# 1) HMM with tied mixtures:-

## (1) Parameters of HMM model:

A typical HMM model has 3-types of parameters:

a) Start probability : $P(X_1)$

b) Transition probability : $P(X_{t+1}|X_t)$

c) Emission probability : $P(O_t|X_t)$

For our case we have tied-mixture of HMM so we can write the parameters as:-

a) Start probability :-

Lets consider start probability denoted as: $S_j$ where $j \in (1 \ldots M)$ then;

$$\boxed{\begin{array}{l} \text{(i)} \quad S_j = P(X_1 = j) \\ \text{(ii)} \quad \sum_{j=1}^{M} S_j = 1 \end{array}}$$

b) Transition probability :-

Lets consider transition probability denoted as: $T_{ij}$ ; $i, j \in \{1 \ldots M\}$ then; (P.T.O)

(i) $T_{ij} = P(X_{t+1} = j \mid X_t = i)$

(ii) $\sum_{j=1}^{M} T_{ij} = 1$

c) Emission Probability :-

Lets consider, mixing coefficients = $w_{jk}$

mean = $\mu_k$

covariances = $\Sigma_k$

for $j \in \{1 \cdots M\}$ and $k \in \{1 \cdots K\}$

then,

(i) $P(O_t \mid X_t = j) = \sum_{k=1}^{K} w_{jk} \mathcal{N}(O_t \mid \mu_k, \Sigma_k)$

(ii) $\sum_{k=1}^{K} w_{jk} = 1$

(Answer)

(P.T.O)

1)

(2) E-step derivation and parameter estimation:-

Forward Algo: Let $a_t$ = probability of seeing observation $O_t$

$$a_1^{jk} = P(O_1, X_1 = j, Z_1 = k)$$

$$\Rightarrow a_1^{jk} = P(O_1 | Z_1 = k)\, P(Z_1 = k | X_1 = j)\, P(X_1 = j)$$

$$\Rightarrow a_1^{jk} = N(O_1 | \mu_k, \Sigma_k)\, w_{jk}\, S_j$$ [by using emission and start probability]

——— (1)

$$A_1^j = P(O_1, X_1 = j)$$

$$\Rightarrow A_1^j = \sum_{k=1}^{K} P(O_1, X_1 = j, Z_1 = k)$$

$$\Rightarrow A_1^j = \sum_{k=1}^{K} a_1^{jk}$$ ——— (2)

(P.T.O)

Generalized form,

$$\alpha_t^{jk} = P(O_{1:t}, X_t = j, Z_t = k)$$

$$\Rightarrow \alpha_t^{jk} = \sum_{i=1}^{m} P(O_{1:t-1}, X_{t-1} = i, O_t, X_t = j, Z_t = k)$$

$$\Rightarrow \alpha_t^{jk} = \sum_{i=1}^{m} P(O_{1:t-1}, X_{t-1} = i)\, P(O_t, X_t = j, Z_t = K \mid O_{1:t-1}, X_{t-1} = i)$$

[by using chain rule]

$$\Rightarrow \alpha_t^{jk} = P(O_t \mid Z_t = k)\, P(Z_t = K \mid X_t = j) \sum_{i=1}^{m} P(O_{1:t-1}, X_{t-1} = i)\, P(X_t = j \mid X_{t-1} = i)$$

$$\Rightarrow \alpha_t^{jk} = N(O_t \mid \mu_k, \Sigma_k)\, \omega_{jk} \sum_{i=1}^{m} A_{t-1}^{i}\, T_{ij}$$

[by using transition and emission probability]

——— (3)

(P.T.O)

Again,

$$A_t^j = P(O_{1:t}, X_t = j)$$

$$\Rightarrow A_t^j = \sum_{k=1}^{K} P(O_{1:t}, X_t = j, Z_t = k)$$

$$\Rightarrow A_t^j = \sum_{k=1}^{K} a_t^{jk}$$ [using previous equation]

$$\text{———} \quad (4)$$

**Backward Algo:-**

Let $B_t$ be the probability of the ending partial sequence $O_{t+1} \cdots O_T$ then we can calculate:

$$B_T^j = 1 \quad \text{———} \quad (5)$$

$$B_t^j = P(O_{t+1:T} \mid X_t = j)$$

$$\Rightarrow B_t^j = \sum_{i=1}^{m} \sum_{k=1}^{K} P(O_{t+1:T}, X_{t+1} = i, Z_{t+1} = k \mid X_t = j).$$

$$\Rightarrow B_t^j = \sum_{j=1}^{m} T_{ij} B_{t+1}^i \sum_{k=1}^{K} w_{ik} N(O_{t+1} \mid \mu_k, \Sigma_k)$$

[by using transition and emission probabilities]

$$\text{———} \quad (6)$$

(P.T.O)

Then finally we need to update the following probability distributions:-

(i) $\gamma = P(O_{1:T}) = \sum_{i=1}^{M} A_t^i B_t^i, \ \forall t$

(ii) $\nu_t^{jk} = P(X_t = j, Z_t = k \mid O_{1:T})$

$\Rightarrow \nu_t^{jk} = \dfrac{a_t^{jk} B_t^j}{P(O_{1:T})}$ [by using equation 3 and equation 6]

$\Rightarrow \nu_t^{jk} = \dfrac{a_t^{jk} B_t^j}{\gamma}$

(iii) $\sigma_t^j = P(X_t = j \mid O_{1:T})$

$\Rightarrow \sigma_t^j = \sum_{k=1}^{K} \nu_t^{jk}$

$\Rightarrow \sigma_t^j = \dfrac{A_t^j B_t^j}{\gamma}$ [by using equation 4 and 6]

(iv) $\phi_t^k = P(Z_t = k \mid O_{1:T})$

$= \sum_{j=1}^{M} \nu_t^{jk}$

(P.T.O)

$$\phi_t^{k} = \frac{\sum_{j=1}^{M} \alpha_t^{jk} B_t^{j}}{\gamma}$$ [by using equation (3) and (6)] (7)

and

(v) $\xi_t^{ij} = P(X_t = i, X_{t+1} = j \mid O_{1:T})$

$$\Rightarrow \xi_t^{ij} = \frac{A_t^{i} T_{ij} B_{t+1}^{j} \sum_{k=1}^{K} w_{jk} \mathcal{N}(O_{t+1} \mid \mu_k, \Sigma_k)}{\gamma}$$

[by using equation (4), (6) and transition and emission probabilities]

(Answer)

(P.T.O)

1)

③ M-step derivation :-

In M-step, we update the model's parameter $\theta$ which means updating the below parameters:

(i) start probability:

$$S_j = \sigma_1^j$$

(ii) Transition probability:

$$T_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t^{ij}}{\sum_{t=1}^{T-1} \sigma_t^i}$$

(iii) Mixing coefficients :-

$$w_{jk} = \frac{\sum_{t=1}^{T-1} \gamma_t^{jk}}{\sum_{t=1}^{T-1} \sigma_t^i}$$

(iv) mean :-

$$\mu_k = \frac{\sum_{t=1}^{T} \phi_t^k o_t}{\sum_{t=1}^{T} \phi_t^k}$$

(v) covariances :-

$$\Sigma_k = \frac{\sum_{t=1}^{T} \phi_t^k (o_t - \mu_k)(o_t - \mu_k)^T}{\sum_{t=1}^{T} \phi_t^k}$$

(Answer)