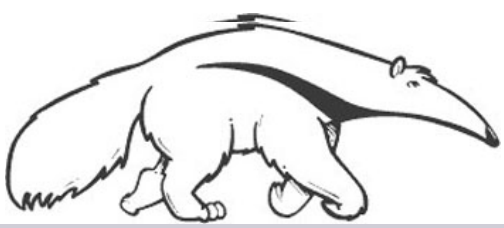
Markov Models

Introduction to Graphical Models

Prof. Alexander Ihler



BREN:ICS INFORMATION AND COMPUTER SCIENCES

UNIVERSITY of CALIFORNIA IRVINE

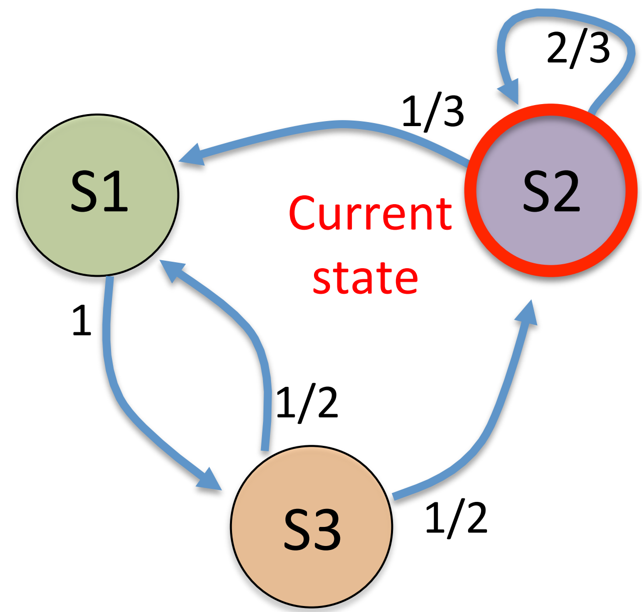
Markov system

- System has d states, $s_1 \ldots s_d$
- Discrete time intervals, t=0,1,…,T
- At time t, system is in state x_t

  $$x_0 \sim p(x_0)$$

- At each t, system transitions to another state according to

“State transition diagram”

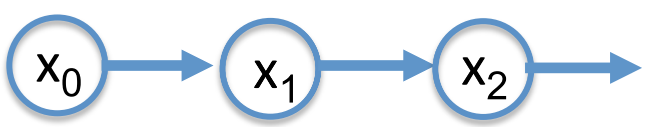


d = 3
t = 0
x_0 = 2

$$p(x_{t+1} \,|\, x_t) =$$

$$p(x_{t+1} = s_1 \,|\, x_t = s_1) = 0 \quad p(x_{t+1} = s_1 \,|\, x_t = s_2) = 0.33 \quad p(x_{t+1} = s_1 \,|\, x_t = s_3) = 0.5$$
$$p(x_{t+1} = s_2 \,|\, x_t = s_1) = 0 \quad p(x_{t+1} = s_2 \,|\, x_t = s_2) = 0.66 \quad p(x_{t+1} = s_2 \,|\, x_t = s_3) = 0.5$$
$$p(x_{t+1} = s_3 \,|\, x_t = s_1) = 1 \quad p(x_{t+1} = s_3 \,|\, x_t = s_2) = 0 \quad p(x_{t+1} = s_3 \,|\, x_t = s_3) = 0$$

- Bayes Net on states x over time: a “Markov chain”



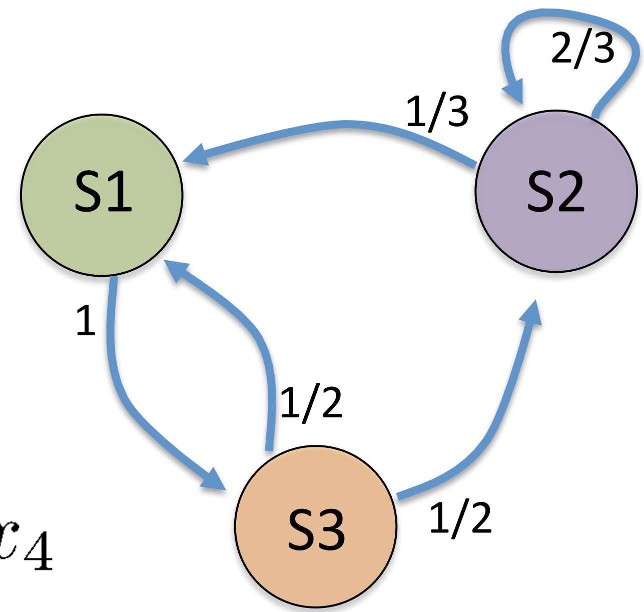
  – Each conditional probability distribution is identical (“homogeneous”)

