Question 1

Given the stem "snow", give one example each of words that would result from the following kinds of morphological processes involving snow:

a) **Compounding:**

b) **Inflection:**

c) **Derivation:**

Question 2

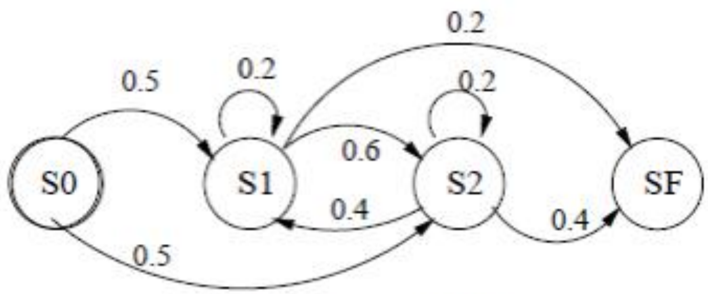
Question 1 (10 points) We define a PCFG where non-terminal symbols are $\{S, A, B\}$, the terminal symbols are $\{a, b\}$, and the start non-terminal (the non-terminal always at the root of the tree) is S. The PCFG has the following rules:

Rule	Probability
$S \rightarrow S\ S$	0.3
$S \rightarrow A\ S$	0.2
$S \rightarrow B\ B$	0.5
$A \rightarrow a$	0.2
$A \rightarrow b$	0.8
$B \rightarrow a$	0.4
$B \rightarrow b$	0.6

For the input string *abab*, show two possible parse trees under this PCFG, and show how to calculate their probability.

Question 3

2. (19 points) Consider the HMM below where the transition probabilities are shown in the graph and the observation probabilities (where V={A,B}) are in the tables below each state.



$b1(A) = 0.8$ $\quad$ $b2(A) = 0.2$

$b1(B) = 0.2$ $\quad$ $b2(B) = 0.8$

Use the Forward algorithm to compute the probability of generating the short output string: "A B".
Show the values computed for each of $\alpha_t(j)$ parameters as they are computed, showing your work.

Question 4

5. **(5 points) True/False:** Given a sentence S, the following equation adequately describes the computation performed by the probabilistic CKY algorithm given in the text.

$$P(S) = \sum_{T\, s.t.\, S = Yield(T)} P(T)$$

Question 5

6. **(10 points)** Describe the purpose of the computation occurring in the **If statement** in the following code for PCKY. That is, describe what its doing and why its being done.

```
function PROBABILISTIC-CKY(words,grammar) returns most probable parse
                                              and its probability
  for j ← from 1 to LENGTH(words) do
     for all { A | A → words[j] ∈ grammar }
        table[j−1, j, A] ← P(A → words[j])
     for i ← from j−2 downto 0 do
        for k ← i+1 to j−1 do
              for all { A | A → BC ∈ grammar,
                            and table[i,k,B] > 0 and table[k, j,C] > 0 }
                    if (table[i,j,A] < P(A → BC) × table[i,k,B] × table[k,j,C]) then
                       table[i,j,A] ← P(A → BC) × table[i,k,B] × table[k,j,C]
                       back[i,j,A] ← {k,B,C}
     return BUILD_TREE(back[1, LENGTH(words), S]), table[1, LENGTH(words), S]
```

Question 6

2. Assume in the context of HMMs, that *S* refers to a sequence of states, *O* refers to a sequence of observations, and that *M* is a particular HMM model. ***Name*** the algorithms that compute the following values:

a) **(5 points)** argmax P(S|O,M)

Question 7

4. Consider the following (inelegant) grammar rules from the Penn Treebank:

VP → VBD PP
VP → VBD PP PP
VP → VBD PP PP PP
VP → VBD PP PP PP PP
...

a) **(5 points)** Give two rules that can be used to replace these four (and all the ensuing ... rules).

Question 8

Question 2 (8 points) Write down a PCFG such that:

