50.007 Machine Learning

Simriti Bundhoo 1006281

Homework 4

1

The parameters associated with the HMM are:

1. **Transition Probability** ($a_{(u,v)}$) which is the chances of moving from one state to another. It is calculated by:

$$a_{(u,v)} = \frac{count(u,v)}{count(u)}$$

2. **Emission Probability** ($b_u(o)$) which is the likelihood of observing a particular piece of data (like a word or symbol) given that the model is in a certain state. It is given by:

$$b_u(o) = \frac{count(u \to o)}{count(u)}$$

$count(X) = 6$ $count(Y) = 6$ $count(Z) = 6$

Transition Probability

$u \backslash v$	X	Y	Z	STOP
START	0.4	0	0.6	0
X	0	0.5	0.33	0.17
Y	0.17	0	0.17	0.66
Z	0.5	0.5	0	0

$a_{(START,X)} = \frac{2}{5} = 0.4$ $a_{(X,X)} = \frac{0}{6} = 0$ $a_{(Y,X)} = \frac{1}{6} = 0.17$ $a_{(Z,X)} = \frac{3}{6} = 0.5$
$a_{(START,Y)} = \frac{0}{5} = 0$ $a_{(X,Y)} = \frac{3}{6} = 0.5$ $a_{(Y,Y)} = \frac{0}{6} = 0$ $a_{(Z,Y)} = \frac{3}{6} = 0.5$
$a_{(START,Z)} = \frac{3}{5} = 0.6$ $a_{(X,Z)} = \frac{2}{6} = 0.33$ $a_{(Y,Z)} = \frac{1}{6} = 0.17$ $a_{(Z,Z)} = \frac{0}{6} = 0$
$a_{(START,STOP)} = \frac{0}{5} = 0$ $a_{(X,STOP)} = \frac{1}{6} = 0.17$ $a_{(Y,STOP)} = \frac{4}{6} = 0.66$ $a_{(Z,STOP)} = \frac{0}{6} = 0$

Emission Probability

$u \backslash o$	a	b	c
X	0.17	0.5	0.33
Y	0.33	0	0.67
Z	0.17	0.33	0.5

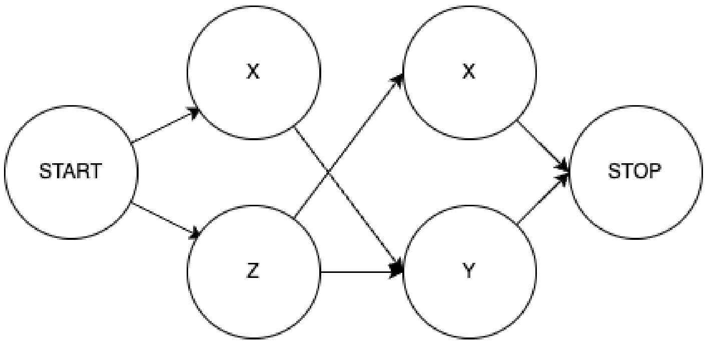
$b_X(a) = \frac{1}{6} = 0.17$ $b_Y(a) = \frac{2}{6} = 0.33$ $b_Z(a) = \frac{1}{6} = 0.17$
$b_X(b) = \frac{3}{6} = 0.5$ $b_Y(b) = \frac{0}{6} = 0$ $b_Z(b) = \frac{2}{6} = 0.33$
$b_X(c) = \frac{2}{6} = 0.33$ $b_Y(c) = \frac{4}{6} = 0.67$ $b_Z(c) = \frac{3}{6} = 0.5$

2

Using the parameters obtained, the following observations can be made:

- We cannot move directly from START to STOP. ($a_{(START,STOP)} = 0$)
- We cannot move directly from X to X. ($a_{(X,X)} = 0$)
- We cannot move directly from Y to Y. ($a_{(Y,Y)} = 0$)
- We cannot move directly from Z to Z. ($a_{(Z,Z)} = 0$)
- We cannot move directly from Z to STOP. ($a_{(Z,STOP)} = 0$)

The possible paths are depicted in the picture:



We thus have the following three possible paths:

1. $Start \rightarrow X \rightarrow Y \rightarrow STOP$
2. $Start \rightarrow Z \rightarrow Y \rightarrow STOP$
3. $Start \rightarrow Z \rightarrow X \rightarrow STOP$

Using backtracking,

- the second last node in the path can be X or Y. But, the observation sequence is **(b,b)** and $b_Y(b) = 0$. Therefore the second last node should be X.
- If the second last node is X, the second node (from the START) is Z.

We thus have this only possible path: $Start \rightarrow Z \rightarrow X \rightarrow STOP$. The score for this path is:

$$\begin{aligned}\textbf{Last Path} &= a_{(START,Z)} \times b_Z(b) \times a_{(Z,X)} \times b_X(b) \times a_{(X,STOP)} \\ &= 0.6 \times 0.33 \times 0.5 \times 0.5 \times 0.17 \\ &= 0.00833\end{aligned}$$

Therefore the path that follows that observation sequence is: $Start \rightarrow Z \rightarrow X \rightarrow STOP$

3

To modify the algorithm to incorporate the prior knowledge $y_i \neq V$ for a specific observation x_i, we need to set the transition probabilities from tag V to other tags to be very low, to discourage the sequence from including tag V after observing x_i. Here is how the algorithm can be implemented:

1. Initialise the Viterbi matrix with the emission probabilities for the first observation x_1 and the initial probabilities for the tags.

2. For each subsequent observation x_i, calculate the emission probabilities for each tag y_k based on the observation x_i.

3. If x_i is known to not have tag V, set the transition probabilities from tag V to other tags to be very low.

4. Run the Viterbi algorithm as per usual. The most probable sequence of tags will thus be outputted accroding to the prior knowledge.

4

Using soft EM, the following can be deduced from the diagram:

- Z is in a state of transition.
- X and Y are emitted from Z.

Therefore,

- $a_{(u,v)} = \frac{count(u,v)}{count(u)}$ where u, v are Z elements.
- $b_v(X_i) = \frac{count(X_i \to o)}{count(X)}$
- $b_v(Y_i) = \frac{count(Y_i \to o)}{count(Y)}$

Using the above, we can compute the forward and the backward probabilities.

Forward Probability:
α_i: The sum of the scores of all paths from START to node u at j.

The base case: $\alpha_u(1) = 1$

The general case: $\alpha_u(i+1) = \sum_v \alpha_v(i) a_{(u,v)} b_v(X_i) b_v(Y_i)$

Backward Probability:
βi: The sum of the scores of all paths from node u at j to STOP.

The base case: $\beta u(n) = a_{(u,STOP)} b_u(X_n) b_u(Y_n)$

The general case: $\beta u(i+1) = \sum_v \beta v(i+1) a_{(u,v)} b_u(X_i) b_u(Y_i)$

Forward Probability Time Complexity:
The time complexity for computing α for a single state at a single time step is O(T). So, computing the forward scores for each state and each time step, the overall time complexity of the forward algorithm is O($n \times T^2$).

Forward Probability Time Complexity: The overall time complexity of the backward algorithm is similar to the forward algorithm: O($n \times T^2$).

Overall Time Complexity: O($n \times T^2$)