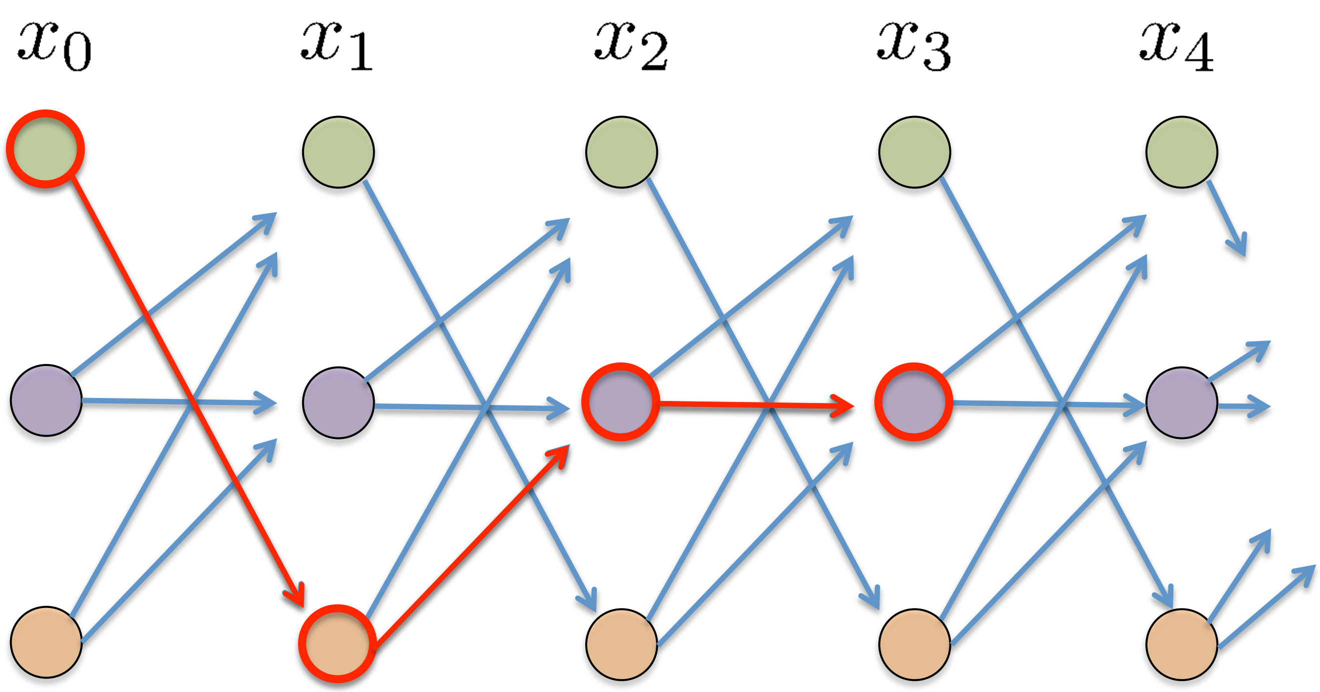
Markov system

- Another view: “lattice of states”
- State sequence = path in lattice

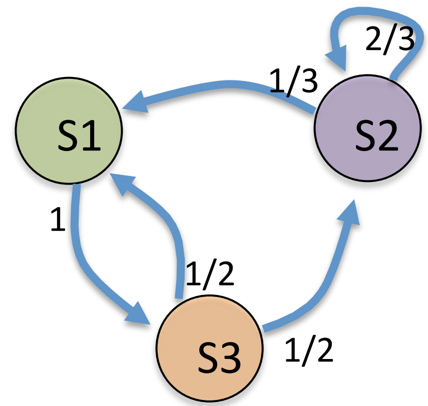
$$[x_0, x_1, x_2, x_3, \ldots] = [s_1, s_3, s_2, s_2, \ldots]$$

“State transition diagram”





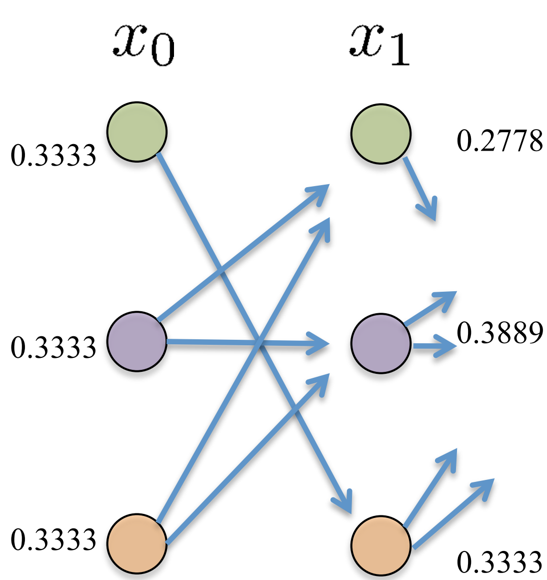
Computing probabilities

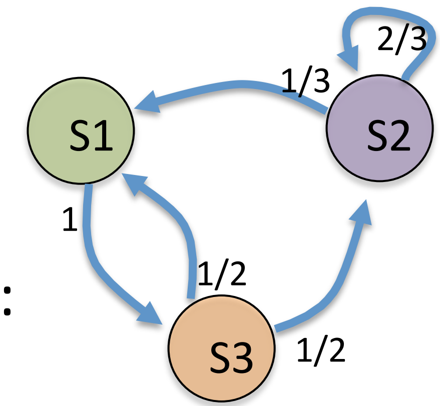


- We can compute the state distribution at time t:

$$p(x_1) = \sum_{x_0} p(x_0) \cdot p(x_1 \,|\, x_0)$$

$$\begin{pmatrix} 0.2778 \\ 0.3889 \\ 0.3333 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 0.333 \\ 0.333 \\ 0.333 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 0.0000 & 0.0000 & 1.0000 \\ 0.3333 & 0.6667 & 0.0000 \\ 0.5000 & 0.5000 & 0.0000 \end{pmatrix}$$





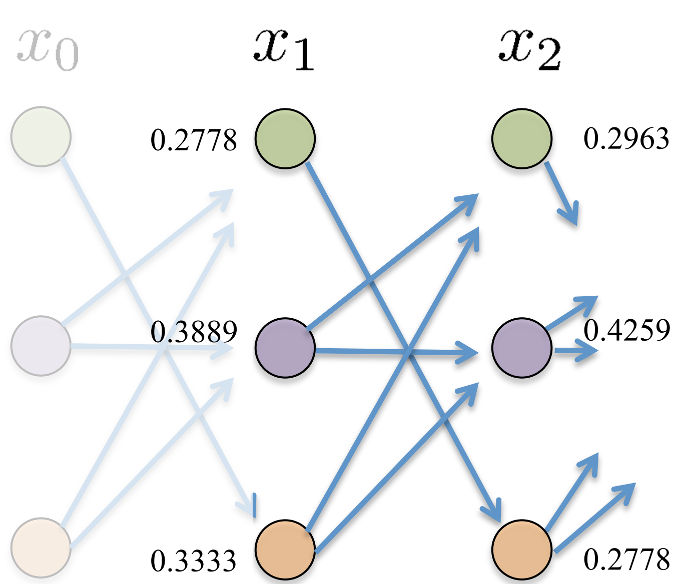
Computing probabilities

- We can compute the state distribution at time t:

$$p(x_1) = \sum_{x_0} p(x_0) \cdot p(x_1 \mid x_0)$$

$$p(x_2) = \sum_{x_1} p(x_1) \cdot p(x_2 \mid x_1)$$

$$\begin{pmatrix} 0.2963 \\ 0.4259 \\ 0.2778 \end{pmatrix}^T = \begin{pmatrix} 0.2778 \\ 0.3889 \\ 0.3333 \end{pmatrix}^T \begin{pmatrix} 0.0000 & 0.0000 & 1.0000 \\ 0.3333 & 0.6667 & 0.0000 \\ 0.5000 & 0.5000 & 0.0000 \end{pmatrix}$$



What's the state occupancy distribution in the far future?

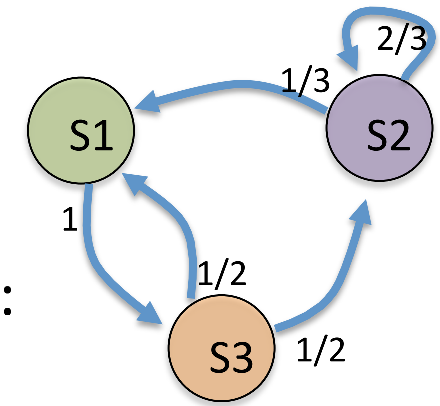
$$\lim_{t \to \infty} p(x_t) = ?$$

Does it depend on x_0?

