LECTURE 10

Let us consider a HMM:

$\Sigma \triangleq (\pi_0, \hat{P}, \hat{Q})$ with: $\pi_0 \triangleq [P(X_0=e_1) \; P(X_0=e_2) \dots P(X_0=e_N)]^T$ - $N\times 1$ vector of initial condition probabilities

They do not depend on k

$\hat{P} \triangleq \{P(X_k=e_i | X_{k-1}=e_j)\}_{ij}$ - $N\times N$ transition probability matrix

$\hat{Q} \triangleq \{P(Y_k=q_i | X_k=e_j)\}_{ij}$ - $N\times L$ output probability matrix

* Evaluation problem

Σ is known

$(y_1, y_2, \dots, y_n)$ - a sequence of output values is collected

$\Rightarrow$ What is the conditional probability $P(y_1, y_2, \dots, y_n | \Sigma)$?

If we knew the actual sequence of states $(x_1, x_2, \dots, x_n)$ corresponding to the given observations, we would write:

$$P(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n, \Sigma) = \prod_{k=1}^{n} \underbrace{P(y_k | x_k)}_{\text{a specific entry in } \hat{Q}}$$

On the other hand, a generic sequence of states $(x_1, x_2, \dots, x_n)$ may occur with the following probability:

$$P(x_1, x_2, \dots, x_n | \Sigma) = \underbrace{P(X_1 = x_1)}_{\text{a specific entry in } \pi_0} \prod_{k=2}^{n} \underbrace{P(X_k = x_k | X_{k-1} = x_{k-1})}_{\text{a specific entry in } \hat{P}} \quad (*)$$

(*) we assume that the initial condition occurs at $k=1$ instead of $k=0$ here

Therefore, the joint probability of $(y_1, y_2, \dots, y_n)$ and $(x_1, x_2, \dots, x_n)$ would be:

$$P(y_1, y_2, \dots, y_n, x_1, x_2, \dots, x_n | \Sigma) = P(X_1 = x_1) P(y_1 | x_1) \prod_{k=2}^{n} P(y_k | x_k) P(X_k = x_k | X_{k-1} = x_{k-1})$$

②

We know that the exact value of $P(y_1, y_2, \dots, y_n | \Sigma)$ would be:

$$P(y_1, y_2, \dots, y_n | \Sigma) = \sum_{\substack{\text{all} \\ (x_1,\dots,x_n) \in S^n}} P(y_1, y_2, \dots, y_n, x_1, x_2, \dots, x_n | \Sigma) \qquad (1)$$

However, the formula (1) would be unfeasible even for small values of n:

$$\left.\begin{array}{l} \text{No of combinations } (x_1, x_2, \dots, x_n) = N^n \\ \text{No of operations in } (1) = 2n \cdot N^n \end{array}\right\} \Rightarrow \begin{array}{l} \text{If } N=2 \text{ and } n=100, \text{ we would have:} \\ 200 \cdot 2^{100} \cong 2.5 \cdot 10^{32} \text{ operations!} \end{array}$$

A solution to this problem is given by the Forward-Backward Procedure:

$\alpha_k(i) \triangleq P(y_1, y_2, \dots, y_k, X_k = e_i | \Sigma)$ - Probability of partial observation sequence up to $k < n$ with state $X_k = e_i$ at time k

Note that $\alpha_k(i)$ satisfies the conditions:

Forward Procedure

a) $\alpha_1(i) = \pi_{0,i} \cdot P(Y_1 = y_1 | X_1 = e_i) = \pi_{0,i} \cdot q_{i,y_1} \qquad 1 \le i \le N$

b) $\alpha_{k+1}(i) = \sum_{j=1}^{N} P(X_{k+1} = e_i | X_k = e_j)\, \alpha_k(j)\, P(Y_{k+1} = y_{k+1} | X_{k+1} = e_i)$

$\qquad = \left[\sum_{j=1}^{N} p_{ji}\alpha_k(j)\right] q_{i,y_{k+1}} \qquad k < n, \quad 1 \le i \le N$

c) $P(y_1, y_2, \dots, y_n | \Sigma) = \sum_{i=1}^{N} \alpha_n(i)$

By using formula a)-c) one is able to estimate $P(y_1, y_2, \dots, y_n | \Sigma)$ with only $n \cdot N^2$ operations. For instance, $N=2$ and $n=100 \Rightarrow 4 \cdot 100 = 400$ operations!

