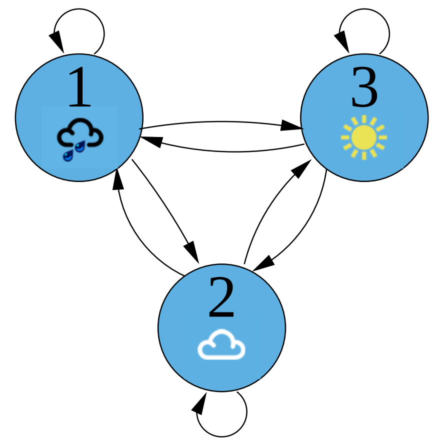
The Hidden Markov Model

Dr Philip Jackson

- Markov models
- State topology diagrams
- Hidden Markov models
  - Likelihood calculation
  - Recognition & training



Conclusion of Dynamic Time Warping

DTW computes scores efficiently with some flexibility in the alignment, treating templates as deterministic patterns with residual noise.

Problems:

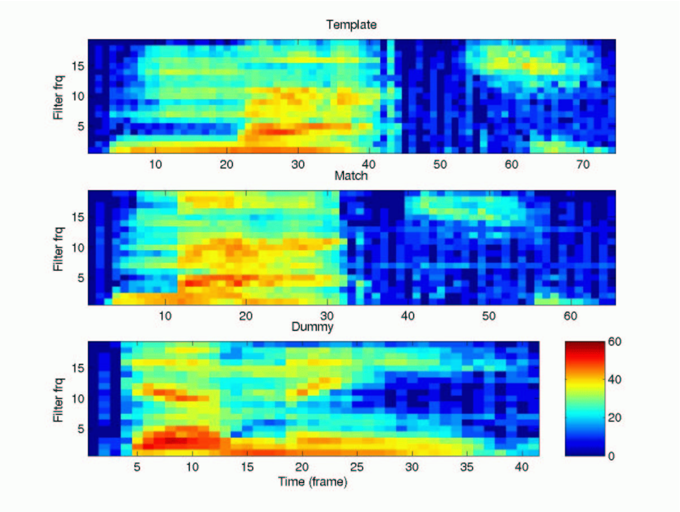
1. How much flexibility should we allow?
2. How should we penalise any warping?
3. How do we determine a fair distance metric?
4. How many templates should we register?
5. How do we select the best ones?

Approach:

- Learn from the statistics of speech data. . .

Characteristics of the desired model

1. sequence evolution is not deterministic
2. observations are coloured by their state
3. the state is not observed directly
4. stochastic sequence + stochastic observations

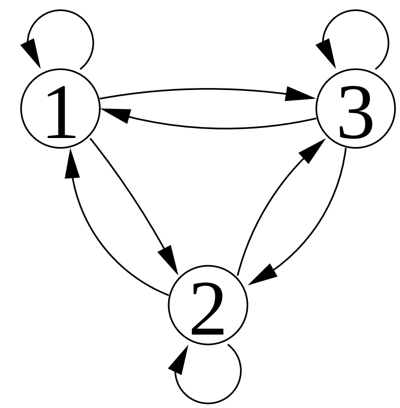


Applications:

- automatic speech recognition
- optical character recognition
- protein and DNA sequencing
- speech synthesis
- noise-robust data transmission
- crytoanalysis
- machine translation
- image classification, etc.

Introduction to Markov Models

We can model stochastic sequences of discrete states with a Markov chain; the state transitions have probabilities



For 1st-order Markov chains, the state transition probability depends only on the previous state (Rabiner, 1989):

$$P(x_t = j | x_{t-1} = i, x_{t-2} = h, \ldots) \approx P(x_t = j | x_{t-1} = i) \quad (1)$$

