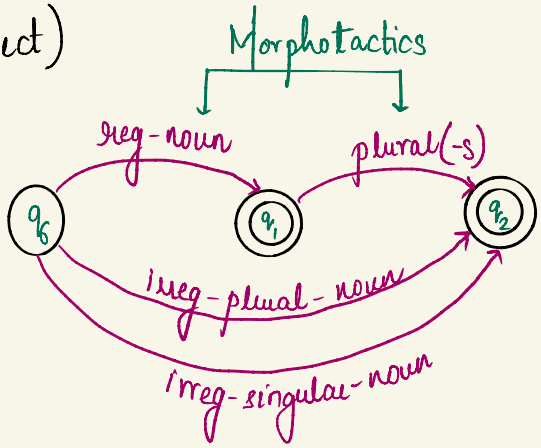
Sample/practice exam 2019, answers

Natural language processing (Vellore Institute of Technology)

Ans-1

(a) (Both are Correct)



(3 marks)

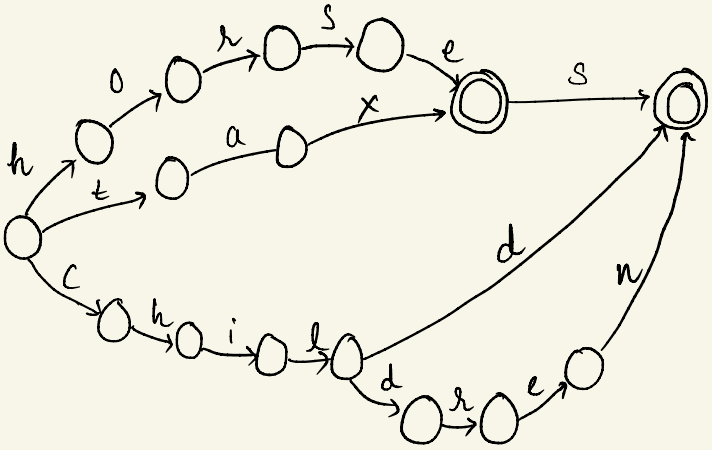
reg-noun = { horse, tax }

irreg-plural-noun = { taxes, children }

irreg-singular-noun = { tax, child }

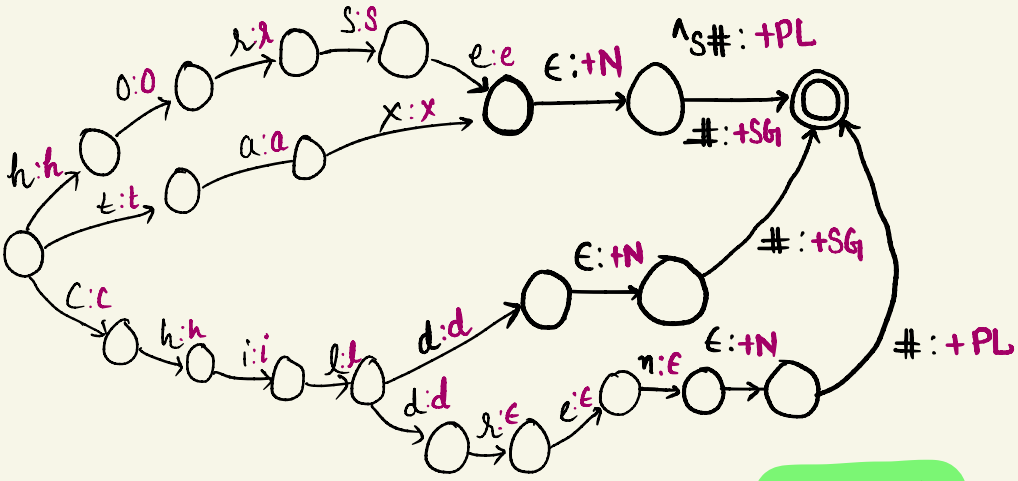
(1 mark)

"OR"



(4 Mark)

(b)



(6 Marks)

(input):(output)

Ans 2) Given Sentence :-

a busy market

Sentence with potential tags :- a/DET busy/NN market/NN
busy/ADJ market/VB
busy/VB

1 mark

Observations : $w_1 = a$, $w_2 = busy$, $w_3 = market$

$t_1 = DET$, $t_2 = NN$, $t_3 = ADJ$, $t_4 = VB$

Viterbi Algorithm :-

Step 1: Create path probability matrix V of dimensions N x T. Here, N = 4 and T = 3.