The key idea is that, regardless of the specific sequence of states from step 1 to step k, the state at step $k+1$ only depends on the state at step k, and this latter state can only have one of N values $\Rightarrow$ The number of operations at step

b) is only N^2 $\forall k$. Analogously, one can define a recursive procedure that moves backward and calculates the probability $P(y_1, y_2, \dots, y_n | \Sigma)$ through the following steps:

$\beta_k(i) \triangleq P(y_{k+1}, y_{k+2}, \dots, y_n | X_k = e_i, \Sigma)$ - Probability of partial observation sequence from $k+1$ to n conditioned on the preceding state being $X_k = e_i$

Backward Procedure

a') $\beta_n(i) = 1$ $\quad 1 \leq i \leq N$ - conventionally, since we do not have further observations after step n regardless of X_n

b')
$$\beta_k(i) = \sum_{j=1}^{N} P(y_{k+1}, y_{k+2}, \dots, y_n, X_{k+1} = e_j | X_k = e_i, \Sigma) =$$
$$= \sum_{j=1}^{N} P(y_{k+1} | X_{k+1} = e_j, \Sigma)\, P(y_{k+2}, \dots, y_n | X_{k+1} = e_j, \Sigma)\, P(X_{k+1} = e_j | X_k = e_i)$$
$$= \sum_{j=1}^{N} q_{j, y_{k+1}} \cdot \beta_{k+1}(j) \cdot p_{ij} \qquad k < n, \; 1 \leq i \leq n$$

c')
$$P(y_1, y_2, \dots, y_n | \Sigma) = \sum_{j=1}^{N} P(y_1 | X_1 = e_j, \Sigma)\, \beta_1(j) = \sum_{j=1}^{N} q_{j, y_1} \cdot \beta_1(j)$$

Note that, by using the backward formula a')–c') one is able to estimate $P(y_1, y_2, \dots, y_n | \Sigma)$ with only $n \cdot N^2$ operations, similarly to the case of forward formula a)–c) □

* Evaluation Problem

Σ is known
$(y_1, y_2, \dots, y_n)$ - a sequence of output values is collected
$\Rightarrow$ What is the sequence of states $(x_1^*, x_2^*, \dots, x_n^*)$ that best explains the sequence of output values?

The solution to this problem actually depends on the definition of the optimality criterion (i.e., in what sense is $(x_1^*, x_2^*, \ldots, x_n^*)$ the "best"?). For instance, one may search for the sequence $(x_1^*, x_2^*, \ldots, x_n^*)$ such that:

$$x_k^* = \arg\max_{e \in S} P\left(X_k = e \mid (y_1, y_2, \ldots, y_n), \Sigma\right)$$

$S \triangleq \{e_1, e_2, \ldots, e_N\}$, i.e., each state in the sequence is optimal in the sense that it is the most likely at its time k, given the entire sequence of observations.

In this case, we can solve for x_k^* $1 \leq k \leq n$ by introducing:

$$\gamma_k(i) \triangleq P\left(X_k = e_i \mid (y_1, y_2, \ldots, y_n), \Sigma\right)$$

$$\Downarrow$$

$$\gamma_k(i) = \frac{P(y_1, y_2, \ldots, y_n, X_k = e_i \mid \Sigma)}{P(y_1, y_2, \ldots, y_n \mid \Sigma)} =$$

$$= \frac{P(y_{k+1}, y_{k+2}, \ldots, y_n \mid X_k = e_i, \Sigma)\, P(y_1, y_2, \ldots, y_k, X_k = e_i \mid \Sigma)}{\sum_{j=1}^{N} P(y_{k+1}, \ldots, y_n \mid X_k = e_j, \Sigma)\, P(y_1, y_2, \ldots, y_k, X_k = e_j \mid \Sigma)}$$

$$= \frac{\beta_k(i)\,\alpha_k(i)}{\sum_{j=1}^{N} \beta_k(j)\,\alpha_k(j)} \qquad 1 \leq i \leq N \qquad 1 \leq k \leq n$$

and $\sum_{i=1}^{N} \gamma_k(i) = 1 \quad \forall k$

Therefore, the solution x_k^* can be obtained as:

$$x_k^* = \arg\max_{1 \leq i \leq N} \gamma_k(i) \qquad (**)$$

Problem (**) can be solved by (1) applying the forward and backward procedures to calculate all the needed values $\alpha_k(i)$, $\beta_k(i)$, respectively; (2) computing $\gamma_k(i)$ for $1 \leq i \leq N$; and (3) picking the value i at each k that maximize $\gamma_k(\cdot)$

