Hidden Markov Models

Slides based on the tutorial by Eric Fosler Lussier

http://www.di.ubi.pt/~jpaulo/competence/tutorials/hmm-tutorial-1.pdf

Markov Models

Three types of weather sunny, rainy, and foggy

- Assume: the weather lasts all day (it doesn't change from rainy to sunny in the middle of the day)
- Weather prediction is all about trying to guess what the weather will be like tomorrow based on a history of observations of weather

$$P(w_n \mid w_{n-1}, w_{n-2}, \ldots, w_1)$$

- For example if we knew that the weather for the past three days was *{sunny, sunny foggy}* in chronological order the probability that tomorrow would be rainy is given by:

$$P(w_4 = Rainy \mid w_3 = Foggy, w_2 = Sunny, w_1 = Sunny)$$

- For example if we knew that the weather for the past three days was *{sunny, sunny foggy}* in chronological order the probability that tomorrow would be rainy is given by:

$$P(w_4 = Rainy \mid w_3 = Foggy, w_2 = Sunny, w_1 = Sunny)$$

- The larger n is the more statistics we must collect. Suppose that n=5 then we must collect statistics for $3^5 = 243$

Markov Assumption

In a sequence $\{w_1, w_2, \ldots, w_n\}$:

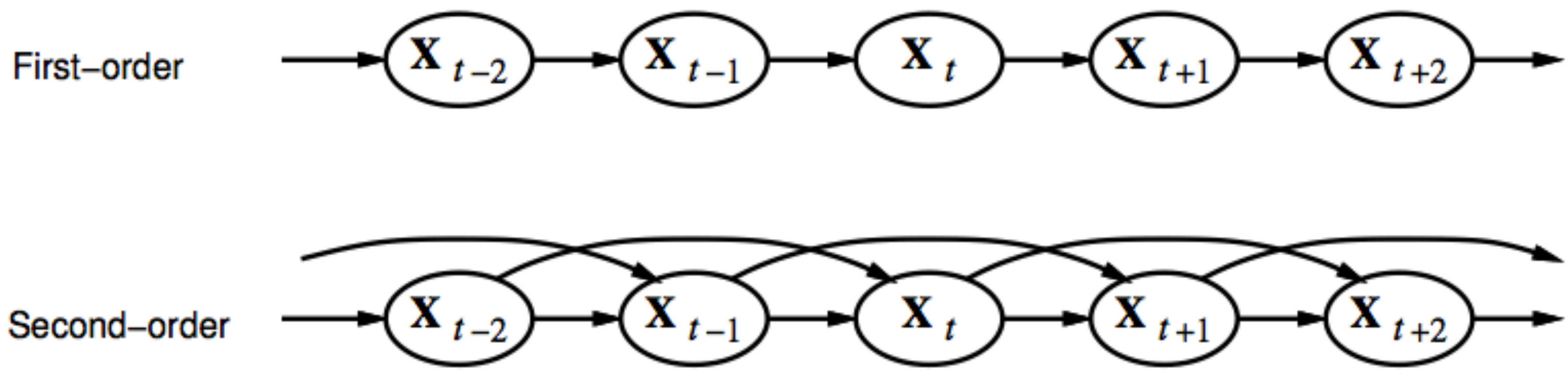
$$P(w_n \mid w_{n-1}, w_{n-2}, \ldots, w_1) \approx P(w_n \mid w_{n-1})$$

- This is called a first order Markov assumption since we say that the probability of an observation at time *n* only depends on the observation at time *n -1*

Markov Assumption

Markov assumption: $\mathbf{X}_t$ depends on **bounded** subset of $\mathbf{X}_{0:t-1}$

First-order Markov process: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$
Second-order Markov process: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-2}, \mathbf{X}_{t-1})$



