Problem set #4 ZIYAO YANG 9490544195

1 (1) The lagrangian of the problem is

$$L(\vec{q}, \lambda, \lambda_1 \cdots \lambda_K) = \sum_{k=1}^{K} a_k \ln q_k + \sum_{k=1}^{K} \lambda_k q_k + \lambda \left(\sum_{k=1}^{K} q_k - 1 \right)$$

$(\lambda_k \geq 0)$

stationary condition implies.

$$a_k \frac{1}{q_k^*} + \lambda_k + \lambda = 0, \quad \text{for each } k$$

$$\Rightarrow \quad q_k^* = -\frac{a_k}{\lambda_k + \lambda} \neq 0 \quad (\text{since } a_k > 0)$$

The complementary slackness condition implies.

$$\lambda_k q_k^* = 0 \Rightarrow \lambda_k = 0 \quad \text{for each } k$$

feasibility implies

$$\sum_{k=1}^{K} q_k^* = \sum_{k=1}^{K} -\frac{a_k}{\lambda} = 1 \quad \Rightarrow \quad \lambda = -\sum_{k=1}^{K} a_k$$

$$\text{so} \quad q_k^* = \frac{a_k}{\sum_{k=1}^{K} a_k}$$

(2) The lagrangian of the problem is

$$L(\vec{q}, \lambda, \lambda_1 \cdots \lambda_K) = \sum_{k=1}^{K} (q_k b_k - q_k \ln q_k) + \lambda \left(\sum_{k=1}^{K} q_k - 1\right) + \sum_{k=1}^{K} \lambda_k q_k$$

$(\lambda_k \geq 0)$

The stationary condition implies

$$b_k - \left(q_k \cdot \frac{1}{q_k} + \ln q_k\right) + \lambda + \lambda_k = 0$$

$$\Rightarrow q_k^* = \exp(b_k - 1 + \lambda + \lambda_k) > 0 \quad \text{for each } k$$

complementary condition implies $q_k^* \lambda_k = 0 \Rightarrow \lambda_k = 0$ for each k.

feasibility implies

$$\sum_{k=1}^{K} q_k = \sum_{k=1}^{K} \exp(b_k + \lambda - 1) = 1$$

$$\Rightarrow \exp(K\lambda) = \sum_{k=1}^{K} \exp(1 - b_k)$$

$$\Rightarrow \lambda = \frac{1}{K} \sum_{k=1}^{K} (1 - b_k)$$

$$q_k^* = \exp\left(b_k - 1 + \frac{1}{K} \sum_{k=1}^{K} (1 - b_k)\right) = \frac{\exp(b_k)}{\sum_{k=1}^{K} \exp(b_k)}$$

2.1 To find w^*, just solve

$$\arg\max \sum_n \sum_k \gamma_{nk} \ln w_k.$$
$$\text{s.t. } w_k \geq 0$$
$$\sum_{k=1}^{K} w_k = 1$$

from the Problem 1-1. $\alpha_k = \sum_n \gamma_{nk}$

we could get $w^* = \dfrac{\sum_n \gamma_{nk}}{\sum_k \sum_n \gamma_{nk}} = \dfrac{\sum_n \gamma_{nk}}{\sum_n 1} = \dfrac{\sum_n \gamma_{nk}}{N}$

To find μ_k and σ_k, we solve $(\Sigma_k = \sigma_k^2 I)$

$$\arg\max_{\mu_k, \Sigma_k} \sum_n \gamma_{nk} \ln\left[\frac{1}{(\sqrt{2\pi}\sigma_k)^D} \exp\left(-\frac{1}{2\sigma_k^2} \|x_n - \mu_k\|^2\right)\right]$$

$$= \arg\max_{\mu_k, \Sigma_k} \sum_n \gamma_{nk} \left(-D \ln \sigma_k - \frac{\|x_n - \mu_k\|^2}{2\sigma_k^2}\right) = \arg\max_{\mu_k, \Sigma_k} F$$

$$\frac{\partial F}{\partial \mu_k} = \frac{1}{\sigma_k^2} \sum_n \gamma_{nk} (x_n - \mu_k) = 0$$

$$\Rightarrow \mu_k^* = \frac{\sum \gamma_{nk} \cdot x_n}{\sum_n \gamma_{nk}}.$$

$$\frac{\partial F}{\partial \sigma_k} = \sum_n \gamma_{nk} \left(-\frac{D}{\sigma_k} + \frac{\|x_n - \mu_k\|^2}{\sigma_k^3}\right) = 0$$

$$\Rightarrow (\sigma_k^*)^2 = \frac{\sum_n \gamma_{nk} \|x_n - \mu_k^*\|^2}{D \sum_n \gamma_{nk}}$$

