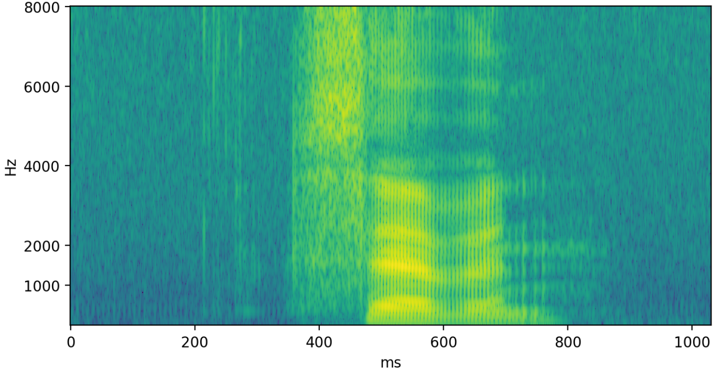
Hidden Markov Models

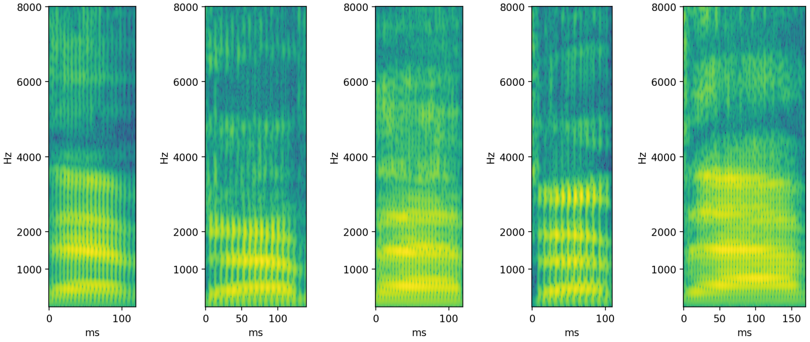
Hao Tang

Automatic Speech Recognition—ASR Lecture 4
27 January 2022

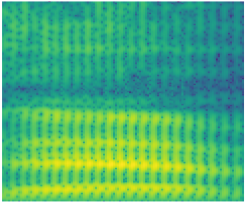
Acoustic Cues



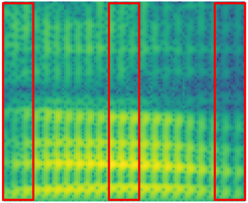
Variable Lengths



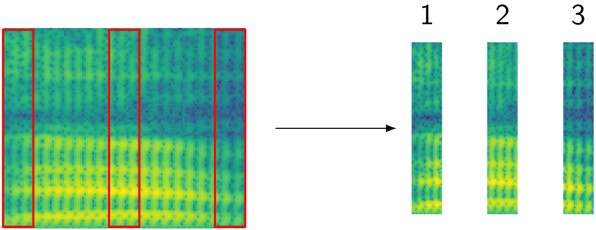
Three Parts



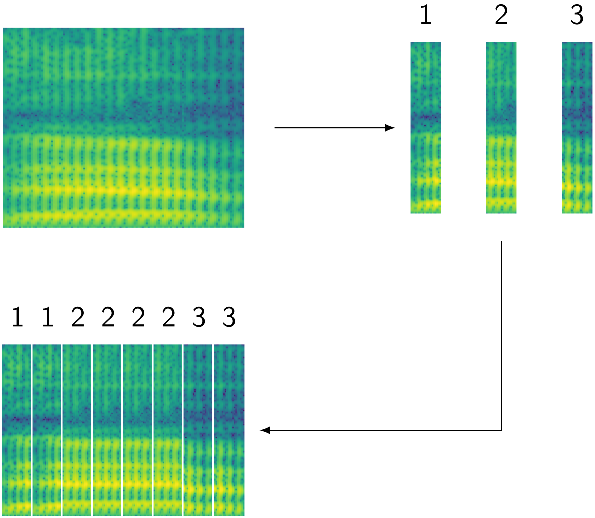
Three Parts



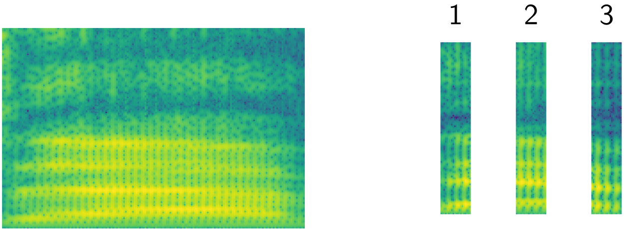
Three Parts



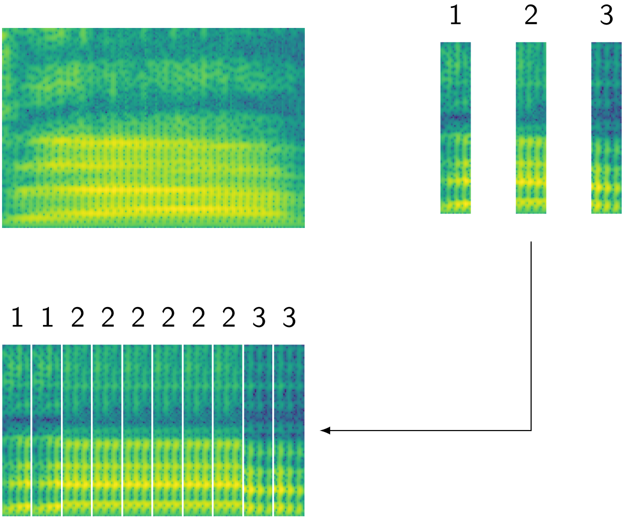
Three Parts

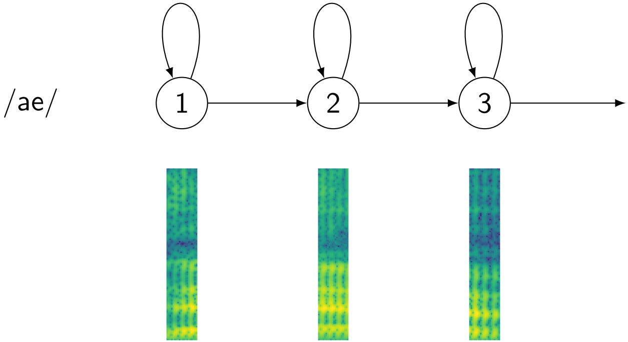


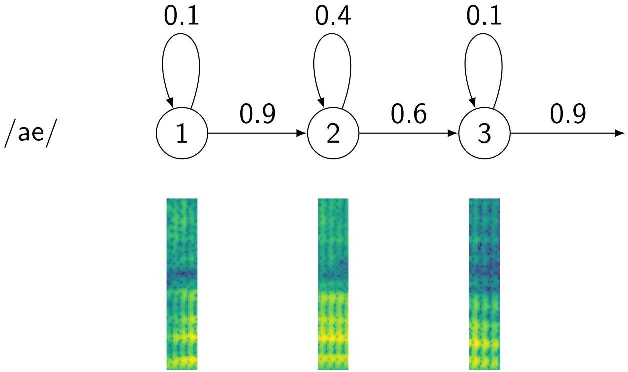
Three Parts



Three Parts

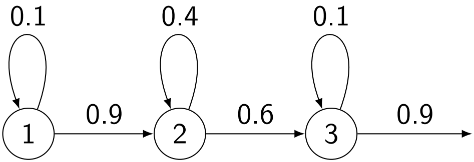



/ae/
1
2
3



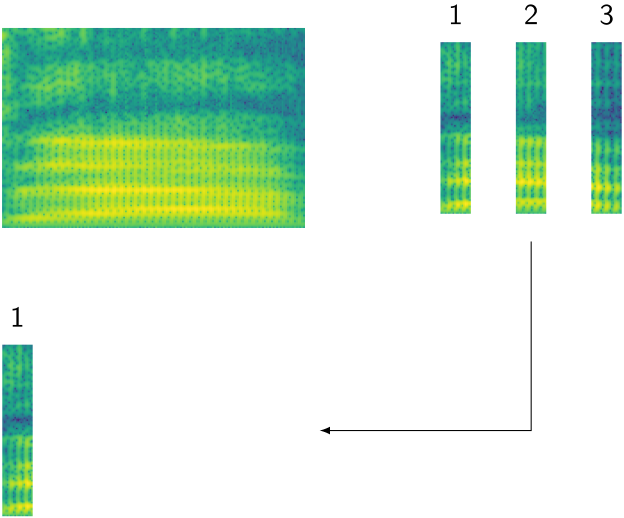
/ae/
0.1
0.4
0.1
0.9
0.6
0.9
1
2
3

Bakis Model

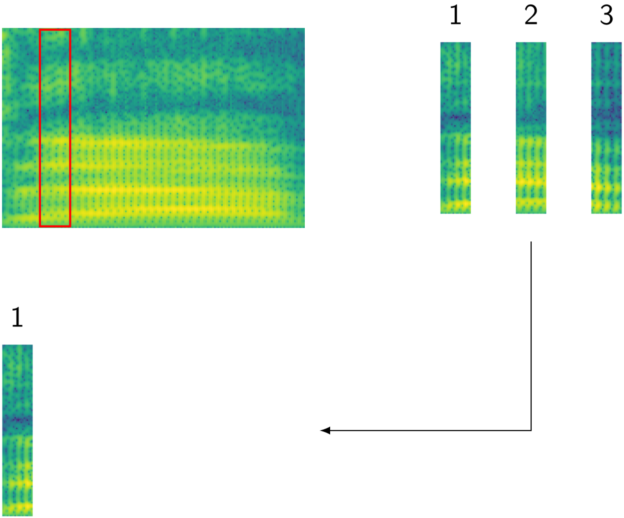


- Acoustic cues
- Variable lengths
- Three states
