Big picture of EM for HMM

1. goal: maximize the likelihood $p(X|\theta) = \sum_Z p(X, Z|\theta)$

2. decompose: $p(X|\theta) = L(q, \theta) + KL(q\|p)$

$$L(q, \theta) = \sum_Z q(Z) \ln \frac{p(X, Z|\theta)}{q(Z)}$$

$$KL(q\|p) = -\sum_Z q(Z) \ln \frac{p(Z|X, \theta)}{q(Z)} \geq 0$$

E step: given θ^{old}, maximize the lower bound $L(q, \theta)$

by choosing the $q(Z)$ such that $KL(q\|p) = 0$

i.e. $q^*(Z) = \underset{q(Z)}{\arg\max}\, L(q, \theta) = \underset{q(Z)}{\arg\min}\, KL(q\|p)$

$= p(Z|X, \theta^{old})$

then $L(q^*, \theta) = \underbrace{\sum_Z p(Z|X, \theta^{old}) \ln p(X, Z|\theta)}_{} - \sum_Z q(Z) \ln q(Z)$

$Q(\theta, \theta^{old})$: expected complete log-likelihood.

M step: maximize the lower bound w.r.t. θ

$$\theta^{new} = \underset{\theta}{\arg\max}\, Q(\theta, \theta^{old})$$

use Lagrangian method as usual.

complete likelihood of HMM

$$p(X, Z|\theta) = p(z_1|\theta) \prod_{t=2}^{T} p(z_t|z_{t-1}, \theta) \prod_{t=1}^{T} p(x_t|z_t, \theta)$$

expected log of complete likelihood

$$Q(\theta, \theta^{old}) = \sum_{Z} p(Z|X, \theta^{old}) \ln p(X, Z|\theta)$$

$$= \sum_{Z} p(Z|X, \theta^{old}) \left[\ln p(z_1|\theta) + \sum_{t=2}^{T} \ln p(z_t|z_{t-1}, \theta) + \sum_{t=1}^{T} \ln p(x_t|z_t, \theta) \right]$$

$$= \sum_{z_1} \underbrace{p(z_1|X, \theta^{old})}_{\gamma_1(w_i) = p(z_1 = w_i|X, \theta^{old})} \ln \underbrace{p(z_1|\theta)}_{\pi_i = p(z_1 = w_i|\theta)}$$

$$+ \sum_{t=2}^{T} \sum_{z_{t-1}, z_t} \left[\underbrace{p(z_{t-1}, z_t|X, \theta^{old})}_{\xi_{t-1}(w_i, w_j) = p(z_{t-1} = w_i, z_t = w_j|X, \theta^{old})} \ln p(z_t|z_{t-1}, \theta) \right]$$

$$+ \sum_{t=1}^{T} \sum_{z_t} \left[\underbrace{p(z_t|X, \theta^{old})}_{\gamma_t(w_i) = p(z_t = w_i|X, \theta^{old})} \ln \underbrace{p(x_t|z_t, \theta)}_{b_{ik} = p(x_t = v_k|z_t = w_i, \theta)} \right]$$

$$= \sum_{i=1}^{c} \gamma_1(w_i) \ln \pi_i + \sum_{t=2}^{T} \sum_{i,j} \xi_{t-1}(w_i, w_j) \ln a_{ij} + \sum_{t=1}^{T} \sum_{i=1}^{c} \gamma_t(w_i) \ln b_{i x_t}$$

constraints :

$$\sum_{i} \pi_i = 1 \quad , \quad \sum_{j=1} a_{ij} = 1 \ , \ i = 1, \cdots, c \quad , \quad \sum_{k=1}^{K} b_{ik} = 1 \ , \ i = 1, \cdots, c$$

M step is to maximize $Q(\theta, \theta^{old})$, $\theta = [\pi_i, a_{ij}, b_{ik}]$

i.e.

