Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division
Second Semester 2020-2021
M.Tech (Data Science and Engineering)
Mid-Semester Test (EC-2 Regular)

Course No.	: DSECL ZG565	No. of Pages = 2
Course Title	: Natural Language Processing	No. of Questions = 6
Nature of Exam	: Open Book	
Weightage	: 30%	
Date of Exam	: 1st November, 2020	

Question 1. [3+2+3=8 Marks]

a) For each of the following sentences, please identify whether they are lexically, syntactically semantically and pragmatically correct

Solution:

1. Eats Ice-cream I in summer. - lexically correct
2. The fruits are flying in the blue sky. - lexically and syntactically correct
3. The baby is eating the chocolate wrapper. Lexically, syntactically and semantically correct

b) How many trigrams phrases can be generated from the following sentence, after replacing punctuations by a single space?

"Natural Language processing is very interesting, though not easy."

Solution: (Any one from 2 options correct)

Number of trigrams=8

<s> Natural Language, Natural Language processing, Language processing is, processing is very, is very interesting, very interesting though, interesting though not, though not easy

OR

Number of trigrams=9

<s> Natural Language, Natural Language processing, Language processing is, processing is very, is very interesting, very interesting though, interesting though not, though not easy, not easy </s>

c) Write the formulae to calculate the unigram, bigram and trigram probabilities of the below sentence

"Life should be great rather than long".

Solution:

Unigram

P (**"Life should be great rather than long"**)

=P(Life)P(should)P(be)P(great)P(rather)P(than)P(long)

Bigram

P (**"Life should be great rather than long"**)

=P(Life|<s>)P(should|Life))P(be|should)P(great|be)P(rather|great)P(than|rather)P(long|than)

Trigram

P (**"Life should be great rather than long"**)

=P(Life|<s>,<S>)P(should|Life,<s>))P(be|should,life)P(great|be, should)P(rather|great,be) P(than|rather,great)P(long|than,rather)

Question 2. [2+5+3 =10 Marks]

a) For an HMM MODEL with N hidden states, V observations, what are the dimensions of parameter matrices A, B, and π? A: Transition matrix, B: Emmision matrix, and π: Initial Probability matrix.

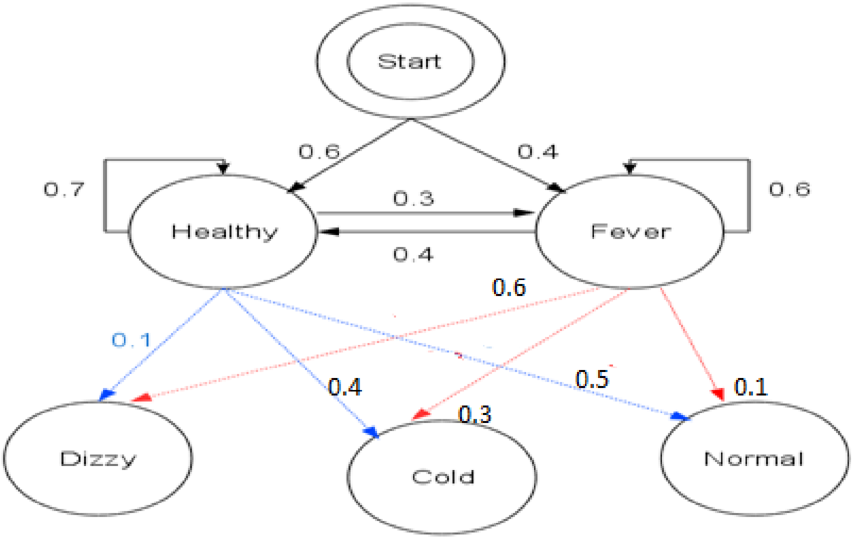
Soln 1 A: Transition matrix
dimension (A) = N × N
B: Emmision matrix
dimension of B = N × V
π: initial probability matrix
dimension of π = 1×N