Markov Assumption

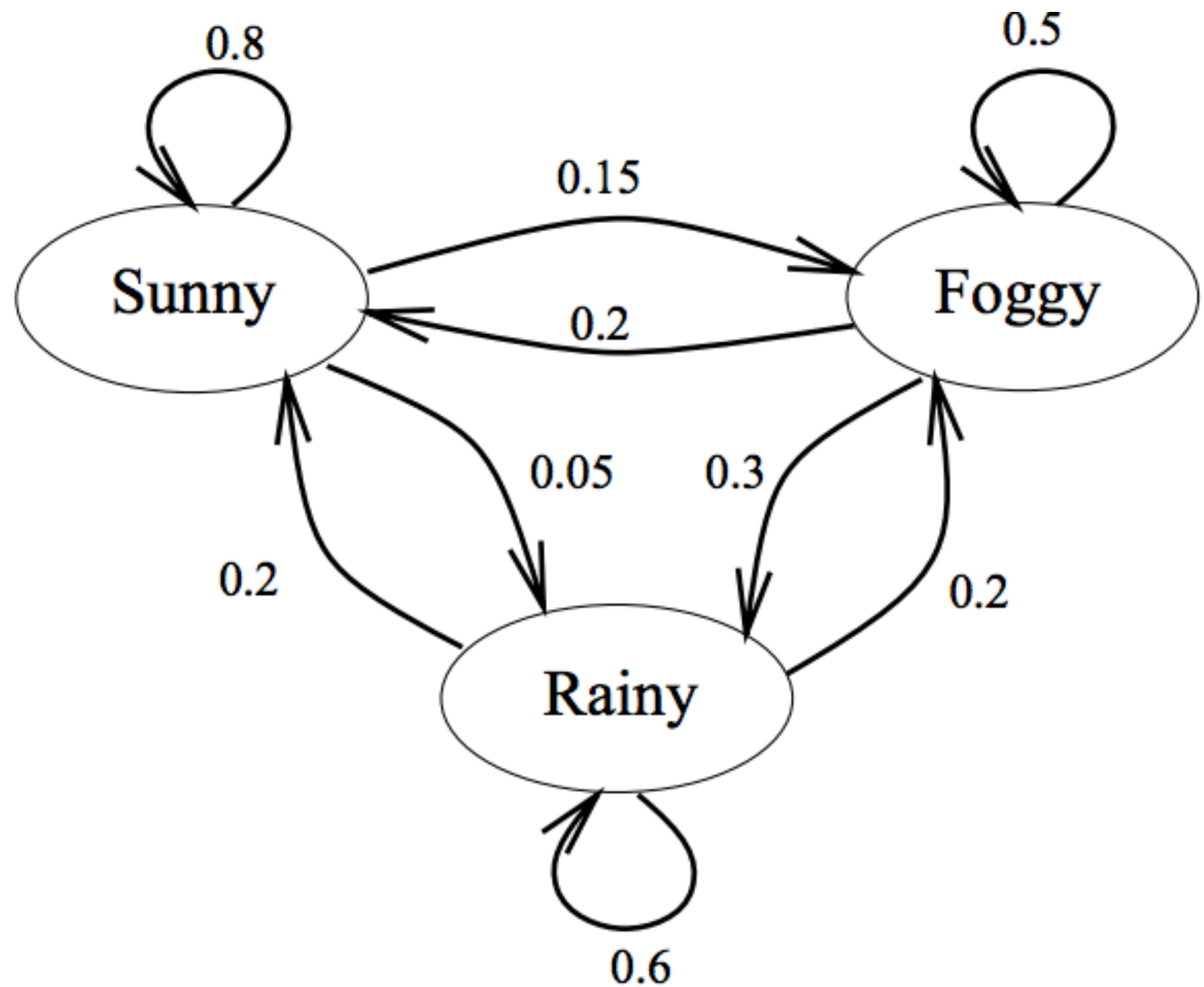
- We can express the joint distribution using the Markov assumption:

