Tutorial 04 — Answers

(Version 2.0; exercises from Ann Copestake)

1. This question is a variant of the fraudulent croupier scenario discussed in the lecture. In this version, the croupier has three six-sided dice: a fair (F) dice, and two loaded ones (L1 and L2). The croupier again secretly switches between dice. You have some training data sequences, of different lengths, with the observed dice rolls paired with the type of dice.

   (a) Assuming a first-order HMM, show what the two HMM probability matrices would look like for this scenario, explaining the notation you use.

   The A matrix is the state transition probability matrix. The notation is a_{nm} where n and m are names of hidden states, which here correspond to the actual dice used (F, 1, 2). These correspond to the probability of moving from state n to state m. We also have a start state, 0, and an end state, E. There is a 4-by-4 matrix, with rows corresponding to states 0, F, 1 and 2, and columns corresponding to states F, 1, 2 and E.

$$A = \begin{bmatrix} a_{0F} & a_{01} & a_{02} & - \\ a_{FF} & a_{F1} & a_{F2} & a_{FE} \\ a_{1F} & a_{11} & a_{12} & a_{1E} \\ a_{2F} & a_{21} & a_{22} & a_{2E} \end{bmatrix}$$

$$a_{ij} = P(X_t = s_j | X_{t-1} = s_i)$$

$$\forall_i \sum_{j=0}^{n+1} a_{ij} = 1$$

   B is the emission probability matrix. Terms correspond to the probability of each observation being associated with a particular hidden state.

$$b_i(k_j) = P(O_t = k_j | X_t = s_i)$$

   Here, the B matrix has three rows, corresponding to the three hidden states (i.e., the three dice), and 6 columns, corresponding to the numbers on the dice. The row corresponding to the fair dice will have values $1/6$ for each cell.

$$B = \begin{bmatrix} b_F(1) & b_F(2) & b_F(3) & b_F(4) & b_F(5) & b_F(6) \\ b_1(1) & b_1(2) & b_1(3) & b_1(4) & b_1(5) & b_1(6) \\ b_2(1) & b_2(2) & b_2(3) & b_2(4) & b_2(5) & b_2(6) \end{bmatrix}$$

   (b) Draw an HMM model and illustrate the different types of probabilities involved (you do not have to label all the arcs).

There are 3 real hidden states, F, 1 and 2, all fully interconnected, all with transitions to themselves, and all emitting observations (i.e., dice rolls, 1, 2, 3 etc.). The transition from F to 1 is labelled a_{F1}, the transition from 1 to F is labelled a_{1F} and the transition from 1 to 1 is labelled a_{11}. There is a start state 0, with transitions to each of F, 1 and 2, and an end state E, with transitions leading to it from F, 1 and 2. The distinction between hidden and observed should be made clear in the diagram.

(c) Suppose the croupier is acting according to certain rules which determine when the dice is switched. For each of the following rules, describe the effect in terms of the parameters of the HMM, and discuss whether the behaviour would be modelled by an HMM.

i. The croupier never switches directly from F to L2.

This is straightforwardly captured with the HMM: it just corresponds to a zero probability state transition: $a_{F2} = 0$ and $P(X_t = s_2|X_{t-1} = s_F) = 0$.
(However, a rule of this type would not lead to a 0 probability in an HMM as usually used, because of smoothing).

ii. The croupier knows in advance how many dice throws the sequence will contain and makes sure that the dice is always F on the last roll.

We cannot deduce much about a_{FE}, which corresponds to $P(X_t = s_E|X_{t-1} = s_F)$. Nevertheless, the counts in the training data will reflect the rule described: there will be no cases of transition between 1 and E or between 2 and E, and hence the counts for these events will be 0.
$a_{1E} = 0$ and $P(X_t = s_E|X_{t-1} = s_1) = 0$
$a_{2E} = 0$ and $P(X_t = s_E|X_{t-1} = s_2) = 0$

iii. The croupier never rolls L2 more than twice in a row.

This rule is equivalent to saying that: $P(X_t = s_2|X_{t-1} = s_2, X_{t-2} = s_2) = 0$. A second-order HMM would capture this. This doesn't correspond to anything in the first order HMM, however. All we can say is that, for the un-smoothed count: $P(X_t = s_2|X_{t-1} = s_2) \leq 0.5$ and $a_{22} \leq 0.5$. This follows because the most L2 transitions under this condition will occur if the dice is thrown exactly twice if it is thrown at all. In this case, there will be one L2 L2 transition and one L2 to not-L2 transition each time. This would correspond to: $P(X_t = s_2|X_{t-1} = s_2) = 0.5$. Hence the observed probability has to be less than that.

iv. The croupier always switches dice after rolling a 6.

We have no direct access to this probability in the HMM since the transitions concern hidden states only. The condition could affect the HMM probabilities however, depending on the extent to which the loaded dice probabilities are skewed in favour of 6. For instance, take the extreme case where L1 always gives a 6: the croupier will always switch after throwing L1, and hence: $P(X_t = s_1|X_{t-1} = s_1) = 0$.

2. Viterbi is an algorithm that allows you to process the input observation in time that is linear to the observation sequence. With a first order HMM, we keep N (number of states) maximum probabilities per observation at each step.

   (a) How many states do we need to keep for an N order HMM?

   We need to keep N^o states, where o is the order of the HMM.

   (b) What are the implications for the asymptotic complexity of Viterbi?

   Instead of $O(N^2T)$, we now have $O(N^{o+1}T)$. For small number of states N, this is OK as we mainly care about T, the length of the sequence, and the algorithm's linear behaviour in that respect is not affected.

Version list

- Version 1.0, February 4th 2020.
- Version 2.0, January 31st 2021.