Computing probabilities in python

```
>> import numpy as np

>> T = np.matrix( [[0.0,0.0,1.0],[.33,.67,0.0],[0.5,0.5,0.0]] )
>> p0 = np.matrix([.33,.33,.33])

>> p0 * T
matrix([[ 0.2739,  0.3861,  0.33  ]])

>> p0 * T * T
matrix([[ 0.292413,  0.423687,  0.2739  ]])

>> p0 * T * T * T
matrix([[ 0.27676671,  0.42082029,  0.292413  ]])

>> v = p0 * (T**20)
matrix([[ 0.28163911,  0.4267249 ,  0.28163599]])
>> v * T
matrix([[ 0.28163721,  0.42672368,  0.28163911]])
```

Computing probabilities



- We can compute the state distribution at time t:

$$p(x_1) = \sum_{x_0} p(x_0) \cdot p(x_1 \,|\, x_0)$$

$$p(x_2) = \sum_{x_1} p(x_1) \cdot p(x_2 \,|\, x_1)$$

Notes:

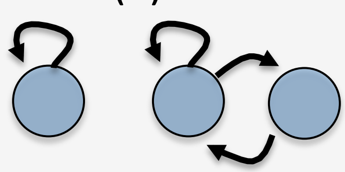
Stationary distribution: $s(x) \quad : \quad s(x_{t+1}) = \sum_{x_t} p(x_{t+1} \,|\, x_t) s(x_t)$

s(x) exists & is unique, so that $p(x_t)$ becomes independent of $p(x_0)$, if:

(a) p(.|.) is irreducible: $\forall i, j \; \exists t \; : \; \Pr[x_t = s_i \,|\, x_0 = s_j] > 0$

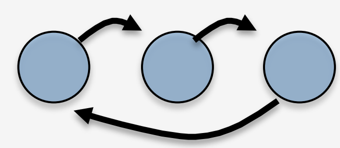
(b) p(.|.) is acyclic: $\gcd\{t \; : \; \Pr[x_t = s_i \,|\, x_0 = s_i] > 0\} = 1$

Ex: if not (a):



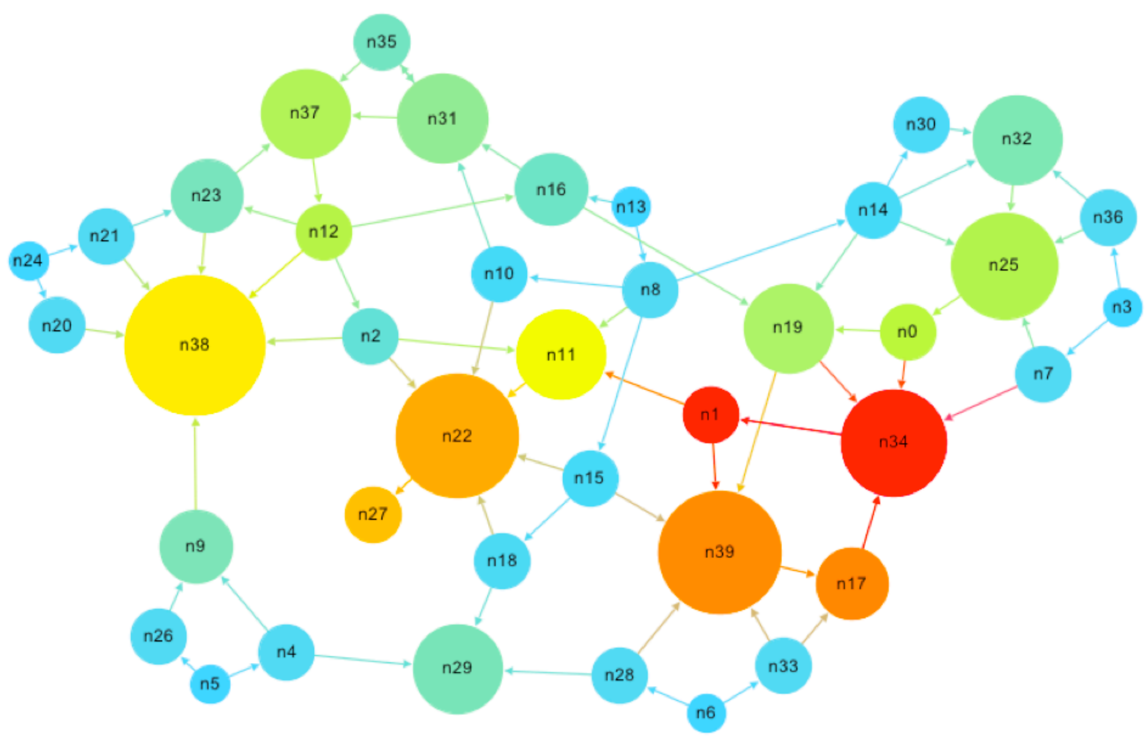
(Long-term prob will depend on initial state dist)

Ex: if not (b):



Stationary distributions

- PageRank is a stationary distribution
  - Small probability of jumping anywhere
  - Otherwise, follow outgoing link uniformly at random

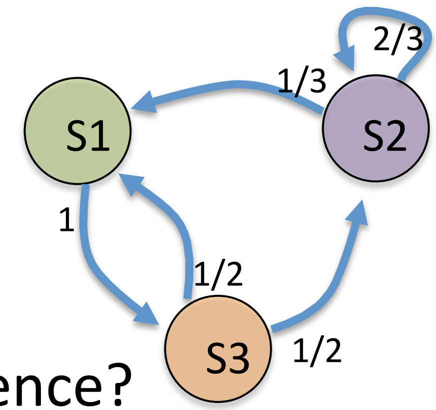


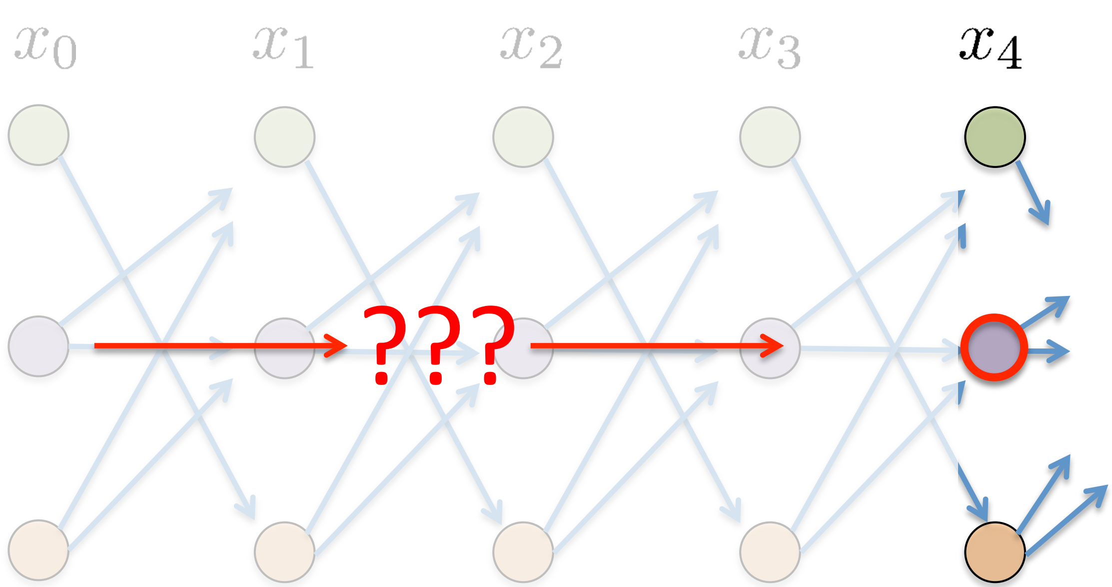
From http://computationalculture.net/article/what_is_in_pagerank