However, this solution does NOT guarantee that two consecutive optimal states x_k^*

and x_{k+1}^* can actually be sequentially reached, i.e., what if $P(X_{k+1}=x_{k+1}^* \mid X_k = x_k^*)=0$? The solution would be unfeasible $\Rightarrow$ This happens because, in maximizing $\gamma_k(i)$, we do not take into consideration the state chosen at step k-1 (if we move forward) or the state chosen at step k+1 (if we move backward)

An alternative solution is obtained if we search for a single best sequence (a.k.a., a "path") instead of a sequence of individually-chosen best states $\Rightarrow$ We search for:

$$(x_1^*, x_2^*, \dots, x_n^*) = \arg\max_{(x_1, x_2, \dots, x_n) \in S^n} P(y_1, y_2, \dots, y_n, x_1, \dots, x_n \mid \Sigma)$$

$S \triangleq \{e_1, e_2, \dots, e_N\}$ - set of state values. The problem is solved by using the VITERBI algorithm:

$$\delta_k(i) \triangleq \max_{(x_1, \dots, x_{k-1}) \in S^{k-1}} P(x_1, x_2, \dots, x_{k-1}, X_k = e_i, y_1, y_2, \dots, y_k \mid \Sigma)$$

Best probability along a single path including the first k steps and terminating in the i-th state

The function $\delta_k(i)$ satisfies the following conditions:

$$\delta_{k+1}(i) = \max P(x_1, \dots, x_k, X_{k+1} = e_i, y_1, \dots, y_{k+1} \mid \Sigma) =$$

$$= \max_{x_k} \left\{ \max_{(x_1, \dots, x_{k-1})} P(x_1, \dots, x_{k-1}, X_k = x_k, X_{k+1} = e_i, y_1, \dots, y_{k+1} \mid \Sigma) \right\}$$

$$= \max_{x_k} \left\{ \max_{(x_1, \dots, x_{k-1})} P(x_1, \dots, x_{k-1}, X_k = x_k, y_1, \dots, y_k \mid X_{k+1} = e_i, y_{k+1}, \Sigma) \cdot P(y_{k+1} \mid X_{k+1} = e_i, \Sigma) P(X_{k+1} = e_i \mid X_k = x_k, \Sigma) \right\}$$

Probability of a path does not depend on the future

$$= \max_{x_k} \left\{ \delta_k(x_k) P_{x_k i} \right\} q_{i, y_{k+1}}$$

Hence, given that $x_k \in S$, we can write:

$$\delta_{k+1}(i) = \max_{1 \le j \le N} \{ \delta_k(j)\, p_{ji} \}\, q_{i, y_{k+1}}$$

Let us define: $\psi_{k+1}(i) \triangleq \arg\max_{1 \le j \le N} \{ \delta_k(j)\, p_{ji} \}$ ← It is the best state to move from at stage k if we want to reach e_i at stage k+1

We can now formulate the algorithm:

V1) $\psi_1(i) = 0$ - conventionally $\quad 1 \le i \le N$] Initialization
$\delta_1(i) = P(X_1 = e_i, y_1 | \Sigma) = \pi_{0,i}\, q_{i, y_1}$

V2) $\psi_{k+1}(i) = \arg\max_{1 \le j \le N} \{ \delta_k(j)\, p_{ji} \}$ $\quad 1 \le i \le N,\ k < n$] Recursion
$\delta_{k+1}(i) = \delta_k(\psi_{k+1}(i))\, p_{\psi_{k+1}(i), i}\, q_{i, y_{k+1}}$

V3) $x_n^* = \{ e_i : i = \arg\max_{1 \le j \le N} \delta_n(j) \}$] Termination
$p^* \triangleq \max_{1 \le j \le N} \delta_n(j)$
↑ Probability of the optimal path

By the time step V3 is completed, we have that, for each possible value of X_k, the best previous state has been determined. Moreover, the optimal final state x_n^* has been calculated ⇒ We can move backward and reconstruct the optimal path:

V4) $i^* \triangleq$ index of the state value e_{i^*} : $x_{k+1}^* = e_{i^*}$

$x_k^* = \psi_{k+1}(i^*) \qquad 1 \le k < n$

□

* Estimation Problem

$(y_1, y_2, \dots, y_n)$ - known $\Rightarrow$ $\begin{cases} \text{What is the HMM } \Sigma^* = (\pi_0^*, \hat{P}^*, \hat{Q}^*) \text{ that} \\ \text{best explains the sequence of output values?} \end{cases}$

In general, an exact solution to the problem does not exist. However, a good approximation can be obtained:

$$\xi_k(i,j) \triangleq P(X_k = e_i, X_{k+1} = e_j \mid y_1, y_2, \dots, y_n, \Sigma)$$

$$\Downarrow$$

$$\xi_k(i,j) = \frac{P(y_1, y_2, \dots, y_k, X_k = e_i \mid \Sigma) P(y_{k+1} \mid X_{k+1} = e_j) P(y_{k+2}, \dots, y_n \mid X_{k+1} = e_j, \Sigma) P(X_{k+1} = e_j \mid X_k = e_i)}{P(y_1, y_2, \dots, y_n \mid \Sigma)}$$

$$= \frac{\alpha_k(i) q_{j, y_{k+1}} \beta_{k+1}(j) \cdot P_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_k(i) q_{j, y_{k+1}} \beta_{k+1}(j) \cdot P_{ij}}$$

Note that: $\gamma_k(i) = \sum_{j=1}^{N} \xi_k(i,j)$ - Probability of having $X_k = e_i$ given the sequence of observations and the model

$\sum_{k=1}^{n-1} \gamma_k(i) \triangleq$ Expected number of transitions from e_i

$\sum_{k=1}^{n-1} \xi_k(i,j) \triangleq$ Expected number of transitions from e_i to e_j

Therefore, given $n > 1$ observations, we can provide the estimation:

(***)

- $\bar{\pi}_{0,i} = \gamma_1(i)$ $\quad 1 \leq i \leq N$ - Estimation of $\pi_{0,i}$ $\forall i$
- $\bar{P}_{ij} = \dfrac{\sum_{k=1}^{n-1} \xi_k(i,j)}{\sum_{k=1}^{n-1} \gamma_k(i)}$ $\quad 1 \leq i,j \leq N$ - Estimation of the (i,j)-th element of the transition probability matrix $\hat{P}$ $\forall i,j$

8

(***) • Denoted with $V \triangleq \{v_1, v_2, \dots, v_L\}$ the alphabet of the HMM, we have:

$$\bar{q}_{ij} = \frac{\sum_{k=1}^{n} P(X_k = e_i \mid Y_k = v_j)}{\sum_{k=1}^{n} \gamma_k(i)}$$ - Estimation of the (i,j)-th element of the matrix $\hat{Q}$ $\forall i, j$

$1 \leq i \leq N, \quad 1 \leq j \leq L$

Note that, since we have only one sequence $(y_1, y_2, \dots, y_n)$, we have:

$$P(X_k = e_i \mid Y_k = v_j) = P(X_k = e_i \mid y_1, y_2, \dots, y_{k-1}, y_k = v_j, y_{k+1}, \dots, y_n, \Sigma) =$$

$$= \begin{cases} \gamma_k(i) & \text{if } y_k = v_j \\ 0 & \text{otherwise} \end{cases}$$

Hence, we can write: $$\bar{q}_{ij} = \frac{\sum_{k \in N_j} \gamma_k(i)}{\sum_{k=1}^{n} \gamma_k(i)} \qquad 1 \leq i \leq N, \; 1 \leq j \leq L$$

where $N_j \triangleq \{ 1 \leq k \leq n : \; y_k = v_j \}$

Note, though, that the estimation $\bar{\Sigma} \triangleq (\bar{\pi}_0, \bar{P}, \bar{Q})$, where $\bar{\pi}_0 \triangleq [\bar{\pi}_{01} \dots \bar{\pi}_{0N}]^T$, $\bar{P} \triangleq \{\bar{p}_{ij}\}$ $1 \leq i, j \leq N$, $\bar{Q} \triangleq \{\bar{q}_{ij}\}$ $1 \leq i \leq N$, $1 \leq j \leq L$, is obtained by using the functions $\gamma_k(i)$, which are defined based on some other model $\Sigma \triangleq (\pi_0, \hat{P}, \hat{Q}) \Rightarrow$ We have done a re-estimation of the model.

