EE4R – Automatic Spoken Language Processing

Question Sheet 2 – answers

February 2015

Multimodal Interaction Lab

b ɜː m ɪ ŋ ə m

UNIVERSITY OF BIRMINGHAM

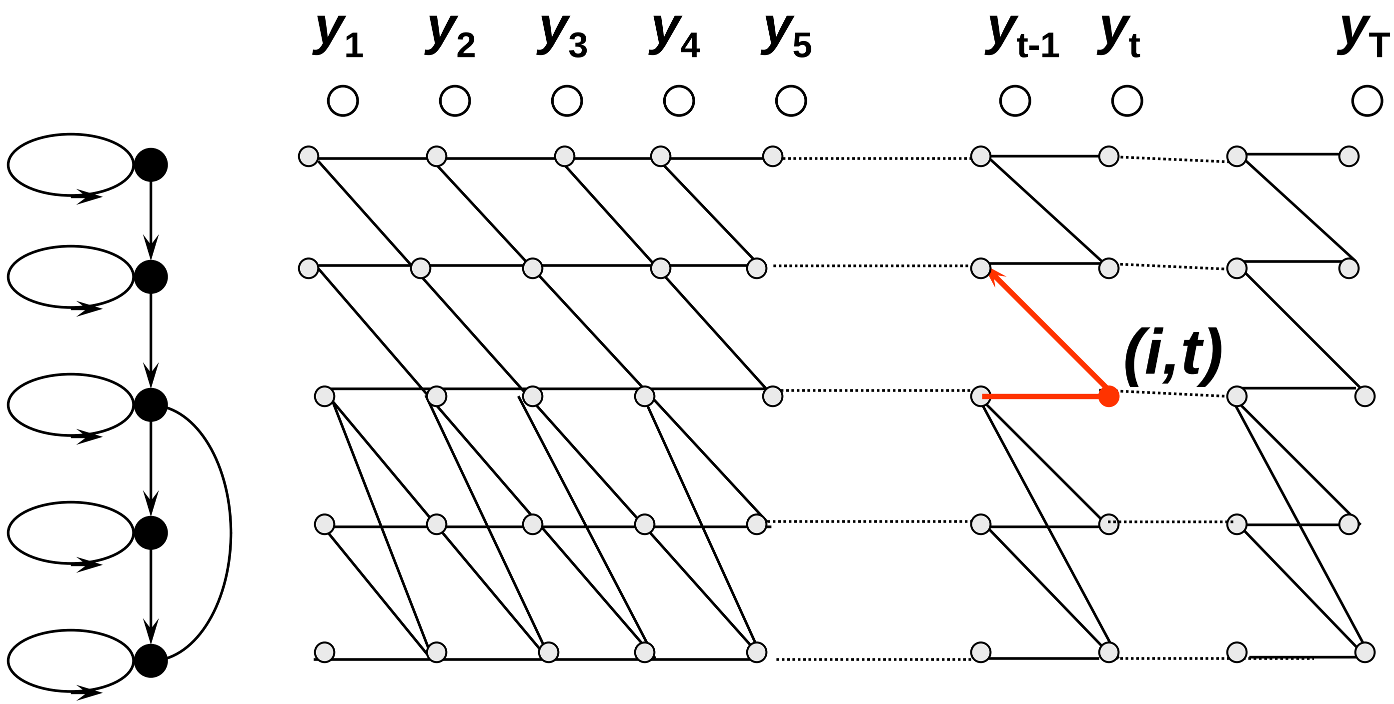
Question 1 – HMM decoding

- Let $X = \{x_1, \ldots, x_T\}$ be a state sequence of length T
- The joint probability of Y and X is given by:

$$p(Y, X) = b_{x_1}(y_1) \prod_{t=2}^{T} a_{x_{t-1} x_t} b_{x_t}(y_t)$$

- i.e. the product of the state-output and state transition probabilities along the state sequence
- $p(Y)$ is the sum of $P(Y,X)$ over all sequences X
- $P(Y, \hat{X})$ is the probability of an observation sequence Y and the optimum state sequence $\hat{X}$

