Natural Language Processing

Assignment 4

Type of Question: MCQ

Number of Questions: 8   Total Marks: $(6\times1)+(2\times2)=10$

---

1. Baum-Welch algorithm is an example of - **[Marks 1]**
    A) Forward-backward algorithm
    B) Special case of the Expectation-maximization algorithm
    C) Both A and B
    D) None

Answer: A

Solution: Theory.

---

2. Once a day (e.g. at noon), the weather is observed as one of
state 1 : rainy   state 2: cloudy   state 3: sunny
The state transition probabilities are :

0.4	0.3	0.3
0.2	0.6	0.2
0.1	0.1	0.8

Given that the weather on day 1 (t = 1) is sunny (state 3), what is the probability that the weather for the next 7 days will be "sun-sun-rain-rain-sun-cloudy-sun"?
[Marks 2]

A) $1.54 * 10^{-4}$
B) $8.9 * 10^{-2}$
C) $7.1 * 10^{-7}$
D) $2.5 * 10^{-10}$

Answer: A

Solution:

O = {S3, S3, S3, S1, S1, S3, S2, S3}

P(O | Model)

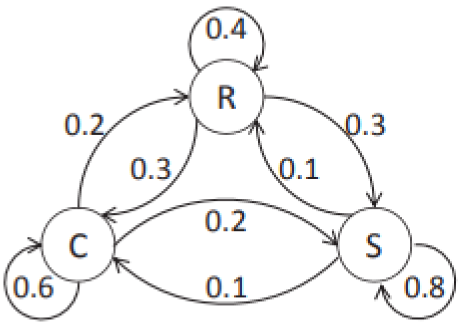
= P(S3, S3, S3, S1, S1, S3, S2, S3 | Model)

= P(S3) P(S3|S3) P(S3|S3) P(S1|S3) P(S1|S1) P(S3|S1) P(S2| S3) P(S3|S2)

= Q3 · a33 · a33 · a31 · a11 · a13 · a32 · a23

= (1)(0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2)

= 1.536 × 10-4



---

3. In the above question, the expected number of consecutive days of sunny weather is:

A) 2
B) 3
C) 4
D) 5 **[Marks 1]**

Answer: D

Solution:

Exp(i) = 1/(1-p_{ii}) So for sunny the exp = 1/(1-0.8) = 5

---

4. Let us define an HMM Model with K classes for hidden states and T data points as observations. The dataset is defined as X = {x1, x2, . . . , xT } and the corresponding hidden states are Z = {z1, z2, . . . , zT }. Please note that each xi is an observed variable and each zi can belong to one of classes for hidden state. What will be the size of the state transition matrix, and the emission matrix, respectively for this example. **[Marks 1]**

A) $K \times K$, $K \times T$
B) $K \times T$, $K \times T$
C) $K \times K$, $K \times K$
D) $K \times T$, $K \times K$

Answer: A

Solution: Since there are K hidden states, the state transition matrix will be of size $K \times K$. The emission matrix will be of size $K \times T$, as it defines the probability of emitting an observed state from a hidden state.

---

5. You are building a model distribution for an infinite stream of word tokens. You know that the source of this stream has a vocabulary of size 1000. Out of these 1000 words you know of 100 words to be stop words each of which has a probability of 0.0019. With only this knowledge what is the maximum possible entropy of the modelled distribution. (Use log base 10 for entropy calculation) **[Marks 2]**

A) 5.079
B) 0
C) 2.984
D) 12.871

Answer: C

Solution: There are 100 stopwords with each having an occurrence probability of 0.0019. Hence,
$P(\text{Stopwords}) = 100 * 0.0019 = 0.19$
$P(\text{non} - \text{stopwords}) = 1 - 0.19 = 0.81$

For maximum entropy, the remaining probability should be uniformly distributed.
For every non-stopword w, $P(w) = 0.81/(1000 - 100) = 0.81/900 = 0.0009$. Finally, the value of the entropy would be,

$$
\begin{aligned}
H &= E(\log(1/p)) \\
&= -100(0.0019 * \log(0.0019)) - 900(0.0009 \log(0.0009)) \\
&= -2.9841
\end{aligned}
$$

---

6. For an HMM model with N hidden states, V observable states, what are the dimensions of parameter matrices A,B and π? A: Transition matrix, B: Emission matrix, π: Initial Probability matrix. **[Marks 1]**

A) $N \times V$, $N \times V$, $N \times N$
B) $N \times N$, $N \times V$, $N \times 1$
C) $N \times N$, $V \times V$, $N \times 1$
D) $N \times V$, $V \times V$, $V \times 1$

Answer: B

Solution: Matrix A contains all the transition probabilities and have dimension $N \times N$. Similarly, matrix B contains all the emission probabilities and dimension $N \times V$. Similarly, π contains initial probability for all hidden states and have dimension $N \times 1$.

---

7. Suppose you have the input sentence “Death Note is a great anime”.
And you know the possible tags each of the words in the sentence can take.
• Death: NN, NNS, NNP, NNPS
• Note: VB, VBD, VBZ
• is: VB
• a: DT
• great: ADJ
• anime: NN, NNS, NNP
How many possible hidden state sequences are possible for the above sentence and States? **[Marks 1]**

A) $4 \times 3 \times 3$
B) $4^{3\wedge 3}$
C) $2^4 \times 2^3 \times 2^3$
D) $2^{4\times 3\times 3}$

Answer: A

Solution: Each possible hidden sequence can take only one POS tag for each of the words. Hence the total possibility will be a product of the number of candidates for each word.

---

8. In Hidden Markov Models or HMMs, the joint likelihood of an observed sequence O with a hidden state sequence Q, is written as P(O, Q; θ). In many applications, like POS tagging, one is interested in finding the hidden state sequence Q, for a given observation sequence, that maximizes P(O, Q; θ). What is the time required to compute the most likely Q using an exhaustive search? The required notations are, N: possible number of hidden states, T: length of the observed sequence. **[Marks 1]**

A) Of the order of TN^T
B) Of the order of N^2T
C) Of the order of T^N
D) Of the order of N^2

Answer: A

Solution: We will need to compute P(O, Q|θ) for all possible Q. There are a total of N^T possible hidden sequences Q for a sequence of length T. Each individual probability calculation also requires T multiplications.