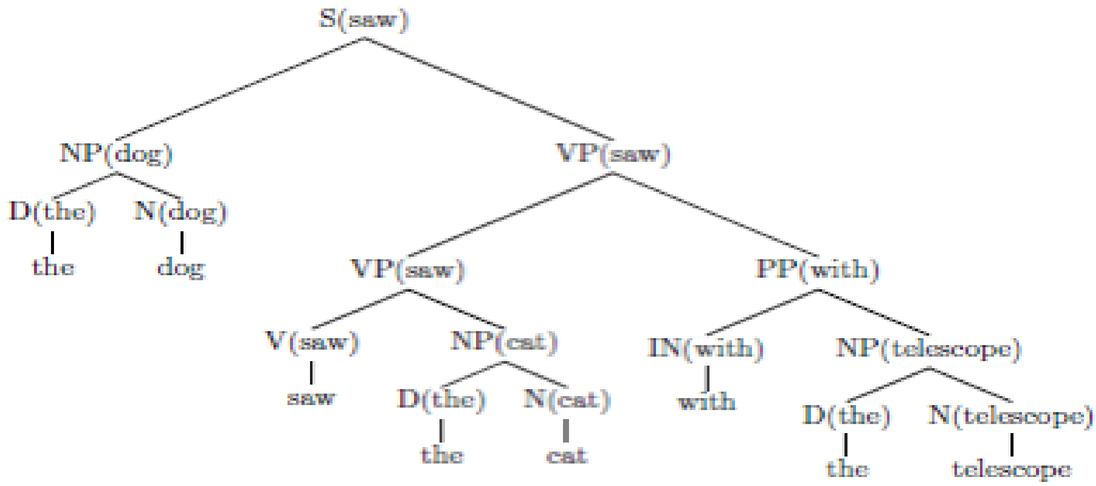
1. Any sentence consisting of the word *the* n times in a row, where $n \geq 1$, has probability
$$0.4 \times 0.6^{n-1}$$

2. Any other sentence has probability 0.

Question 9

Part #2 10 points

Consider the lexicalized tree below:



Question 3 (10 points) Complete the head-finding rules below that would give this lexicalized tree:

For S $\rightarrow$ NP VP choose VP as the head

For NP $\rightarrow$ D N choose as the head

For VP $\rightarrow$ VP PP choose as the head

For VP $\rightarrow$ V NP choose as the head

For PP $\rightarrow$ IN NP choose as the head

Question 10

5.2 Use the Penn Treebank tagset to tag each word in the following sentences from Damon Runyon's short stories. You may ignore punctuation. Some of these are quite difficult; do your best.

a. It is a nice night.
b. This crap game is over a garage in Fifty-second Street…
c. …Nobody ever takes the newspapers she sells …
d. He is a tall, skinny guy with a long, sad, mean-looking kisser, and a mournful voice.
e. …I am sitting in Mindy's restaurant putting on the gefillte fish, which is a dish I am very fond of, …
f. When a guy and a doll get to taking peeks back and forth at each other, why there you are indeed.

Question 11

For each of the following linguistic and computational phenomena give an example sentence and explain in a paragraph or two why it presents a challenge to natural language processing systems:

(a) (4 points) garden path sentence

(b) (4 points) inflectional morphology

(c) (4 points) coordination ambiguity

Question 12

Give an example of each of the following instances of morphological derivation: verb → noun; adjective → noun; adjective → adverb; adjective → verb.

Question 13

Write a simple context-free grammar that can produce the following sentences:

1. Joe ate dinner.

2. Did Maureen eat dinner?

3. When did Joe eat dinner?

4. Eat Dinner!

Question 14

Give an example sentence for each of the following verb subcategorization phrases.

1. NP

2. NP NP

3. Ø

4. P_{to}

5. S

Question 15

What is transformation-based part-of-speech tagging? What makes it different from the other major techniques in POS tagging?

Question 16

How does one deal with long-distance dependencies in statistical parsing? Give an example.

Question 17

Rewrite the grammar below in Chomsky Normal Form:

A → B C D

A → B h

A → a | F

B → b

C → c

D → d | e

F → f|g

Where A, B, C, D, F are non-terminals, A is the start symbol, and a, b, c, d, e, f, g, h are terminals.

Question 18

What is the worst-case time complexity of the CKY-CFG parsing algorithm used to find all parses for a sentence.

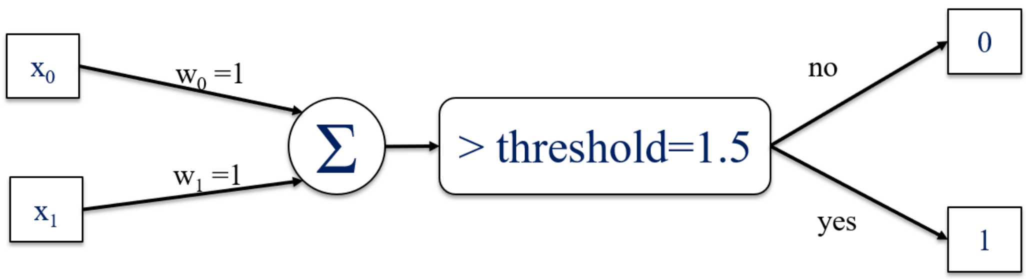
Question 19

Write a CCG lexical entry for a transitive verb such as "watch" in the following sentence: "Akira watched the movie".