- Transition probabilities

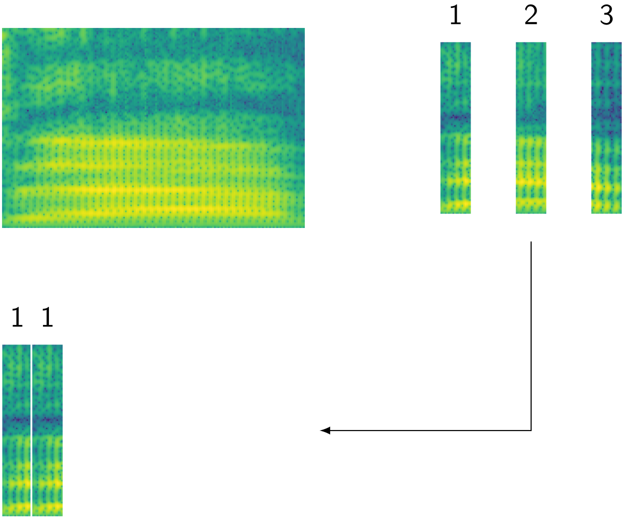
Matching



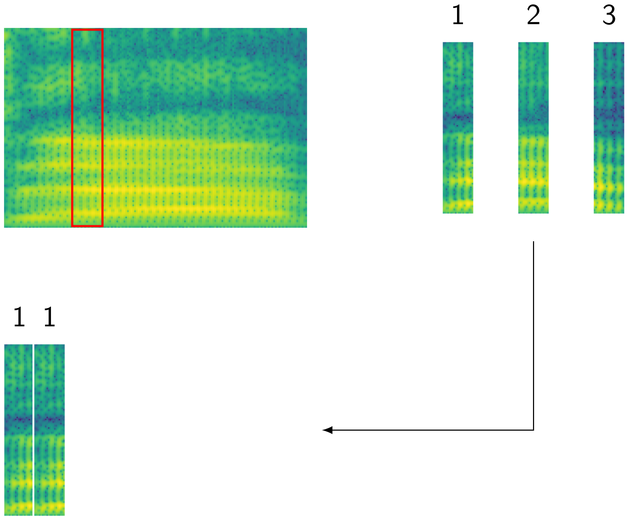
Matching



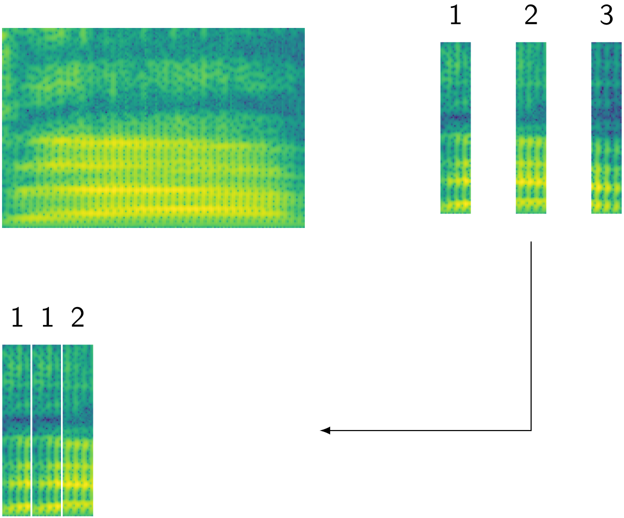
Matching



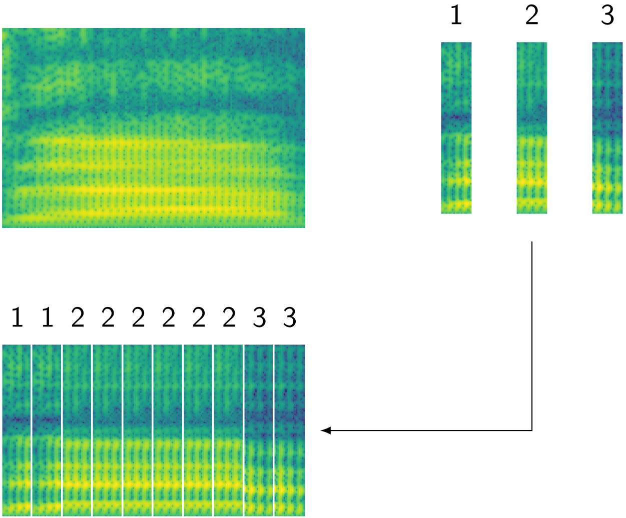
Matching



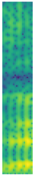
Matching



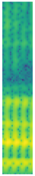
Matching



Spectral Distortion



x

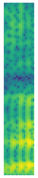


y

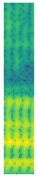
$$d(x, y) = \|x - y\|_2^2$$

Likelihood

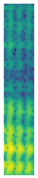
1



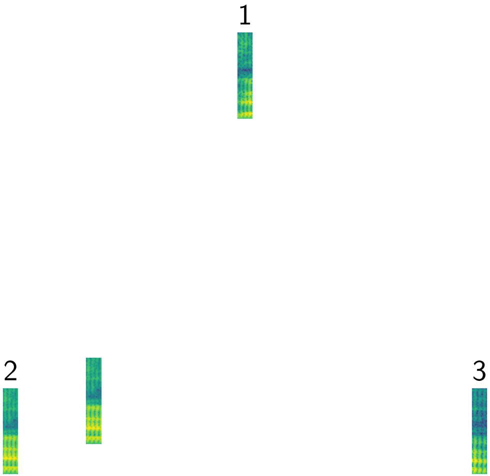
2



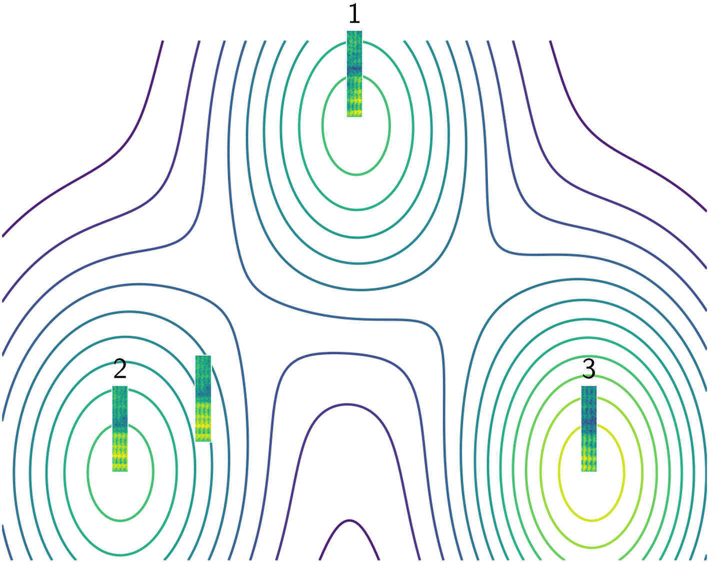
3



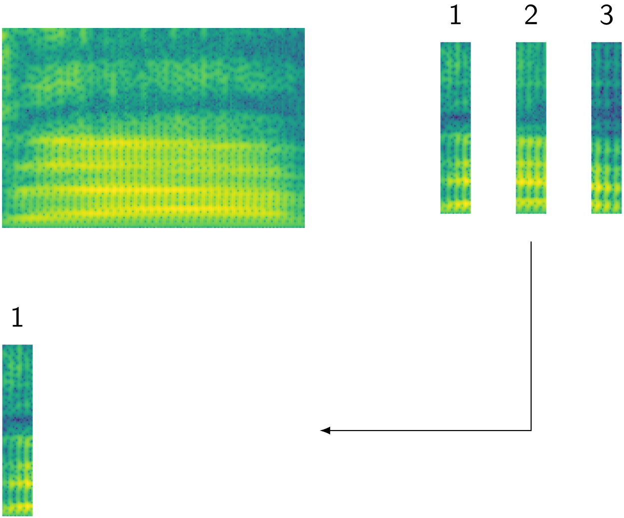
Likelihood



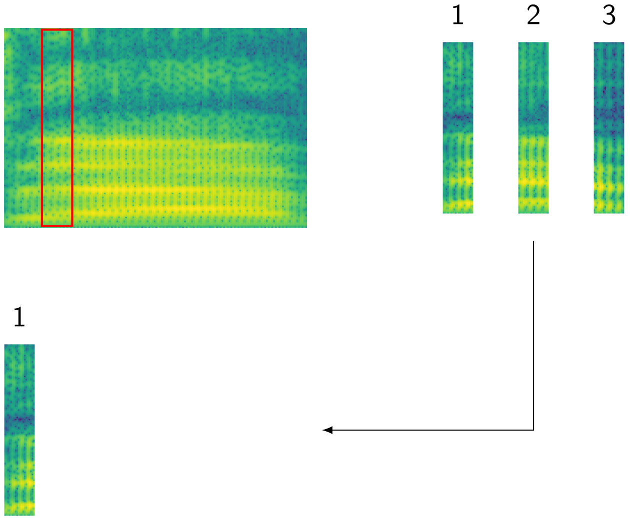
Likelihood



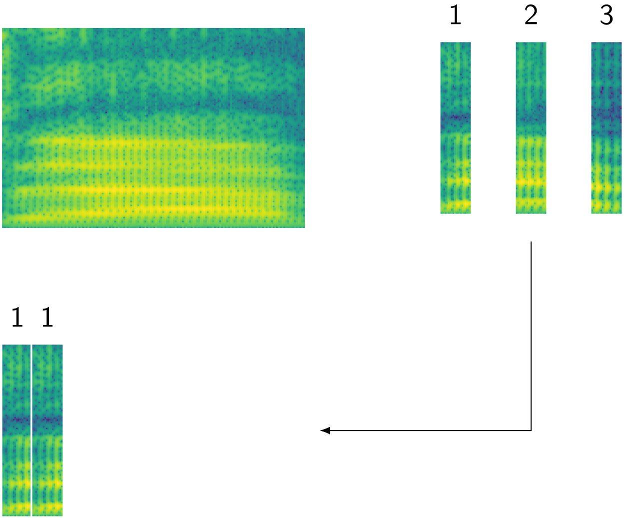
Likelihood



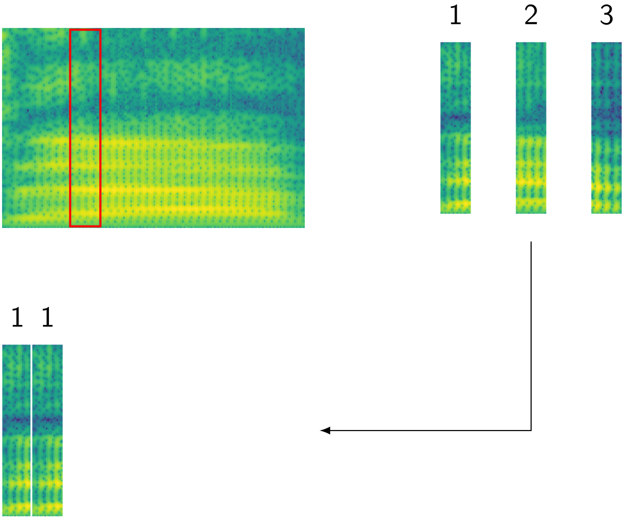
Likelihood



