MSA101/MVE187 2017 Lecture 12

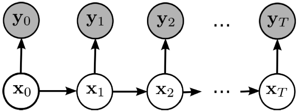
Petter Mostad

Chalmers University

October 5, 2017

Example: The Forward-Backward algorithm

Message passing applied to the following Bayesian Network: A *Hidden Markov Model*



Objective: Compute the marginal posterior distribution of every x_i given data $y_0, \ldots, y_T$: Use $\pi(x_i \mid y_0 \ldots, y_T) \propto \pi(y_{i+1}, \ldots, y_T \mid x_i)\pi(x_i \mid y_0, \ldots, y_i)$ and

1. Forward: For $i = 0, \ldots, T$ compute $\pi(x_i \mid y_0, \ldots, y_i)$ using

$$\begin{aligned}
\pi(x_i \mid y_0, \ldots, y_i) &\propto \pi(y_i \mid x_i)\pi(x_i \mid y_0, \ldots, y_{i-1}) \\
&= \pi(y_i \mid x_i) \int \pi(x_i \mid x_{i-1})\pi(x_{i-1} \mid y_0, \ldots, y_{i-1})\, dx_{i-1}
\end{aligned}$$

2. Backward: For $i = T-1, \ldots, 0$ compute $\pi(y_{i+1}, \ldots, y_T \mid x_i)$ using

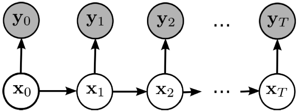
$$\pi(y_{i+1}, \ldots, y_T \mid x_i) = \int \pi(y_{i+2}, \ldots, y_T \mid x_{i+1})\pi(y_{i+1} \mid x_{i+1})\pi(x_{i+1} \mid x_i)\, dx_{i+1}$$

Finding the maximum aposteriori (MAP) in graphical models

- A basic example: The Viterbi algorithm, for HMMs.
- A more advanced example: The Baum-Welch algorithm, for estimating parameters in HMMs.
- The ideas of the two examples above can be generalized into handling graphical models where the graph does not contain loops.
- When the graph contains loops, the problem may be very hard (NP) and approximate solutions may be necessary.

The Viterbi algorithm

We start with an HMM where the x_i have a finite state space $\{1, \ldots, k\}$:

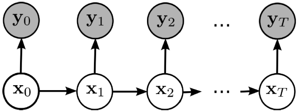


Objective: Compute the vector $x_0, \ldots, x_T$ which gives the highest posterior, for given fixed values of the y_i. IDEA: Compute and store, recursively, for $i = 0, \ldots, T$, the following:

- For $j = 1, \ldots, k$:
  - The vector $(\hat{x}_1, \ldots, \hat{x}_i)$ maximizing $\pi(\hat{x}_1, \ldots, \hat{x}_i, y_1, \ldots, y_i)$ with $\hat{x}_i = j$. NOTE: Only the value of $\hat{x}_{i-1}$ needs to be stored!
  - The value of this maximum.
- Because of indepenencies, the first $i - 1$ values of $(\hat{x}_1, \ldots, \hat{x}_i)$ will always correspond to those considered at the $i - 1$'th step.
- At any point, $(\hat{x}_1, \ldots, \hat{x}_i)$ can be reconstructed tracing backwards through stored information.
- The recursion step consists in considering all possible combinations of x_{i-1} and x_i.

The Baum-Welch algorithm

We start with an HMM where all the nodes have a finite state spaces



but where the parameters of the distributions $\pi(X_0)$, $\pi(X_i \mid X_{i-1})$, and $\pi(Y_i \mid X_i)$ are unknown. Objective: Given fixed values for the y_i, find maximum likelihood estimates for the parameters in the model.

- Note: By adding nodes representing the unknown parameters, and assuming flat priors, the problem becomes that of computing a MAP.
- Idea: Use the EM algorithm, with the values of the x_i as the augmented data.
- The E step of the EM algorithm is computed using the Forward-Backward algorithm.