However, it can be proved that:

- Given Σ, the re-estimation $\bar{\Sigma}$ maximizes the function:

$$\eta(\Sigma, \Gamma) \triangleq \sum_{(x_1, \dots, x_n) \in S^n} P(x_1, x_2, \dots, x_n \mid y_1, y_2, \dots, y_n, \Sigma) \cdot \log\left(P(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n \mid \Gamma)\right)$$

i.e., $\bar{\Sigma} = \arg\max_{\Gamma} \eta(\Sigma, \Gamma)$

- Denoted with: $\mathcal{L}(\Gamma) \triangleq P(y_1, y_2, \ldots, y_n | \Gamma)$ - likelihood of the sequence of observations given the model Γ, $\bar{\Sigma}$ satisfies:

$$\mathcal{L}(\bar{\Sigma}) \geq \mathcal{L}(\Sigma)$$

Therefore, we can define an iterative procedure with the following steps:

BW1) Initialization: $\Sigma = \Sigma_0$ with Σ_0 - tentative model to be chosen

BW2) Estimation: Solve $\bar{\Sigma} = \arg\max_{\Gamma} \eta(\Sigma, \Gamma)$ by using (***)

BW3) If $\mathcal{L}(\bar{\Sigma}) > \mathcal{L}(\Sigma)$, then set $\Sigma = \bar{\Sigma}$ and repeat BW2)
If $\mathcal{L}(\bar{\Sigma}) = \mathcal{L}(\Sigma)$, stop $\Rightarrow \Sigma$ is the MLE of the model we are seeking

Baum-Welch (BW) Algorithm.

Note that $\eta(\Sigma, \Gamma)$, in general, may have many local maxima and the estimation (***) may lead to a local maximum $\Rightarrow$ The BW algorithm provides an approximated solution to the original estimation problem $\Rightarrow$ The choice Σ_0 becomes relevant $\square$

* More Sophisticated Variants of HMM

So far, we have considered HMMs in the form $\Sigma = (\pi_0, \hat{P}, \hat{Q})$ ($\Rightarrow N + N^2 + NL$ parameters must be estimated). However, the concept of HMM can be generalized to continuous observation processes:

Case 1: $Y_k \sim f_Y(y | X_k = e_i) = \sum_{m=1}^{M} c_{im} N(y | \mu_{im}, cov_{im})$ - Gaussian Mixture

$1 \leq i \leq N$, where the number M of mixture functions is given and the parameters:

$$\left.\begin{aligned} &c_{im} - \text{gain} \\ &\mu_{im} - \text{vector of mean values} \\ &cov_{im} - \text{covariance matrix} \end{aligned}\right\} \text{ depend on the state } e_i \,, \quad 1 \le i \le N$$

$\Rightarrow$ The HMM is defined as $\Sigma_1 \triangleq \left(\pi_0, \hat{P}, \{c_{im}, \mu_{im}, cov_{im}\}_{\substack{1 \le i \le N \\ 1 \le m \le M}}\right)$

Note that, while the assumption of Gaussianity can be relaxed (i.e., any log-concave function can be considered), it is necessary that:

$$\left.\begin{aligned} &c_{im} \ge 0 \quad \forall i, m \\ &\sum_{m=1}^{M} c_{im} = 1 \quad \forall i \end{aligned}\right\} \Rightarrow \int f_Y(y \mid X_k = e_i)\, dy = 1 \quad \forall i$$

It can be shown that the BW algorithm can be extended to this class of HMMs by updating the estimation formula (***) as follows:

$$\gamma_k(i,j) \triangleq \frac{\alpha_k(i)\beta_k(i)}{\sum_{\ell=1}^{N}\alpha_k(\ell)\beta_k(\ell)} \cdot \frac{c_{ij}\, N(y_k, \mu_{ij}, cov_{ij})}{\sum_{\ell=1}^{M} c_{i\ell}\, N(y_k, \mu_{i\ell}, cov_{i\ell})}$$

- Generalization of $\gamma_k(i)$ to the j-th mixture

- $$\bar{c}_{ij} = \frac{\sum_{k=1}^{n} \gamma_k(i,j)}{\sum_{k=1}^{n}\left(\sum_{\ell=1}^{M} \gamma_k(i,\ell)\right)}$$

- $$\bar{\mu}_{ij} = \frac{\sum_{k=1}^{n} \gamma_k(i,j) \cdot y_k}{\sum_{k=1}^{n} \gamma_k(i,j)}$$