Viterbi Decoding

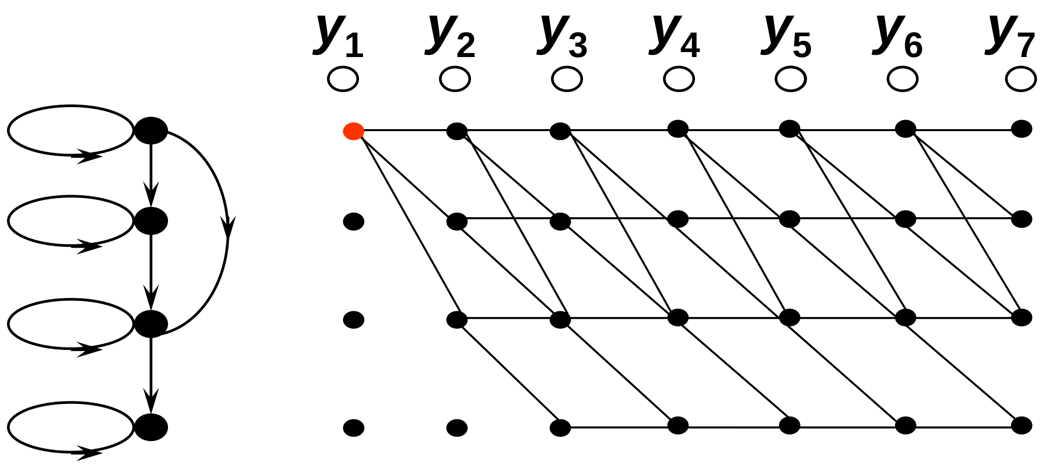


$$p_t(i) = \text{Prob}(y_1, \ldots, y_t \text{ , opt sequence to } (i,t))$$

$$p_t(i) = \max\{p_{t-1}(i-1)a_{i-1,i} \text{ , } p_{t-1}(i)a_{i,i}\}b_i(y_t)$$

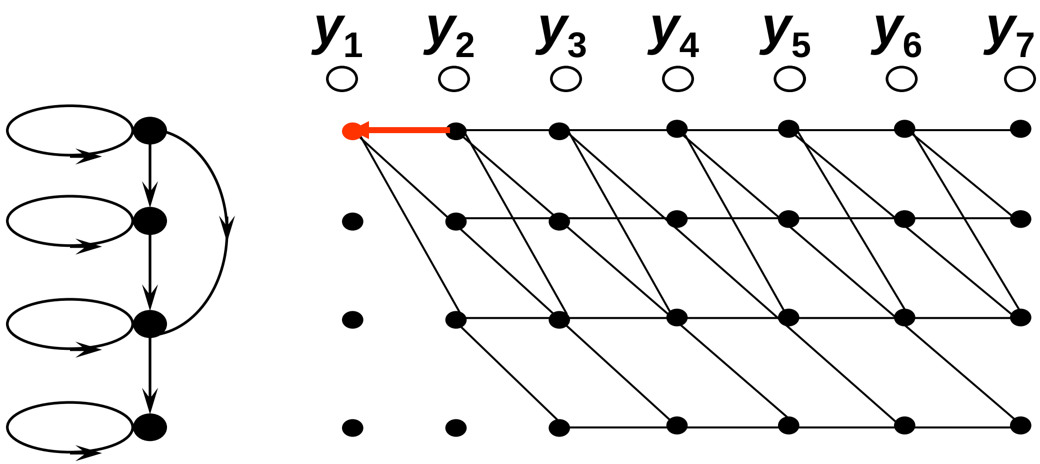