Dynamic programming

- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$

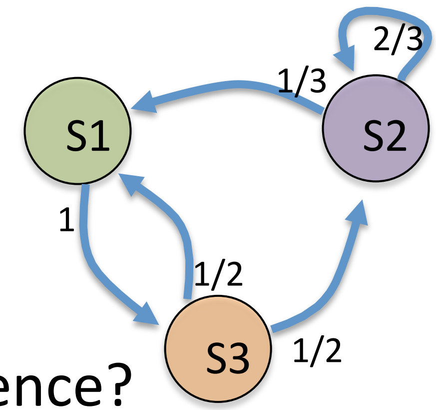




$$r(x_4) = \delta(x_4 = s_2)$$

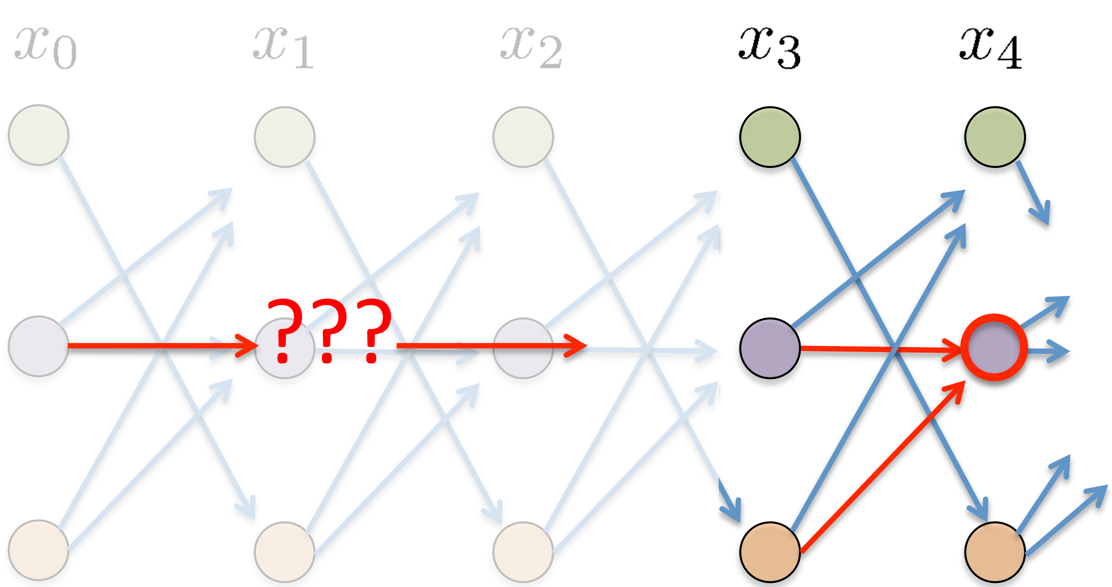
$$\begin{pmatrix} 0.0000 \\ 1.0000 \\ 0.0000 \end{pmatrix}$$

Dynamic programming



- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\dots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$

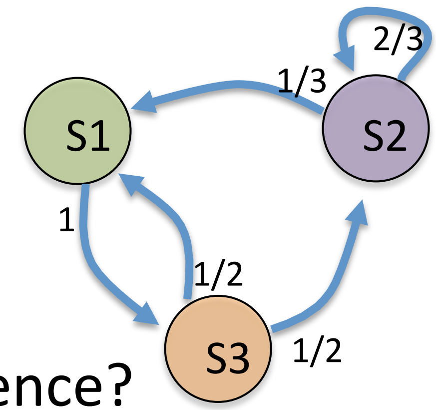


$$r(x_4) = \delta(x_4 = s_2)$$

$$r(x_3) = \max_{x_4} p(x_4 \mid x_3) r(x_4)$$

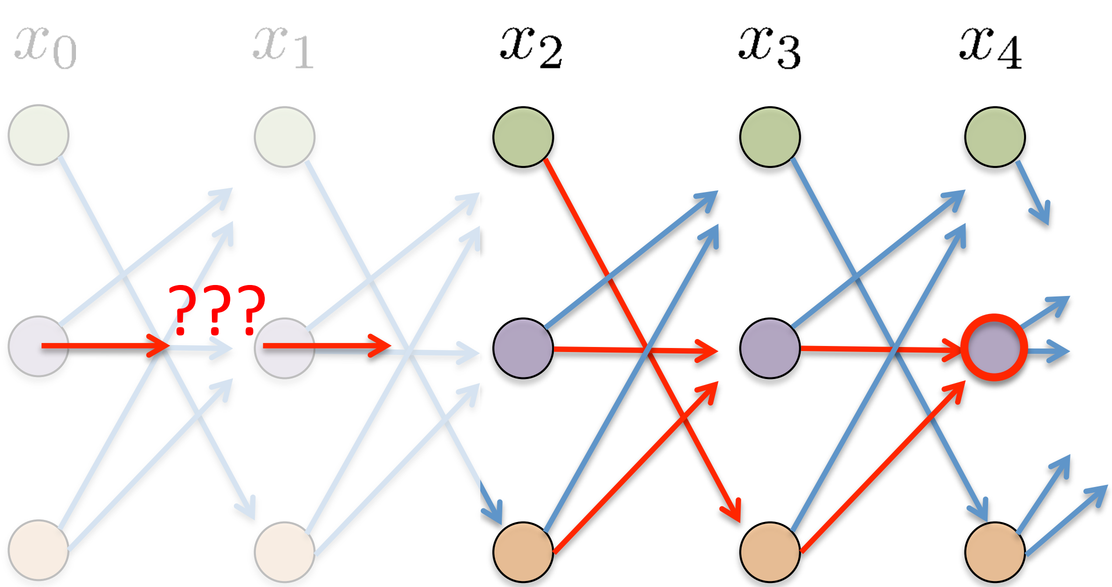
$$\begin{pmatrix} 0.0000 \\ 0.6667 \\ 0.5000 \end{pmatrix}$$