$$\max_{\theta} \ Q(\theta, \theta^{old})$$

$$\text{s.t.} \quad \sum_{i=1}^{c} \pi_i = 1$$

$$\sum_{j=1}^{c} a_{ij} = 1 \ , \ i = 1, \cdots, c$$

$$\sum_{k=1}^{K} b_{ik} = 1 \ , \ i = 1, \cdots, c$$

define the Lagrangian

$$L(\theta, \lambda_1, \vec{\lambda}_2, \vec{\lambda}_3)$$

$$= Q(\theta, \theta^{old}) - \lambda_1\left[\sum_{i=1}^{c}\pi_i - 1\right] - \sum_{i=1}^{c}\lambda_{2i}\left[\sum_{j=1}^{c} a_{ij} - 1\right] - \sum_{i=1}^{c}\lambda_{3i}\left[\sum_{k=1}^{K} b_{ik} - 1\right]$$

$$= \sum_{i=1}^{c}\gamma_1(w_i)\ln\pi_i + \sum_{t=2}^{T}\sum_{i,j}\xi_{t-1}(w_i, w_j)\ln a_{ij}$$

$$+ \sum_{t=1}^{T}\sum_{i=1}^{c}\gamma_t(w_i)\ln b_{i x_t}$$

$$- \lambda_1\left[\sum_{i=1}^{c}\pi_i - 1\right] - \sum_{i=1}^{c}\lambda_{2i}\left[\sum_{j=1}^{c} a_{ij} - 1\right] - \sum_{i=1}^{c}\lambda_{3i}\left[\sum_{k=1}^{K} b_{ik} - 1\right]$$

compute $\hat{\pi}_i$:

$$\frac{\partial L}{\partial \pi_i} = \frac{\gamma_1(w_i)}{\pi_i} - \lambda_1 = 0 \quad \Rightarrow \pi_i = \frac{\gamma_1(w_i)}{\lambda_1}$$

$$\because \sum_{i=1}^{c}\pi_i = 1 \quad \therefore \sum_{i=1}^{c}\frac{\gamma_1(w_i)}{\lambda_1} = 1 \quad \Rightarrow \lambda_1 = \sum_{i=1}^{c}\gamma_1(w_i)$$

$$\therefore \hat{\pi}_i = \frac{\gamma_1(w_i)}{\sum_{j=1}^{c}\gamma_1(w_j)}$$

compute $\hat{a}_{ij}$:

$$\frac{\partial L}{\partial a_{ij}} = \sum_{t=2}^{T}\xi_{t-1}(w_i, w_j)\cdot\frac{1}{a_{ij}} - \lambda_{2i} = 0 \quad \Rightarrow a_{ij} = \frac{\sum_{t=2}^{T}\xi_{t-1}(w_i, w_j)}{\lambda_{2i}}$$

$$\because \sum_{j=1}^{c} a_{ij} = \frac{1}{\lambda_{2i}}\sum_{t=2}^{T}\sum_{j=1}^{c}\xi_{t-1}(w_i, w_j) = 1 \quad \Rightarrow \lambda_{2i} = \sum_{t=2}^{T}\sum_{j=1}^{c}\xi_{t-1}(w_i, w_j)$$

$$\therefore \hat{a}_{ij} = \frac{\sum_{t=2}^{T}\xi_{t-1}(w_i, w_j)}{\sum_{t=2}^{T}\sum_{j'=1}^{c}\xi_{t-1}(w_i, w_{j'})}$$

compute $\hat{b}_{ik}$:

$$\frac{\partial L}{\partial b_{ik}} = \sum_{x_t=v_k} \gamma_t(w_i) \cdot \frac{1}{b_{ik}} - \lambda_{3i} = 0 \quad \Rightarrow \quad b_{ik} = \frac{\sum_{x_t=v_k} \gamma_t(w_i)}{\lambda_{3i}}$$

$$\because \sum_{k=1}^{K} b_{ik} = 1 \qquad \cancel{\therefore \frac{1}{\lambda_{3i}}\sum_{k=1}^{K}} \therefore \frac{1}{\lambda_{3i}} \sum_{k=1}^{K} \sum_{x_t=v_k} \gamma_t(w_i) = 1 \quad \Rightarrow \lambda_{3i} = \sum_{t=1}^{T} \gamma_t(w_i)$$

$$\therefore \quad \hat{b}_{ik} = \frac{\sum_{x_t=v_k} \gamma_t(w_i)}{\sum_{t=1}^{T} \gamma_t(w_i)}$$