Viterbi Decoding

- State-time trellis



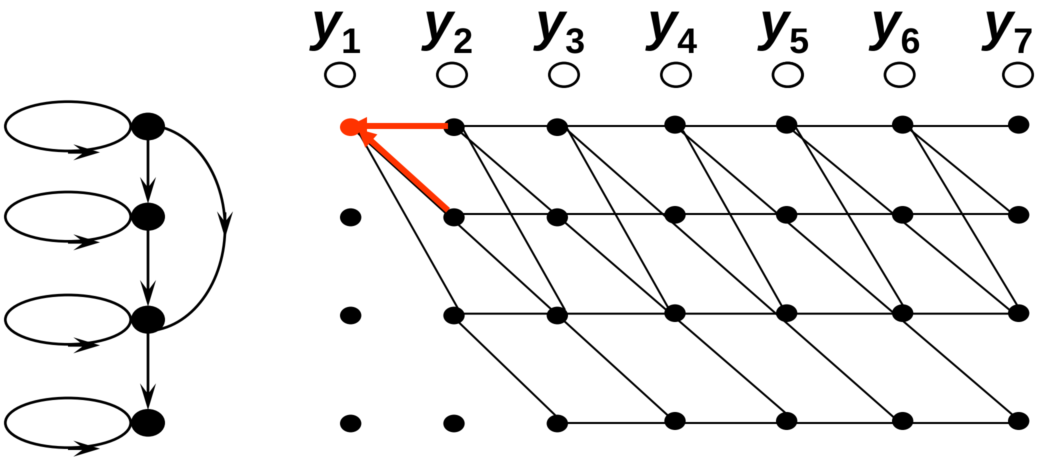
$$\alpha_1(1) = b_1(y_1) = 0.6$$

Viterbi Decoding



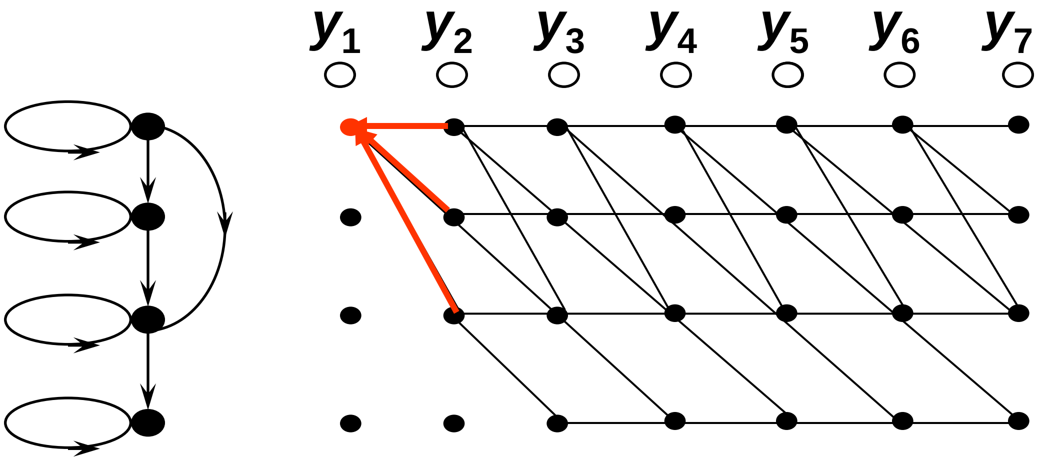
$$\alpha_2(1)=\alpha_1(1)a_{11}b_1(y_2) \quad =0.6*0.5*0.2=0.06$$