2.2

According to 1.2 with $b_k = \ln p(\vec{x_n}, z_n = k; \theta^{(t)})$.

$$q_{nk}^* = \frac{P(\vec{x_n}, z_n = k; \theta^{(t)})}{\sum_{k=1}^{K} P(\vec{x_n}, z_n = k; \theta^{(t)})}$$

$$= \frac{P(\vec{x_n}, z_n = k; \theta^{(t)})}{P(\vec{x_n}; \theta^{(t)})} = P(z_n = k; \vec{x_n}, \theta^{(t)})$$

2.3 when $w_k = \frac{1}{K}$ for each k. and $\Sigma = 0$,

GMM could be reduced to k-mean.

$$\gamma_{nk} = P(z_n = k \mid x_n) = \frac{P(z_n = k, x_n)}{P(x_n)}$$

$$= \frac{P(z_n = k) \cdot N(x_n \mid \mu_n, \Sigma_k)}{\sum_{j=1}^{k} P(z_n = j)\, N(x_n \mid \mu_j, \Sigma_j)}$$

$$= \frac{w_k \exp(-\frac{1}{2\sigma^2} \| x_n - \mu_k \|^2)}{\sum_{j=1}^{K} w_j \exp(-\frac{1}{2\sigma^2} \| x_n - \mu_j \|^2)}$$

As $\sigma^2 \to 0$, the denominator will be denommated by the term with the smallese $\| x_n - \mu_j \|^2$.

$$\gamma_{nj} \approx \frac{w_j \exp\{-\frac{1}{2\sigma^2} \| x_n - \mu_j \|^2\}}{w_j \exp\{-\frac{1}{2\sigma^2} \| x_n - \mu_j \|^2\}} = 1$$

for $l \neq j$, $\| x_n - \mu_l \|^2 > \| x_n - \mu_j \|^2$.

$\gamma_{nj} \approx 0$.

Thus, for $\Sigma_k = \sigma^2 I$ and $\sigma^2 \to 0$, GMM reduces to KMM

3.1 $P(Z_{T+1}=s \mid X_{1:T}=x_{1:T})$

$$= \frac{P(Z_{T+1}=s,\ X_{1:T}=x_{1:T})}{P(X_{1:T}=x_{1:T})}$$

$$= \frac{\sum_{s'} P(Z_{T+1}=s,\ Z_T=s',\ X_{1:T}=x_{1:T})}{\sum_s P(Z_T=s,\ X_{1:T}=x_{1:T})}$$

$$= \frac{\sum_{s'} P(Z_T=s',\ X_{1:T}=x_{1:T})\, P(Z_{T+1}=s \mid Z_T=s',\ X_{1:T}=x_{1:T})}{\sum_s P(Z_T=s,\ X_{1:T}=x_{1:T})}$$

$$= \frac{\sum_{s'} P(Z_T=s',\ X_{1:T}=x_{1:T})\, P(Z_{T+1}=s \mid Z_T=s')}{\sum_s P(Z_T=s,\ X_{1:T}=x_{1:T})} \quad \text{(Markov Property)}$$

$$= \frac{\sum_{s'} \alpha_{s'}(T)\, a_{s',s}}{\sum_s \alpha_s(T)}$$

3.2 $t=1$, $\alpha_A(1) = 0.7 \times 0.4 = 0.28$

$\alpha_B(1) = 0.3 \times 0.7 = 0.21$

$t=2$ $\alpha_A(2) = 0.6 \times (0.7 \times 0.28 + 0.2 \times 0.28) = 0.1218$

$\alpha_B(2) = 0.3 \times (0.3 \times 0.21 + 0.7 \times 0.28) = 0.0714$

$$P(Z_3=A \mid X_{1:2}=x_{1:2}) = \frac{\sum_{s'} \alpha_{s'}(2)\, a_{s',A}}{\sum_s \alpha_s(2)}$$

$$= \frac{\alpha_A(2)\, a_{A,A} + \alpha_B(2)\, a_{B,A}}{\alpha_A(2) + \alpha_B(2)}$$

$$\approx 0.3848$$

Similarly, $P(Z_3=B \mid X_{1:2}=x_{1:2}) \approx 0.5152$

$P(Z_3=\text{end} \mid X_{1:2}=x_{1:2}) = 0.1$

Therefore $t=3$, the observed sequcece should be B

7.2 (b) According to Viterbi

$t=1$ $\quad \delta_A(1) = 0.7 \times 0.4 = 0.28$

$\delta_B(1) = 0.3 \times 0.7 = 0.21$

$t=2$ $\quad \delta_A(2) = 0.6 \times \max\{0.7 \times 0.21,\ 0.2 \times 0.28\} = 0.0882$