Dynamic programming



- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$



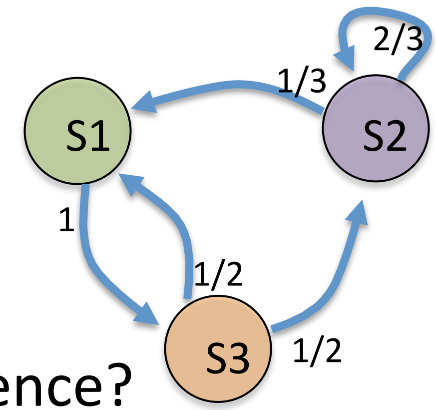
$$r(x_4) = \delta(x_4 = s_2)$$

$$r(x_3) = \max_{x_4} p(x_4 \mid x_3) r(x_4)$$

$$r(x_2) = \max_{x_3} p(x_3 \mid x_2) r(x_3)$$

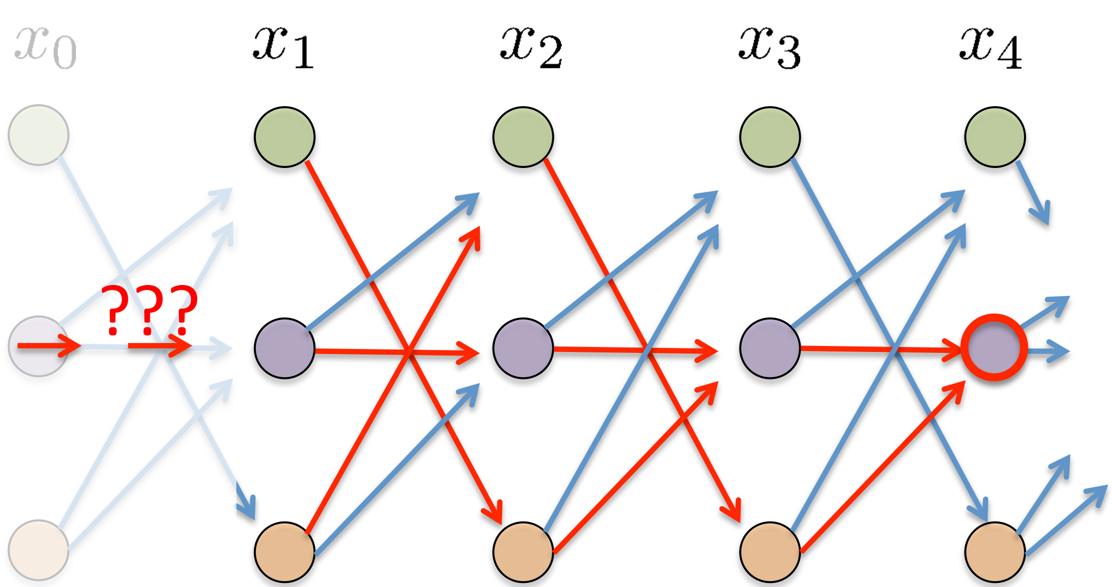
$$\begin{pmatrix} 0.5000 \\ 0.4444 \\ 0.3333 \end{pmatrix}$$

Dynamic programming



- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$



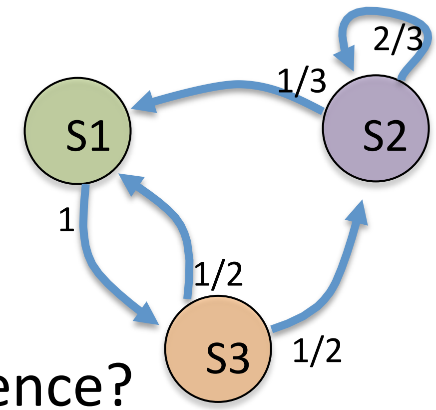
$$r(x_4) = \delta(x_4 = s_2)$$

$$r(x_3) = \max_{x_4} p(x_4 \mid x_3) r(x_4)$$

$$r(x_2) = \max_{x_3} p(x_3 \mid x_2) r(x_3)$$

$$r(x_1) = \max_{x_2} p(x_2 \mid x_1) r(x_2)$$

$$\begin{pmatrix} 0.3333 \\ 0.2963 \\ 0.2500 \end{pmatrix}$$

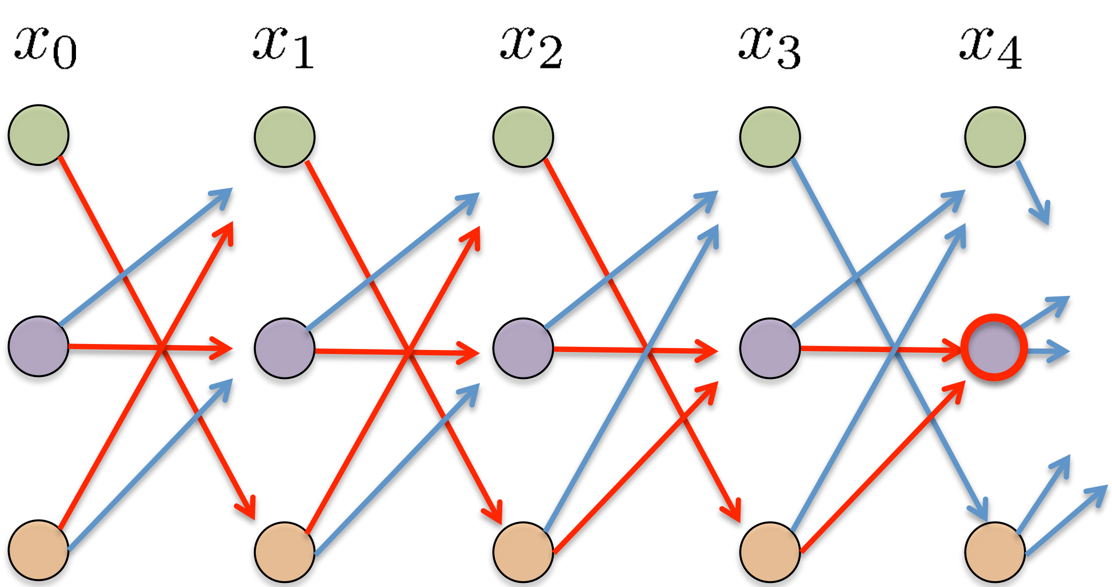


Dynamic programming

- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$

$$= 0.0833$$



$$r(x_4) = \delta(x_4 = s_2)$$

$$r(x_3) = \max_{x_4} p(x_4 \mid x_3) r(x_4)$$

$$r(x_2) = \max_{x_3} p(x_3 \mid x_2) r(x_3)$$

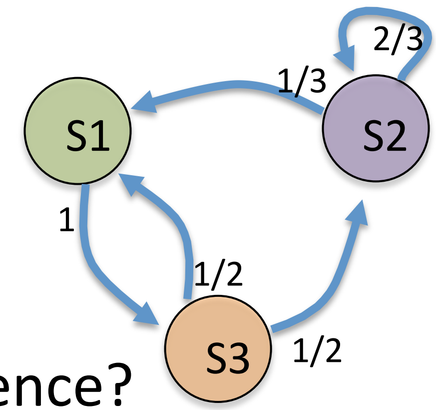
$$r(x_1) = \max_{x_2} p(x_2 \mid x_1) r(x_2)$$

$$r(x_0) = \max_{x_1} p(x_1 \mid x_0) r(x_1)$$

$$r^* = \max_{x_0} p(x_0) r(x_0)$$

$$\begin{pmatrix} 0.0833 \\ 0.0658 \\ 0.0556 \end{pmatrix} = \begin{pmatrix} 0.3333 \\ 0.3333 \\ 0.3333 \end{pmatrix} \begin{pmatrix} 0.2500 \\ 0.1975 \\ 0.1667 \end{pmatrix}$$

