Hidden Markov models (continued)

$$\Pr\left(x^{i+1}, y_{i+1}\right)$$

$$= \Pr\left(x_{i+1}, x^{i}, y_{i+1}\right)$$

$$= \sum_{y_i} \Pr\left(x_{i+1}, x^{i}, y_{i+1}, y_i\right)$$

$$= \sum_{y_i} \Pr\left(x_{i+1}, y_{i+1} \mid x^{i}, y_i\right) \Pr\left(x^{i}, y_i\right)$$

$$= \sum_{y_i} \Pr\left(x_{i+1} \mid y_{i+1}, x^{i}, y_i\right) \Pr\left(y_{i+1} \mid x^{i}, y_i\right) \Pr\left(x^{i}, y_i\right)$$

$$= \sum_{y_i} \Pr\left(x_{i+1} \mid y_{i+1}\right) \Pr\left(y_{i+1} \mid y_i\right) \Pr\left(x^{i}, y_i\right)$$

$$= \sum_{y_i} \Pr\left(x^{i}, y_i\right) p_{x_{i+1} y_{i+1}} \rho_{y_i y_{i+1}}$$

We initialize the algorithm by setting

$$\Pr(x_1, y_1) = \Pr(x_1 \mid y_1)\Pr(y_1) = p_{x_1 y_1}\Pr(y_1)$$

where $\Pr(y_1)$ has to be set by the user.

At the end of the sequence of length N,

$$\Pr\left(x^N\right) = \sum_{y_N} \Pr\left(x^N, y_N\right).$$

This strategy for calculating $\Pr\left(x^N\right)$ is known as the *Forward* algorithm.

Viterbi algorithm

The Viterbi algorithm finds the most probable “path” of states though the hidden Markov model (HMM).

In our notation, the Viterbi algorithm finds the path $y = y_1 y_2 \ldots y_N$ that maximizes $\Pr(x, y)$.

The Viterbi algorithm returns the probability $\Pr(x, y)$ associated with the path.

Notice that the path that maximizes $\Pr(x, y)$ must be identical to the path that maximizes

$$\Pr(y \mid x) = \frac{\Pr(x, y)}{\Pr(x)}$$

The path through the hidden Markov model that maximizes $\Pr(x, y)$ will be denoted by ν.

Consider all paths that travel through hidden states $1, 2, \ldots, i$ **and that end with** $y_i = j$. Among these paths, let $\nu^i(j)$ be the path that maximizes $\Pr\left(x^i, y^{i-1}, y_i = j\right)$.

$$\Pr\left(x^{i+1}, \nu^{i+1}(j)\right)$$
$$= \max_k \Pr\left(x_{i+1}, x^i, y_{i+1} = j, \nu^i(k)\right)$$
$$= \max_k \Pr\left(x_{i+1}, y_{i+1} = j \mid x^i, \nu^i(k)\right)\Pr\left(x^i, \nu^i(k)\right)$$
$$= \max_k \Pr\left(x_{i+1} \mid y_{i+1} = j, x^i, \nu^i(k)\right)\Pr\left(y_{i+1} = j \mid x^i, \nu^i(k)\right)\Pr\left(x^i, \nu^i(k)\right)$$
$$= \max_k \Pr\left(x_{i+1} \mid y_{i+1} = j\right)\Pr\left(y_{i+1} = j \mid \nu^i(k)\right)\Pr\left(x^i, \nu^i(k)\right)$$
$$= \max_k \Pr\left(x_{i+1} \mid y_{i+1} = j\right)\Pr\left(y_{i+1} = j \mid y_i = k\right)\Pr\left(x^i, \nu^i(k)\right)$$

$$\Pr\left(x^N, \nu\right) = \max_j \Pr\left(x^N, \nu^N(j)\right)$$

(path ν can be recovered by keeping track of which choice achieved the maximum at every step)

Backward Algorithm

Sometimes we want to infer a specific hidden Markov model state rather than to jointly infer all states in the hidden Markov model.

In this case, we care about

$$\Pr(y_i \mid x) = \frac{\Pr(x, y_i)}{\Pr(x)}$$

Let χ^i represent the subsequence $x_i, x_{i+1}, x_{i+2}, \ldots, x_N$.

With this notation, x has the same meaning as x^i together with χ^{i+1}.

So,

$$\Pr(y_i \mid x) = \frac{\Pr(x^i, \chi^{i+1}, y_i)}{\Pr(x)}$$

$$\begin{aligned}\Pr(x^i, \chi^{i+1}, y_i) &= \Pr(\chi^{i+1} \mid x^i, y_i)\Pr(x^i, y_i) \\ &= \Pr(\chi^{i+1} \mid y_i)\Pr(x^i, y_i)\end{aligned}$$

$\Pr(x^i, y_i)$ is determined by the forward algorithm.

The backward algorithm starts with

$$\Pr(\chi^N \mid y_{N-1}) = \Pr(x_N \mid y_{N-1}) = \sum_{y_N} \Pr(x_N, y_N \mid y_{N-1})$$

$$= \sum_{y_N} \Pr(x_N \mid y_N, y_{N-1})\Pr(y_N \mid y_{N-1})$$

$$= \sum_{y_N} \Pr(x_N \mid y_N)\Pr(y_N \mid y_{N-1})$$

$$= \sum_{y_N} p_{x_N y_N} \rho_{y_{N-1} y_N}$$

The backward algorithm continues with

$$\Pr\left(\chi^{i+1} \mid y_i\right)$$

$$= \Pr\left(x_{i+1}, \chi^{i+2} \mid y_i\right)$$

$$= \sum_{y_{i+1}} \Pr\left(x_{i+1}, \chi^{i+2}, y_{i+1} \mid y_i\right)$$

$$= \sum_{y_{i+1}} \Pr\left(x_{i+1}, \chi^{i+2} \mid y_i, y_{i+1}\right) \Pr\left(y_{i+1} \mid y_i\right)$$

$$= \sum_{y_{i+1}} \Pr\left(x_{i+1} \mid y_i, y_{i+1}, \chi^{i+2}\right) \Pr\left(\chi^{i+2} \mid y_i, y_{i+1}\right) \Pr\left(y_{i+1} \mid y_i\right)$$

$$= \sum_{y_{i+1}} \Pr\left(x_{i+1} \mid y_{i+1}\right) \Pr\left(\chi^{i+2} \mid y_{i+1}\right) \Pr\left(y_{i+1} \mid y_i\right)$$

$$= \sum_{y_{i+1}} p_{x_{i+1} y_{i+1}} \Pr\left(\chi^{i+2} \mid y_{i+1}\right) \rho_{y_i y_{i+1}}$$