$$P(w_1, \ldots, w_n) = \Pi_{i=1}^{n} P(w_i \mid w_{i-1})$$

- Now, the number of statistics that we need to collect is $3^2 = 9$

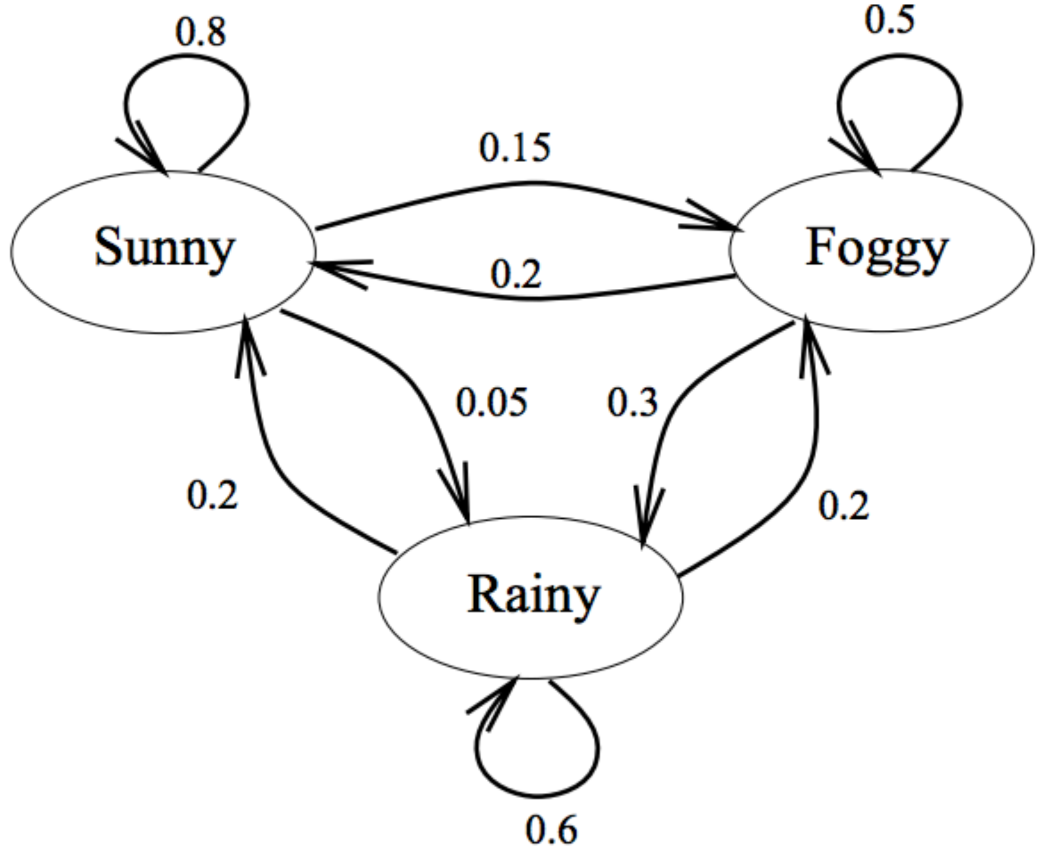
		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

Table 1: Probabilities of Tomorrow's weather based on Today's Weather



		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

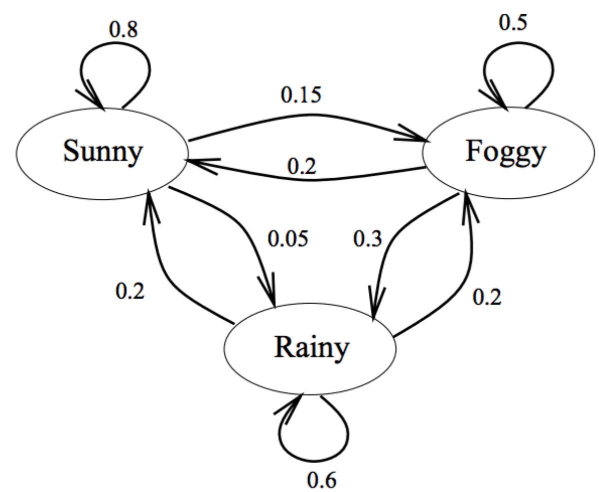
- Given that today is sunny what is the probability that tomorrow is sunny and the day after is rainy?