So, if we assume the RHS of eq. 1 is independent of time, we can write the **state-transition probabilities**

$$a_{ij} = P(x_t = j | x_{t-1} = i), \qquad 1 \leq i, j \leq N \quad (2)$$

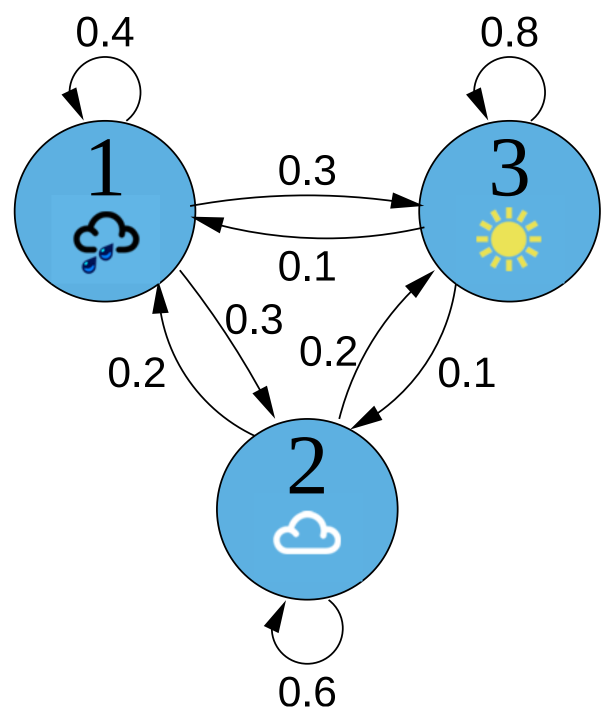
with the usual properties of probabilities

$$a_{ij} \geq 0 \qquad \text{and} \qquad \sum_{j=1}^{N} a_{ij} = 1 \qquad \forall\ i, j \in 1..N$$

Weather prediction example

We represent the state of the weather by a 1st-order, fully-connected Markov model, $\mathcal{M}$:

state 1: raining
state 2: cloudy
state 3: sunny



with state-transition probabilities expressed in matrix form:

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \quad (3)$$

Weather predictor probability calculation

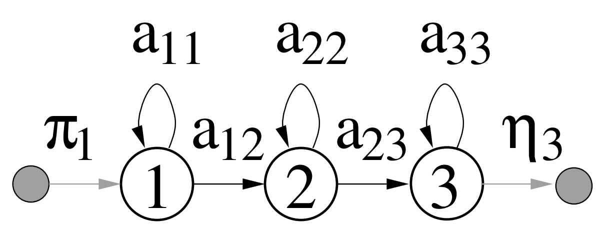
Given today's weather what is the probability of directly observing the sequence of weather states "rain-sun-sun" with model $\mathcal{M}$?

$$A = \begin{array}{r c} & \begin{array}{ccc} \text{rain} & \text{cloud} & \text{sun} \end{array} \\ \begin{array}{r} \text{rain} \\ \text{cloud} \\ \text{sun} \end{array} & \left[\begin{array}{ccc} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{array} \right] \end{array}$$

$$\begin{aligned} P(X|\mathcal{M}) &= P(X = \{1,3,3\}|\mathcal{M}) \\ &= P(x_1 = \text{rain}|\text{today}) \times P(x_2 = \text{sun}|x_1 = \text{rain}) \\ &\quad \times P(x_3 = \text{sun}|x_2 = \text{sun}) \\ &= a_{\ \ 1}\, a_{13}\, a_{33} \\ &= \quad \times 0.3 \times 0.8 \\ &= \end{aligned}$$

Start and end of a state sequence

Null states deal with the start and end of sequences, as in the state topology of this left-right Markov model:



Entry probabilities at t=1 for each state j are defined

$$\pi_j = P(x_1 = j) \qquad 1 \leq j \leq N \tag{4}$$

with the properties $\pi_j \geq 0$, and $\sum_{j=1}^{N} \pi_j = 1$ for $j \in 1..N$