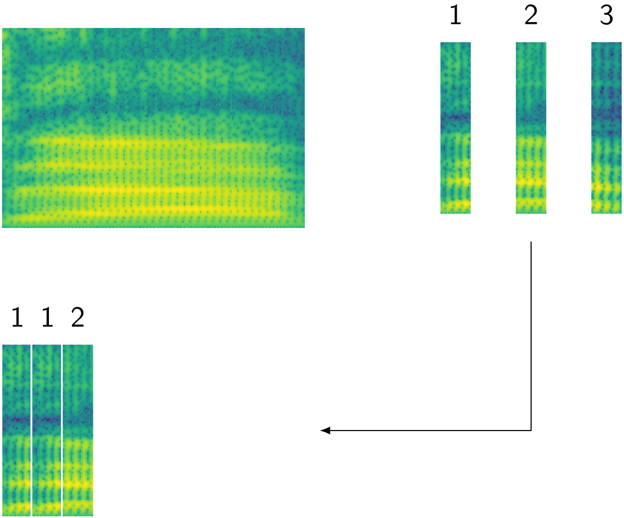
Likelihood



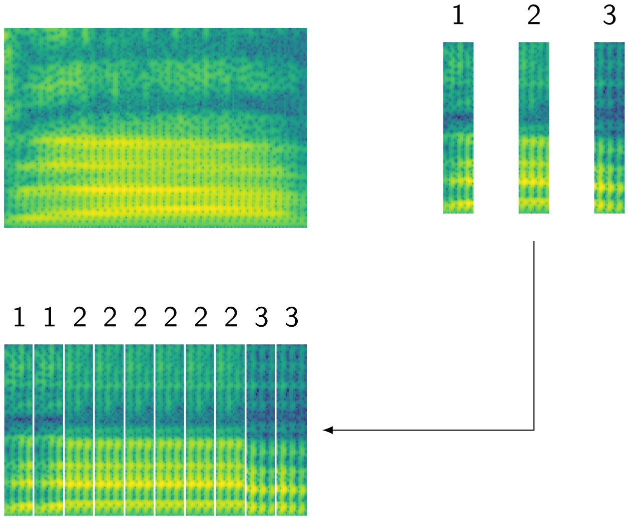
Likelihood



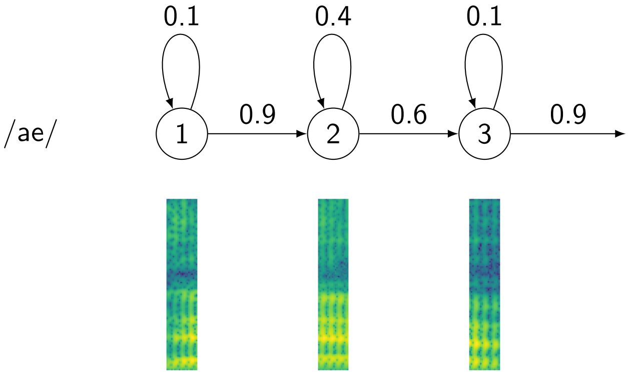
Likelihood

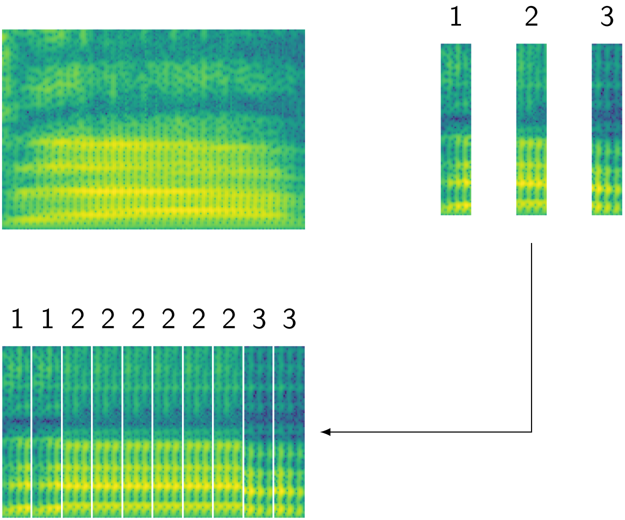


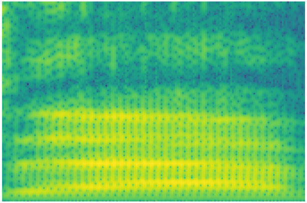
Likelihood



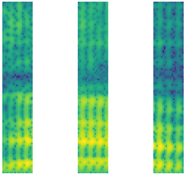
(Probabilistic) Bakis Model




1
2
3
1 1 2 2 2 2 2 2 3 3

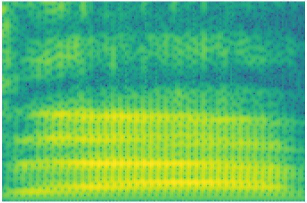


1 2 3

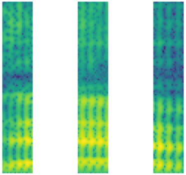


1 1 2 2 2 2 2 2 3 3

frames

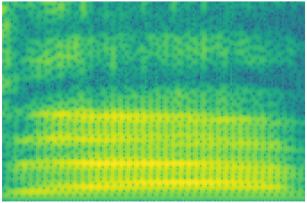


1 1 2 2 2 2 2 2 3 3

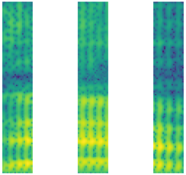


$x_1 \; x_2 \; x_3 \; x_4 \qquad \cdots \qquad x_T$

frames



1 2 3



$q_1 \; q_2 \; q_3 \; q_4 \qquad \cdots \qquad q_T$

Hidden Markov Models

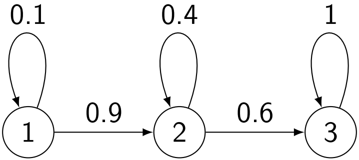
- States S, e.g., $\{1, 2, 3\}$ ($J = |S|$)
- Prior probabilities π, e.g., $[1, 0, 0]$