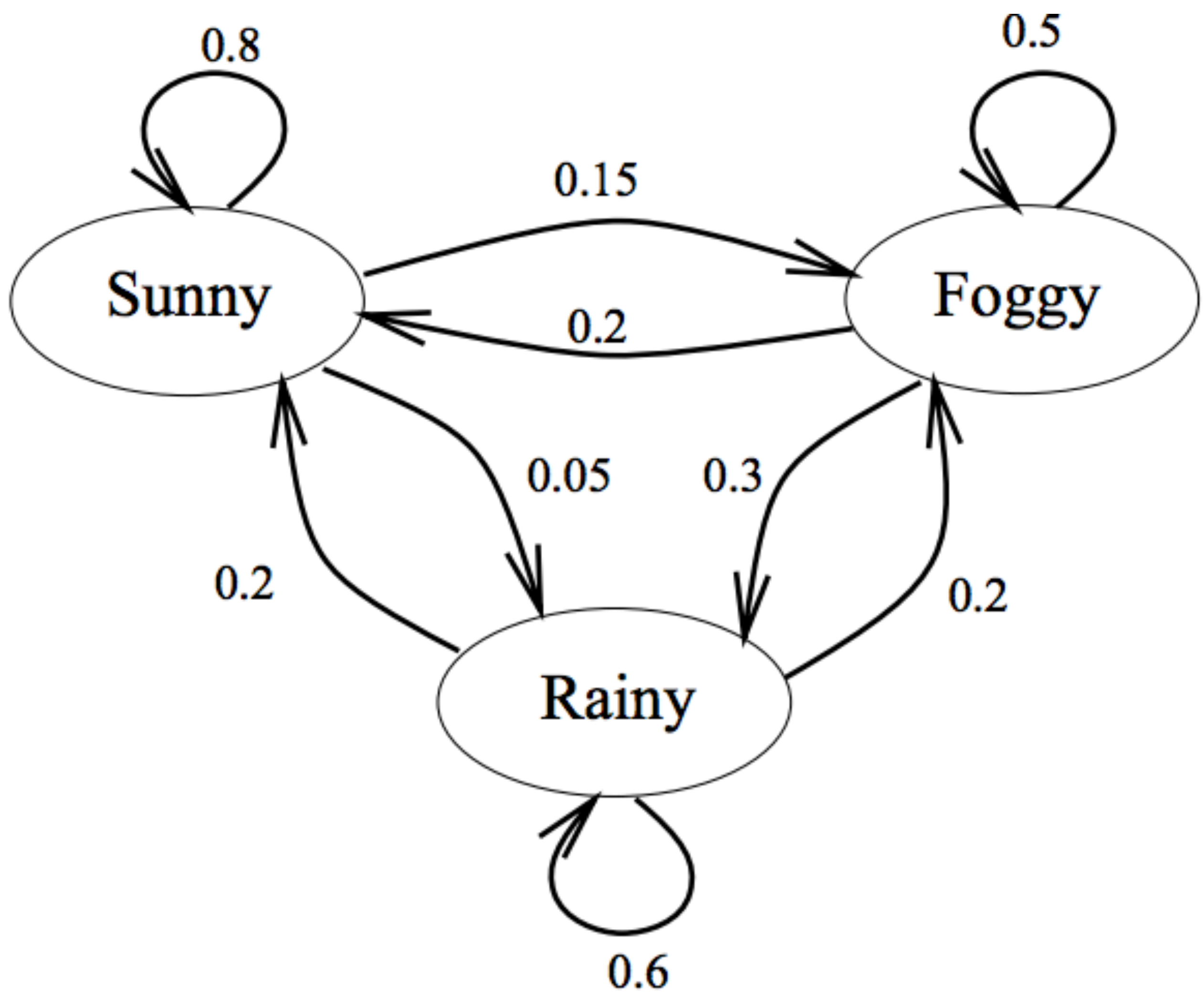
- Given that today is sunny what is the probability that tomorrow is sunny and the day after is rainy?

$$P(w_2 = \text{Sunny}, w_3 = \text{Rainy} | w_1 = \text{Sunny}) \quad =$$

$$\begin{aligned}
&= \quad P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny}, w_1 = \text{Sunny}) * \\
&\qquad P(w_2 = \text{Sunny} \mid w_1 = \text{Sunny}) \\
&= \quad P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny}) * \\
&\qquad P(w_2 = \text{Sunny} \mid w_1 = \text{Sunny}) \\
&= \quad (0.05)(0.8) \\
&= \quad 0.04
\end{aligned}$$