(2 marks)

Step 2: for each state s = 1 to N (viz. 4) do

$$V[s,1] \leftarrow \pi_s * P(w_1 | t_s)$$

Given :- All tags have equal probabilities to appear in the beginning of the sentence.

We have,

$$\pi_1 = \pi_2 = \pi_3 = \pi_4 = 1/4$$

So,

$$V[1,1] = \pi_1 * P(w_1|t_1) = 1/4 * 0.4 = 0.1$$
$$V[2,1] = \pi_2 * P(w_1|t_2) = 1/4 * 0 = 0$$
$$V[3,1] = \pi_3 * P(w_1|t_3) = 1/4 * 0 = 0$$
$$V[4,1] = \pi_4 * P(w_1|t_4) = 1/4 * 0 = 0$$

$BP[1,1] = 0$
$BP[2,1] = 0$
$BP[3,1] = 0$
$BP[4,1] = 0$

(BP : Back Pointer)

Step 3:- for each step $t = 2$ to T (viz. 3) do

for each state $s = 1$ to 4 do

$$V[s,t] \leftarrow \max_{s'=1\ldots N} \left(V[s',t-1] * a_{s's} * b_s(w_t) \right)$$

$$BP[s,t] \leftarrow \arg\max_{s'=1\ldots N} \left(V[s',t-1] * a_{s's} * b_s(w_t) \right)$$

$t = 2$ (w_2 = busy) (2 marks)

$s = 1$

$V[1,2] = 0$, $BP[1,2] = 1$

$s = 2$

$V[2,2] = 4 \times 10^{-5}$; $BP[2,2] = 1$

$s = 3$

$V[3,2] = 0.016$; $BP[3,2] = 1$

$s = 4$

$V[4,2] = 3 \times 10^{-9}$; $BP[4,2] = 1$

$t = 3$, w_3 = market (2 marks)

$s = 1$

$V[1,3] = 0$; $BP[1,3] = 1$

$s = 2$

$V[2,3] = 0.00144$ $BP[2,3] = 3$

$s = 3$

$V[3,3] = 0$; $BP[3,3] = 3$

$s = 4$

$V[4,3] = 0.00048$ $BP[4,3] = 3$

Step 4:- $\text{bestpathprob} \leftarrow \max_{s=1..4} V[s,T] = \max_{s=1..4} V[s,3] = 0.00144$ (2 marks)

$\text{bestBP} \leftarrow \operatorname{argmax}_{s=1:4} V[s,3] = 2$

Step 5:- Backtrack from bestBP

$BP[BP[\text{bestpath},3],2] \leftarrow BP[\text{bestpath},3] \leftarrow \text{bestpath}=2$

$(=1)$ $(=3)$

So, the best tag sequence is $\langle t_1, t_3, t_2 \rangle$

ie., $\langle$ DET, ADJ, NN $\rangle$ (1 mark)

Qns 3) Given the sentence,

a fair deal

a/DET fair/ADJ /ADV /NN /VB deal/VB /NN

(a) a/DET (0.5 mark)

fair/ADJ, fair/ADV, fair/NN, fair/VB (1 mark)