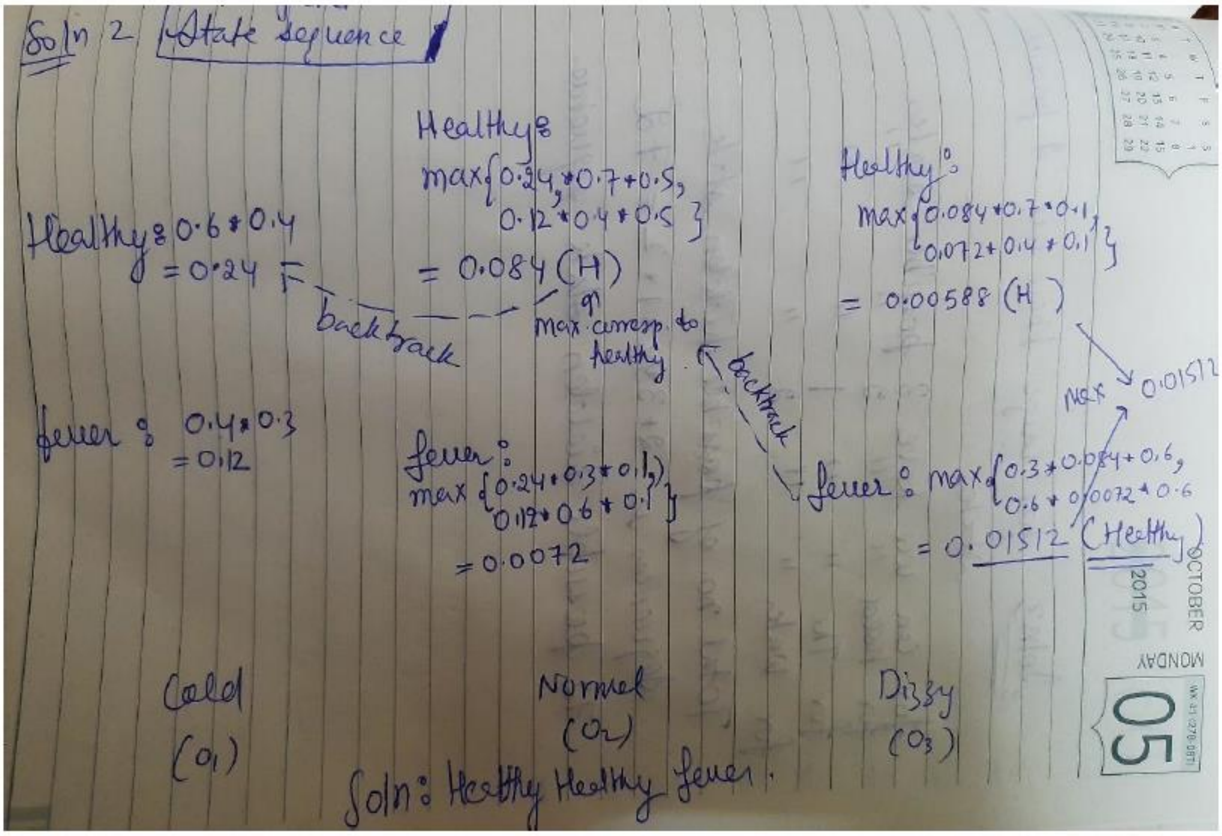
b) Consider an apartment where all residents are either healthy or have a fever and only the doctor can determine whether each has a fever. The doctor diagnoses fever by asking patients how they feel. The residents may only answer that they feel normal, dizzy, or cold.

The doctor believes that the health condition of his patients operates as a discrete Markov chain. There are two states, "Healthy" and "Fever", but the doctor cannot observe them directly; they are hidden from him. On each day, there is a certain chance that the patient will tell the doctor he is "normal", "cold", or "dizzy", depending on their health condition.

Set of Observations: {Dizzy, cold, Normal}, Set of states={Healthy, fever}

If the observation sequence is [cold normal dizzy]. Use Viterbi Algorithm to compute the corresponding state sequence.





Soln 2 State sequence

Healthy: $0.6 * 0.4 = 0.24$

fever: $0.4 * 0.3 = 0.12$

Healthy: $\max\{0.24 * 0.7 * 0.5,\ 0.12 * 0.4 * 0.5\} = 0.084$ (H)

backtrack — max corresp. to healthy

fever: $\max\{0.24 * 0.3 * 0.1,\ 0.12 * 0.6 * 0.1\} = 0.0072$

Healthy: $\max\{0.084 * 0.7 * 0.1,\ 0.0072 * 0.4 * 0.1\} = 0.00588$ (H)

backtrack

fever: $\max\{0.3 * 0.084 * 0.6,\ 0.6 * 0.0072 * 0.6\} = 0.01512$ (Healthy)

max 0.01512

Cold (O_1) Normal (O_2) Dizzy (O_3)

Soln: Healthy Healthy fever.

c) Suppose you have a sentence “Large can can hold the water”. And you know the possible tags for each word in the sentence.

Large: N, V

Can: V, Aux, N

Hold: N, V

The: article

Water: V, N

How many possible hidden state sequences are possible for the above sentence?

Soln3 for large there are 2 possible states
for can we have 3 possible states
for hold " " 2 " "
for The " " 1 " "
for water " " 2 " "
Total no. of possible hidden state sequences is 2*3*2*1*2 = 72
72 possible hidden state sequences

Question 3. [Marks 3+5+4=12 marks]

a) Given the grammar and lexicon below dervie the parse tree using top down parsing method for the sentence [3 marks]

S :The guy ate pizza

S->NP VP

VP->VNP

NP->Det N

N->pizza

N->guy ,Det ->the

V->ate

Solution:

1The 2 guy 3 ate 4 the 5pizza 6

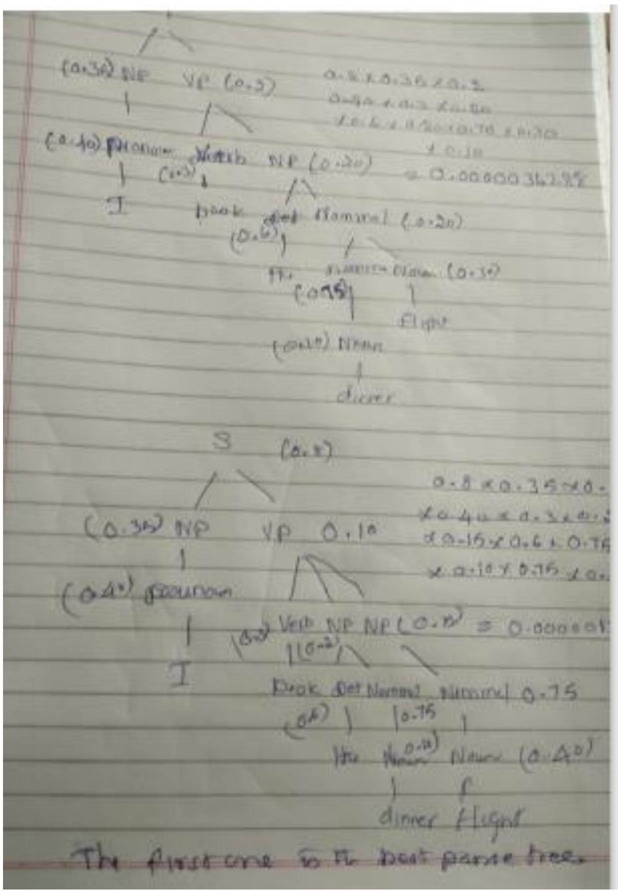
State	Backup State	Action
1.((S) 1)		
2.((NP VP) 1)		
3.(DT N VP) 1)		matches the
4.((N VP) 2)		matches guy
5.((VP)3)		
6.((V NP) 3)		matches ate
7.((Det N) 4)		matches the
8.((N))5		matches pizza

b) Given the grammar and lexicon below find the probabilty of the best parse tree using PCFG for the below sentence [5 marks]

S: I book the dinner flight

Grammar		Lexicon
$S \rightarrow NP\ VP$	[.80]	$Det \rightarrow that$ [.10] | a [.30] | the [.60]
$S \rightarrow Aux\ NP\ VP$	[.15]	$Noun \rightarrow book$ [.10] | $flight$ [.30]
$S \rightarrow VP$	[.05]	| $meal$ [.15] | $money$ [.05]
$NP \rightarrow Pronoun$	[.35]	| $flights$ [.40] | $dinner$ [.10]
$NP \rightarrow Proper\text{-}Noun$	[.30]	$Verb \rightarrow book$ [.30] | $include$ [.30]
$NP \rightarrow Det\ Nominal$	[.20]	| $prefer$; [.40]
$NP \rightarrow Nominal$	[.15]	$Pronoun \rightarrow I$ [.40] | she [.05]
$Nominal \rightarrow Noun$	[.75]	| me [.15] | you [.40]
$Nominal \rightarrow Nominal\ Noun$	[.20]	$Proper\text{-}Noun \rightarrow Houston$ [.60]
$Nominal \rightarrow Nominal\ PP$	[.05]	| NWA [.40]
$VP \rightarrow Verb$	[.35]	$Aux \rightarrow does$ [.60] | can [40]
$VP \rightarrow Verb\ NP$	[.20]	$Preposition \rightarrow from$ [.30] | to [.30]
$VP \rightarrow Verb\ NP\ PP$	[.10]	| on [.20] | $near$ [.15]
$VP \rightarrow Verb\ PP$	[.15]	| $through$ [.05]
$VP \rightarrow Verb\ NP\ NP$	[.05]	
$VP \rightarrow VP\ PP$	[.15]	
$PP \rightarrow Preposition\ NP$	[1.0]	

Solution:



c) Give the correct sequence of arc eager parsing operations for the given sentence [2marks]

a) Provide a modified transition sequence where the parser mistakenly predicts the arc cat → slept, but gets the other dependencies right. [2marks]

Solution:

c) **SH,SH,LA,LA,SH,LA,RA**

[ ]	[The lazy cat slept]	[]
[The]	[lazy cat slept]	[Shift]
[The ,lazy]	[cat slept]	[Shift]
[The ,lazy]	[cat slept]	[LA]
[The]	[cat slept]	[LA]
[cat]	[slept]	[SH]
[ ]	[slept]	[LA]
[slept]	[]	[RA]

OR

[ ]	[The lazy cat slept]	[]
[Root,The]	[lazy cat slept]	[Shift]
[Root ,the ,Lazy]	[cat slept]	[Shift]
[Root ,the ,Lazy]	[cat slept]	[LA]
[Root the]	[cat slept]	[LA]
[Root,Cat]	[slept]	[SH]
[ ]	[slept]	[LA]
[Root,Slept]	[]	[RA]
[Root]	[]	[RE]

d)

[ ]	[The lazy cat slept]	[]
[Root,The]	[lazy cat slept]	[Shift]
[Root ,the ,Lazy]	[cat slept]	[Shift]
[Root ,the ,Lazy]	[cat slept]	[LA]
[Root the]	[cat slept]	[LA]
[Root,Cat]	[slept]	[SH]
[Cat]	[]	[RA]
[Root,Cat ,Slept]	[]	[RE]
[Root,cat]	[]	[RE]
[Root]	[]	[RE]