Dynamic programming

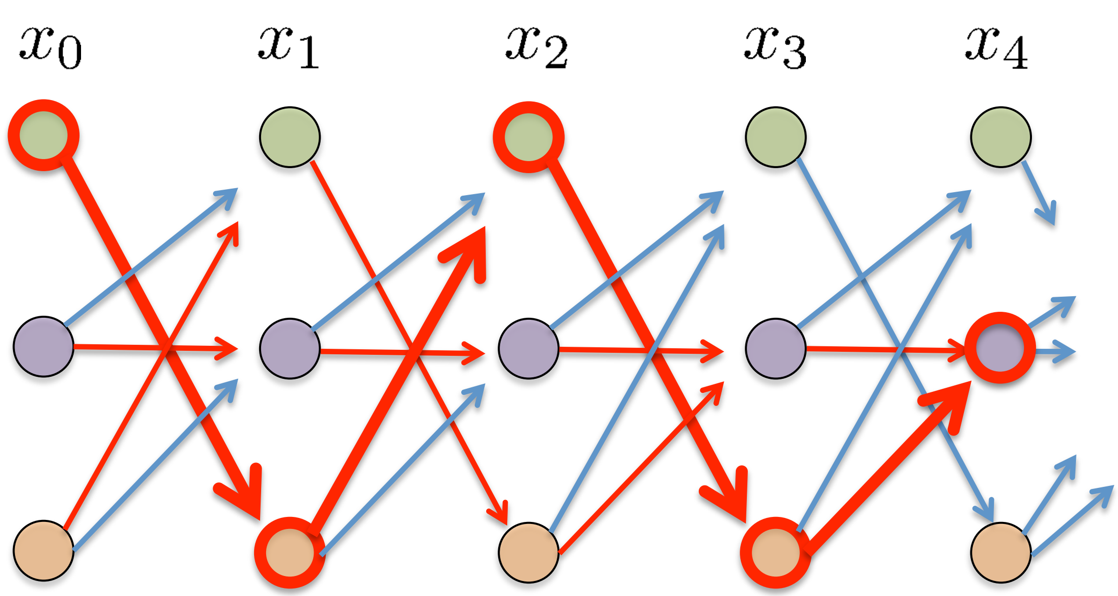


- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$

$$= 0.0833$$

$$\mathbf{x}^* = [s_1,\ s_3,\ s_1,\ s_3,\ s_2]$$



$$r(x_4) = \delta(x_4 = s_2)$$

$$r(x_3) = \max_{x_4} p(x_4 \mid x_3) r(x_4)$$

$$r(x_2) = \max_{x_3} p(x_3 \mid x_2) r(x_3)$$

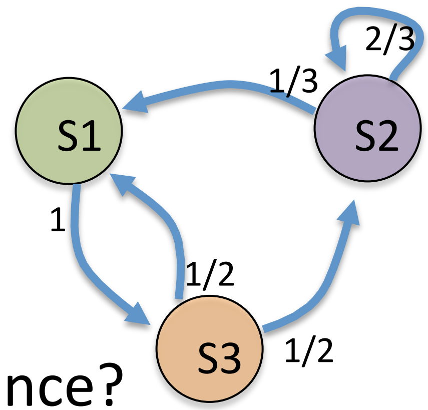
$$r(x_1) = \max_{x_2} p(x_2 \mid x_1) r(x_2)$$

$$r(x_0) = \max_{x_1} p(x_1 \mid x_0) r(x_1)$$

$$r^* = \max_{x_0} p(x_0) r(x_0)$$

$$\begin{pmatrix} 0.0833 \\ 0.0658 \\ 0.0556 \end{pmatrix} = \begin{pmatrix} 0.3333 \\ 0.3333 \\ 0.3333 \end{pmatrix} \begin{pmatrix} 0.2500 \\ 0.1975 \\ 0.1667 \end{pmatrix}$$

Dynamic programming

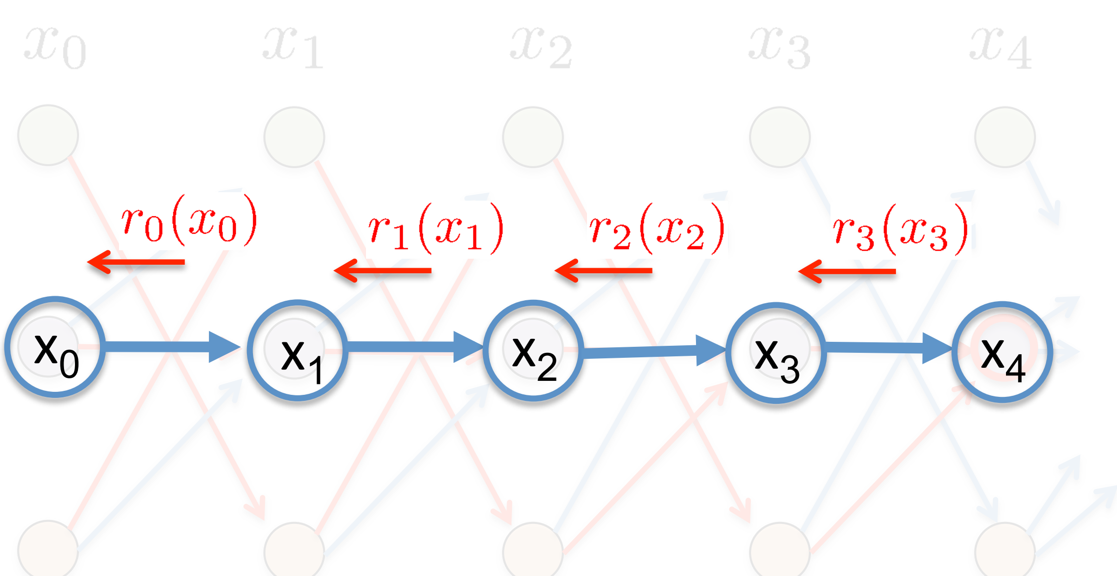


- Observe, say, $x_4 = 2$
- What's the (value of the) most likely state sequence?