Exit probabilities at t=T are similarly defined

$$\eta_i = P(x_T = i) \qquad 1 \leq i \leq N \tag{5}$$

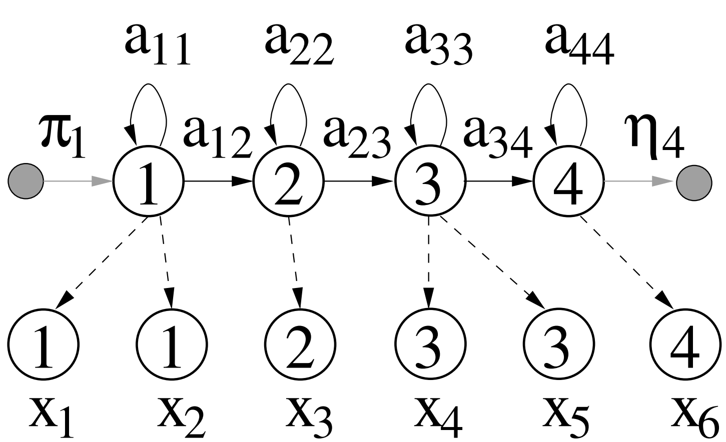
with properties $\eta_i \geq 0$, and $\eta_i + \sum_{j=1}^{N} a_{ij} = 1$ for $i \in 1..N$

Parameters of the Markov Model, $\mathcal{M}$

State transition probabilities,

$A = \{\pi_j, a_{ij}, \eta_i\} = \{P(x_t = j | x_{t-1} = i)\}$ for $1 \leq i, j \leq N$

where N is the number of states

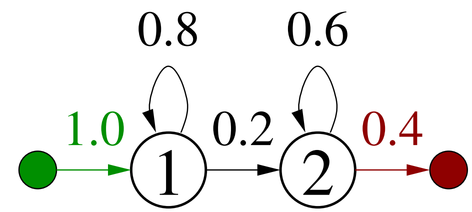


producing a sequence

$X = \{1, 1, 2, 3, 3, 4\}$

Example: probability of MM state sequence

Consider the state topology



state transition probabilities

$$A = \left[\begin{array}{c|cc|c} 0 & 1 & 0 & 0 \\ \hline 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ \hline 0 & 0 & 0 & 0 \end{array}\right]$$

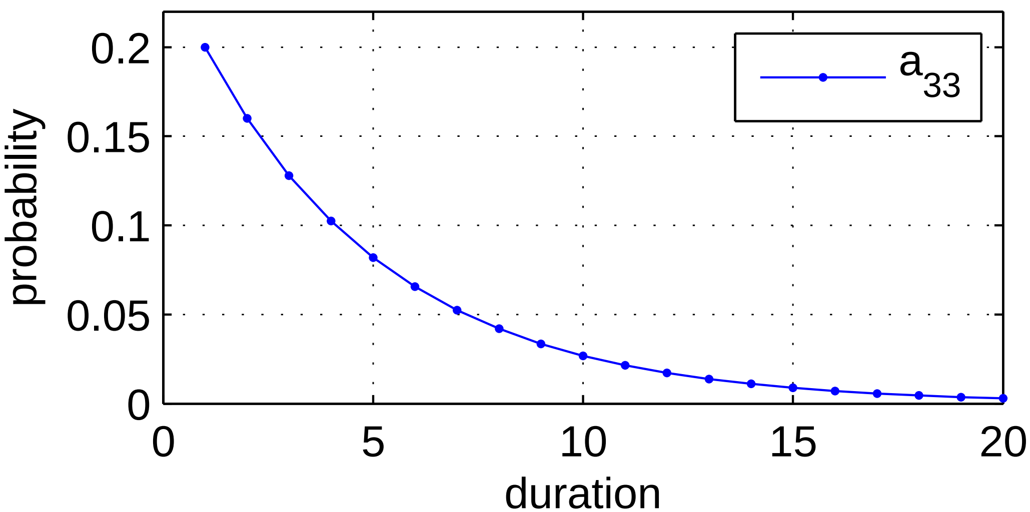
The probability of state sequence $X = \{1, 2, 2\}$ is

$$\begin{aligned} P(X|\mathcal{M}) &= \pi_1 \, a_{12} \, a_{22} \, \eta_2 \\ &= 1 \times 0.2 \times 0.6 \times 0.4 \\ &= 0.048 \end{aligned}$$

