Hidden Markov Models and the Viterbi algorithm

An HMM $H = \big(p_{ij}, e_i(a), w_i\big)$ is understood to have N hidden Markov states labelled by i ($1 \le i \le N$), and M possible observables for each state, labelled by a ($1 \le a \le M$). The state transition probabilies are $p_{ij} = p(q_{t+1} = j \,|\, q_t = i)$, $1 \le i, j \le N$ (where q_t is the hidden state at time t), the emission probability for the observable a from state i is $e_i(a) = p(O_t = a \,|\, q_t = i)$ (where O_t is the observation at time t), and the initial state probabilities are $w_i = p(q_1 = i)$.

Given a sequence of observations $O = O_1\, O_2 \cdots O_T$, and an HMM $H = \big(p_{ij}, e_i(a), w_i\big)$, we wish to find the maximum probability state path $Q = q_1\, q_2 \cdots q_T$. This can be done recursively using the Viterbi algorithm.

Let $v_i(t)$ be the probability of the most probable path ending in state i at time t, i.e.,

$$v_i(t) = \max_{q_1, q_2, \cdots q_{t-1}} P(q_1\, q_2 \cdots q_{t-1}, q_t = i, O_1\, O_2 \cdots O_t \,|\, H) \ ,$$

and let w_i be the initial probabilities of the states i at time $t = 1$. (Note this notation avoids the frightening greek letters δ, π, and λ used in the Rabiner notes, using instead v for Viterbi, w for weights, and H for hidden Markov model. The correspondence with the notation used in the Rabiner notes is $v_i(t) \leftrightarrow \delta_t(i)$, $e_i(a) \leftrightarrow b_i(a)$, $p_{ij} \leftrightarrow a_{ij}$, $w_i \leftrightarrow \pi_i$, $H \leftrightarrow \lambda$.)

Then $v_j(t)$ can be calculated recursively using

$$v_j(t) = \max_{1 \le i \le N} \big[v_i(t-1)\, p_{ij}\big]\, e_j(O_t)$$

together with initialization

$$v_i(1) = w_i\, e_i(O_1) \qquad 1 \le i \le N$$

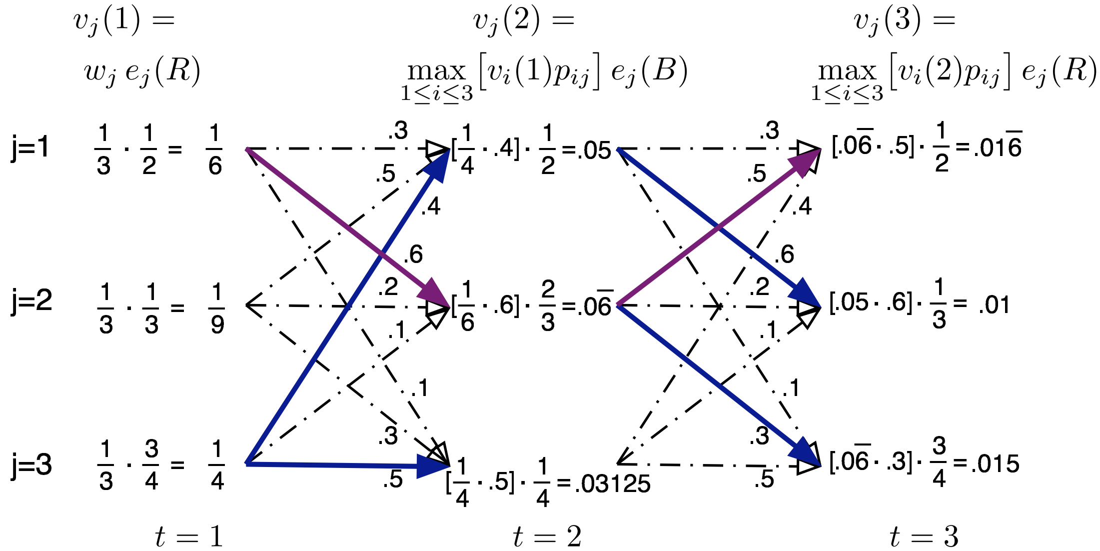
and termination

$$P^* = \max_{1 \le i \le N} [v_i(T)]$$

(i.e., at the end we choose the highest probability endpoint, and then we backtrack from there to find the highest probability path).