- $$\overline{cov}_{ij} = \frac{\sum_{k=1}^{n} \gamma_k(i,j)(y_k - \mu_{ij})(y_k - \mu_{ij})^T}{\sum_{k=1}^{n} \gamma_k(i,j)}$$

These steps replace the last step in (***) and allow to estimate the output probability function of the new model $\bar{\Sigma}$ given the previous on Σ

Case 2: The observation variable Y_k has "d" components that are related via an autoregressive model:

$$Y_k = \underline{y}, \text{ with } \underline{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} \qquad y_h = -\sum_{w=1}^{p} a_w y_{h-w} + \varepsilon_h \qquad \varepsilon_h \sim N(0, \sigma^2)$$

The model captures the dynamics in the data (e.g., each output Y_k is a sequence of sounds, a motif, a curve, etc.)

Note that, if $d > 1$ is large enough, the probability function $f(\underline{y})$ is approximately:

$$f(\bar{y}) \cong \frac{1}{\sqrt{(2\pi\sigma^2)^d}} e^{-\frac{1}{2\sigma^2}\delta(\bar{y})}$$

where:

$$\delta(\bar{y}) \triangleq r_a(0) r(0) + \sum_{\ell=1}^{p} 2 r_a(\ell) r(\ell)$$

$$r_a(\ell) \triangleq \sum_{w=1}^{p-\ell} a_w a_{w+\ell} \qquad 0 \le \ell \le p$$

$$r(\ell) \triangleq \sum_{w=1}^{d-\ell} y_w y_{w+\ell} \qquad 0 \le \ell \le p$$

autocorrelation functions

Based on this, the HMM is defined as follows:

$$f_{Y_k}(\underline{y} \mid X_k = e_i, \Sigma) = \sum_{m=1}^{M} c_{im} f_{im}(\underline{y})$$

$$f_{im}(\bar{y}) = \frac{1}{\sqrt{(2\pi\sigma^2)^d}} \exp\left\{ -\frac{1}{2\sigma^2} \left[r_{a_{im}}(0) r(0) + \sum_{\ell=1}^{p} r_{a_{im}}(\ell) r(\ell) \right] \right\}$$

$$\underline{a_{im}} \triangleq [a_{im_1}\ a_{im_2} \dots a_{im_p}]^T \quad \text{s.t.} \quad y_h = -\sum_{w=1}^{p} a_{im_w} y_{h-w} + \varepsilon_h$$

i.e., we assume that the output model (which is given by the vector of parameters $\underline{a_{im}}$) changes with the state e_i and it is the sum of M autoregressive models (i.e., $1 \le m \le M$) $\Rightarrow$ We have an AUTOREGRESSIVE HMM

In this case, the estimation formula (***) are modified in this way:

$$\gamma_k(i,j) \triangleq \frac{\alpha_k(i)\,\beta_k(i)}{\sum_{\ell=1}^{N}\alpha_k(\ell)\,\beta_k(\ell)} \cdot \frac{c_{ij}\, f_{ij}(\underline{y}_k)}{\sum_{\ell=1}^{M} c_{i\ell}\, f_{i\ell}(\underline{y}_k)}$$

- Generalization of $\gamma_k(i)$ to the j-th mixture

$$r_{ij}(\ell) \triangleq \frac{\sum_{k=1}^{N}\gamma_k(i,j)\, r_k(\ell)}{\sum_{k=1}^{N}\gamma_k(i,j)} \qquad 0 \le \ell \le p$$

- Re-estimation of the autocorrelation function $r(\ell)$ to be used in the pdf of the j-th mixture in the i-th state

From the formula:

$$r_{ij}(\ell) = \sum_{w=1}^{d-\ell} y_w\, y_{w+\ell} = \sum_{w=1}^{d-\ell}\left(-\sum_{h=1}^{p} a_{ij_h}\, y_{w-h}\right)\left(-\sum_{h=1}^{p} a_{ij_h}\, y_{w+\ell-h}\right) \qquad 0 \le \ell \le p$$

one can then derive p equations in the p variables $(a_{ij_1}, a_{ij_2}, \dots, a_{ij_p})$ and determine a re-estimation of the coefficients of the autoregressive model for the j-th mixture in the i-th state. □

References

L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. of the IEEE, vol. 77 (2), pp. 257–286, 1989

A Copy of the article is available on Husky CT

Note: These lecture notes cover up to section IV.B in the article. Please, be aware that a reading of the entire article is requested.