State duration characteristics

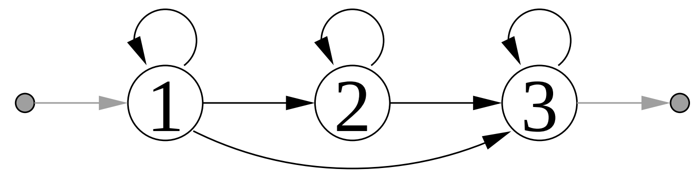
As a consequence of the first-order Markov model, the probability of occupying a state for a given duration, τ, decays exponentially:

$$p(X|x_1 = i, \mathcal{M}) = (a_{ii})^{\tau-1}\,(1 - a_{ii}) \tag{6}$$



Summary of Markov models

State topology diagram:



entry probabilities $\boldsymbol{\pi} = \{\pi_j\} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ and

exit probabilities $\boldsymbol{\eta} = \{\eta_i\} = \begin{bmatrix} 0 & 0 & 0.2 \end{bmatrix}^{\mathsf{T}}$ are combined with state transition probabilities in complete A matrix:

$$A = \left[\begin{array}{c|ccc|c} 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0.6 & 0.3 & 0.1 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

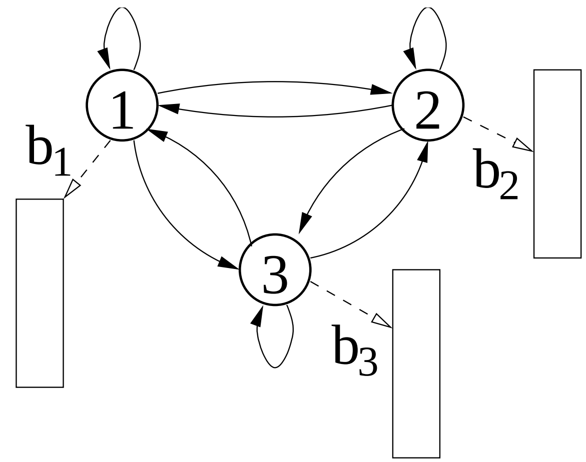
Probability of a given state sequence X:

$$P(X|\mathcal{M}) \;=\; \left(\prod_{t=1}^{T} a_{x_{t-1}x_t}\right)\eta_{x_T} \qquad (7)$$

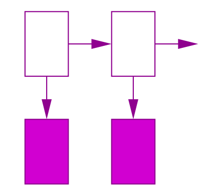
writing the entry probabilities as $a_{x_0x_1} = \pi_{x_1}$

Hidden Markov Models

HMMs use a Markov chain to model stochastic state sequences which then emit stochastic observations, e.g., the state topology of a fully-connected HMM:

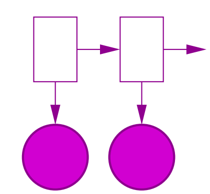


Probability of state i generating **discrete** observation o_t, which has a value from a finite set $k \in 1..K$, is



$$b_i(o_t) = P(o_t = k | x_t = i) \qquad (8)$$

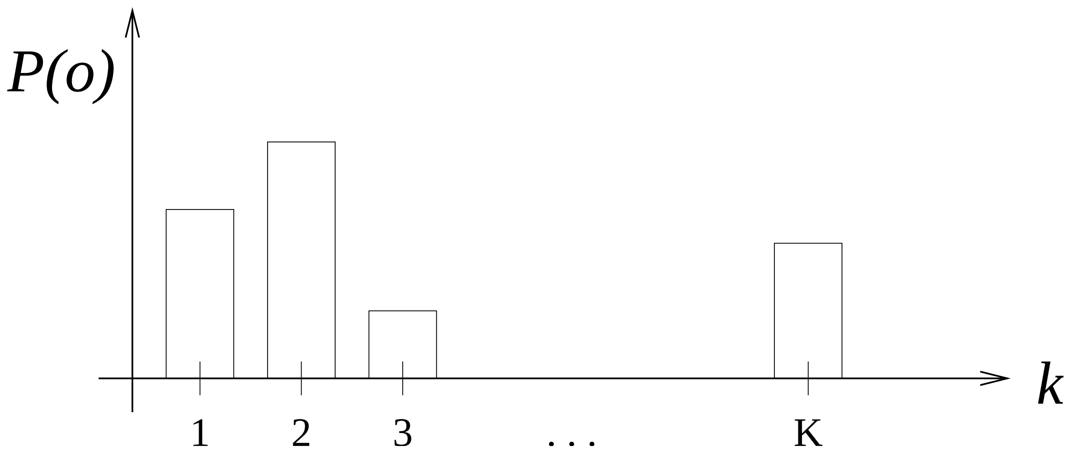
Probability distribution of a **continuous** observation o_t, which has a value from an infinite set, is