$\Delta_A(2) = B$

$\delta_B(2) = 0.3 \times \max\{0.2 \times 0.21,\ 0.7 \times 0.28\} = 0.0588$

$\Delta_B(2) = A$

$t=3$ $\quad \delta_A(3) = 0.6 \times \max\{0.7 \times 0.0588,\ 0.2 \times 0.0882\}$

$= 0.024696$

$\Delta_A(3) = B$

$\delta_B(3) = 0.3 \times \max\{0.2 \times 0.0588,\ 0.7 \times 0.0882\}$

$= 0.018522$

$\Delta_B(3) = A$

Then back tracking

$z_3^* = A,\quad z_2^* = B,\quad z_1^* = A$

3.3 $P(X_2 = s' \mid O_1 = o_1, O_2 = o_2)$

$$= \frac{\sum_s P(X_1 = s, X_2 = s', O_1 = o_1, O_2 = o_2)}{P(O_{1:2})}$$

$$= \frac{\sum_s P(O_1 = o_1, O_2 = o_2 \mid X_1 = s, X_2 = s') \cdot P(X_1 = s, X_2 = s')}{P(O_1 = o_1, O_2 = o_2)}$$

$$= \frac{\sum_s P(O_1 = o_1 \mid X_1 = s) \cdot P(O_2 = o_2 \mid X_2 = s') \cdot P(X_2 = s' \mid X_1 = s) \cdot P(X_1 = s)}{P(O_1 = o_1, O_2 = o_2)}$$

We set $P(X_{t+1} \mid X_t) = P(X_{t+1})$.

$$= \frac{\sum_s P(O_1 = o_1 \mid X_1 = s) \cdot P(X_1 = s) \cdot P(O_2 = o_2 \mid X_2 = s') \cdot P(X_2 = s')}{P(O_1 = o_1, O_2 = o_2)}$$

$$= \frac{P(O_2 = o_2, X_2 = s') \cdot \sum_s P(O_1 = o_1, X_1 = s)}{P(O_1 = o_1, O_2 = o_2)}$$

$$= P(O_2 = o_2, X_2 = s') \cdot \frac{P(O_1 = o_1)}{P(O_1 = o_1, O_2 = o_2)}$$

$$P(O_1 = o_1) = \sum_s \alpha_s(1)$$

$$P(O_1 = o_1, O_2 = o_2) = \sum_{s'} \alpha_{s'}(2)$$

$$\alpha_s(1) = \pi_s \cdot b_{s,o_1}$$

$$\alpha_{s'}(2) = b_{s',o_2} \sum_s a_{s,s'} \alpha_s(1)$$

$$= b_{s',o_2} \cdot P(X_2 = s') \cdot \sum_s \alpha_s(1) = b_{s',o_2} \cdot P(X_2 = s') \cdot P(O_1 = o_1)$$

So $P(O_1 = o_1, O_2 = o_2) = \sum_{s'} b_{s'o_2} \cdot P(X_2 = s') \cdot P(O_1 = o_1)$

$$P(X_2=s' \mid O_1=o_1, O_2=o_2)$$

$$= P(O_2=o_2, X_2=s') \cdot \frac{P(O_1=o_1)}{P(O_1=o_1, O_2=o_2)}$$

$$= \frac{P(O_2=o_2, X_2=s')}{\sum_{s'} b_{s'o_2} \cdot P(X_2=s')}$$

$$= \frac{P(O_2=o_2, X_2=s')}{\sum_{s'} P(O_2=o_2 \mid X_2=s') P(X_2=s')}$$

$$= \frac{P(O_2=o_2, X_2=s')}{P(O_2=o_2)} = P(X_2=s' \mid O_2=o_2)$$

from problem 2-2

we know $q^*_{n=2, k=s'} = P(z_2=s' \mid \vec{x_2}, \theta^{(t)})$.

$$\frac{\text{replace } z \text{ with } X}{\text{replace } X \text{ with } O} \qquad P(X_2=s' \mid O_2)$$

Therefore, when $P(X_1, X_2) = P(X_1)P(X_2)$, or each state are independent, HMM reduce to GMM.

$P(X_2=s' \mid O_1, O_2)$ match $q^*_{n=2, k=s'}$.