Slot A Natural Language Processing (CS6370)

SLOT A **Name:** | **Roll No.:**

Natural Language Processing (CS 6370) 2020 Full Marks:60 Time: 3 hours

==========

1. Using an example of each, identify precisely the distinction between (a) denotational and distributional semantics (b) constructive and discriminative definitions of words. [2]
2. Give a simple example to explain the difference between discriminative and generative classifiers. The example should be based on how humans (and not machines) use both approaches to classify in everyday tasks. [2]
3. We identified the spellcheck problem as one of diagnosis, and used Machine Learning to solve it. Can you give a precise example of a diagnosis task which will NOT need Machine Learning? [1]
4. Can you give a Bayesian justification of why Google works better (in terms of retrieval effectiveness and not time efficiency) than earlier search engines (Altavista/Yahoo!)? [1]
5. What are the precise meanings of the terms "Mean" and "Average" in the context "Mean Average Precision"? [1]
6. In the context of evaluation of parsers, *Labelled Precision* is the number of correct nonterminal constituents found by the parser (summed over all the sentences in the test set) divided by the total number of nonterminal constituents the parser postulated, where non-terminals refer to all symbols (like Part of Speech tags and higher level constituents) other than words themselves (which are treated as terminals, and are leaves of the parse tree). Is this statement correct? If yes, do nothing. If no, justify and correct it. [1.5]
7. In the context of disambiguation using PCFGs, why do researchers prefer parent encoding over Prepositional Phrase mark-up (the latter is analogous to subcategorization)? [1.5]
8. Why is the second eigenvalue important in the context of the PageRank algorithm? [1]
9. Name a smoothing technique that exploits recursion in its definition (trigram model defined in terms of bigram model etc.), and explain briefly the central intuition. [2]
10. Can LSA be viewed as solving an optimization problem? If yes, how? If no, why? Can you derive a correspondence between the three matrices obtained after SVD, to three corresponding analogues in PLSA? [1+1]
11. Identify an advantage and a disadvantage of intrinsic measures over extrinsic ones in evaluation of language models. [2]
13. For an Information Theoretic measure of semantic relatedness to work, there is a special way to estimate frequencies of synsets. What is it, and why is it needed? [2]
14. In the context of Statistical Machine Translation from Hindi to English, we are interested in estimating P(h|e) where h refers to a Hindi sentence which is a candidate translation corresponding to a given English sentence e. Using the Bayesian rewrite, this amounts to estimating P(e|h) and P(h). Given a parallel corpus, what is the motivation for estimating P(e|h) instead of estimating P(h|e) directly? [2]
15. Can the size (NOT number) of documents in a collection have any bearing on the impact of Porter's stemming on IR performance? Justify your answer. [1]
16. The rank of a term document matrix corresponds to the number of underlying concepts. True or false? If true, justify. If false, correct the sentence and justify the corrected version. [1]
17. What is the Systems Reply (refutation) to Searle's Chinese Room Argument? [1]
18. Give two examples each of (a) Rhetorical Relations (b) Hearst patterns. [2]
19. Identify the specific limitation of using conjugate priors that MCMC approaches attempt to fix. [1]
20. How can bootstrapping be used in (a) Information Extraction (b) Word Sense Disambiguation? [1.5+1.5]
21. What is the implication of the Universal Approximation Theorem in the design of Multi Layer Feedforward Networks? Explain briefly the central intuition. [2]

22. A researcher claims that using Dynamic Programming (DP) has speeded up her data-to-text generation system that is designed to generate textual summaries of satellite time series data. Can you speculate the context in which she may have used DP? Can DP speed up Part of Speech tagging as well? If yes, how? [1+1]
23. (a) How would you characterize hardness of NLP problems? (b) Give an example of a hard NLP problem that we do not have currently have the tools to solve. (c) The hardest problems of NLP are at least as hard as the foundational unsolved problems in physical and life sciences. Argue convincingly for or against this statement (make clear any premises or assumptions that are pivotal to your argument). [1+0.5+3.5]
24. Consider the following grammar G.

    S → AB | BC
    A → BA | a
    B → CC | b
    C → AB | a

    (a) Is G in a form that CYK can deal with? If not, transform it. If yes, justify. (b) Consider the string baaba. Show clearly the table of sub-problems in CYK to decide if this string is accepted by G. [2+5]
25. Consider a Machine Translation parallel corpus having two sentence pairs. The first sentence pair is "Eat Something"/"Kuchh Khaao". The second sentence pair is "Eat"/"Khaao". (a) Show how the first few iterations of EM (M_0, E_1, M_1, E_2, M_2) are useful in learning word alignments from this corpus. Make clear any simplifying assumptions (with respect to IBM Model 3) that you use.[6+1]
26. A PCFG has the following rules:
    a. S → NP VP
    b. VP → V NP
    c. VP → V NP PP
    d. NP → NP PP
    e. NP → A dog
    f. NP → a cat
    g. NP → an optical mouse
    h. PP → in the backyard
    i. V → chased

    The corpus has the following two sentences, the first occurring 10 times and the second 20 times:

    A dog chased the cat in the backyard.

    The cat chased an optical mouse.

    (a) Which of these two sentences is/are ambiguous? Show all possible parse trees of these sentences.
    (b) Make an APPROPRIATE initial choice of the rule probabilities. Show the first few steps of the EM algorithm (M_0, E_1, M_1, E_2, M_2) for estimating the parameters of this PCFG.
    (c) Does the PCFG get better at disambiguation with each step? If so, show how.

    [2+4+1]

The End