Viterbi Decoding



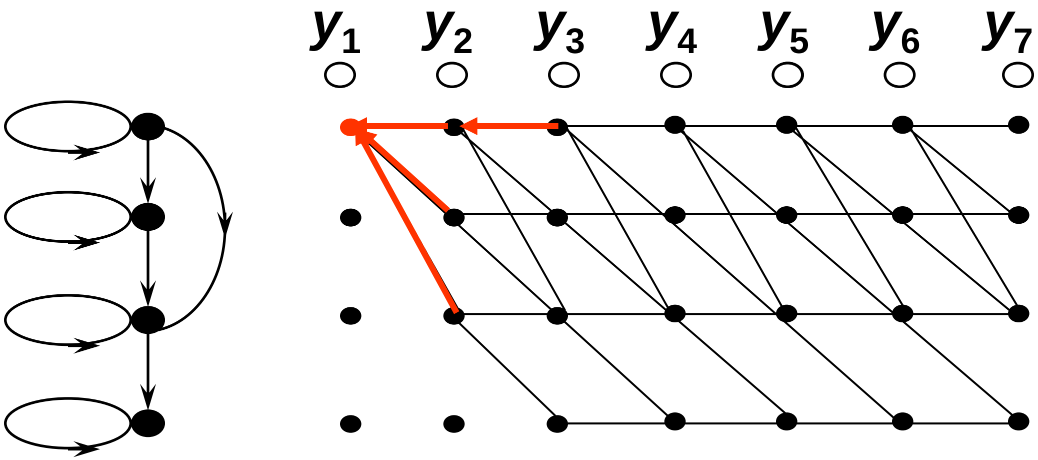
$$\alpha_2(2) = \alpha_1(1) a_{12} b_2(y_2) = 0.6 * 0.2 * 0.7 = 0.084$$

Viterbi Decoding



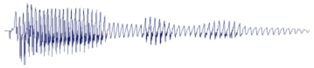
$$\alpha_2(3) = \alpha_1(1) a_{13} b_3(y_2) = 0.6 * 0.3 * 0.4 = 0.072$$

Viterbi Decoding



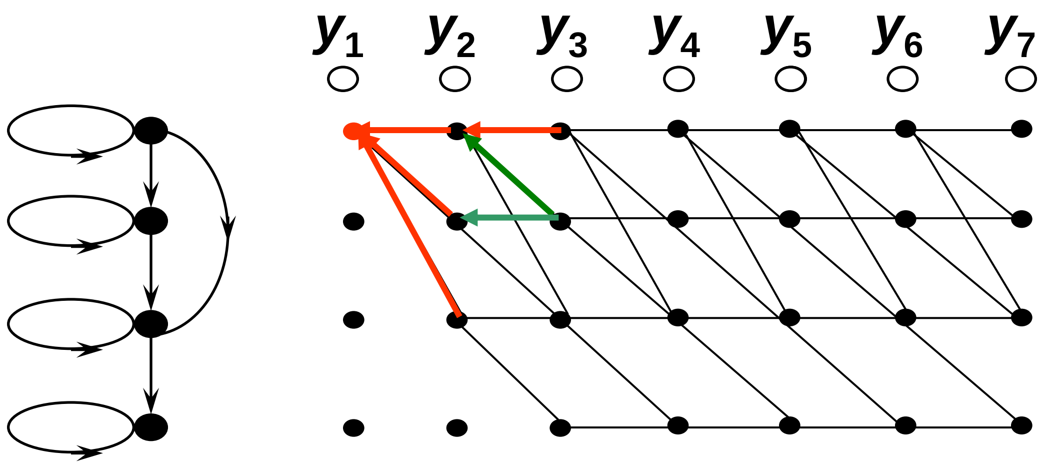
$$\alpha_3(1) = \alpha_2(1) a_{11} b_1(y_3) = 0.06 * 0.5 * 0.6 = 0.018$$

