PES University, Bengaluru
(Established under Karnataka Act No. 16 of 2013)

UE18CS334

MARCH 2021: IN SEMESTER ASSESSMENT B.Tech VI SEMESTER
TEST – 1

UE18CS334 – NATURAL LANGUAGE PROCESSING

Time: 2 Hrs	Answer All Questions	Max Marks: 60

1. a) A Naïve Bayes Classifier has to decide whether the document 'London Paris' is news about the United Kingdom (Class U) or news about Spain (Class S). **[4]**

S.no	Document	Class
1.	London Paris	U
2.	Madrid London	S
3.	London Madrid	U
4.	Madrid Paris	S

a. Estimate the probabilities that are relevant for this decision from the following document collection using Maximum Likelihood estimation (without smoothing). Answer with fractions.

b. Based on the estimated probabilities, which class does the classifier predict and why?

b) Convert the following words in $[C](VC)^m[V]$ form and identify the measure of the word: **[4]**

1. ANALYZES
2. BAYWOODS
3. CALYCEAL
4. Hmmmm

c) Identify the type of ambiguities in the following sentences with reason: **[2]**

a. I shot a chimpanzee wearing my glasses.

b. London had snow yesterday. It fell to a depth of a meter.

2. a) Calculate the minimum edit distance between the two words "Kitten" and "Sitting". Make the complete table showing all the entries and also identify the operations that are required to convert Kitten to Sitting. Cost of every operation is 1. **[6]**

b) *Professor Jane, a noted Mathematician, noticed that when his wife wanted to boil water for their tea, she took their kettle from a kitchen cabinet, filled it with water, and put it on the burner.* **[4]**

Identify the tokens and types in the above sentence. Also mention 3 content and 3 function words.

3. a) Consider the following sentences in a training corpus: **[5]**

Training corpus:
<s> I am from Vellore </s>
<s> I am a teacher </s>
<s> students are good and are from various cities</s>
<s> students from Vellore do engineering</s>

Test data: *<s> students are from Vellore </s>*

Compute the Bigram probability and perplexity of the given test sentence. What does low value of perplexity show?

b) The Transition and Emission probabilities corresponding to an HMM are given below: (AUX is Auxillary Verb, VB is Verb, N is Noun) — 5

Transition Probabilities	Emission Probabilities
P(N||Pronoun)=0.001, P(Pronoun|<s>)=0.5, P(N|<s>)=0.5, P(VB|Aux)=0.5, P(Aux|Pronoun)=0.2, P(Aux|Noun)=0.1, P(N|VB)=0.2, P(N|NP)=0.0001,	P(Shuarya|N)=0.1, P(she|Pronoun)=0.1, P(can|AUX)=0.2, P(can|Noun)=0.01, P(cook|VB)=0.3, P(food|Noun)=0.2.

Extract the graph for HMM from above probabilities. Compute the likelihood that the sequence "N AUX VB N" would generate the output "Shuarya can cook food".

Also, State the two assumptions considered by an HMM.

4 (a) Why a MEMM, another probabilistic model is required when HMM is there for POS tagging? What makes MEMM special over an HMM? Illustrate with an example how features are considered in an MEMM. — 6

(b) Suppose in Q3(a), the test sentence is — 4

<s> Students are running</s>.

Now compute the bigram probability of the sentence. And make sure the final answer is not zero. If it is zero then you need to rework for a non-zero probability.

5 (a) Consider the following CFG: — 5

S→NP VP	**V→like**
VP→V NP	**NP→kids**
VP→ VP PP	**NP→picnic**
NP→NP PP	**NP→ friends**
NP→Pronoun	**P→with**
PP→P NP	

Use CKY parsing to fill CKY table corresponding to " Kids like picnic with friends".

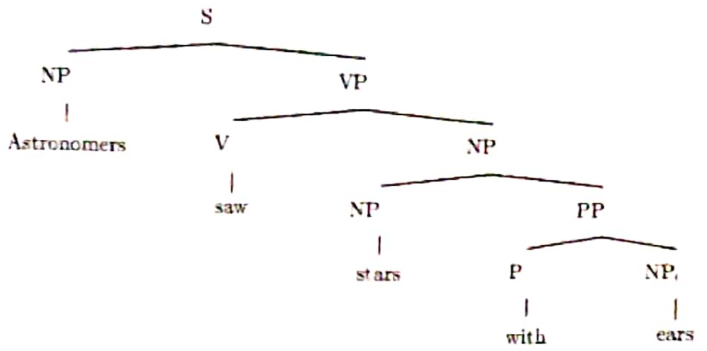
(b) Draw the parse trees for Q5(a). Interpret the two parse trees. Check if you are getting multiple parse trees with CKY parsing then suggest a way to select the right parse tree.. — 5

6 (a) Consider the following Probabilistic PCFG — 2

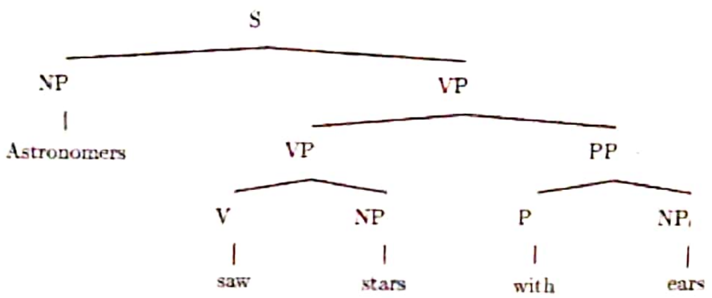
S→NP VP (1.0)	P→ with (1.0)
VP→V NP (0.7)	V→ saw (1.0)
VP→ VP PP (0.3)	NP→ astronomers (0.1)
NP→NP PP (0.4)	NP→ saw (0.04)
PP→P NP (1.0)	NP→ stars (0.18)
	NP→ telescope (0.1)
	NP→ ears (0.18)

And given two parse trees:

Parse Tree1:



Parse Tree 2:



Which amongst Parse Tree 1 and 2 will be considered a better parse tree?

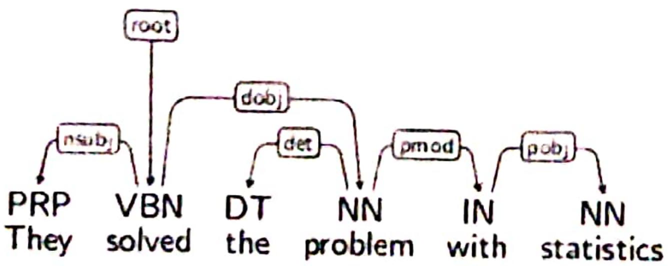
(b) The following is the final chart using Earley Parser for " Mary runs" with three state lists: **3**

```
S0:   [($ → • S),
       (S → • Noun Verb)]
S1:   [(Noun → mary •),
       (S → Noun • Verb)]
S2:   [(Verb → runs •),
       (S → Noun Verb •),
       ($ → S •)]
```

Where \$ is the dummy state. Identify the operators(Predictor, Scanner or Completer) that will called at different states.

For S_0 : $\$ \rightarrow \bullet\ S$. is the Start state. Identify remaining states in S_0, S_1 and S_2

(c) Identify if it is a dependency tree? If yes, why? If no, why? **5**



Is arc between (solved, problem) projective? Justify your answer.

Identify Projective and non-projective arcs in the above structure.