Markov Assumption

$$P(z_{t+1}, z_{t+1} \mid z_t = z_t, \quad , z_1, z_1)$$

$$= P(z_{t+1}, z_{t+1} \mid z_t, z_t)$$

$$z_{t+1} \mid z_t, z_{t-1} \; - - - - \; z_1$$

$$= z_{t+1} \mid z_t$$

Model

$$P(X_1 = x_1, \ldots, X_T = x_T, Z_1 = z_1, \ldots, Z_T = z_T)$$

$$= P(Z_1 = z_1) \prod_{t=1}^{T-1} P(Z_{t+1} = z_{t+1} \mid Z_t = z_t)$$

$$\cdot \prod_{t=1}^{T} P(X_t = x_t \mid Z_t = z_t)$$

$$\boxed{\text{Conditional Independence Assumptions}}$$

$$Z_{t+1} \mid Z_t, \ldots, Z_1 = Z_{t+1} \mid Z_t$$

$$X_t \mid X_1, \ldots, X_{t-1}, X_{t+1}, \ldots, X_T, Z_1, \ldots, Z_t, \ldots Z_T$$

$$= X_t \mid Z_t$$

$$z_t \in \{s_1, \ldots, s_N\}$$

$$x_t \in \{v_1, \ldots, v_M\}$$

$$P(z_{t+1} = s_j \mid z_t = i) = a_{ij}$$

$$\sum_{j=1}^{N} a_{ij} = 1.$$

$$(A)_{ij} = a_{ij}$$

$$P(x_t = v_l \mid z_t = s_i) = b_{il}$$

$$(B)_{il} = b_{il} \qquad \sum_{l=1}^{M} b_{il} = 1$$

$$P(Z_1 = s_i) = c_i \qquad \sum_{i=1}^{N} c_i = 1$$

$\theta = (A, B, C)$ are the parameters

Notation

If $x_t = v_l$ then $b_{i(x_t)} = b_{il}$

$$P(X_1 = x_1, \ldots, X_T = x_T \mid \theta)$$

$$= ?$$

N^T computations

$$\alpha_t(i) = P(X_1 = x_1, \dots, X_t = x_t, Z_t = s_i \mid \theta)$$

$$\alpha_{t+1}(j)$$

$$= P(X_1 = x_1, \dots, X_t = x_t, X_{t+1} = x_{t+1}, Z_{t+1} = s_j \mid \theta)$$

$$= \sum_{i=1}^{N} P(X_1 = x_1, \dots, X_t = x_t, X_{t+1} = x_{t+1}, Z_t = s_i, Z_{t+1} = s_j \mid \theta)$$

$$= \sum_{i=1}^{N} P(X_1 = x_1, \dots, X_t = x_t, Z_t = s_i) \; P(X_{t+1} = x_{t+1}, Z_{t+1} = s_j \mid Z_t = s_i)$$

$$= \sum_{i=1}^{N} \alpha_t(i) \; P(X_{t+1} = x_{t+1} \mid Z_{t+1} = s_j) \; P(Z_{t+1} = s_j \mid Z_t = s_i)$$

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i)\, b_{j(x_{t+1})}\, a_{ij}$$

$$\alpha_1(i) = P(X_1 = x_1, Z_1 = s_i \mid \theta)$$

$$= P(Z_1 = s_i \mid \theta)\, P(X_1 = x_1 \mid Z_1 = s_i)$$

$$= c_i\, b_{i(x_1)}$$

$$\alpha_T(j) = P(X_1 = x_1, \ldots, X_T = x_T, Z_T = j)$$

$$\sum_{j=1}^{N} \alpha_t(j) = P(X_1 = x_1, \ldots, X_T = x_T \mid \theta)$$

α is a matrix of size $T \times N$. (computation is $\underline{\underline{N^2 T}}$)

Backward Algorithm

$$P(X_{t+1}=x_{t+1}, X_{t+2}=x_{t+2}, \ldots, X_T=x_T \mid Z_t=s_i) = \beta_t(i)$$

$$\beta_{t-1}(i) = P(X_t=x_t, X_{t+1}=x_{t+1}, \ldots, X_T=x_T \mid Z_{t-1}=s_i)$$

$$= \sum_{j=1}^{N} P(Z_t=s_j, X_t=x_t, X_{t+1}=x_{t+1}, \ldots, X_T=x_T \mid Z_{t-1}=s_i)$$

$$= \sum_{j=1}^{N} P(Z_t=s_j \mid Z_{t-1}=s_i)\, P(X_t=x_t \mid Z_t=s_i)\, P(X_{t+1}=x_{t+1}, \ldots, X_T=x_T \mid X_t=x_t, Z_t=s_i)$$

