CS281A/STAT241A Homework 5

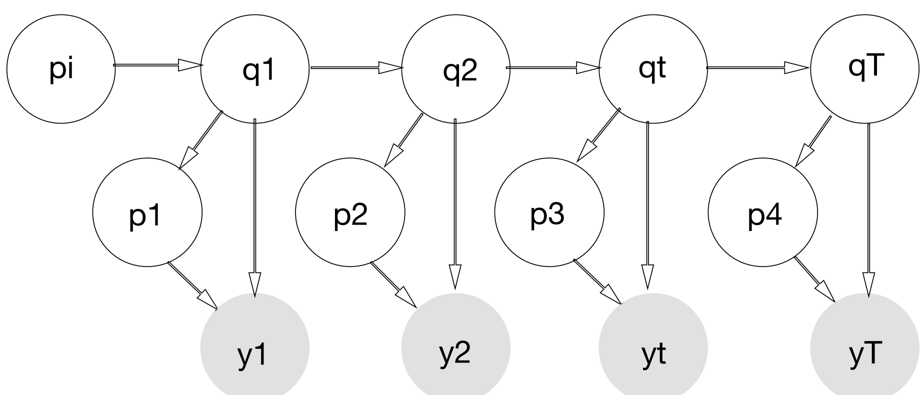
Thibault Doutre

November 12, 2015

1 HMM

1.1 Graph

Here is the graphical model of the HMM with mixtures of Poissons.



$$p_t = \begin{pmatrix} 0 \\ | \\ 0 \\ 1 \\ 0 \\ | \\ 0 \end{pmatrix} \in R^K, \qquad q_t = \begin{pmatrix} 0 \\ | \\ 0 \\ 1 \\ 0 \\ | \\ 0 \end{pmatrix} \in R^m$$

1.2 E-step

Here is the complete log-likelihood:

$$log(p(q, y|\theta)) = log(\pi_{q_0}) + \sum_{t=1}^{T-1} log(a_{q_t} a_{q_t}) + \sum_{t=1}^{T} log(p(y_t|q_t)) \tag{1}$$

The expected complete log-likelihood can be obtained with computing the following statistics:

$$E_1 = \sum_{i=1}^{m} p(q_0^i = 1|y) log(\pi_i) \tag{2}$$

$$E_2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{t=1}^{T-1} p(q_t^i q_t^{i+1} = 1|y) log(a_{ij}) \tag{3}$$

$$\begin{aligned} E_3 = & \sum_{i=1}^{m} \sum_{k=1}^{K} \sum_{t=1}^{T} p(q_t^i p_t^k = 1|y) log(\frac{e^{-\lambda_{ik}} \lambda_{ik}^{y_t}}{y_t!}) \\ & - \sum_{i=1}^{m} \sum_{k=1}^{K} \sum_{t=1}^{T} p(q_t^i p_t^k = 1|y) log(\lambda_{ik}) \\ & + \sum_{i=1}^{m} \sum_{k=1}^{K} \sum_{t=1}^{T} y_t p(q_t^i p_t^k = 1|y) log(\lambda_{ik}) \\ & - \sum_{t=1}^{T} log(y_t!) \end{aligned} \tag{4}$$

Let define the following objects:

$$\gamma_0^i = p(q_0^i|y) \tag{5}$$

$$A_{ij} = \sum_{t=1}^{T-1} p(q_t^i q_{t+1}^j = 1|y) \tag{6}$$

$$B_{ik} = \sum_{t=1}^{T} y_t p(q_t^i p_t^k = 1|y) \tag{7}$$

$$C_{ik} = \sum_{t=1}^{T} p(q_t^i p_t^k = 1|y) \tag{8}$$

Using these objects above, he expected complete log-likelihood can be written as:

$$\begin{aligned} E(l_c(\theta, y, q)|y, \theta) = &\sum_{i=1}^{m} \gamma_0^i log(\pi_i) + \sum_{i=1}^{m}\sum_{j=1}^{m} A_{ij} log(a_{ij}) \\ &- \sum_{i=1}^{m}\sum_{k=1}^{K} C_{ik}\lambda_{ik} + \sum_{i=1}^{m}\sum_{k=1}^{K} B_{ik} log(\lambda_{ik}) \\ &- \sum_{t=1}^{T} log(y_t!) \end{aligned} \tag{9}$$