- Transition probability $p(q' = j | q = i) = a_{ij}$

$$A = \begin{bmatrix} 0.1 & 0.9 & 0 \\ 0 & 0.4 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}$$

- Emission probability $p(x | q = j) = b_j(x)$
- Observations $X = x_{1:T} = x_1, x_2, \dots, x_T$

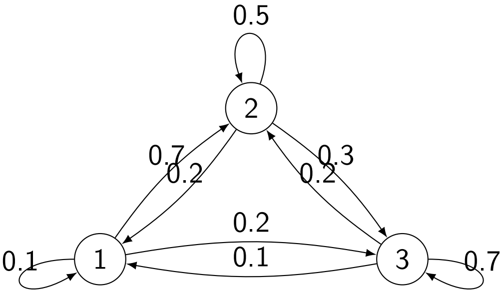
Transition Probabilities and State Machines

$$\begin{bmatrix} 0.1 & 0.9 & 0 \\ 0 & 0.4 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}$$



Transition Probabilities and State Machines

$$\begin{bmatrix} 0.1 & 0.7 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$$



Joint Probability

$$p(X, Q) =$$

Joint Probability

$$p(X, Q) = p(q_1)$$

Joint Probability

$$p(X, Q) = p(q_1)p(x_1|q_1)$$

Joint Probability

$$p(X, Q) = p(q_1)p(x_1|q_1)p(q_2|q_1)p(x_2|q_2)$$

Joint Probability

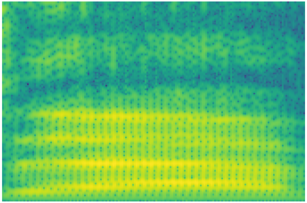
