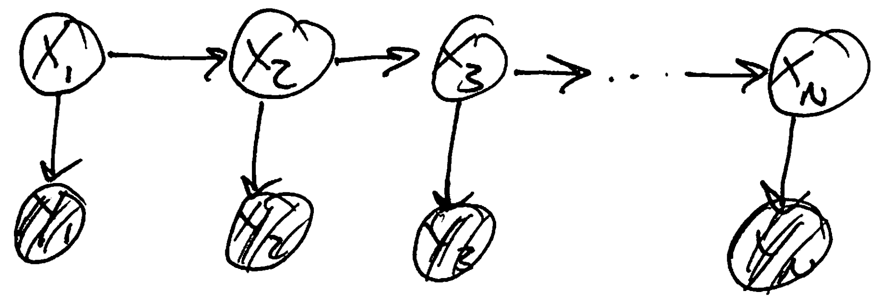
INFERENCE ON HMMS

JASON PACHECO
4/27/11

FORWARD BACKWARD ALGORITHM

Assume we have an HMM with



X_n: Latent state
Y_n: Observation

GOAL: To compute marginal posterior distributions $p(X_n \mid Y_1, \ldots, Y_N)$ for all X_n.

ALGORITHM STATEMENT:

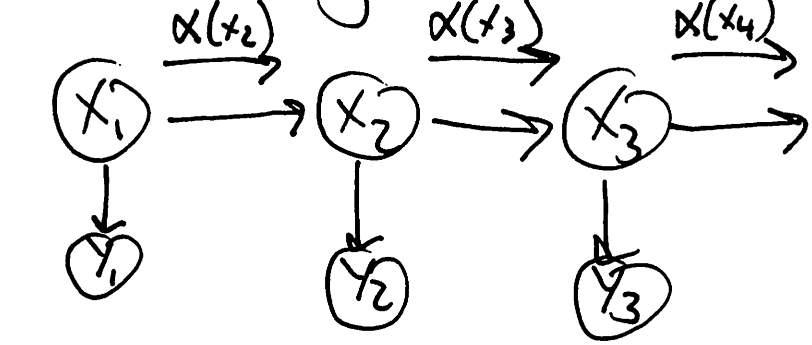
The forward-backward algorithm computes the quantities

$$\alpha(X_n) = p(X_n, Y_1, \ldots, Y_n) \qquad (\text{"forward message"})$$
$$\beta(X_n) = p(Y_{n+1}, \ldots, Y_N \mid X_n) \qquad (\text{"backward message"})$$

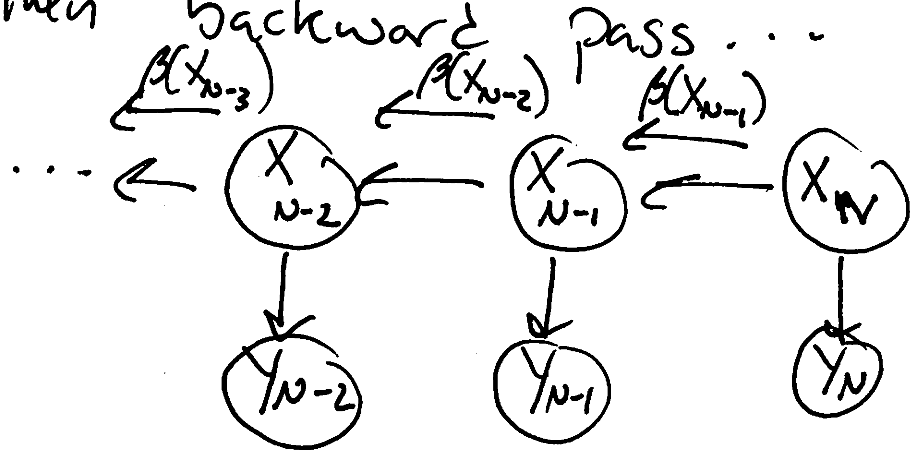
Once these are computed the marginal probabilities are easily obtained as.

$$p(X_n \mid Y_1, \ldots, Y_N) \propto \alpha(X_n) \cdot \beta(X_n)$$

Algorithm begins with forward pass...



Then backward pass...



