Hidden Markov Models

- System is in one *state* at a time
- Discrete time

Markov Property

Probability of being in one state depends only on the immediately previous state

- State transitions
- Every state transition has a constant probability

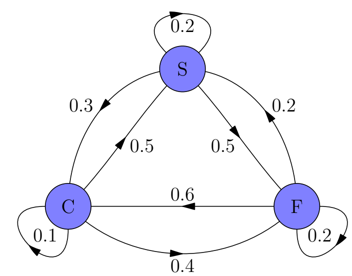
Transition Matrix

$a_{i,j}$ — probability of being in state j when you were in state i in the preceding step

Halloween Example:
Our local witch

1. Cooking potion (C)
2. Practising spells (S)
3. Out flying (F)

$C : 1 \qquad S : 2 \qquad F : 3$



$$A = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.3 & 0.2 & 0.5 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

The transition matrix can be used for "prediction"

- If we know the probability distribution at time t: π_s
- Distribution at time $t+1$ is then $A^T\vec{\pi}$
- Distribution at time $t+n$ is then $(A^T)^n\vec{\pi}$

Hidden Markov Model

Markov model where the states are not (directly) observable

Witch example

- We don't actually *know* what the witch is doing
- We can only observe if smoke is coming from the chimney
- We must *infer* what she is probably doing

Emission Matrix

$b_{i,k}$ — probability of observing k when you are in state i

Witch example

$$\textbf{States}: C:1 \quad S:2 \quad F:3$$

$$\textbf{Observations}: \text{NoSmoke}:1 \quad \text{Smoke}:2$$

$$A = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.3 & 0.2 & 0.5 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} \quad B = \begin{bmatrix} 0.1 & 0.9 \\ 0.5 & 0.5 \\ 1.0 & 0.0 \end{bmatrix} \quad \pi = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$$

Hidden Markov Model algorithms

- What it the probability for a specific observation sequence?
- Which is the most likely state sequence behind an observation sequence?
- How to measure the values for A and B from many observation sequences?

What is the probability for a specific observation sequence, $o_1, o_2, \ldots, o_T$?

Forward Algorithm

Efficient algorithm to calculate $P(o_1, o_2, \ldots, o_T | A, B)$

Combining probabilities

- x **and** y must happen: $P = P(x) \cdot P(y)$
- x **or** y can happen: $P = P(x) + P(y)$

Forward Algorithm

Key idea: Compute $\alpha_{s,t} \equiv$ Probability of being in state s at time t and making the observations $o_1, o_2, \ldots, o_t$

- Base case

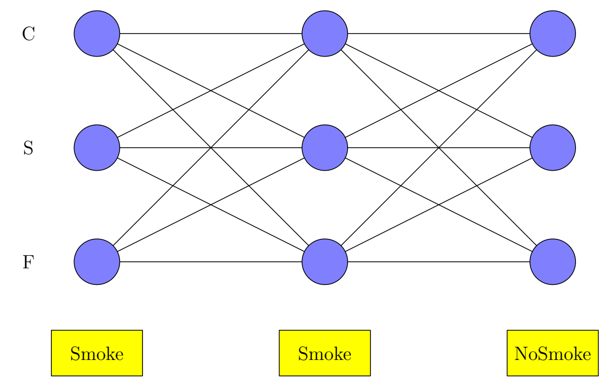
$$\alpha_{s,1} = \pi_s b_{s,o_1}$$

- Recursive computation

$$\alpha_{s,t} = b_{s,o_t} \sum_i a_{i,s} \alpha_{i,t-1}$$

- Termination

$$P(o_1, o_2, \ldots, o_T | A, B) = \sum_i \alpha_{i,T}$$



Backward Algorithm

Alternative algorithm calculating from the end

$\beta_{s,t}$ — Probability of being in state s at time t and then making the observations $o_{t+1}, o_{t+2}, \ldots, o_T$

Which is the most likely state sequence behind a specific observation sequence?

Viterbi Algorithm

Solution via Dynamic Programming

- It is sufficient to keep track of the best path to every state for each time
- Key idea: compute $v_{s,t} \equiv$ probability of most likely path which ends in state s at time t

$v_{s,t} \equiv$ probability of most likely path which ends in state s at time t

- Initialization
$$v_{s,1} = \pi_s b_{s,o_1}$$
- Forward pass
$$v_{s,t} = \max_i v_{i,t-1} a_{i,s} b_{s,o_t}$$
$$p_{s,t} = \arg\max_i v_{i,t-1} a_{i,s}$$
- Unwinding the most likely state sequence
$$q_T = \arg\max_i v_{i,T}$$
$$q_t = p_{q_{t+1},t+1}$$

Baum-Welch Algorithm

Expectation-Maximization (EM) method to adjust model parameters $(a_{i,j}, b_{i,k})$ from a large number of observation sequences $(o_1, o_2, \ldots, o_T)$

- Expectation
Use variant of forward and backward algorithms to estimate probable transitions and emissions
- Maximization
Use these transitions and emissions to update the model

- $\alpha_{s,t}$ — based on observations up to time t
- $\beta_{s,t}$ — based on observations from time t

Key idea: compute $\gamma_{i,j,t} \equiv$ Probability that we, at time t, made the transition from i to j

$$\gamma_{i,j,t} = \frac{\alpha_{i,t} a_{i,j} b_{i,o_k} \beta_{j,t+1}}{\sum_i \sum_j \alpha_{i,t} a_{i,j} b_{i,o_k} \beta_{j,t+1}}$$

Update model parameters

$$a_{i,j} = \frac{\sum_t \gamma_{i,j,t}}{\sum_t \sum_m \gamma_{i,m,t}}$$

$$b_{i,k} = \frac{\sum_{t,o_t=k} \sum_m \gamma_{i,m,t}}{\sum_t \sum_m \gamma_{i,m,t}}$$