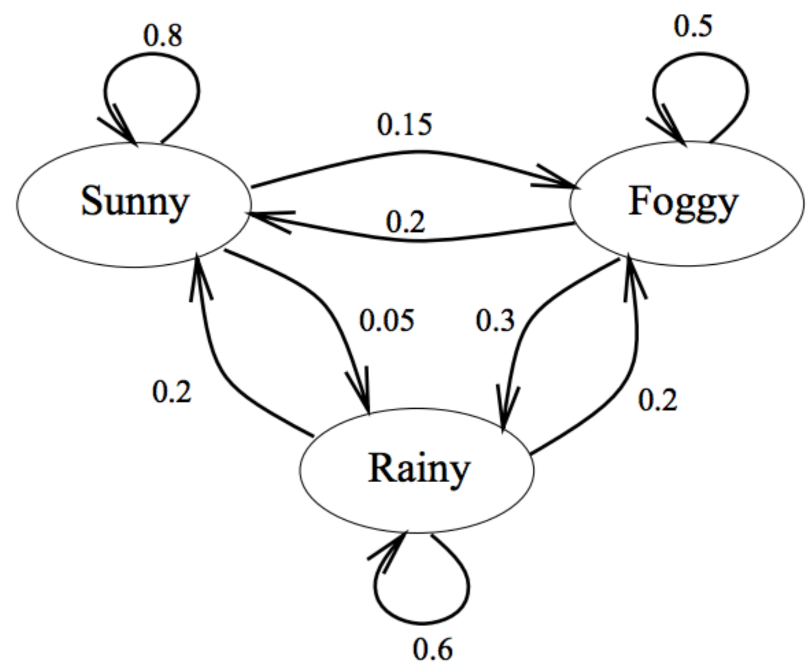
Given that today is foggy what's the probability that it will be rainy two days from now?



Given that today is foggy what's the probability that it will be rainy two days from now?

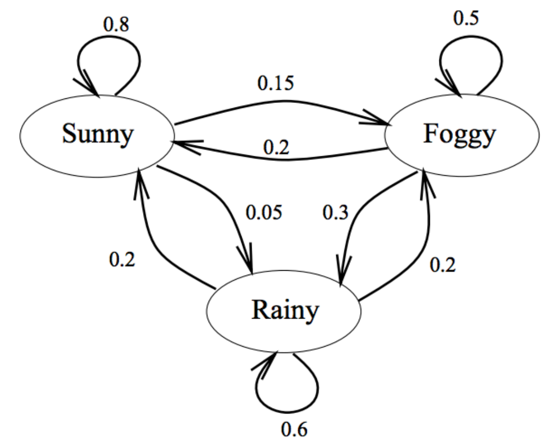
There are three ways to get from foggy today to rainy two days from now {foggy, foggy, rainy} {foggy, rainy, rainy} and {foggy, sunny, rainy}

Therefore we have to sum over these paths



Given that today is foggy what's the probability that it will be rainy two days from now?

{foggy, foggy, rainy} {foggy, rainy, rainy} {foggy, sunny, rainy}



$$
\begin{aligned}
& P(w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) = \\
= \quad & P(w_2 = \text{Foggy}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
& P(w_2 = \text{Rainy}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
& P(w_2 = \text{Sunny}, w_3 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
= \quad & P(w_3 = \text{Rainy} \mid w_2 = \text{Foggy}) P(w_2 = \text{Foggy} \mid w_1 = \text{Foggy}) + \\
& P(w_3 = \text{Rainy} \mid w_2 = \text{Rainy}) P(w_2 = \text{Rainy} \mid w_1 = \text{Foggy}) + \\
& P(w_3 = \text{Rainy} \mid w_2 = \text{Sunny}) P(w_2 = \text{Sunny} \mid w_1 = \text{Foggy}) \\
= \quad & (0.3)(0.5) + (0.6)(0.3) + (0.05)(0.2) \\
= \quad & 0.34
\end{aligned}
$$