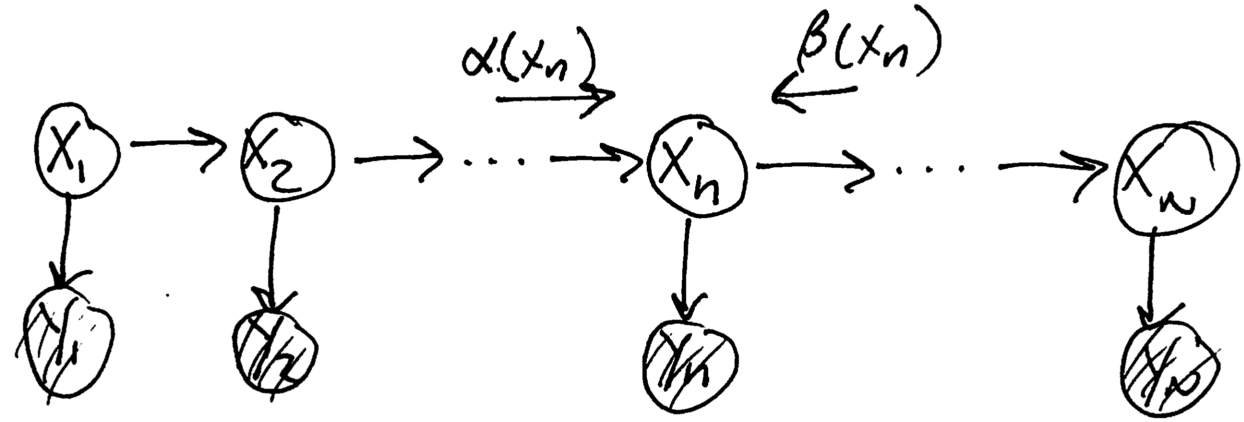
Messages are updated recursively as,

$$\alpha(x_n) = p(y_n | x_n) \sum_{x_{n-1}} \alpha(x_{n-1})\, p(x_n | x_{n-1})$$

$$\beta(x_n) = \sum_{x_{n+1}} \beta(x_{n+1})\, p(y_{n+1} | x_{n+1})\, p(x_{n+1} | x_n)$$

These come from independence properties of the graph and simple algebra. See book for detailed derivation.

~~Derivation for backward message follows similarly.~~ $\cancel{p(x_n|y_1^N) \propto \alpha(x_n)\beta(x_n)}$



With marginal posterior distributions we can find the most probable x_n, given the data $y_1, \ldots, y_N$ for all x_n.

This is not the same as finding $x_1, \ldots, x_n$ which is jointly most probable, e.g.

$$\max_{x_1, x_2, \ldots, x_N} p(x_1, x_2, \ldots, x_N, y_1, \ldots, y_N)$$

For this we use the Viterbi Algorithm.

VITERBI ALGORITHM

Consists of a forward and backward pass similar to forward-backward algorithm.

Recall forward messages represent

$$\alpha(x_n) = p(y_1, \cdots, y_n, x_n) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{n-1}} p(x_1, \cdots, x_n, y_1^m)$$

We introduce a similar quantity known as the <u>max-marginals</u>

$$\omega(x_n) = \max_{x_1, \cdots, x_{n-1}} p(y_1, \ldots, y_n, x_1, \ldots, x_n)$$

The update equation for $\alpha(x_n)$ was,

$$\alpha(x_n) = p(y_n | x_n) \sum_{x_{n-1}} \alpha(x_{n-1}) p(x_n | x_{n-1})$$

We replace the summation with max

$$p(y_n | x_n) \max_{x_{n-1}} \alpha(x_{n-1}) p(x_n | x_{n-1})$$

and take the logarithm

$$\omega(x_n) = \log p(y_n | x_n) + \max_{x_{n-1}} \left\{ \omega(x_{n-1}) + \log p(x_n | x_{n-1}) \right\}$$

This defines the forward recursion.

At final time N we have that $\omega(x_N)$ yields the optimal value for x_N^{max}. We then backtrack to find the full assignment

$$x_n^{max} = \underset{x_n}{\arg\max} \left[\omega(x_n) \, p(x_{n+1}^{max} \mid x_n) \right]$$

This can be viewed as backtracking through the trellis diagram (see example slides)