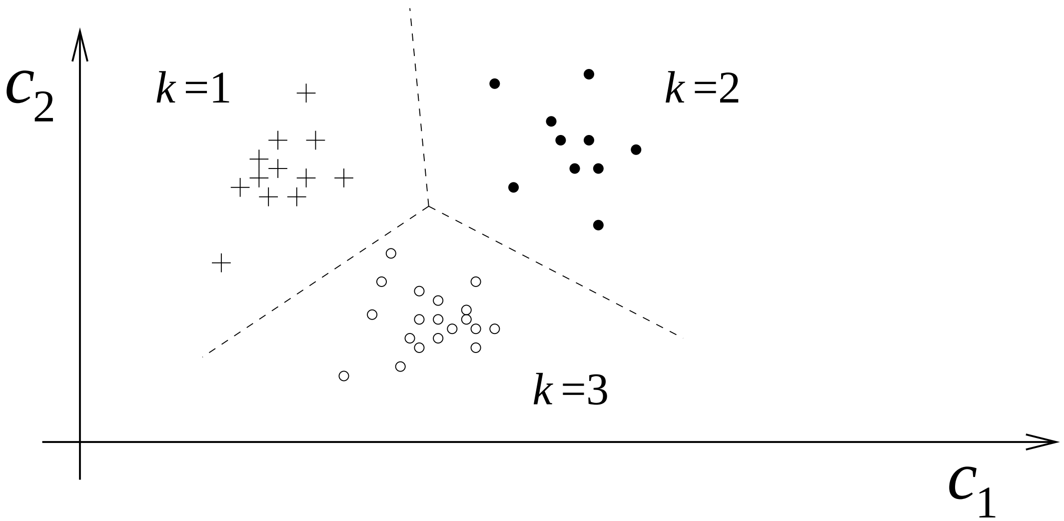
$$b_i(o_t) = p(o_t | x_t = i) \qquad (9)$$

We begin by considering only discrete observations.

Discrete output probabilities



Observations in discretised feature space



Parameters of a discrete HMM, λ

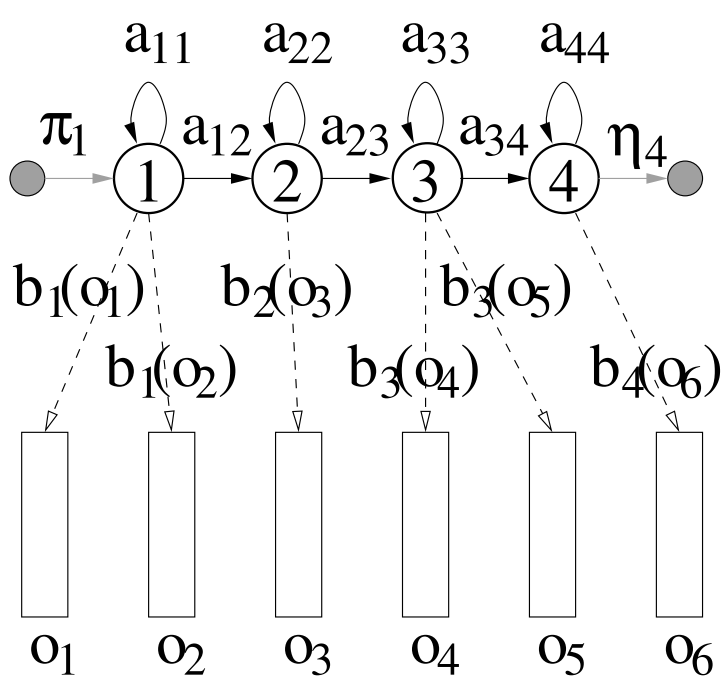
State transition probabilities,

$$A = \{\pi_j, a_{ij}, \eta_i\} = \{P(x_t = j | x_{t-1} = i)\} \qquad \text{for } 1 \leq i, j \leq N$$