$$p(X, Q) = p(q_1)p(x_1|q_1)p(q_2|q_1)p(x_2|q_2) \cdots p(q_T|q_{T-1})p(x_T|q_T)$$

Joint Probability

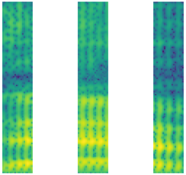
$$
\begin{aligned}
p(X, Q) &= p(q_1)p(x_1|q_1)p(q_2|q_1)p(x_2|q_2)\cdots p(q_T|q_{T-1})p(x_T|q_T) \\
&= p(q_1)p(x_1|q_1)\prod_{t=2}^{T} p(q_t|q_{t-1})p(x_t|q_t)
\end{aligned}
$$

$x_1\ x_2\ x_3\ x_4\ \cdots\ x_T$

frames



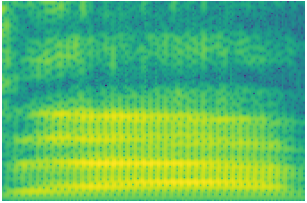
1 2 3



1 1 2 2 2 2 2 2 3 3

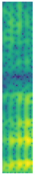
$x_1\ x_2\ x_3\ x_4 \quad \cdots \quad x_T$

frames

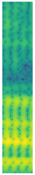


$q_1\ q_2\ q_3\ q_4 \quad \cdots \quad q_T$

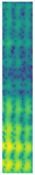
1



2



3



Viterbi Algorithm

$$\underset{Q}{\operatorname{argmax}}\, p(Q|X)$$

Viterbi Algorithm

$$\underset{Q}{\operatorname{argmax}}\, p(Q|X) = \underset{Q}{\operatorname{argmax}}\, \frac{p(X, Q)}{p(X)}$$