Multimodal Interaction Lab



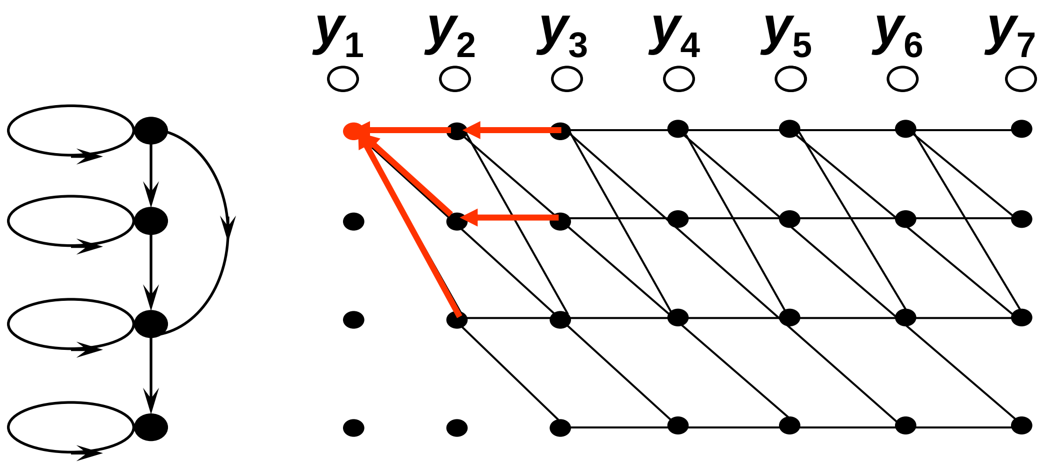
UNIVERSITY OF BIRMINGHAM

Viterbi Decoding



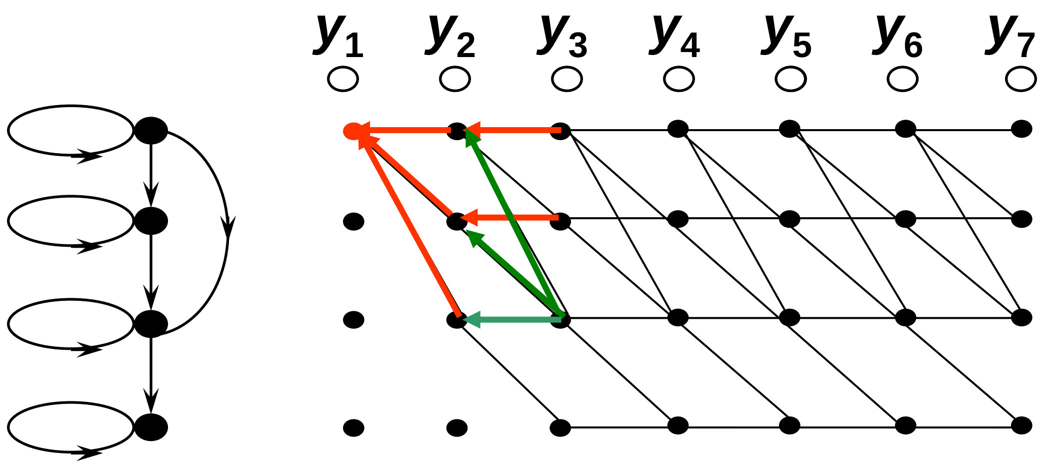
$$\alpha_3(2) = \max \begin{cases} \cancel{\alpha_2(1) a_{12} b_2(y_3) = 0.06 * 0.2 * 0.2 = 2.4 * 10^{-3}} \\ \alpha_2(2) a_{22} b_2(y_3) = 0.084 * 0.6 * 0.2 = 0.01008 \end{cases}$$

Viterbi Decoding



$$\alpha_3(2) = \max \begin{cases} \cancel{\alpha_2(1) a_{12} b_2(y_3) = 0.06 * 0.2 * 0.2 = 2.4 * 10^{-3}} \\ \alpha_2(2) a_{22} b_2(y_3) = 0.084 * 0.6 * 0.2 = 0.01008 \end{cases}$$

