Querstion1)

Matching trigram =2 { the cat was, under the bed}
Total trigram in the actual sentence: 4
Total trigram in predicted answer=5

precision=> 2 / 5 Recall=> 2/4

Both Correct: 4 marks, single correct 2 marks

Question 2)

1/50 * sum(1/1 + ½ + ……..+ 1 /40)

ans=>> .085

$$\mathrm{MRR} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\mathrm{rank}_i}.$$

4 marks for correct expression, else 0

Question 3)

Ans:

$$\mathrm{PMI}(X,Y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

---- 1mark

$$p_{ij} = \frac{f_{ij}}{\sum_{i=1}^{W}\sum_{j=1}^{C} f_{ij}}$$

-------1 mark

P(data)=0.37
P(information)=0.58
P(information, data)=0.32----1 mark

$$pmi(information,data) = \log_2 (.32 / (.37*.58)) = .58$$

------ 1 mark

Question4)

Column dimension has to be **NLP, IR, Math, Interesting, else -1**

Question5)

★ **SOLUTION:** The Viterbi algorithm predicts that the most likely sequence of states is A, A, A. The relevant computations are:

$V_1^A = 0.99 \times 0.8 = 0.792$
$V_1^B = 0.01 \times 0.1 = 0.001$
$V_2^A = 0.2(0.792)(0.99) = 0.156816$
$V_2^B = 0.9(0.792)(0.01) = 0.007128$
$V_3^A = 0.8(0.156816)(0.99) = 0.1241983$
$V_3^B = 0.1(0.007128)(0.99) = 0.000705672$

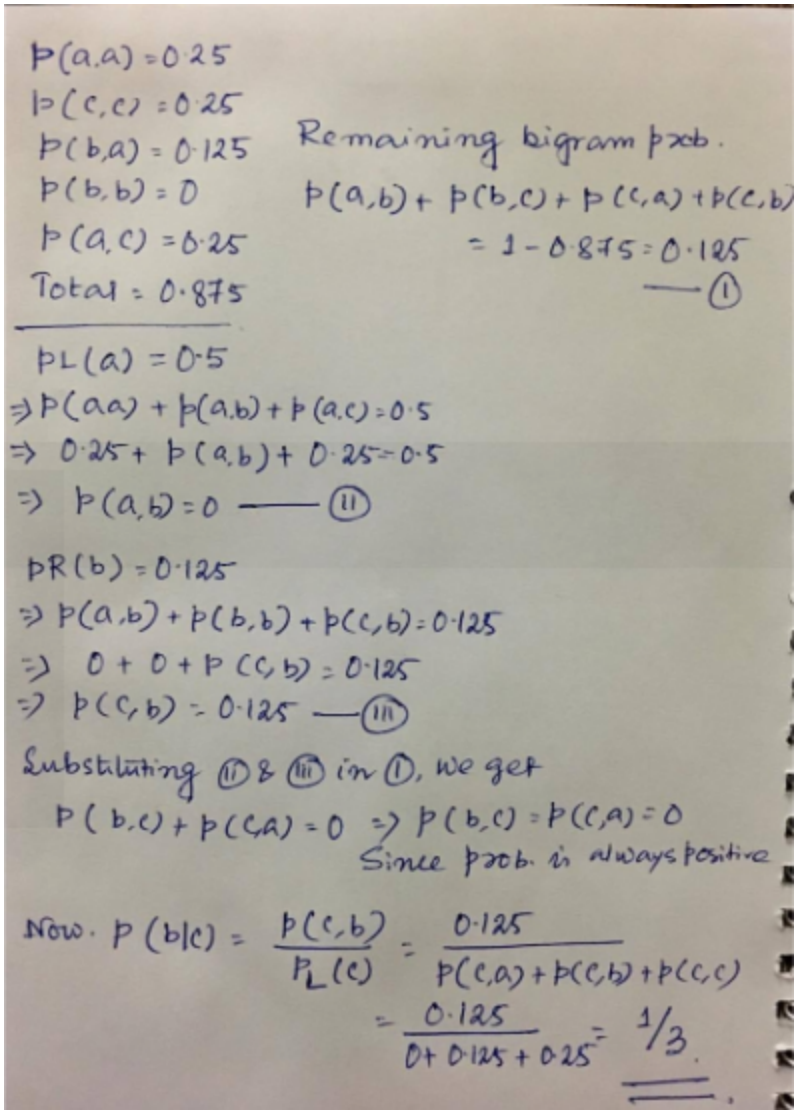
Correct state sequence and relevant calculations (6 marks)