Viterbi Algorithm

$$\begin{aligned}\operatorname*{argmax}_Q p(Q|X) &= \operatorname*{argmax}_Q \frac{p(X, Q)}{p(X)} \\ &= \operatorname*{argmax}_Q p(X, Q) = \operatorname*{argmax}_{q_{1:T}} p(x_{1:T}, q_{1:T})\end{aligned}$$

Viterbi Algorithm

$$\operatorname*{argmax}_{Q} p(Q|X) = \operatorname*{argmax}_{Q} \frac{p(X, Q)}{p(X)}$$
$$= \operatorname*{argmax}_{Q} p(X, Q) = \operatorname*{argmax}_{q_{1:T}} p(x_{1:T}, q_{1:T})$$

- How many sequence should we consider in the max? $|S|^T = J^T$

A Useful Recurrsion

$$p(x_{1:t}, q_{1:t}) = p(q_1)p(x_1|q_1)\prod_{\tau=2}^{t} p(q_\tau|q_{\tau-1})p(x_\tau|q_\tau)$$

A Useful Recurrsion

$$
\begin{aligned}
p(x_{1:t}, q_{1:t}) &= p(q_1)p(x_1|q_1)\prod_{\tau=2}^{t} p(q_\tau|q_{\tau-1})p(x_\tau|q_\tau) \\
&= p(q_1)p(x_1|q_1)\prod_{\tau=2}^{t-1} p(q_\tau|q_{\tau-1})p(x_\tau|q_\tau)\left[p(q_t|q_{t-1})p(x_t|q_t)\right]
\end{aligned}
$$

A Useful Recurrsion

$$
\begin{aligned}
p(x_{1:t}, q_{1:t}) &= p(q_1)p(x_1|q_1)\prod_{\tau=2}^{t} p(q_\tau|q_{\tau-1})p(x_\tau|q_\tau) \\
&= p(q_1)p(x_1|q_1)\prod_{\tau=2}^{t-1} p(q_\tau|q_{\tau-1})p(x_\tau|q_\tau)\left[p(q_t|q_{t-1})p(x_t|q_t)\right] \\
&= p(x_{1:t-1}, q_{1:t-1})\left[p(q_t|q_{t-1})p(x_t|q_t)\right]
\end{aligned}
$$

Viterbi Algorithm

$$\max_{q_{1:t-1}} p(x_{1:t}, q_{1:t})$$

Viterbi Algorithm

$$\max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) = \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t}, q_{1:t})$$

Viterbi Algorithm

$$\max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) = \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t}, q_{1:t})$$
$$= \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) p(q_t | q_{t-1}) p(x_t | q_t)$$

Viterbi Algorithm

$$
\begin{aligned}
\max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) &= \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t}, q_{1:t}) \\
&= \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) p(q_t | q_{t-1}) p(x_t | q_t) \\
&= \max_{q_{t-1}} p(q_t | q_{t-1}) p(x_t | q_t) \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1})
\end{aligned}
$$

Viterbi Algorithm

$$V_{q_t}(t) = \max_{q_{1:t-1}} p(x_{1:t}, q_{1:t})$$

Viterbi Algorithm

$$\begin{aligned} V_{q_t}(t) &= \max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) \\ &= \max_{q_{t-1}} p(q_t | q_{t-1}) p(x_t | q_t) \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) \end{aligned}$$

Viterbi Algorithm

$$
\begin{aligned}
V_{q_t}(t) &= \max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) \\
&= \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t) \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) \\
&= \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t) V_{q_{t-1}}(t-1)
\end{aligned}
$$

Viterbi Algorithm

$$
\begin{aligned}
V_{q_t}(t) &= \max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) \\
&= \max_{q_{t-1}} p(q_t|q_{t-1}) p(x_t|q_t) \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) \\
&= \max_{q_{t-1}} p(q_t|q_{t-1}) p(x_t|q_t) V_{q_{t-1}}(t-1)
\end{aligned}
$$

$$
V_{q_1}(1) = p(q_1)
$$

Viterbi Algorithm

$$V_{q_t}(t) = \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)V_{q_{t-1}}(t-1)$$
$$V_{q_1}(1) = p(q_1)$$

Viterbi Algorithm

$$V_{q_t}(t) = \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)V_{q_{t-1}}(t-1)$$
$$V_{q_1}(1) = p(q_1)$$

$$V_j(t) = \max_{i=1,\ldots,J} p(q_t = j|q_{t-1} = i)p(x_t|q_t = j)V_i(t-1)$$

Viterbi Algorithm

$$V_{q_t}(t) = \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)V_{q_{t-1}}(t-1)$$
$$V_{q_1}(1) = p(q_1)$$

$$\begin{aligned} V_j(t) &= \max_{i=1,\ldots,J} p(q_t = j|q_{t-1} = i)p(x_t|q_t = j)V_i(t-1) \\ &= \max_{i=1,\ldots,J} a_{ij}b_j(x_t)V_i(t-1) \end{aligned}$$

Viterbi Algorithm

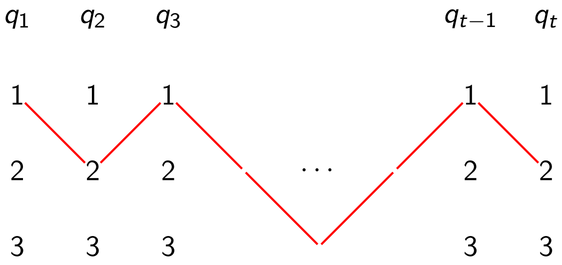
$$V_{q_t}(t) = \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)V_{q_{t-1}}(t-1)$$