Viterbi Decoding



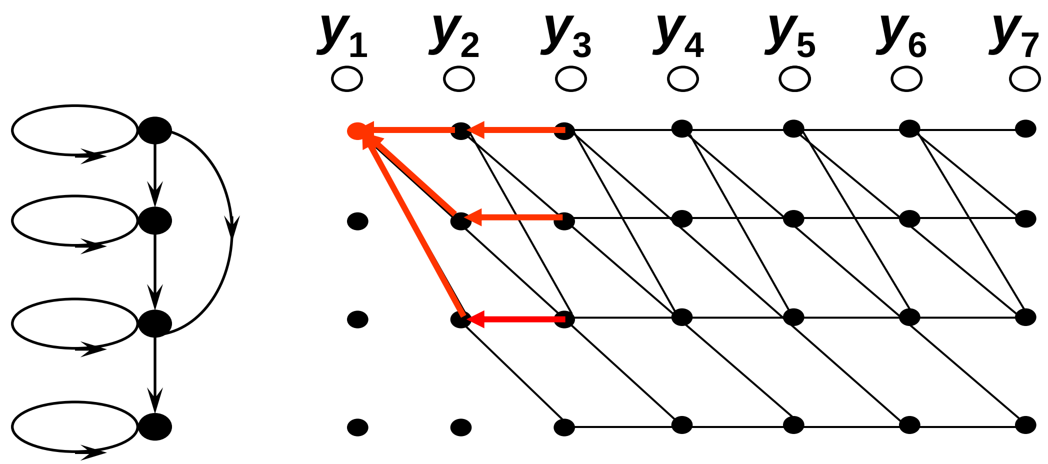
$$\alpha_3(3) = \max \begin{cases} \alpha_2(1) a_{13} b_3(y_3) & = 0.06 * 0.3 * 0.2 = 3.6 * 10^{-3} \\ \alpha_2(2) a_{23} b_3(y_3) & = 0.084 * 0.4 * 0.2 = 6.72 * 10^{-3} \\ \alpha_2(3) a_{33} b_3(y_3) & = 0.072 * 0.6 * 0.2 = 8.64 * 10^{-3} \end{cases}$$

Multimodal Interaction Lab

b ɜː m ɪ ŋ ə m

UNIVERSITY OF BIRMINGHAM

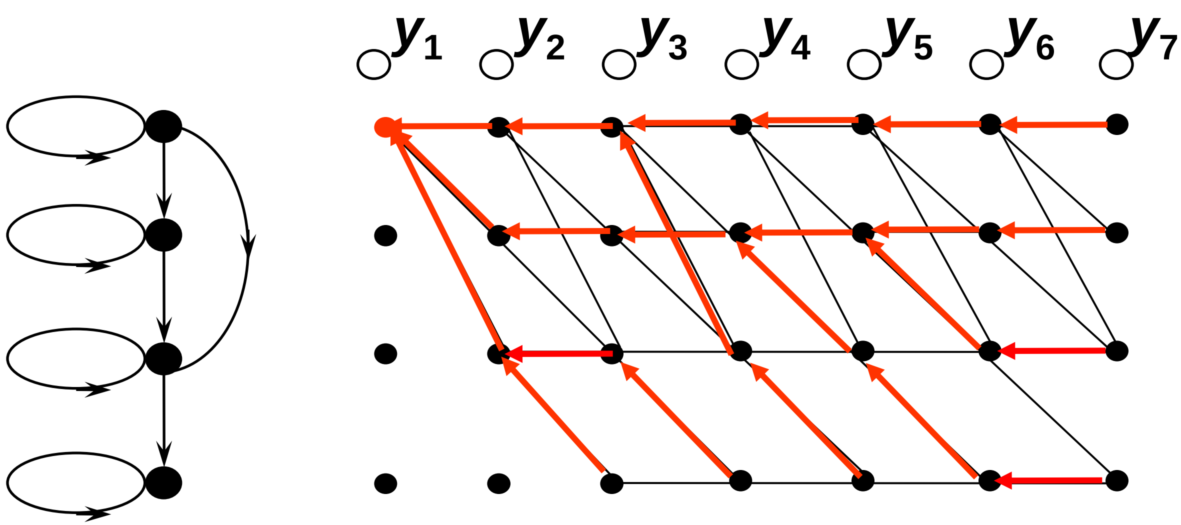
Viterbi Decoding



$$\alpha_3(3) = \max \begin{cases} \cancel{\alpha_2(1) a_{13} b_3(y_3) = 0.06 * 0.3 * 0.2 = 3.6 * 10^{-3}} \\ \cancel{\alpha_2(2) a_{23} b_3(y_3) = 0.084 * 0.4 * 0.2 = 6.72 * 10^{-3}} \\ \alpha_2(3) a_{33} b_3(y_3) = 0.072 * 0.6 * 0.2 = 8.64 * 10^{-3} \end{cases}$$

