, great, have}. Now, let us create a simple word vector – the one-hot vector – for each of the words in V. For this, represent every word as vector in $\mathscr{R}^{|V|x1}$ (column vector with 5 rows), with the components all 0s and one 1 at the index of that word in the sorted vocabulary. As an example, the one-hot vector for `have =` [0,0,0,0,1]' (' for the transpose operation).

a) Write the one-hot vectors for the words `a, day, good` and `great.` (5 points)

`a = , day = , good = , great =`

b) One of the desired properties of word embeddings is that they should be able to capture similarity in meaning, i.e., related words should have 'similar' vectors and unrelated words 'dissimilar' vectors. Suppose we use **cosine similarity** to compare words. Does the one-hot encoding satisfy the desired property of word embeddings? If yes, demonstrate why; if not, show why not. (5 points)

B) Another semantic property of word embeddings is their ability to capture relational meanings. For example, it was found that for word2vec, $v_{king} - v_{man} \approx v_{queen} - v_{woman}$

Which of the following would you also expect to hold for a good word embedding? Check all that are correct. (5 points)

a. $v_{boy} - v_{girl} \approx v_{brother} - v_{sister}$
b. $v_{boy} - v_{girl} \approx v_{sister} - v_{brother}$
c. $v_{boy} - v_{brother} \approx v_{girl} - v_{sister}$
d. $v_{boy} - v_{brother} \approx v_{sister} - v_{girl}$