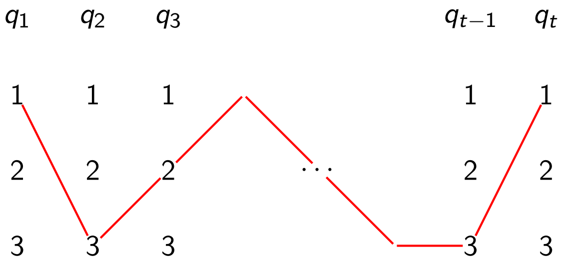
$$V_{q_1}(1) = p(q_1)$$

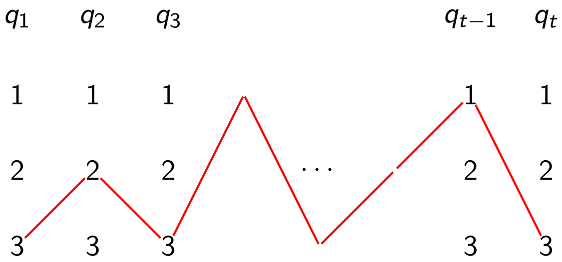
$$\begin{aligned} V_j(t) &= \max_{i=1,\ldots,J} p(q_t = j|q_{t-1} = i)p(x_t|q_t = j)V_i(t-1) \\ &= \max_{i=1,\ldots,J} a_{ij}b_j(x_t)V_i(t-1) \end{aligned}$$

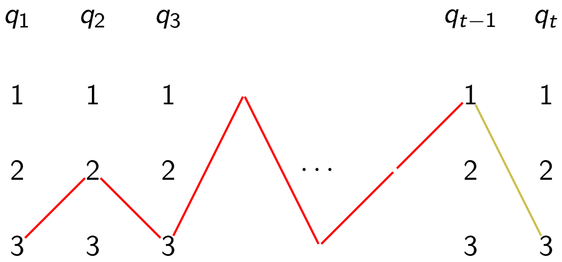
$$V_i(1) = p(q_1 = i) = \pi_i$$

q_1	q_2	q_3		q_{t-1}	q_t
1	1	1		1	1
2	2	2	$\cdots$	2	2
3	3	3		3	3

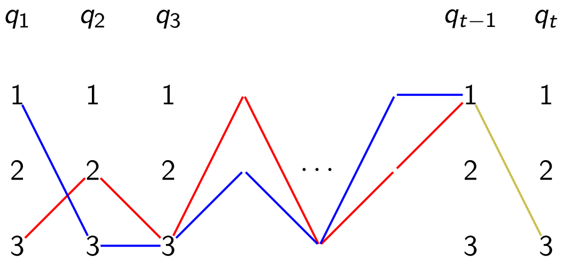

q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 ... 2 2
3 3 3 3 3


q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 ... 2 2
3 3 3 3 3

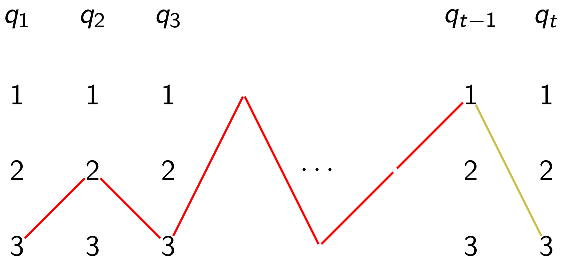

q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 ⋯ 2 2
3 3 3 3 3



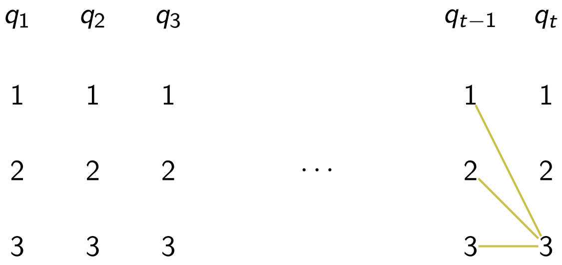
q_1
q_2
q_3
q_{t-1}
q_t
1 1 1 1 1
2 2 2 · · · 2 2
3 3 3 3 3

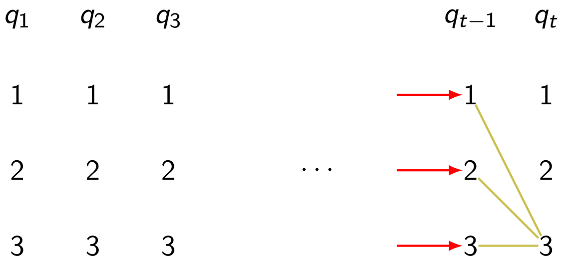


q_1
q_2
q_3
q_{t-1}
q_t
1
1
1
1
1
2
2
2
$\cdots$
2
2
3
3
3
3
3



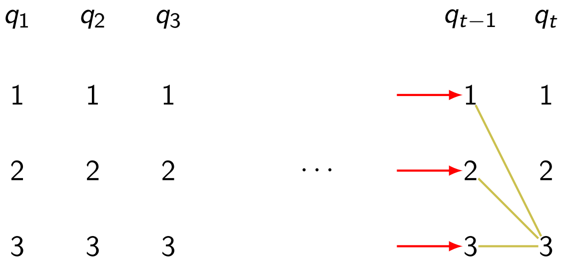
q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 ··· 2 2
3 3 3 3 3


q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 … 2 2
3 3 3 3 3

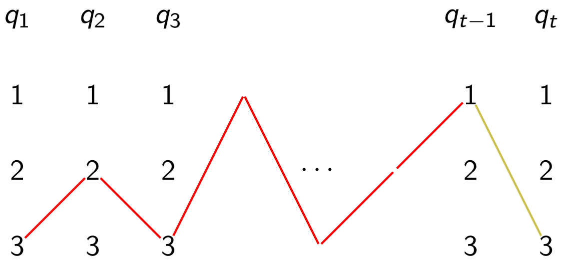

q_1
q_2
q_3
q_{t-1}
q_t
1 1 1 1 1
2 2 2 ⋯ 2 2
3 3 3 3 3

Viterbi Algorithm

$$V_{q_t}(t) = \max_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)V_{q_{t-1}}(t-1)$$


q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 · · · 2 2
3 3 3 3 3


q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 ⋯ 2 2
3 3 3 3 3


q_1 q_2 q_3 q_{t-1} q_t
1 1 1 1 1
2 2 2 ... 2 2
3 3 3 3 3

Backtracking

$$
\begin{aligned}
B_{q_t}(t) &= \operatorname*{argmax}_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) \\
&= \operatorname*{argmax}_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t) \operatorname*{argmax}_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) \\
&= \operatorname*{argmax}_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t) B_{q_{t-1}}(t-1)
\end{aligned}
$$