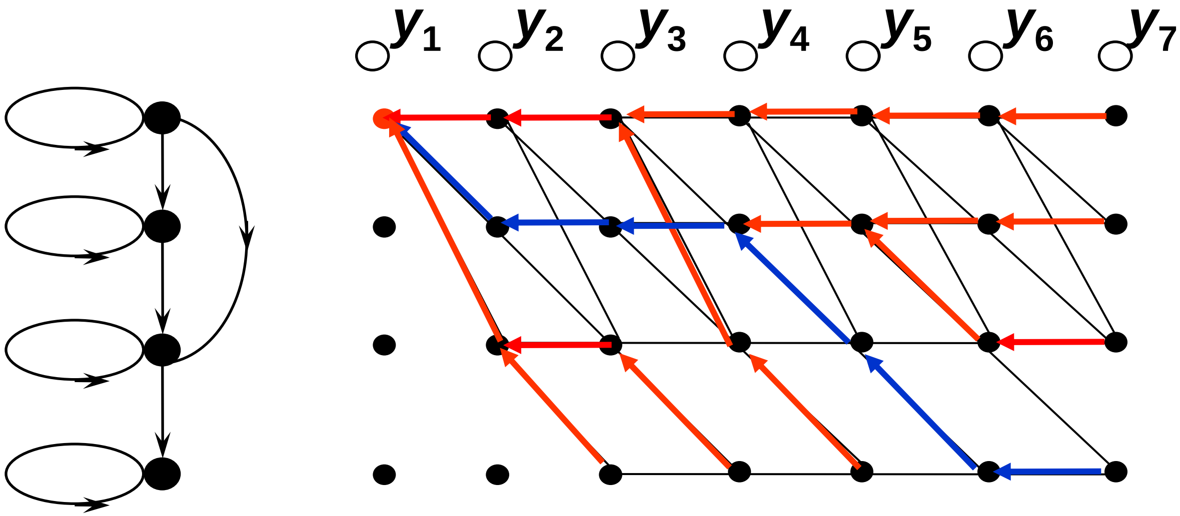
Viterbi Decoding

- Continue in a similar manner
- Final overall probability $P(Y,\hat{X})=\alpha_7(4)=1.73*10^{-4}$



Viterbi Decoding

- Continue in a similar manner
- Final overall probability $P(Y,\hat{X}) = \alpha_7(4) = 1.73*10^{-4}$



Question 2 – classification (GMM)

- Calculation of the probabilities

$$P(Y \mid C_i) = P(y_1, \dots, y_T \mid C_i) = \prod_{t=1}^{T} P(y_t \mid C_i)$$

$$P(y_t \mid C_i) = \sum_{m=1}^{M} w_m P(y_t \mid m, C_i) \qquad P(y_t \mid m, C_i) = \prod_{d=1}^{D} Gaussian(1D)$$

$$P(y_1 \mid C_1) = \sum_{m=1}^{3} w_m P(y_1 \mid m, C_1) = 0.5 \cdot 0.0449 + 0.3 \cdot 0.0385 + 0.2 \cdot 0.0117 = 0.0363$$

$$P(y_2 \mid C_1) = 0.5 \cdot 0.0396 + 0.3 \cdot 0.0560 + 0.2 \cdot 0.0141 = 0.0394$$

$$P(y_3 \mid C_1) = 0.0270 \qquad P(y_4 \mid C_1) = 0.0033 \qquad P(y_5 \mid C_1) = 0.0357$$

$$P(Y \mid C_1) = 4.5938 \cdot 10^{-9}$$

Question 2 – classification (GMM)

- Calculation of the probabilities

$$P(y_1 \mid C_2) = 0.0249$$

$$P(y_2 \mid C_2) = 0.0250$$

$$P(y_3 \mid C_2) = 0.0397$$

$$P(y_4 \mid C_2) = 0.0122$$

$$P(y_5 \mid C_2) = 0.0210$$

$$P(Y \mid C_1) = 4.5938 \cdot 10^{-9} \qquad P(Y \mid C_2) = 6.3475 \cdot 10^{-9}$$