$$= \sum_{j=1}^{N} a_{ij}\, b_i(x_t)\, \beta_t(j)$$

$$\beta_{T-1}(i) = P(X_T=x_T \mid Z_{T-1}=s_i)$$

$$= \sum_{j=1}^{N} P(X_T=x_T, Z_T=s_j \mid Z_{T-1}=s_i)$$

$$= \sum_{j=1}^{N} b_j(x_{t+1})\, a_{ij}$$

Learning $\theta = \{A, B, C\}$ from data

$\mathcal{D} = \{x^{(1)}, \quad , x^{(n)}\}$

$x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_T^{(i)}\}$

Model

$$P(X_1 = x_1, \quad , X_T = x_T, Z_1 = z_1, \dots, Z_T = z_T)$$

$$= P(Z_1 = z_1) \prod_{t=1}^{T-1} P(Z_{t+1} = z_{t+1} \mid Z_t = z_t)$$

$$\cdot \prod_{t=1}^{T} P(X_t = x_t \mid Z_t = z_t)$$

$$\delta(x, a) = 1 \quad x = a$$
$$= 0 \quad x \neq a$$

$$P(Z_1 = z_1 | \theta) = \prod_{i=1}^{N} c_i^{\delta(z_1, s_i)}$$

$$P(Z_{t+1} = z_{t+1} | Z_t = z_t, \theta)$$
$$= \prod_{j=1}^{N} \prod_{i=1}^{N} a_{ij}^{\delta(z_t, s_i)\,\delta(z_{t+1}, s_j)}$$

$$P(X_t = x_t | Z_t = z_t)$$
$$= \prod_{i=1}^{N} \prod_{l=1}^{M} \delta(x_t, v_l)\,\delta(z_t, s_i)$$
$$\log b_{il}$$

$$\log p(X, Z \mid \theta)$$

$$= \sum_{i=1}^{N} \delta(z_1, s_i) \log c_i$$

$$+ \sum_{t=1}^{T-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(z_{t+1}, s_j) \delta(z_t, s_i) \log a_{ij}$$

$$+ \sum_{t=1}^{T} \sum_{i=1}^{N} \delta(z_t, s_i) \delta(x_t, v_l)$$

$$\log b_{il}$$

$$E_{z|x,\theta} \log P(X, x, Z \mid \theta)$$

$$E_{z|x,\theta}\, \delta(z_1, s_i) = P(z_1 = s_i \mid \theta)$$

$$E_{z|x,\theta}\, \delta(z_t, s_i)\, \delta(z_{t+1}, s_j)$$

$$= P(z_t = s_i, z_{t+1} = s_j \mid x, \theta)$$

$$E_{z|x,\theta}\, \delta(z_t, s_i) = P(z_t = s_i \mid x, \theta)$$

$$P(Z_t = s_i \mid X = x, \theta)$$

$$= \frac{P(X = x, Z_t = s_i \mid \theta)}{P(X = x \mid \theta)}$$

$$= \frac{P(X_1 = x_1, \ldots, X_{t-1} = x_{t-1}, X_t = x_t, Z_t = s_i, X_{t+1} = x_{t+1}, X_T = x_T \mid \theta)}{P(X = x \mid \theta)}$$

$$= \frac{\alpha_t(i)\, \beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\, \beta_t(i)}$$

$$P(Z_t = s_i, Z_{t+1} = s_j \mid \theta)$$

$$= \frac{P(X_1 = x_1, \ldots, X_t = x_t, Z_t = s_i, Z_{t+1} = s_j, X_{t+1} = x_{t+1}, \ldots, X_T = x_T \mid \theta)}{P(X = x \mid \theta)}$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_i(x_{t+1})\, \beta_{t+1}(j)}{\sum_i \alpha_t(i)\, \beta_t(j)}$$

$$\tilde{c}_i = \frac{\alpha_1(i)\,\beta_1(j)}{\sum_{i'} \alpha_1(i')\,\beta_1(i')}$$

$$\tilde{a}_{ij} = \sum_{t=1}^{T-1} \frac{\alpha_t(i)\, a_{ij}\, b_i(x_{t+1})\, \beta_{t+1}(j)}{\sum_{i} \alpha_t(i)\, \beta_t(j)}$$

$$\tilde{b}_{il} = \sum_{t=1}^{T} \delta(x_t, v_l) \left(\frac{\alpha_t(i)\, \beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\, \beta_t(i)} \right)$$

$$= \sum_{i=1}^{N} \tilde{c}_i \log c_i$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{a}_{ij} \log a_{ij}$$

$$+ \sum_{i=1}^{N} \sum_{l=1}^{M} \tilde{b}_{il} \log b_{il}$$

$$\sum_{i=1}^{N} c_i = 1, \quad c_i \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1, \quad a_{ij} \geq 0$$

$$\sum_{l=1}^{N} b_{il} = 1, \quad b_{il} \geq 0$$

$$\theta^* = \arg\max_{\theta} E_{Z|X,\theta} \log P(X, x, Z \mid \theta)$$

$$c_i^* = \frac{\tilde{c}_i}{\sum_{j=1}^{N} \tilde{c}_i} \qquad \sum_i c_i^* = 1$$

$$a_{ij}^* = \frac{\tilde{a}_{ij}}{\sum_{j=1}^{N} \tilde{a}_{ij}} \qquad \sum_j a_{ij}^* = 1$$

$$b_{il}^* = \frac{\tilde{b}_{il}}{\sum_{l=1}^{M} \tilde{b}_{il}} \qquad \sum_l b_{il}^* = 1$$