Discrete output probabilities,

$$B = \{b_i(k)\} = \{P(o_t = k | x_t = i)\} \qquad \text{for } 1 \leq i \leq N, \quad 1 \leq k \leq K$$

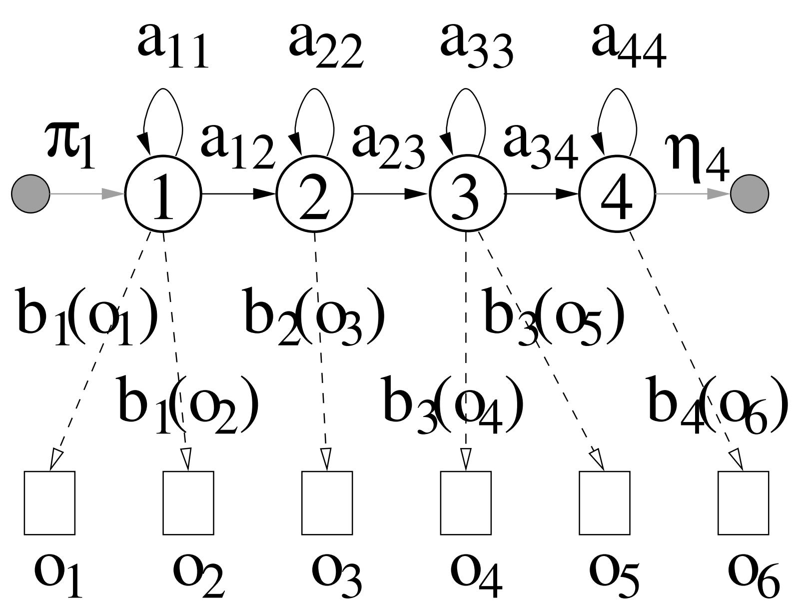
with N states, K observation types



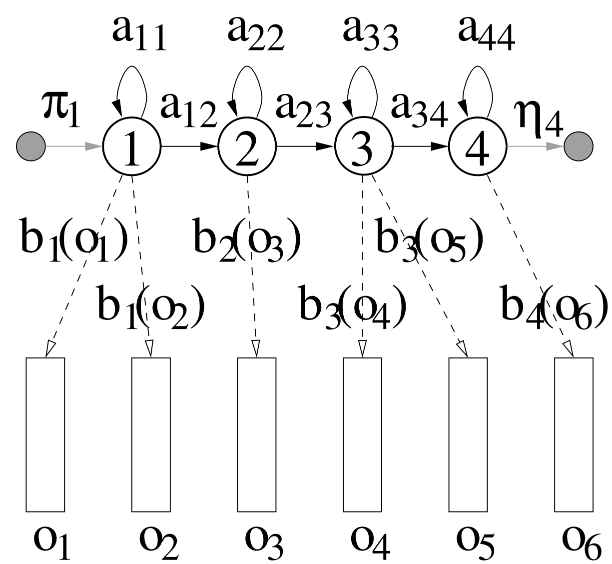
generating
a state sequence
$X = \{1, 1, 2, 3, 3, 4\}$
and observations
$\mathcal{O} = \{o_1, o_2, \ldots, o_6\}$

Procedure for generating an observation sequence

1. For $t = 1$, choose state $x_t = j$ using entry probability π_j
2. Select $o_t = k$ according to $b_{x_t}(k)$
3. Transit according to a_{ij} and η_i, then respectively:
   (a) increment t, set $x_t = j$ and repeat from 2, or
   (b) terminate the sequence, $t = T$.