$$r^* = \max_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$

$$= 0.0833$$

$$\mathbf{x}^* = [s_1,\ s_3,\ s_1,\ s_3,\ s_2]$$



$$r(x_4) = \delta(x_4 = s_2)$$

$$r(x_3) = \max_{x_4} p(x_4 \mid x_3) r(x_4)$$

$$r(x_2) = \max_{x_3} p(x_3 \mid x_2) r(x_3)$$

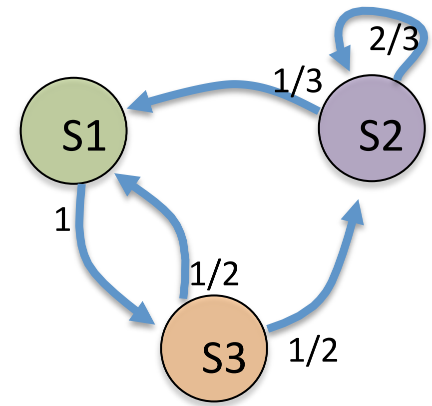
$$r(x_1) = \max_{x_2} p(x_2 \mid x_1) r(x_2)$$

$$r(x_0) = \max_{x_1} p(x_1 \mid x_0) r(x_1)$$

$$r^* = \max_{x_0} p(x_0) r(x_0)$$

$$\begin{pmatrix} 0.0833 \\ 0.0658 \\ 0.0556 \end{pmatrix} = \begin{pmatrix} 0.3333 \\ 0.3333 \\ 0.3333 \end{pmatrix} \begin{pmatrix} 0.2500 \\ 0.1975 \\ 0.1667 \end{pmatrix}$$

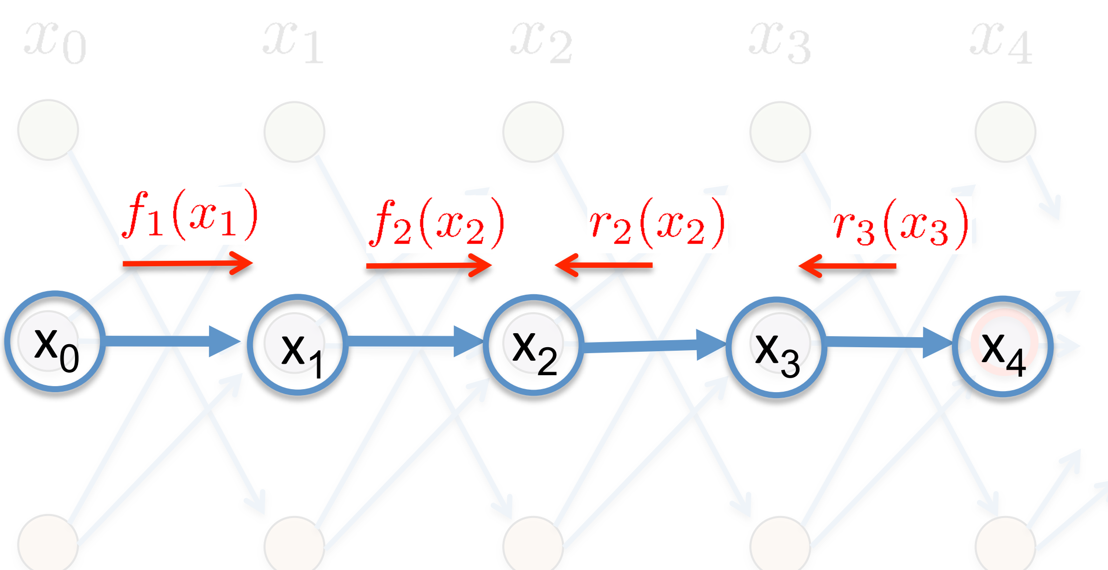
Dynamic programming



- Observe, say, $x_4 = 2$
- Similar algorithm for computing marginals:

$$p(x_2 | x_4 = s_2) \propto \sum_{x_0,\ldots,x_4} p(x_0, x_1, x_2, x_3, x_4)\delta(x_4 = s_2)$$

$$\propto f_2(x_2) \cdot r_2(x_2)$$



$$r_4(x_4) = \delta(x_4 = s_2)$$

$$r_3(x_3) = \sum_{x_4} p(x_4 \,|\, x_3) r(x_4)$$

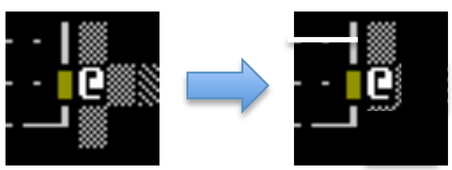
$$r_2(x_2) = \sum_{x_3} p(x_3 \,|\, x_2) r(x_3)$$

$$f_1(x_1) = \sum_{x_0} p(x_1 \,|\, x_0) p(x_0)$$

$$f_2(x_2) = \sum_{x_1} p(x_2 \,|\, x_1) f_1(x_1)$$

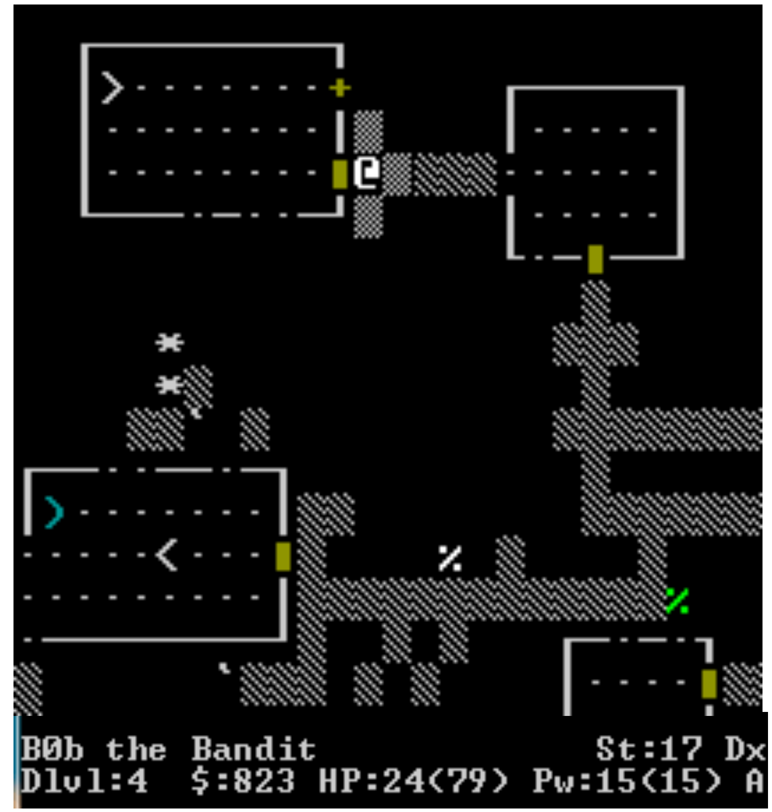
Hidden Markov Model

- Consider a robot
  - State = position in world
  - Randomly moves R/L/U/D
- Where is it located? p(Xt)
- Sensors observe the world
  - But, noisy: p(Ot | Xt)

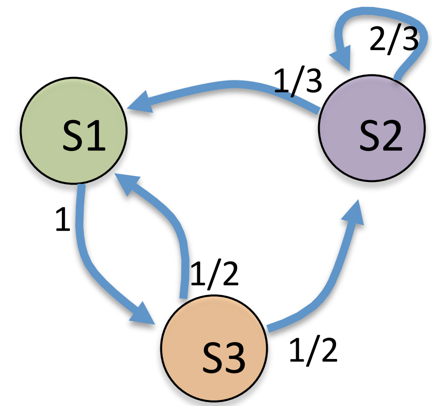


- Given a sequence of observations, where is it located?