Backtracking

$$B_{q_t}(t) = \operatorname*{argmax}_{q_{1:t-1}} p(x_{1:t}, q_{1:t})$$
$$= \operatorname*{argmax}_{q_{t-1}} p(q_t|q_{t-1}) p(x_t|q_t) \operatorname*{argmax}_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1})$$
$$= \operatorname*{argmax}_{q_{t-1}} p(q_t|q_{t-1}) p(x_t|q_t) B_{q_{t-1}}(t-1)$$

$$B_j(t) = \operatorname*{argmax}_{i=1,\dots,J} p(q_t = j|q_{t-1} = i) p(x_t|q_t = j) B_i(t-1)$$
$$= \operatorname*{argmax}_{i=1,\dots,J} a_{ij} b_j(x_t) B_i(t-1)$$

Backtracking

$$\hat{q}_T = \operatorname*{argmax}_{q_T} V_{q_T}(T)$$

$$\hat{q}_{t-1} = B_{\hat{q}_t}(t) \qquad \text{for } t = T, \ldots, 2$$

Backtracking

$$\hat{q}_T = \operatorname*{argmax}_{q_T} V_{q_T}(T)$$

$$\hat{q}_{t-1} = B_{\hat{q}_t}(t) \qquad \text{for } t = T, \ldots, 2$$

$$\hat{q}_T = \operatorname*{argmax}_{i=1,\ldots,J} V_i(T)$$

$$\hat{q}_{t-1} = B_{\hat{q}_t}(t) \qquad \text{for } t = T, \ldots, 2$$

Viterbi Algorithm

- Recurrsion

$$p(x_{1:t}, q_{1:t}) = p(x_{1:t-1}, q_{1:t-1})\left[p(q_t|q_{t-1})p(x_t|q_t)\right]$$

- Optimal path argument

  If a path is optimal, then all of its subpaths are optimal. Otherwise, we can always replace the non-optimal subpaths to get a better path.

$$
\begin{aligned}
\max_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) &= \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t}, q_{1:t}) \\
&= \max_{q_{t-1}} \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) p(q_t | q_{t-1}) p(x_t | q_t) \\
&= \max_{q_{t-1}} p(q_t | q_{t-1}) p(x_t | q_t) \max_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1})
\end{aligned}
$$

$$
\begin{aligned}
\sum_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) &= \sum_{q_{t-1}} \sum_{q_{1:t-2}} p(x_{1:t}, q_{1:t}) \\
&= \sum_{q_{t-1}} \sum_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1}) p(q_t | q_{t-1}) p(x_t | q_t) \\
&= \sum_{q_{t-1}} p(q_t | q_{t-1}) p(x_t | q_t) \sum_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1})
\end{aligned}
$$

Forward Algorithm

$$\alpha_{q_t}(t) = \sum_{q_{1:t-1}} p(x_{1:t}, q_{1:t})$$
$$= \sum_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t) \sum_{q_{1:t-2}} p(x_{1:t-1}, q_{1:t-1})$$
$$= \sum_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)\alpha_{q_{t-1}}(t-1)$$

$$\alpha_{q_1}(1) = p(q_1)$$

Forward Algorithm

$$\alpha_{q_t}(t) = \sum_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)\alpha_{q_{t-1}}(t-1)$$

$$\alpha_{q_1}(1) = p(q_1)$$

Forward Algorithm

$$\alpha_{q_t}(t) = \sum_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)\alpha_{q_{t-1}}(t-1)$$

$$\alpha_{q_1}(1) = p(q_1)$$

$$\alpha_j(t) = \sum_{i=1}^{J} p(q_t = j|q_{t-1} = i)p(x_t|q_t = j)\alpha_i(t-1)$$

Forward Algorithm

$$\alpha_{q_t}(t) = \sum_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)\alpha_{q_{t-1}}(t-1)$$
$$\alpha_{q_1}(1) = p(q_1)$$

$$\alpha_j(t) = \sum_{i=1}^{J} p(q_t = j|q_{t-1} = i)p(x_t|q_t = j)\alpha_i(t-1)$$
$$= \sum_{i=1}^{J} a_{ij} b_j(x_t)\alpha_i(t-1)$$

Forward Algorithm

$$\alpha_{q_t}(t) = \sum_{q_{t-1}} p(q_t|q_{t-1})p(x_t|q_t)\alpha_{q_{t-1}}(t-1)$$
$$\alpha_{q_1}(1) = p(q_1)$$

$$\alpha_j(t) = \sum_{i=1}^{J} p(q_t = j|q_{t-1} = i)p(x_t|q_t = j)\alpha_i(t-1)$$
$$= \sum_{i=1}^{J} a_{ij}b_j(x_t)\alpha_i(t-1)$$
$$\alpha_i(1) = p(q_1 = i) = \pi_i$$

Marginal Probability

$$\alpha_{q_t}(t) = \sum_{q_{1:t-1}} p(x_{1:t}, q_{1:t})$$

Marginal Probability

$$\alpha_{q_t}(t) = \sum_{q_{1:t-1}} p(x_{1:t}, q_{1:t}) = p(x_{1:t}, q_t)$$

Marginal Probability

$$\alpha_{q_t}(t) = \sum_{q_{1:t-1}} p(x_{1:t}, q_{1:t})$$
$$= p(x_{1:t}, q_t)$$

$$p(X) = p(x_{1:T}) = \sum_{q_T} \alpha_{q_T}(T) = \sum_{i=1}^{J} \alpha_i(T)$$

Hidden Markov Models

- hidden, because $Q = q_{1:T}$ is not observed
- Markov property

$$\begin{aligned} p(x_t|q_{1:t}) &= p(x_t|q_1, q_2, \dots, q_t) \\ &= p(x_t|q_t) \end{aligned}$$

Further Reading

- Chapter 6, Rabiner and Juang, "Fundamentals of Speech Recognition," 1993