HMM probability calculation



The joint likelihood of state and observation sequences is

$$P(\mathcal{O}, X|\lambda) = P(X|\lambda)\,P(\mathcal{O}|X,\lambda)$$

For example, state sequence $X = \{1, 1, 2, 3, 3, 4\}$ produces the set of observations $\mathcal{O} = \{o_1, o_2, \ldots, o_6\}$:

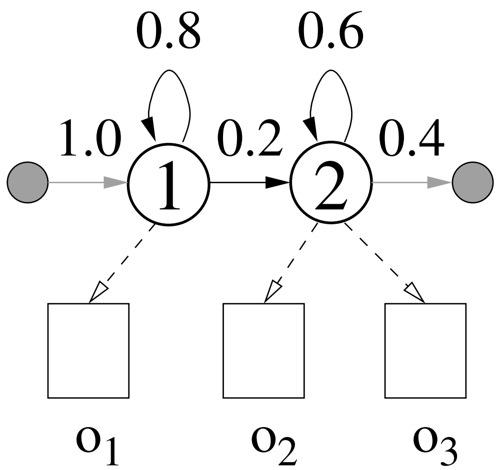
$$P(X|\lambda) = \pi_1\, a_{11}\, a_{12}\, a_{23}\, a_{33}\, a_{34}\, \eta_4 = \left(\prod_{t=1}^{T} a_{x_{t-1}x_t}\right)\eta_{x_T}$$

$$\begin{aligned} P(\mathcal{O}|X,\lambda) &= b_1(o_1)\, b_1(o_2)\, b_2(o_3)\, b_3(o_4)\, b_3(o_5)\, b_4(o_6) \\ &= \prod_{t=1}^{T} b_{x_t}(o_t) \end{aligned}$$

$$P(\mathcal{O}, X|\lambda) = \left(\prod_{t=1}^{T} a_{x_{t-1}x_t} b_{x_t}(o_t)\right)\eta_{x_T} \qquad (10)$$

Example: probability of HMM state sequence

Consider state topology and state transition matrix:



$$A = \left[\begin{array}{c|cc|c} 0 & 1 & 0 & 0 \\ \hline 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ \hline 0 & 0 & 0 & 0 \end{array}\right]$$

Output probabilities:

$$B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{array}{c} \begin{array}{ccc} \mathsf{R} & \mathsf{G} & \mathsf{B} \end{array} \\ \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0 & 0.9 & 0.1 \end{bmatrix} \end{array}$$

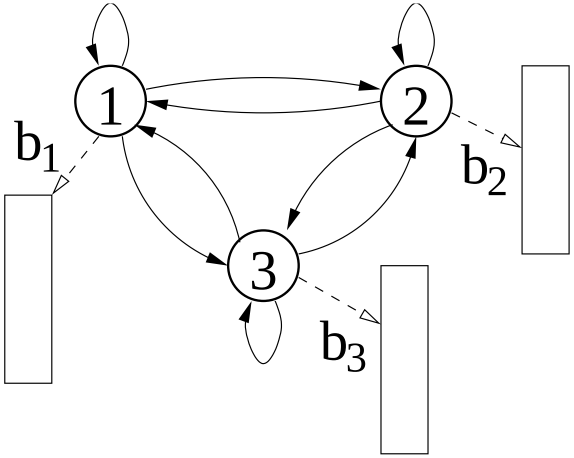
Probability of observations with state sequence $X = \{1, 2, 2\}$:

$$\begin{aligned} P(\mathcal{O}, X|\lambda) &= \left(\prod_{t=1}^{T} a_{x_{t-1}x_t} b_{x_t}(o_t) \right) \eta_{x_T} \\ &= \pi_1 b_1(o_1)\, a_{12} b_2(o_2)\, a_{22} b_2(o_3)\, \eta_2 \\ &= 1 \times \\ &= \end{aligned}$$

HMM summary

- Markov models
  - sequence of directly observable states
  - state topology diagram

- Hidden Markov models (HMMs)
  - hidden state sequence
  - generation of observations
  - likelihood calculation



HMM Recognition & Training

Three tasks within HMM framework

1. Compute likelihood of a set of observations for a given model, $P(\mathcal{O}|\lambda)$
2. Decode a test sequence by calculating the most likely path, X^*
3. Optimise pattern templates by training the model parameters, $\Lambda = \{\lambda\}$