Note that the maximally likely path is not the only possible optimality criterion, for example choosing the most likely state at any given time requires a different algorithm and can give a slightly different result. But the overall most likely path provided by the Viterbi algorithm provides an optimal state sequence for many purposes.

To illustrate this, consider a three state HMM, with R or B emitted by each state (e.g., three urns, each with red or blue balls) with emission probabilities $e_1(R) = 1/2$, $e_2(R) = 1/3$, and $e_3(R) = 3/4$ (and correspondingly $e_1(B) = 1/2$, $e_2(B) = 2/3$, and $e_3(B) = 1/4$), state transition matrix $p_{ij} = \begin{pmatrix} .3 & .6 & .1 \\ .5 & .2 & .3 \\ .4 & .1 & .5 \end{pmatrix}$, and initial state probabilities $w_i = 1/3$. Suppose we observe the sequence RBR, then we can find the "optimal" state sequence to explain this sequence of observations by running the Viterbi algorithm by hand:



In the first step, we initialize the probabilities at $t = 1$ to $v_j(t = 1) = w_j\, e_j(R)$ for each $j = 1, 2, 3$. These are given in the first column to the left, as 1/6, 1/9, 1/4, respectively.

In the second step, $t = 2$, we determine first $v_1(t = 2)$ by considering the three quantities $v_i(1)p_{i1}$ for $i = 1, 2, 3$. They are respectively $(1/6)\cdot .3$, $(1/9)\cdot .5$, and $(1/4)\cdot .4$. The third one is the largest, so according to the algorithm we set $v_1(2) = [(1/4)\cdot .4]\cdot(1/2) = .05$, and remember that the maximum probability path to state $j = 1$ at time $t = 2$ came from state $j = 3$ at time $t = 1$ (blue line). Similarly, to determine $v_2(2)$ we consider the three quantities $v_i(1)p_{i2}$ for $i = 1, 2, 3$, respectively $(1/6)\cdot .6$, $(1/9)\cdot .2$, $(1/4)\cdot .1$, and the first is the largest, so we set $v_2(2) = [(1/6)\cdot .6]\cdot(2/3) = .0\bar{6}$. Finally, to determine $v_3(2)$ we consider the three quantities $v_i(1)p_{i3}$ for $i = 1, 2, 3$, respectively $(1/6)\cdot .1$, $(1/9)\cdot .3$, $(1/4)\cdot .5$, and the third is the largest, so we set $v_3(2) = [(1/4)\cdot .5]\cdot(1/4) = .03125$.

In the third step, $t = 3$, we determine first $v_1(t = 3)$ by considering the three quantities $v_i(2)p_{i1}$ for $i = 1, 2, 3$. They are respectively $.05\cdot .3$, $.0\bar{6}\cdot .5$, and $.03125\cdot .4$. The second is the largest, so according to the algorithm we set $v_1(3) = [.0\bar{6}\cdot .5]\cdot(1/2) = .01\bar{6}$, and remember that the maximum probability path to state $j = 1$ at time $t = 3$ came from state $j = 2$ at time $t = 2$ (blue line). Similarly, to determine $v_2(3)$ we consider the three quantities $v_i(2)p_{i2}$ for $i = 1, 2, 3$, respectively $.05\cdot .6$, $.0\bar{6}\cdot .2$, $.03125\cdot .1$, and the first is the largest, so we set $v_2(3) = [.05\cdot .6]\cdot(1/3) = .01$. Finally, to determine $v_3(2)$ we consider the three quantities $v_i(1)p_{i3}$ for $i = 1, 2, 3$, respectively $.05\cdot .1$, $.0\bar{6}\cdot .3$, $.03125\cdot .5$, and the second is the largest, so we set $v_3(3) = [.0\bar{6}\cdot .3]\cdot(3/4) = .015$.

Since there are only three observations, we can now use the termination step to determine that the maximum probability for the observations $O = RBR$ is $P^* = .01\bar{6}$ with state path $Q = 1, 2, 1$ (purple lines).