  P(Xt | O1 … Ot) = ?



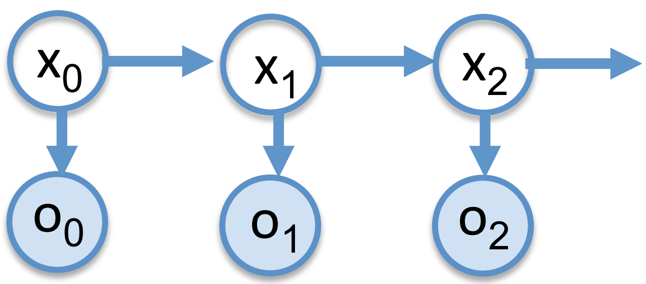
Hidden Markov Model



- In addition to the Markov state variables x_t
- We also have "emission" variables, o_t
- Model is specified by

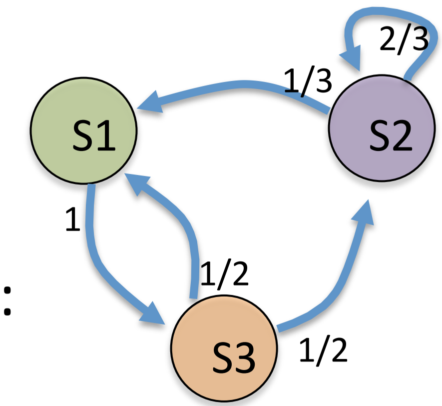
$$x_0 \sim p(x_0) \qquad x_{t+1} \sim p(x_{t+1} \,|\, x_t)$$
$$o_t \sim p(o_t \,|\, x_t)$$

- Bayes Net on states x_t and observations o_t over time t



- Typically, we'll observe the values of the o's (shaded)
  - Induces a model over the x's, and use this to answer queries about x's

HMM example



- Use previous state transitions; define emissions:

$$p(o_t \mid x_t) =$$

$$p(o_t = A \mid x_t = s_1) = 1.0 \quad p(o_t = A \mid x_t = s_2) = 0.5 \quad p(o_t = A \mid x_t = s_3) = 0$$
$$p(o_t = B \mid x_t = s_1) = 0 \quad p(o_t = B \mid x_t = s_2) = 0.5 \quad p(o_t = B \mid x_t = s_3) = 1.0$$

- Observe O=[0,1]?

$p(X_0) =$ [0.33 0.33 0.33]

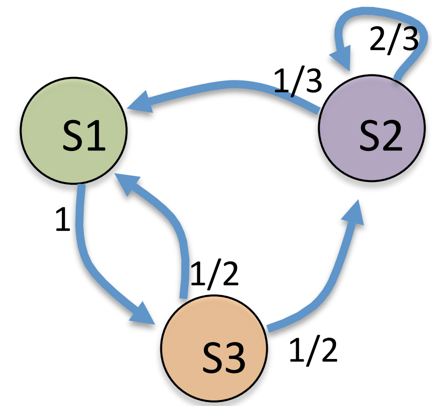
$p(X_0, O_0 = A) =$ [0.33 * 1.0 0.33 * 0.5 0.33 * 0] = [0.33 0.17 0]

$p(X_0 | O_0 = A) =$ [0.66 0.33 0]

$p(X_0, X_1 | O_0 = A) =$ [0 0 0.66]
[0.11 0.22 0]
[0 0 0]

$p(X_1 | O_0 = A) =$ [0.11 0.22 0.66]

HMM example



$$x_{t+1} \sim p(x_{t+1} \,|\, x_t)$$

- Let's use our previous state transition model
- Define "emission" variables, o_t

$$p(o_t \,|\, x_t) =$$

$$p(o_t = A \,|\, x_t = s_1) = 1.0 \qquad p(o_t = A \,|\, x_t = s_2) = 0.5 \qquad p(o_t = A \,|\, x_t = s_3) = 0$$

$$p(o_t = B \,|\, x_t = s_1) = 0 \qquad p(o_t = B \,|\, x_t = s_2) = 0.5 \qquad p(o_t = B \,|\, x_t = s_3) = 1.0$$

- Observe O=[A,B]?

$p(X_1 | O_0 = A) =$ [0.11 0.22 0.66]

$p(X_1, O_1 = B | O_0 = A) =$ [0.11 * 0 0.22 * 0.5 0.66 * 1.0] = [0 0.11 0.66]

$p(X_1 | O_0 = A, O_1 = B) =$ [0 0.14 0.86]

Dynamic programming

- Can similarly use dynamic programming on an HMM:

Forward messages:

$$f_t(x_t) = p(x_t \mid o_0, \ldots, o_t)$$
$$= \frac{1}{Z_t} \sum_{x_{t-1}} p(x_t, x_{t-1}, o_t | o_0, \ldots, o_{t-1})$$
$$= \frac{1}{Z_t} p(o_t|x_t) \sum_{x_{t-1}} p(x_t|x_{t-1}) f_{t-1}(x_{t-1})$$

Reverse messages:

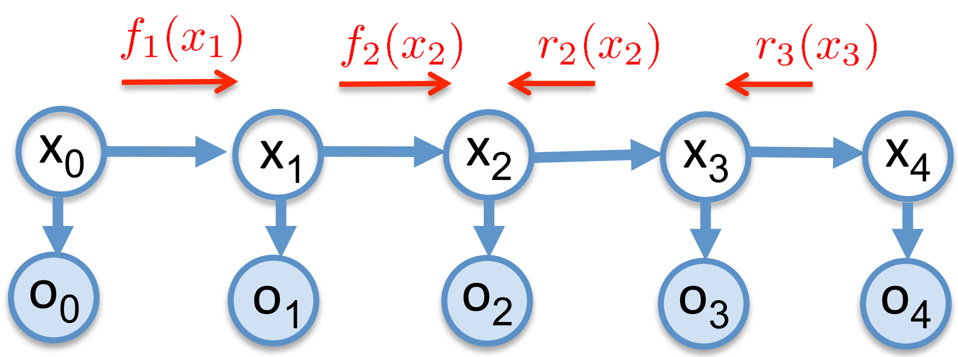
$$r_t(x_t) \propto p(o_{t+1}, \ldots, o_T \mid x_t)$$
$$\propto \sum_{x_{t+1}} p(x_{t+1}|x_t)\, p(o_{t+1}|x_{t+1}\, r_{t+1}(x_{t+1})$$

Z_t is the scalar that normalizes $f_t(x_t)$:

$$Z_t = p(o_t|o_0, \ldots, o_{t-1})$$

Observation likelihood:

$$p(O = o) = p(o_0)\, p(o_1|o_0) \; \ldots$$
$$= \prod_t Z_t$$



Marginal probabilities:

$$p(x_2|O = o) = \frac{1}{p(O = o)} \sum_{\mathbf{x} \setminus x_2} p(\mathbf{x}, O = o) \propto f_2(x_2) \cdot r_2(x_2)$$