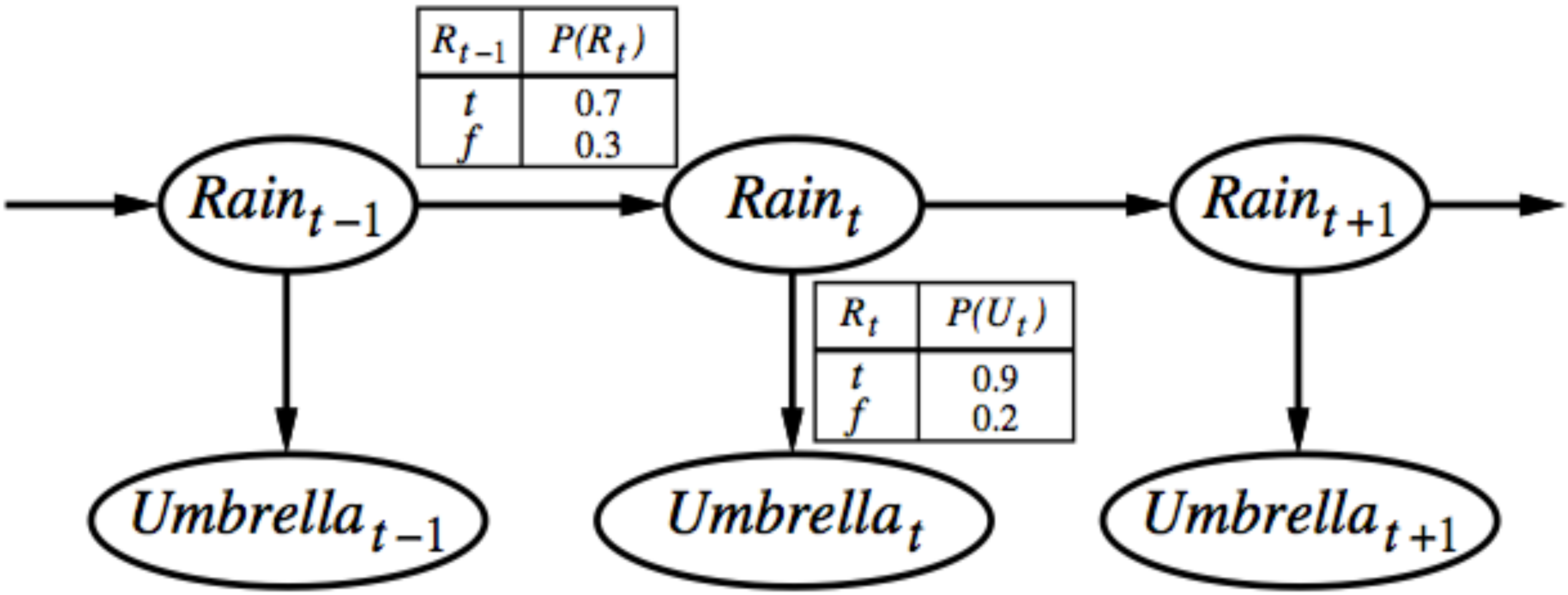
- Note that you have to know where you start from.
- Usually Markov models start with a null start state and have transitions to other states with certain probabilities.
- In the previous problems you can just add a start state with a single arc with probability 1 to the initial state (sunny in problem 1 and foggy in problem 2)

Hidden Markov Models

What makes a Hidden Markov Model?

- Suppose you were locked in a room for several days and you were asked about the weather outside. The only piece of evidence you have is whether the person who comes into the room carrying your daily meal is carrying an umbrella or not

Example



R_{t-1}	$P(R_t)$
t	0.7
f	0.3

R_t	$P(U_t)$
t	0.9
f	0.2

HMM

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probabilities of Seeing an Umbrella Based on the Weather

HMM

- The equation for the weather Markov process before you were locked in the room was:

$$P(w_1, \ldots, w_n) = \Pi_{i=1}^{n} P(w_i \mid w_{i-1})$$

- Now we have to factor in the fact that the actual weather is hidden from you. We do that by using Bayes Rule:

$$P(w_1, \ldots, w_n \mid u_1, \ldots, u_n) = \frac{P(u_1, \ldots, u_n | w_1, \ldots, w_n) P(w_1, \ldots, w_n)}{P(u_1, \ldots, u_n)}$$

HMM

- The probability $P(w_1, \ldots, w_n \mid u_1, \ldots, u_n)$ can be estimated as $\Pi_{i=1}^{n} P(u_i|w_i)$
  – if you assume that for all i given wi, ui is independent of all uj and wj for all j ≠ I

$P(u_1, \ldots, u_n)$ is the prior probability of seeing a particular sequence of umbrella events eg {True False True}

HHM Question 1

- Suppose the day you were locked in it was sunny.
- The next day the caretaker carried an umbrella into the room.
  - Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5
- What is the probability that the second day was rainy?

