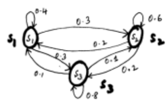
Hidden Markov

- In one afternoon the weather is one of the following
- State 1: Rain
- State 2: Cloud
- State 3: Sunny
- We postulate that the weather on day t is
  - characterized one of the three state.
  - State transition probability
  - Weather on day1 (t=1) is sunny (state 3)
- A = {aij}



.4	.3	.3
.2	.6	.2
.1	.1	0.8

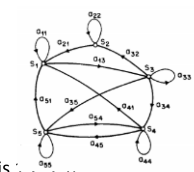
What is the probability that weather for
Next seven days
sun sun rain rain sun cloudy rain?

- Observation sequence = O =
  - S3 S3, S3, S1, S1, S3, S2, S1.
- Time t =
  - t1 t2 t3 t4 t5 t6 t7 t8
- We want to compute the probability given the model.

P(O |M)=P(S3 S3, S3, S1, S1, S3, S2, S1 | Model)
= P(s3) p(s3|s3) p(S3|s3) p(s1|s3) p(s1|s1)
p(s3|s1) p(s2|s3) p(s1|s2) = α_3 a_{33} a_{33} a_{31} a_{11} a_{13} a_{32} a_{21} = 0.8 * 0.8 * 0.1 * 0.4 * 0.3 * 0.1 * 0.2

- We can also ask if the state is known today what is the probability that it will be in same state for next n days.
  - $S_{i1}, S_{i2}, S_{i3} \ldots S_{id}\ S_{j(d+1)}$ i≠j
  - P(o|Model, q1=si) = $(a_{ii})^d (1-a_{ii}) = p_i (d)$

Discrete Markov process



- 5 distinct state of model
  - $S_1 \ldots S_5$
- Systems is going through change
  - Same state or different state
  - The change of state is subject to
    - Probability at that state
  - Time associated with each change is 1,2,3....
- Discrete first order Markov chain is sequence of states
  - It depends on current state and previous state
  - $P(q_t = s_i \mid q_{t-1} = s_j)$ state transition coefficient $a_{ij} \geq 0$

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

Discrete Markov process

- Applications where successive instances are dependent.
  - Such processes where there is a *sequence* of observations
    - letters in a word, h after t not after x (the)
    - Base pairs in a DNA sequence—cannot be modeled as simple probability distributions.
    - speech utterances are composed of speech primitives called phonemes
      - Certain sequences of phonemes are allowed as per the words of the language.

Discrete Markov process

- A sequence can be characterized as being generated by a *parametric random process*.
  - how this modeling is done
  - how the parameters of such a model can be learned from a training sample of example sequences
- *N* distinct states: $S_1, S_2, \ldots, S_N$
- *State at time t =>* q_t $t = 1, 2, \ldots,$
- *State at time t is Si =>* $q_t = S_i$ (space, position)

- The system moves to a state j with a probability (dependent on previous state)
  - $P(q_{t+1} = S_j | q_t = S_i, q_{t-1} = S_k, \cdots)$
- First order Markov model
  - $P(q_{t+1} = S_j | q_t = S_i)$
  - $a_{ij} \equiv P(q_{t+1} = S_j | q_t = S_i)$, $a_{ij} \geq 0$ *and* $\sum a_{ij} = 1$
  - Transition probability is independent of time
  - Move state S_i to S_j with probability.
  - Require Initial probability, $\pi_i \equiv P(q_1 = S_i)$

Observable Markov model

- The states are observable
  - we know q_t, At any time t
  - as the system moves from one state to another, we get an observation sequence that is a sequence of states.
  - The O/P of process
    - the set of states at each instant of time
      - each state corresponds to a physical observable event
  - *Observation sequence O =*
    - Maps to State Sequence $Q = \{q_1 q_2 \cdots q_T\}$

Observable Markov model

- $P(O = Q | A, \Pi) = P(q_1) \prod_{t=1}^{n} P(qt | q_{t-1})$
  $= \pi_{q1} a_{q1q2} \cdots a_{qT\text{-}1qT}$

N urns each urn contains balls of only one color. So there is

- an urn of red balls, another of blue balls, green
- S1 = Red, S2 = Blue, S3 = Green
- q_t denote the color of the ball drawn at time t
- Initial probability = $\pi_{q1} = [0.5, 0.2, 0.3]^T$
- A = Transition Matrix = a_{ij} = $\begin{matrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{matrix}$

Observable Markov model

- Given π and A, it is easy to generate *K* random sequences each of length *T*
  - “red, red, green, green
  - $O = \{S1, S1, S3, S3\}$.

$P(O| \pi, A) = (PS_1)\, P(S_1|S_1)\, P(S_3|S_1)\, P(S_3|S_3)$
$= \pi_1 * a_{11} * a_{13} * a_{33} = 0.5 * 0.4 * 0.3 * 0.8$

We have K sequences of length T, How to find π, A ?

π_i = #{sequences starting with Si}/ #{sequences}

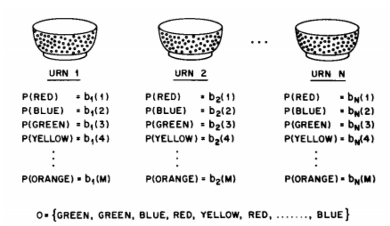
a_{ij} = #{transitions from S_i to S_j} / #{transitions from S_i}

Hidden Markov Models

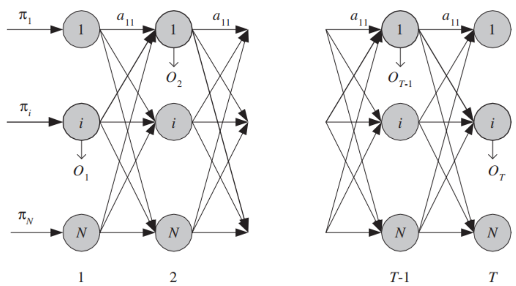
- The states are not observable
- we visit a state
  - an observation is recorded that is a probabilistic function of the state
  - a discrete observation in each state from the set $\{v1, v2, \ldots, v_M\}$
  - *Observation probability* = $b_j(m) \equiv P(O_t = v_m | q_t = S_j)$
    - Emission Probability
  - We have only observation sequence. We need to infer the state from observation based on probability.
  - We want the state sequence with maximum likelihood of generating the observation.

Hidden Markov Models

- Two sources of randomness.
  - Randomly moving from one state to another,
  - The observation in a state is also random.



Hidden Markov Models



Hidden Markov Models

- Elements of HMM
1. *N*: Number of states in the model $S = \{S1, S2, \ldots, SN\}$
2. *M*: Number of distinct observation symbols in the state $V = \{v1, v2, \ldots, vM\}$
3. State transition probabilities
   $A = [aij]$ $aij \equiv P(qt+1 = Sj | qt = Si)$
4. Observation probabilities
   $B = [bj(m)]$ $bj(m) \equiv P(Ot = vm | qt = Sj)$
5. Initial state probabilities: $\Pi = [\pi i]$
   $\pi i \equiv P(q1 = Si)$

HMM Model has $\lambda = (A, B, \Pi)$ implicit definition of N,M

Hidden Markov Models

- If we have, HMM Model, $\lambda = (A, B, \Pi)$
  - the model can be used to generate an arbitrary number of observation sequences of arbitrary length
- We need to estimate model parameters given a training set of sequences.
  - $X\{O^k\}_k$
  - We want λ^* = maximum $P(X \mid \lambda)$

Reference

- Introduction to Machine Learning, Third Edition, Ethem Alpaydın, The MIT Press
  - Chapter 15.
  -