COS513: Sequence Models II

Dominic Kao

November 24, 2010

1 Final Project Ideas

Jonathan Cohen's schizophrenic data

Kenneth Norman's fmri data

Online text data (i.e. news sites)

kaggle.com is a site for participants to compete against each other to produce the best models for data sets

movielens.org provides data sets that contain movie preferences (see grouplens.org/node/73)

2 Hidden Markov Models

Recall from the previous class that a hidden markov model is a generalization of the finite mixture model, with $z_{1:T}$ representing the indices of clusters associated with $x_{1:T}$, the data.

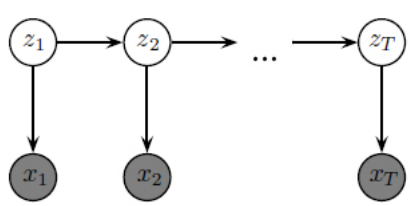

Figure 1: Graphical Model for HMMs

The transition probability is $p(z_t|z_{t-1})$ or equivalently a_{z_{t-1},z_t}. The emission probability is $p(x_t|z_t)$ or equivalently $p(x_t|\theta_{z_t})$.

Recall that EM is broken down into two steps:

1. Compute conditional expectations of the hidden variables.
2. Maximize the complete log likelihood using the expectations in 1.

$$E[z_t^i|x_{1:T}] = p(z_t = i|x_{1:T}) \tag{1}$$

Equation 1 is the expectation that the indicator vector z, for some observation t, takes on the mixture component i, conditioned on the data. From a simplified speech recognition perspective, this is the probability that, for some piece of a waveform t, we recognize it as a particular word i.

Figure 2: Speech Recognition Example

Similarly,

$$E[z_{t-1}^i z_t^j|x_{1:T}] = p(z_{t-1} = i, z_t = j|x_{1:T}) \tag{2}$$

Recall that we defined alpha and beta as follows (each is a vector of K elements):

$$\alpha(z_t) \triangleq p(x_{1:t}, z_t) \tag{3}$$

$$\beta(z_t) \triangleq p(x_{t-1:T}|z_t) \tag{4}$$

And the expectations (from Equations 1 and 2):

$$E[z_t|x_{1:T}] = \frac{\alpha(z_t)\beta(z_t)}{p(x_{1:T})} \tag{5}$$

$$E[z_{t-1}z_t|x_{1:T}] = \frac{\alpha(z_{t-1})p(z_t|z_{t-1})p(x_t|z_t)\beta(z_t)}{p(x_{1:T})} \tag{6}$$

We can define $p(x_{1:T})$ as follows:

$$\begin{aligned} p(x_{1:T}) &= \sum_{z_{1:T}} p(x_{1:T}|z_{1:T})p(z_{1:T}) \\ &= \sum_{z_{1:T}} (\prod_{t=1}^{T} p(x_t|z_t))(p(z_1)\prod_{t=2}^{T} p(z_t|z_{t-1})) \end{aligned} \tag{7}$$

We can evaluate $p(x_{1:T})$ by the following:

$$p(z_t|x_{1:T}) = \frac{\alpha(z_t)\beta(z_t)}{p(x_{1:T})} \tag{8}$$

$$\begin{aligned} \sum_{z_t} \frac{\alpha(z_t)\beta(z_t)}{p(x_{1:T})} = 1 \\ \sum_{z_t} \alpha(z_t)\beta(z_t) = p(x_{1:T}) \end{aligned} \tag{9}$$

Then all that's left is to update $\alpha(z_t)$ and $\beta(z_t)$. For the base case $\alpha(z_1)$, we recall the graphical model:

Figure 3: Graphical Model for z_1 and x_1

$$
\begin{aligned}
\alpha(z_1) &= p(x_1, z_1) \\
&= p(z_1)p(x_1|z_1) \\
&= \pi_{z_1} p(x_1|\theta_{z_1})
\end{aligned} \tag{10}
$$

Now we can assume that we have computed $\alpha(z_t)$ and find a recursive formula $\alpha(z_{t+1})$:

$$
\begin{aligned}
\alpha(z_{t+1}) &= p(x_{1:t+1}, z_{t+1}) \\
&= p(x_{1:t+1}|z_{t+1})p(z_{t+1})
\end{aligned} \tag{11}
$$

Our graphical model again:

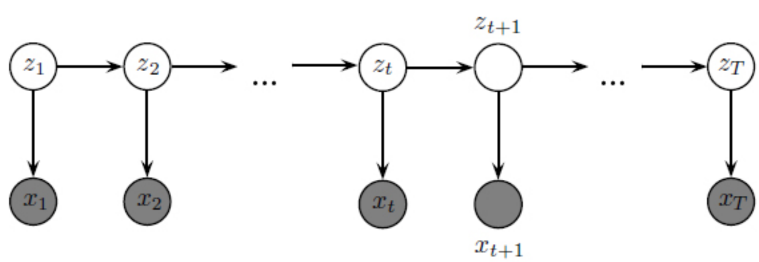

Figure 4: Graphical Model for HMMs

Using Bayes Ball, the following independence holds:

$$
x_{t+1} \perp\!\!\!\perp x_{1:t}|z_{t+1} \tag{12}
$$

Then, from Equations 11 and 12:

$$\begin{aligned}
\alpha(z_{t+1}) &= p(x_{1:t}|z_{t+1})p(x_{t+1}|z_{t+1})p(z_{t+1}) \\
&= p(x_{1:t}, z_{t+1})p(x_{t+1}|z_{t+1}) \\
&= \sum_{z_t} p(x_{1:t}, z_t, z_{t+1})p(x_{t+1}|z_{t+1}) \\
&= \sum_{z_t} p(x_{1:t}, z_{t+1}|z_t)p(x_{t+1}|z_{t+1})p(z_t) \\
&= \sum_{z_t} p(x_{1:t}|z_t)p(z_{t+1}|z_t)p(z_t)p(x_{t+1}|z_{t+1}) \\
&= \sum_{z_t} p(x_{1:t}, z_t)p(z_{t+1}|z_t)p(x_{t+1}|z_{t+1})
\end{aligned} \tag{13}$$

In the last line of Equation 13, the first term corresponds to $\alpha(z_t)$, the second is the transition probability $a_{z_t,z_{t+1}}$, and the last is the emission probability $p(x_{t+1}|\theta_{z_{t+1}})$. Computing $\alpha(z_{t+1})$ from $\alpha(z_t)$ is $O(K^2)$. Computing all of the alphas is $O(TK^2)$.

We now develop a recursive equation for $\beta(z_t)$:

$$\begin{aligned}
\beta(z_T) &\triangleq \vec{1} \\
\beta(z_t) &= p(x_{t+1:T}|z_t) \\
&= \sum_{z_{t+1}} p(x_{t+1:T}, z_{t+1}|z_t) \\
&= \sum_{z_{t+1}} p(x_{t+1:T}|z_{t+1}, z_t)p(z_{t+1}|z_t) \\
&= \sum_{z_{t+1}} p(x_{t+1:T}|z_{t+1})p(z_{t+1}|z_t) \\
&= \sum_{z_{t+1}} p(x_{t+2:T}|z_{t+1})p(x_{t+1}|z_{t+1})p(z_{t+1}|z_t)
\end{aligned} \tag{14}$$

In the last line of Equation 14, the first term corresponds to $\beta(z_{t+1})$, the second is the emission probability $p(x_{t+1}|\theta_{z_{t+1}})$, and the third the transition probability $a_{z_t,z_{t+1}}$.