- P(Y|C1) < P(Y|C2) -> classified as C2

Question 3 – GMM training

- Component 1:

$$P(y_t \mid m_1) = \frac{1}{\sqrt{2\pi\sigma_1^{\,2}}} \exp\left(-0.5 \frac{(y_t - \mu_1)^2}{\sigma_1^{\,2}}\right)$$

$$P(y_1 \mid m_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-0.5 \cdot \frac{(7-6)^2}{1}\right) = 0.2420$$

$$P(y_2 \mid m_1) = 0.2420$$

$$P(y_3 \mid m_1) = 0.00013383$$

$$P(y_4 \mid m_1) = 0.0044$$

$$P(y_5 \mid m_1) = 0.0540$$

Question 3 – GMM training

- Component 2:

$$P(y_t \mid m_2) = \frac{1}{\sqrt{2\pi\sigma_2{}^2}} \exp\left(-0.5 \frac{(y_t - \mu_2)^2}{\sigma_2{}^2} \right)$$

$$P(y_1 \mid m_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-0.5 \cdot \frac{(7-8)^2}{1} \right) = 0.2420$$

$$P(y_2 \mid m_2) = 0.0044$$

$$P(y_3 \mid m_2) = 0.0540$$

$$P(y_4 \mid m_2) = 0.2420$$

$$P(y_5 \mid m_2) = 0.00013383$$

Question 3 – GMM training

- Posterior probabilities:

$$P(m_1 \mid y_t) = \frac{P(y_t \mid m_1)P(m_1)}{P(y_t \mid m_1)P(m_1) + P(y_t \mid m_2)P(m_2)}$$

$$P(m_1 \mid y_1) = \frac{0.2420}{0.2420 + 0.2420} = 0.5 \qquad P(m_2 \mid y_1) = 0.5$$

$$P(m_1 \mid y_2) = \frac{0.2420}{0.2420 + 0.0044} = 0.982 \qquad P(m_2 \mid y_2) = 0.01785$$

$$P(m_1 \mid y_3) = 0.00247 \qquad P(m_2 \mid y_3) = 0.99753$$

$$P(m_1 \mid y_4) = 0.01785 \qquad P(m_2 \mid y_4) = 0.982$$

$$P(m_1 \mid y_5) = 0.99753 \qquad P(m_2 \mid y_5) = 0.00247$$

Question 3 – GMM training

- New estimates:

$$\mu_1^{new} = \frac{0.5\cdot 7 + 0.982\cdot 5 + 0.00247\cdot 10 + 0.01785\cdot 9 + 0.99753\cdot 4}{0.5 + 0.982 + 0.00247 + 0.01785 + 0.99753} = 5.034$$

$$\mu_2^{new} = \frac{0.5\cdot 7 + 0.01785\cdot 5 + 0.99753\cdot 10 + 0.982\cdot 9 + 0.00247\cdot 4}{0.5 + 0.01785 + 0.99753 + 0.982 + 0.00247} = 8.965$$

$$\sigma_1^{2\,new} = \frac{0.5\cdot(7-5.034)^2 + 0.982\cdot(5-5.034)^2 + 0.00247\cdot(10-5.034)^2 + 0.01785\cdot(9-5.034)^2 + 0.99753\cdot(4-5.034)^2}{0.5 + 0.982 + 0.00247 + 0.01785 + 0.99753} = 1.33$$

$$\sigma_2^{2\,new} = \frac{0.5\cdot(7-8.965)^2 + 0.01785\cdot(5-8.965)^2 + 0.99753\cdot(10-8.965)^2 + 0.982\cdot(9-8.965)^2 + 0.00247\cdot(4-8.965)^2}{0.5 + 0.01785 + 0.99753 + 0.982 + 0.00247} = 1.33$$