In case of wrong state sequence:

- (3 marks) State sequence is partially correct and relevant calculation
- (2 marks or 1 marks) Some calculations are correct.
- (0 marks) in case the sequence is completely wrong or question not attempted.

Question 6.

YES---------- 1 mark



1 mark for eq I and II
1 mark for p(b,c)=p(c,a)=0
1 mark for final ans = 1/3

Question7)

Choice of Model -> Choosing Model B over Model A (+2 marks) {calculations were unnecessary as was indicated in the question paper}

Reasons why - (These were the points that we were looking for. Explanations are awarded marks in accordance with these ideas.)

- Occam's Razor, choosing the simplest model available.

- An unbiased model is a mode with higher entropy (**randomness**). We want to reduce the bias in model.

Students who have mentioned these key points in their arguments are awarded **2 marks**.

Some common errors:
- Entropy for MEMM has little relation with the number of features (0.5 marks awarded)
- Entropy of MEMM has effect on the predictions of regressor for sure, but this is not the reason for maximum entropy in a model.

Question 12)
1)
S →

X → aY | bY | ε
Y → X | c
The variable X is nullable; and so therefore is Y .
After elimination of ε, we obtain:
S → aXbX | abX | aXb | ab
X → aY | bY | a | b
Y → X | c

After elimination of the unit production Y → X,
we obtain:
S → aXbX | abX | aXb | ab
X → aY | bY | a | b
Y → aY | bY | a | b | c

Now, break up the RHSs of S; and replace a by A,
b by B and c by C wherever not units:
S → EF | AF | EB | ABP

A'x.ax.x
/home/subho/Downloads/new doc 2018-12-01 19.01.24_20181201190131.pdf
X → AY | BY | a | b
Y → AY | BY | a | b | c
E → AX
F → BX

A → a
B → b
C → c

ii)

S → ASA | aB
A → B | S
B → b | e (epsilon)
After adding the new start symbol S0, we get the following grammar.
(G1)
⇒

S
S → ASA | aB
A → B | S
B → b |
Removing nullable variables In the above grammar, both A and B are the nullable variables. We have the rule S → ASA. Since A is nullable, we need to add S → SA and S → AS and S → S (which is of course a silly rule, so we will not waste our time putting it in). We also have S → aB. Since B is nullable, we need to add S → a. The resulting grammar is the following.
(G2)
⇒ S0 → S
S → ASA | aB | a | SA | AS
A → B | S
B → b
Removing unit rules. The unit pairs for this grammar are {A → B, A → S, S0 → S}. We need to copy the productions for S up to S0, copying the productions for S down to A, and copying the production B → b to A → b.
(G3)
⇒ S0 → ASA | aB | a | SA | AS
S → ASA | aB | a | SA | AS
A → b | ASA | aB | a | SA | AS
B → b
Final restructuring. Now, we can directly patch any places where our grammar rules have the wrong form for CNF. First, if the rule has at least two symbols on its righthand side but some of them are terminals, we introduce new variables which expand into these terminals. For our example, the o ending rules are S0 → aB, S → aB, and A → aB. We can rid these by replacing the a's with a new variable U, and adding a rule U → a.
(G4)

⇒ S0 → ASA | UB | a | SA | AS
S → ASA | UB | a | SA | AS
A → b | ASA | UB | a | SA | AS
B → b
U → a