In order to compute this expected likelihood, I need to define:

$$\begin{aligned} p(q_t q_t|y) &= \frac{\alpha(q_t)p(y_{t+1}|q_{t+1})\beta(q_{t+1})a_{q_t q_t}}{p(y)} \\ &= \frac{\alpha(q_t)p(y_{t+1}|q_{t+1})p(q_{t+1}|y)a_{q_t q_t}}{\alpha(q_{t+1})} \end{aligned} \tag{10}$$

$$\begin{aligned} p(q_t p_t|y) &= p(p_t|q_t y)p(q_t|y) \\ &= \frac{p(p_t, q_t, y)}{p(q_t, y)} \frac{p(q_t, y)}{p(y)} \\ &= p(p_t|q_t y_t)p(q_t|y) \\ &= \frac{p(y_t|q_t p_t)p(p_t|q_t)p(q_t|y)}{p(y_t|q_t)} \end{aligned} \tag{11}$$

The alphas, betas and dzetas are computed using the classical formulas in the textbook.

1.3 M-step

Taking the derivative under constraints of the expected likelihood with respect to each parameter, I obtain the following updates for the M-step:

$$\hat{a}_{ij} = \frac{A_{ij}}{\sum_{j=1}^{m} A_{ij}} \tag{12}$$

$$\hat{\lambda}_{ik} = \frac{B_{ik}}{C_{ik}} \tag{13}$$

$$\hat{p}(p_t^i = 1 | q_t^k = 1) = \frac{C_{ij}}{\sum_{j=1}^{m} C_{ij}} \tag{14}$$

2 Trees

2.1 Alpha-Beta recursion

Here we need to define a new alpha beta recursion. As for the tree model, I chose to include the root i in the subtree y_{D_i}.

$$\begin{aligned}
p(x_i|y) &= p(x_i, y_{D_i}, y_{D_i^c}) \\
&= \frac{p(x_i|y_{D_i})p(y_{D_i^c}|x_i, y_{D_i})}{p(y_{D_i^c}|y_{D_i})} \\
&= \frac{p(y_{D_i}|x_i)p(x_i)p(y_{D_i^c}|x_i)}{p(y_{D_i^c}|y_{D_i})p(y_{D_i})} \\
&= \frac{p(y_{D_i}|x_i)p(y_{D_i^c}, x_i)}{p(y)}
\end{aligned} \tag{15}$$

I then define alpha and beta this way:

$$\alpha(x_i) = p(y_{D_i^c}, x_i) \tag{16}$$

$$\beta(x_i) = p(y_{D_i}|x_i) \tag{17}$$

I can then compute the alpha recursion:

$$\begin{aligned}
\alpha(x_{2i}) &= p(y_{D_{2i}^c}, x_{2i}) \\
&= \sum_{x_i} p(y_{D_i^c}, y_{D_{2i+1}}, y_i, x_{2i}|x_i)p(x_i) \\
&= \sum_{x_i} p(y_{D_i^c}|x_i)p(y_{D_{2i+1}}|x_i)p(y_i|x_i)p(x_{2i}|x_i)p(x_i) \\
&= \sum_{x_i} \alpha(x_i)p(y_{D_{2i+1}}|x_i)p(y_i|x_i)p(x_{2i}|x_i)
\end{aligned} \tag{18}$$

Where:

$$
\begin{aligned}
p(y_{D_{2i+1}}|x_i) &= \sum_{x_{2i+1}} p(y_{D_{2i+1}}|x_i, x_{2i+1}) p(x_{2i+1}|x_i) \\
&= \sum_{x_{2i+1}} p(y_{D_{2i+1}}|x_{2i+1}) p(x_{2i+1}|x_i) \\
&= \sum_{x_{2i+1}} \beta(x_{2i+1}) p(x_{2i+1}|x_i) \qquad (19)
\end{aligned}
$$

Similarly, I have by symmetry:

$$
\alpha(x_{2i+1}) = \sum_{x_i} \alpha(x_i) p(y_{D_{2i}}|x_i) p(y_i|x_i) p(x_{2i+1}|x_i) \qquad (20)
$$

$$
p(y_{D_{2i}}|x_i) = \sum_{x_{2i}} \beta(x_{2i}) p(x_{2i}|x_i) \qquad (21)
$$

As for the beta recursion:

$$
\begin{aligned}
\beta(x_i) &= p(y_{D_i}|x_i) \\
&= \sum_{x_{2i}} \sum_{x_{2i+1}} p(y_{D_i}|x_i, x_{2i+1}, x_{2i}) p(x_{2i+1}, x_{2i}|x_i) \\
&= \sum_{x_{2i}} \sum_{x_{2i+1}} p(y_{D_{2i}}, y_{D_{2i+1}}, y_i|x_i, x_{2i+1}, x_{2i}) p(x_{2i}|x_i) p(x_{2i+1}|x_i) \\
&= \sum_{x_{2i}} \sum_{x_{2i+1}} p(y_{D_{2i}}|x_{2i}) p(y_{D_{2i+1}}|x_{2i+1}) p(y_i|x_i) p(x_{2i}|x_i) p(x_{2i+1}|x_i) \\
&= \sum_{x_{2i}} \sum_{x_{2i+1}} \beta(x_{2i}) \beta(x_{2i+1}) p(y_i|x_i) p(x_{2i}|x_i) p(x_{2i+1}|x_i) \qquad (22)
\end{aligned}
$$

For the M-step I will also have to compute:

$$
\begin{aligned}
p(x_i, x_{2i}|y) &= \sum_{x_{2i+1}} p(x_i, x_{2i}, x_{2i+1}|y) \\
&= \frac{1}{p(y)} \sum_{x_{2i+1}} p(y|x_i, x_{2i}, x_{2i+1}) p(x_i, x_{2i}, x_{2i+1}) \\
&= \frac{1}{p(y)} \sum_{x_{2i+1}} p(y_{D_{2i}}, y_{D_{2i+1}}, y_{D_i^c}, y_i|x_i, x_{2i}, x_{2i+1}) p(x_i, x_{2i}, x_{2i+1}) \\
&= \frac{1}{p(y)} \sum_{x_{2i+1}} p(y_{D_{2i}}|x_i) p(y_{D_{2i+1}}|x_{2i+1}) p(y_{D_i^c}|x_i) p(y_i|x_i) p(x_{2i}|x_i) p(x_{2i+1}|x_i) p(x_i) \\
&= \frac{1}{p(y)} \sum_{x_{2i+1}} \alpha(x_i) \beta(x_{2i}) \beta(x_{2i+1}) p(y_i|x_i) p(x_{2i}|x_i) p(x_{2i+1}|x_i)
\end{aligned}
\tag{23}
$$

Similarly:

$$
p(x_i, x_{2i+1}|y) = \frac{1}{p(y)} \sum_{x_{2i}} \alpha(x_i) \beta(x_{2i}) \beta(x_{2i+1}) p(y_i|x_i) p(x_{2i}|x_i) p(x_{2i+1}|x_i)
\tag{24}
$$

2.2 E-step

Here is the expression of the expected complete log-likelihood:

$$
\begin{aligned}
E(l_c(x, y)|y) = & \sum_{x_i} p(x_i|y) log(p_0^{x_i}) \\
& + \sum_{(i,j) \in E} \sum_{x_i} \sum_{x_j} p(x_i \neq x_j) log(a) + p(x_i = x_j) log(1 - a) \\
& + \sum_{i=1}^{2n+1} p(x_i = 0|y) log(N(y_i|\mu_0, \sigma_0^2)) + p(x_i = 1|y) log(N(y_i|\mu_1, \sigma_1^2))
\end{aligned}
\tag{25}
$$

2.3 M-step

As for the previous part, I take the derivative under constraints of the expected log likelihood and I deduce the following update parameters:

$$
\hat{a} = \frac{1}{2n} \sum_{(i,j) \in E} \sum_{x_i, x_j} p(x_i \neq x_j|y)
\tag{26}
$$

$$\hat{\mu_0} = \frac{\sum_{i=1}^{2n+1} p(x_i = 0|y) y_i}{\sum_{i=1}^{2n+1} p(x_i = 0|y)} \tag{27}$$

$$\hat{\mu_1} = \frac{\sum_{i=1}^{2n+1} p(x_i = 1|y) y_i}{\sum_{i=1}^{2n+1} p(x_i = 1|y)} \tag{28}$$

$$\hat{\sigma_0^2} = \frac{\sum_{i=1}^{2n+1} p(x_i = 0|y)(y_i - \hat{\mu_0})^2}{\sum_{i=1}^{2n+1} p(x_i = 0|y)} \tag{29}$$

$$\hat{\sigma_1^2} = \frac{\sum_{i=1}^{2n+1} p(x_i = 1|y)(y_i - \hat{\mu_1})^2}{\sum_{i=1}^{2n+1} p(x_i = 1|y)} \tag{30}$$

Where $2n + 1$ is the number of nodes, and as the matter of fact $2n$ is the number of edges in the tree.

2.4 Hidden States

To compute the maximum likelihood configuration for the hidden states, we have to substitute the sums by maximums in the beta recursion in order to store the most likely hidden children.