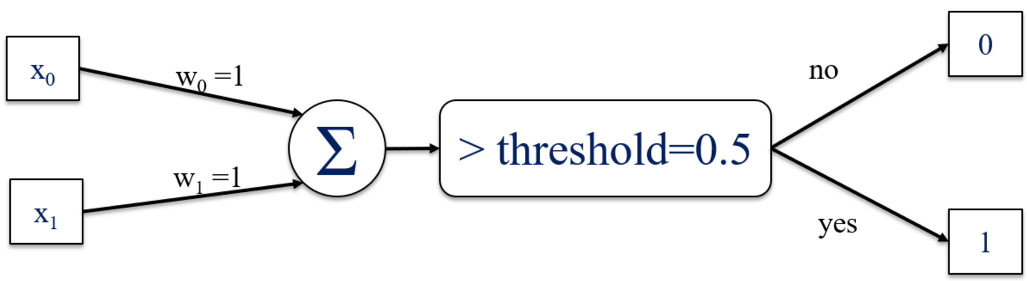
Question 20

(1) Consider the following perceptrons:

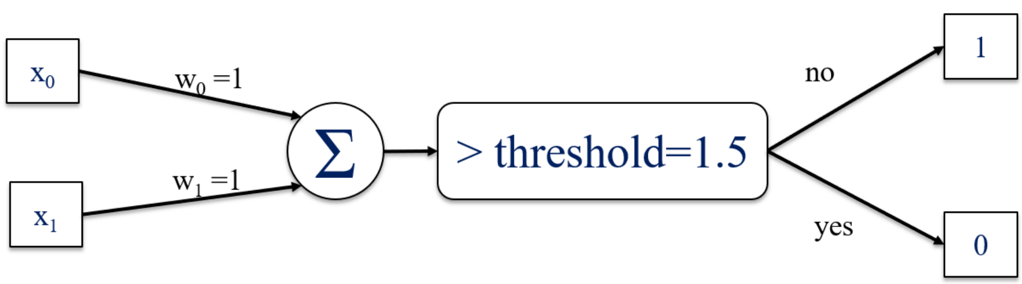
(a)



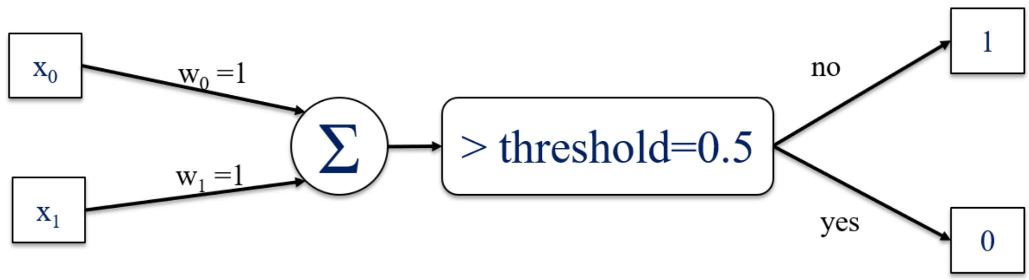
(b)



(c)



(d)



(1) 5 points. Please match the perceptrons above with the logical gates below. One or more of the gates may not match any of the perceptrons, and you can use (e) in this case.

_________ AND gate
_________ XOR gate
_________ OR gate
_________ NAND gate (not and)
_________ NOR gate (not or)

(2) 1 point. If we take the threshold node out, the perceptrons will be equivalent to which one of the following:

A. Linear regression
B. Logistic regression

(3) 1 point. Which of the following methods may help with preventing overfitting (There are may be more than one correct answer):

A. Use a larger dataset
B. Increase the number of hidden layers
C. Penalize the weights which is to use the regularized loss function
D. Normalize input data
E. Train the neural net with batches of examples instead of one at a time

(4) 1 point. Which of the following methods may help with training the neural net faster (There may be more than one correct answer):

A. Penalize the weights which is to use the regularized loss function
B. Normalize input data
C. Increase the learning rate as large as possible

Question 21

The sentence "My aunt collects old clocks and silverware" illustrates what type of ambiguity?

a. Prepositional phrase (PP) ambiguity
b. The sentence is not ambiguous.
c. Dialogue ambiguity
d. Coordinating conjunction (CC) ambiguity

Question 22

Assuming that exactly two constituents get combined at each iteration, a sequence of three nouns can be parenthesized in two different ways:

(a(bc))
((ab)c).

A sequence of four nouns can be parenthesized in five different ways:

((ab)c)d
(a(bc))d
(ab)(cd)
a((bc)d)
a(b(cd)).

In how many ways can a sequence of five nouns be parsed?

a. 28
b. 32
c. 14
d. 16
e. 10

Question 23

Consider an autoencoder with three layers: input layer (X neurons), hidden layer (Y neurons), and output layer (Z neurons).
What is the relationship between the number of neurons in each layer?

a. $X < Y < Z$
b. $X = Y = Z$
c. $X < Y, Y > Z$
d. $X = Z, Y < X$
e. $X > Y > Z$

Question 24

Describe how an LSTM works. Draw a diagram and explain all gates and all forward propagation formulas.