Forward algorithm

Marginalization

\[
\Pr(X_t = j|Y_1, \ldots, Y_{t-1}) = \sum_{i \in S} \Pr(X_t = j|X_{t-1} = i) \Pr(X_{t-1} = i|Y_1, \ldots, Y_{t-1})
\]  

(1)

Marginalization

\[
\Pr(Y_t|Y_1, \ldots, Y_{t-1}) = \sum_{i \in S} \Pr(Y_t = j|X_t = i) \Pr(X_t = i|Y_1, \ldots, Y_{t-1})
\]  

(2)

Bayes Theorem

\[
\Pr(X_t = j|Y_1, \ldots, Y_t) = \frac{\Pr(Y_t|X_t = j) \Pr(X_t = j|Y_1, \ldots, Y_{t-1})}{\Pr(Y_t|Y_1, \ldots, Y_{t-1})}
\]

(3)

We only need the initial distribution \(\Pr(X_0 = i)\) for all \(i\), the transition matrix \(P\) containing \(\Pr(X_t = j|X_{t-1} = i)\) and the emission distributions \(\Pr(Y_t|X_t = j)\) of each state \(j \in S\).
Backward algorithm

Bayes Theorem

\[ \Pr(X_t = j|X_{t+1} = i, Y_1, \ldots, Y_n) = \frac{\Pr(Y_{t+1}, \ldots, Y_n, X_t = j|X_{t+1} = i, Y_1, \ldots, Y_t)}{\Pr(Y_{t+1}, \ldots, Y_n|X_{t+1} = i, Y_1, \ldots, Y_t)} = \Pr(X_t = j|X_{t+1} = i, Y_1, \ldots, Y_t) \]

because the numerator can be written as

\[ \Pr(Y_{t+1}, \ldots, Y_n|X_t = j, X_{t+1} = i, Y_1, \ldots, Y_t) \Pr(X_t = j|X_{t+1} = i, Y_1, \ldots, Y_t) \]

and \( \Pr(Y_{t+1}, \ldots, Y_n|X_t = j, X_{t+1} = i, Y_1, \ldots, Y_t) \) is independent of \( X_t \).

Bayes Theorem

\[ \Pr(X_t = j|X_{t+1} = i, Y_1, \ldots, Y_t) = \frac{\Pr(X_{t+1} = i|X_t = j) \Pr(X_t = j|Y_1, \ldots, Y_t)}{\sum_{j \in S} \Pr(X_{t+1} = i|X_t = j) \Pr(X_t = j|Y_1, \ldots, Y_t)} \]  
(4)

Marginalization

\[ \Pr(X_t = j|Y_1, \ldots, Y_n) = \sum_{i \in S} \Pr(X_t = j|X_{t+1} = i, Y_1, \ldots, Y_n) P(X_{t+1} = i, |Y_1, \ldots, Y_n) \]  
(5)