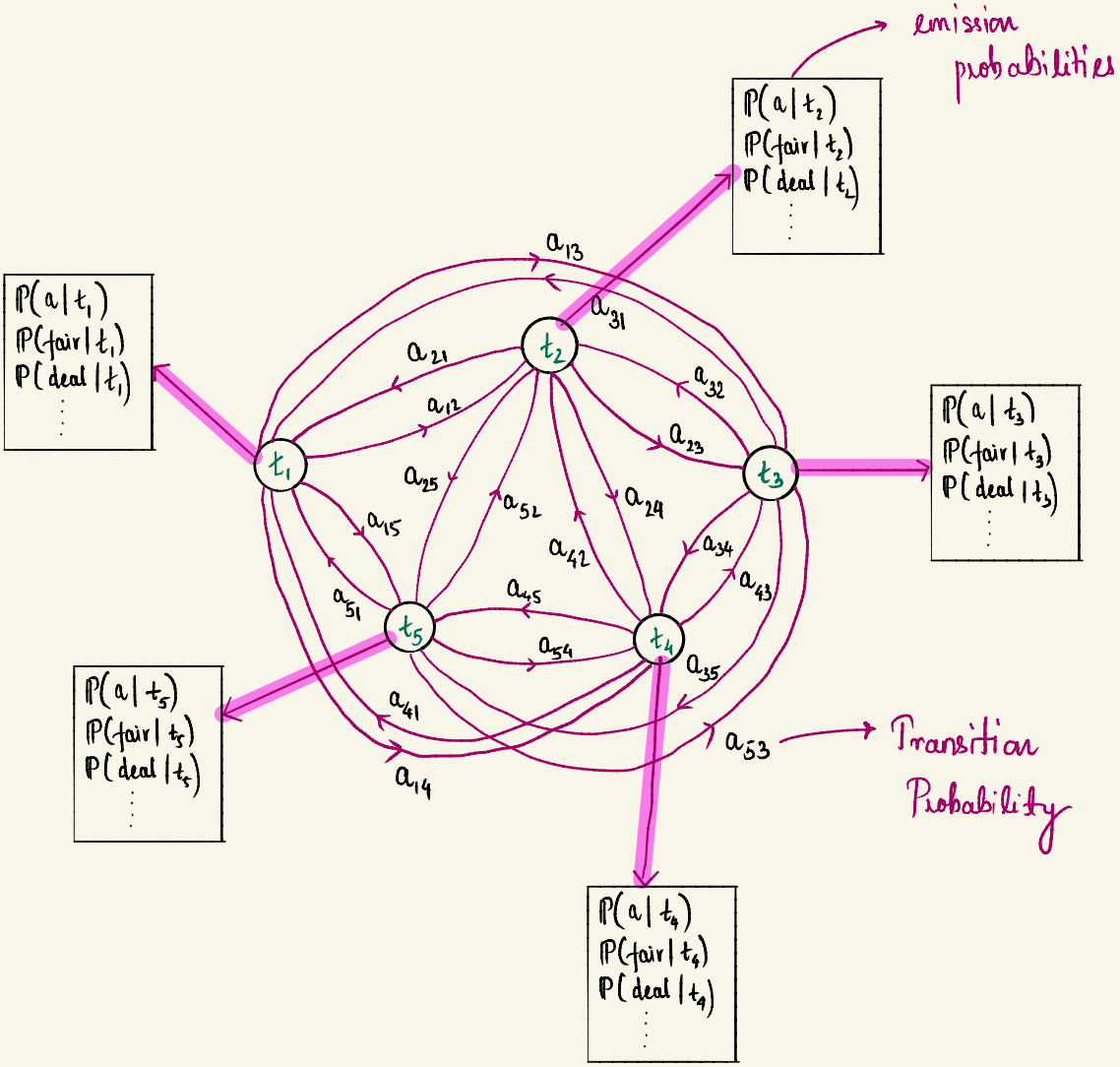
deal/VB, deal/NN (0.5 mark)

(b) tag sequences.

1) ⟨DET, ADJ, VB⟩
2) ⟨DET, ADJ, NN⟩
3) ⟨DET, ADV, VB⟩
4) ⟨DET, ADV, NN⟩
5) ⟨DET, NN, VB⟩
6) ⟨DET, NN, NN⟩
7) ⟨DET, VB, VB⟩
8) ⟨DET, VB, NN⟩

(3 marks)

(c) (5 marks)



$t_1 = \text{DET}$, $t_2 = \text{NN}$, $t_3 = \text{ADJ}$, $t_4 = \text{ADV}$, $t_5 = \text{VB}$

Ans 4)

<s> I am Sam </s>

<s> Sam I am </s>

<s> I am Sam </s>

<s> I do not like green eggs and ham </s>

(a)

