

The Forward Algorithm

Given a hidden Markov model (HMM) with

N hidden states $\{q_1, q_2, \dots, q_N\}$
 M emission states $\{y_1, y_2, \dots, y_M\}$

And where

π_i is the initial probability for hidden state q_i
 a_{ij} is the transition probability for moving from hidden state q_i to q_j
 $b_i(y_j)$ is the emission probability for hidden state q_i and observation y_j

We want to calculate $P(y_1, y_2, \dots, y_T)$, i.e. the probability of a given sequence of observed emissions, y_1, y_2, \dots, y_T .

We derive the forward algorithm, which allows us to calculate the probability of being in hidden state j after the first t observations. We call this value the forward probability:

$$\alpha_t(j) = P(y_1, y_2, \dots, y_t, q_t = j) \quad (1)$$

Which says that the forward probability for state j at time t is the probability of observing the sequence of emissions y_1, y_2, \dots, y_t and being in hidden state j at time t .

When $t = 1$, this is simply

$$\alpha_1(j) = P(y_1, q_1 = j) \quad (2)$$

$$= P(y_1 | q_1 = j) P(q_1 = j) \quad (3)$$

$$= b_j(y_1) \pi_j \quad (4)$$

$$= \pi_j b_j(y_1) \quad (5)$$

(2) \rightarrow (3) by chain rule

(3) \rightarrow (4) by observing that the first value is the emission probability for the first observation y_1 from hidden state j , and the second value is the probability of state j being the first hidden state

(5) rearranges (4) into our base case

$$\alpha_1(j) = \pi_j b_j(y_1)$$

For $t > 1$,

$$\alpha_t(j) = P(y_1, y_2, \dots, y_{t-1} y_t, q_t = j) \quad (6)$$

$$= \sum_{i=1}^N P(y_1, y_2, \dots, y_t, q_{t-1} = i, q_t = j) \quad (7)$$

$$= \sum_{i=1}^N P(y_t | y_1, y_2, \dots, y_{t-1}, q_{t-1} = i, q_t = j) P(y_1, y_2, \dots, y_{t-1}, q_{t-1} = i, q_t = j) \quad (8)$$

$$= \sum_{i=1}^N P(y_t | q_t = j) P(y_1, y_2, \dots, y_{t-1}, q_{t-1} = i, q_t = j) \quad (9)$$

$$\begin{aligned}
&= \sum_{i=1}^N P(y_t | q_t = j) P(q_t = j | y_1, y_2, \dots, y_{t-1}, q_{t-1} = i) P(y_1, y_2, \dots, y_{t-1}, q_{t-1} = i) \quad (10) \\
&= \sum_{i=1}^N P(y_t | q_t = j) P(q_t = j | q_{t-1} = i) P(y_1, y_2, \dots, y_{t-1}, q_{t-1} = i) \quad (11) \\
&= \sum_{i=1}^N b_j y(t) a_{ij} \alpha_{t-1}(i) \quad (12)
\end{aligned}$$

(6)→(7) because the probability of being in state j at time t depends on all possible hidden states at the previous time $t - 1$. We sum over all these N possible hidden states

(7)→(8) by chain rule

(8)→(9) by property of emission independence

(9)→(10) by chain rule

(10)→(11) by Markov property

(12) by rewriting in terms of emission, transmission, and forward properties

Rearranged, this is our recursion step:

$$\alpha_t(j) = \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} b_j(y_t)$$

On termination, we sum the forward probabilities at the last time T to obtain the probability for observing the given sequence,

$$P(y_1, y_2, \dots, y_T) = \sum_{i=1}^N \alpha_T(i)$$