Markov models

Markov models are used to reason about a sequence of random variables. We want to look at the probability of an event happening as the next state in a system given the current state of the system.

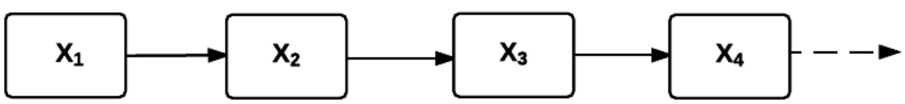
Similar to an MDP, but with no actions and no rewards.

We've introduced another variable, such as time or space into the Bayes net model. We're no longer dealing with a fixed Bayes net. Now, we have a Bayes net that grows over time.

Examples of where Markov models are used
Speech recognition: what is the most likely next sound or word given the current sound or word.

Medical monitoring: what is the health of the patient given the current and previous states.

You can think of a Markov model as a chain structured Bayes net



where X_i is the state of the variables at time *i*.

Each node Xi has the same distribution, which is $P(X_{t+1} \mid X_t)$. The only exception is X_1, which is described by $P(X_1)$. The conditional probability tables that we used for each node in a Bayes net are the $P(X_{t+1} \mid X_t)$ transition probabilities in a Markov model. A transition probability is the probability of going from the current state to the next state.

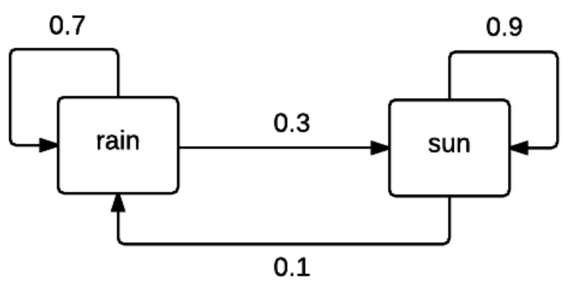
We see the Markov property in that the probabilities depend only on the current state. We don't consider any of the previous steps.

Example:
States X = {rain, sun}
Initial distribution, $P(X_1)$ = 1.0 sun

X_t	X_{t+1}	$P(X_{t+1} \mid X_t)$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

If it's sunny at time *t*, then there's a 0.90 probability that it's sunny at time *t+1*. Another way to think of this is that if it's sunny at the current state, then there's a 90% chance that it's sunny at the next state.

We can also represent the transition probabilities in a graph, where the nodes are the states and the edge weights are the probabilities of transitioning between the states.



What is probability distribution after one step, i.e. what is $P(X_2)$?

$P(X_1 = sun) = 1.0$

$$P(X_2 = sun) = \sum_{X_1} P(X_2, X_1)$$

$$= \sum_{X_1} P(X_2 \mid X_1) P(X_1)$$

$$= P(X_2|sun)P(X_2 = sun) + P(X_2 \mid rain)P(X_2 = rain)$$

P(X1 = rain) = 0. Therefore, the second term in the summation is 0, and the calculation is

0.9 * 1.0 = 0.9.

$P(X_2 = rain) = 1 - P(X_2 = sun) = 0.1$

What is $P(X_3=rain)$?

$$= \sum_{X_2} P(X_3 \mid X_2) P(X_2)$$

$$= P(X_3|sun)P(X_2 = sun) + P(X_3 \mid rain)P(X_2 = rain)$$

= (0.1 * 0.9) + (0.7 * 0.1) = 0.16

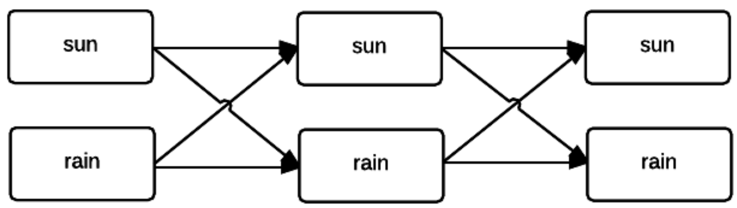
P(X3 = sun) = 0.84

What is $P(X_t)$, for some day *t*?

Start at X_1, using $P(X_1)$, which is given.

$$P(X_t) = \sum_{t-1} P(X_t \mid X_{t-1}) P(X_{t-1}), \forall t$$

This is called the mini-forward algorithm.



We can use the mini-forward algorithm to calculate the probability distribution at time *t* using the distribution at time *t-1*.

What is *P(sun)*, or *P(rain)*?

We don't know this from looking at our probability distributions, what we have is transition probabilities.

If we run the mini-forward algorithm, eventually we get stationarity, or the stationary distribution of sun, rain:

$P(X_1)$	$P(X_2)$	$P(X_3)$	$P(X_4)$	…	$P(X_{inf})$
[1.0, 0]	[0.9, 0.1]	[0.84, 0.16]	[0.804, 0.196]		[0.75, 0.25]

But, what if we start from a different initial condition? We get the same stationary distribution.

$P(X_1)$	$P(X_2)$	$P(X_3)$	$P(X_4)$	...	$P(X_{inf})$
[0, 1.0]	[0.3, 0.7]	[0.48, 0.52]	[0.588, 0.412]		[0.75, 0.25]

The influence of the initial distribution decreases over time.

The joint probability distribution
The joint probability distribution in a Markov chain is the probability of seeing a particular sequence of events, such as rain, rain, sun, sun.

Starting with two events:
$P(x1, x2) = P(x2 \mid x1)P(x1)$
Ex: $P(sun, rain) = .9*1 = .9$

Three events:
$P(x1, x2, x3) = P(x3 \mid x2, x1)P(x2 \mid x1)P(x1)$
Ex: P(sun, rain, rain)

But, due to the Markov property:
P(x3 | x2, x1) = P(x3 | x2)
Ex: P(sun, rain, rain) = .3 * .1 * 1 = .03

To calculate the joint:

$$P(x1, x2, x3, \dots xn) = P(X_1)\prod_{t=1}^{n} P(X_{t+1} | X_t)$$