$P(I \mid <s>) = 3/4$

$P(am \mid I) = 3/4$

$P(Sam \mid am) = 2/3$

$P(</s> \mid Sam) = 2/3$

$P(Sam \mid <s>) = 1/4$

$P(I \mid Sam) = 1/3$

$P(am \mid I) = 3/4$

$P(</s> \mid am) = 1/3$

$P(I \mid <s>) = 3/4$

$P(do \mid I) = 1/4$

$$P(\text{not} \mid \text{do}) = 1$$

$$P(\text{like} \mid \text{not}) = P(\text{green} \mid \text{like}) = P(\text{eggs} \mid \text{green}) = 1$$

$$P(\text{and} \mid \text{eggs}) = P(\text{ham} \mid \text{and}) = P(\langle /s \rangle \mid \text{ham}) = 1$$

(4 marks)

(b) $W = \langle s \rangle$ I am Sam $\langle /s \rangle$

$$P(W) = \prod_{i=1}^{5} P(w_i \mid w_{i-1})$$

(1 marks)

$$PP(W) = \left(P(W)\right)^{-1/5}$$

(2 marks)

$$= \left(\frac{3}{9} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{2}{3}\right)^{-1/5}$$

$$= \left(\frac{1}{4}\right)^{-1/5}$$

(2 marks)

Ans 5) (a)

	i	want	to	eat	chinese	food	lunch	spend
i	0.002	0.33	0	0.0036	0	0	0	0.00079
want	0.0022	0	0.66	0.0011	0.0065	0.0065	0.0054	0.0011
to	0.00083	0	0.0017	0.28	0.00083	0	0.0025	0.087
eat	0	0	0.0027	0	0.021	0.0027	0.056	0
chinese	0.0063	0	0	0	0	0.52	0.0063	0
food	0.014	0	0.014	0	0.00092	0.0037	0	0
lunch	0.0059	0	0	0	0	0.0029	0	0
spend	0.0036	0	0.0036	0	0	0	0	0

→ (2 marks)

$P(\langle s\rangle\ I\ want\ english\ food\ \langle /s\rangle) = 3.1 \times 10^{-5}$

$P(\langle s\rangle\ I\ want\ chinese\ food\ \langle /s\rangle) = 0$

→ (2 marks)

(b)

	i	want	to	eat	chinese	food	lunch	spend
i	6	828	1	10	1	1	1	3
want	3	1	609	2	7	7	6	2
to	3	1	5	687	3	1	7	212
eat	1	1	3	1	17	3	43	1
chinese	2	1	1	1	1	83	2	1
food	16	1	16	1	2	5	1	1
lunch	3	1	1	1	1	2	1	1
spend	2	1	2	1	1	1	1	1

→ (1 mark)

	i	want	to	eat	chinese	food	lunch	spend
i	0.0015	0.21	0.00025	0.0025	0.00025	0.00025	0.00025	0.00075
want	0.0013	0.00042	0.26	0.00084	0.0029	0.0029	0.0025	0.00084
to	0.00078	0.00026	0.0013	0.18	0.00078	0.00026	0.0018	0.055
eat	0.00046	0.00046	0.0014	0.00046	0.0078	0.0014	0.02	0.00046
chinese	0.0012	0.00062	0.00062	0.00062	0.00062	0.052	0.0012	0.00062
food	0.0063	0.00039	0.0063	0.00039	0.00079	0.002	0.00039	0.00039
lunch	0.0017	0.00056	0.00056	0.00056	0.00056	0.0011	0.00056	0.00056
spend	0.0012	0.00058	0.0012	0.00058	0.00058	0.00058	0.00058	0.00058

→ (1 mark)

$P(\langle s\rangle\ I\ want\ english\ food\ \langle /s\rangle) = 1.9635 \times 10^{-5}$

$P(\langle s\rangle\ I\ want\ chinese\ food\ \langle /s\rangle) = 7.92 \times 10^{-7}$

→ (1 mark)

(c) Change of probability can be very abrupt. → (2 Mark)

One can use add-k smoothing. → (1 Mark)