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny}|u_2 = \text{T})} \\
(u_2 \ \ and \ \ w_1 \ \ independent) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny})}
\end{aligned}
$$

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny}|u_2 = \text{T})} \\
(u_2 \ \textit{and} \ w_1 \ \textit{independent}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny})}
\end{aligned}
$$

$$
(Bayes' Rule) = \frac{P(u_2 = \text{T}|w_1 = \text{Sunny}, w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})}
$$

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny} | u_2 = \text{T})}{P(w_1 = \text{Sunny} | u_2 = \text{T})} \\
(u_2 \ and \ w_1 \ independent) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny} | u_2 = \text{T})}{P(w_1 = \text{Sunny})}
\end{aligned}
$$

$$
\begin{aligned}
(Bayes' Rule) &= \frac{P(u_2 = \text{T} | w_1 = \text{Sunny}, w_2 = \text{Rainy}) P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny}) P(u_2 = \text{T})} \\
(Markov \ assumption) &= \frac{P(u_2 = \text{T} | w_2 = \text{Rainy}) P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny}) P(u_2 = \text{T})}
\end{aligned}
$$

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny}|u_2 = \text{T})} \\
(u_2 \;\; and \;\; w_1 \;\; independent) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny})}
\end{aligned}
$$

$$
\begin{aligned}
(Bayes' \, Rule) &= \frac{P(u_2 = \text{T}|w_1 = \text{Sunny}, w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Markov \;\; assumption) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(P(A, B) = P(A|B)P(B)) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}|w_1 = \text{Sunny})P(w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})}
\end{aligned}
$$

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

$$
\begin{aligned}
P(w_2 = \text{Rainy}| \\ w_1 = \text{Sunny}, u_2 = \text{True}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny}|u_2 = \text{T})} \\
(u_2 \ \textit{and} \ w_1 \ \textit{independent}) &= \frac{P(w_2 = \text{Rainy}, w_1 = \text{Sunny}|u_2 = \text{T})}{P(w_1 = \text{Sunny})} \\
(Bayes' \, Rule) &= \frac{P(u_2 = \text{T}|w_1 = \text{Sunny}, w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Markov \ assumption) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}, w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(P(A, B) = P(A|B)P(B)) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}|w_1 = \text{Sunny})P(w_1 = \text{Sunny})}{P(w_1 = \text{Sunny})P(u_2 = \text{T})} \\
(Cancel : P(Sunny)) &= \frac{P(u_2 = \text{T}|w_2 = \text{Rainy})P(w_2 = \text{Rainy}|w_1 = \text{Sunny})}{P(u_2 = \text{T})} \\
&= \frac{(0.8)(0.05)}{0.5} \\
&= .08
\end{aligned}
$$

HMM Question 2

- Suppose the day you were locked in the room it was sunny the caretaker brought in an umbrella on day 2 but not on day 3
- Again assuming that the prior probability of the caretaker bringing an umbrella is 0.5
- What is the probability that it's foggy on day 3?

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

$$\begin{aligned} P(w_3 = \mathrm{F} \mid \quad & = \quad P(w_2 = \mathrm{Foggy}, w_3 = \mathrm{Foggy} \mid \\ w_1 = \mathrm{S}, u_2 = \mathrm{T}, u_3 = \mathrm{F}) & \qquad\qquad w_1 = \mathrm{Sunny}, u_2 = \mathrm{True}, u_3 = \mathrm{False}) + \\ & \quad\;\; P(w_2 = \mathrm{Rainy}, w_3 = \mathrm{Foggy} \mid \ldots) + \\ & \quad\;\; P(w_2 = \mathrm{Sunny}, w_3 = \mathrm{Foggy} \mid \ldots) \end{aligned}$$

HMM Q2

$$
\begin{aligned}
= \quad & \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} + \\
& \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} + \\
& \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)P(w_1 = S)}{P(u_3 = F)P(u_2 = T)P(w_1 = S)} \\
= \quad & \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = F)P(w_3 = F|w_2 = F)P(w_2 = F|w_1 = S)}{P(u_3 = F)P(u_2 = T)} + \\
& \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = R)P(w_3 = F|w_2 = R)P(w_2 = R|w_1 = S)}{P(u_3 = F)P(u_2 = T)} + \\
& \frac{P(u_3 = F|w_3 = F)P(u_2 = T|w_2 = S)P(w_3 = F|w_2 = S)P(w_2 = S|w_1 = S)}{P(u_3 = F)P(u_2 = T)}
\end{aligned}
$$

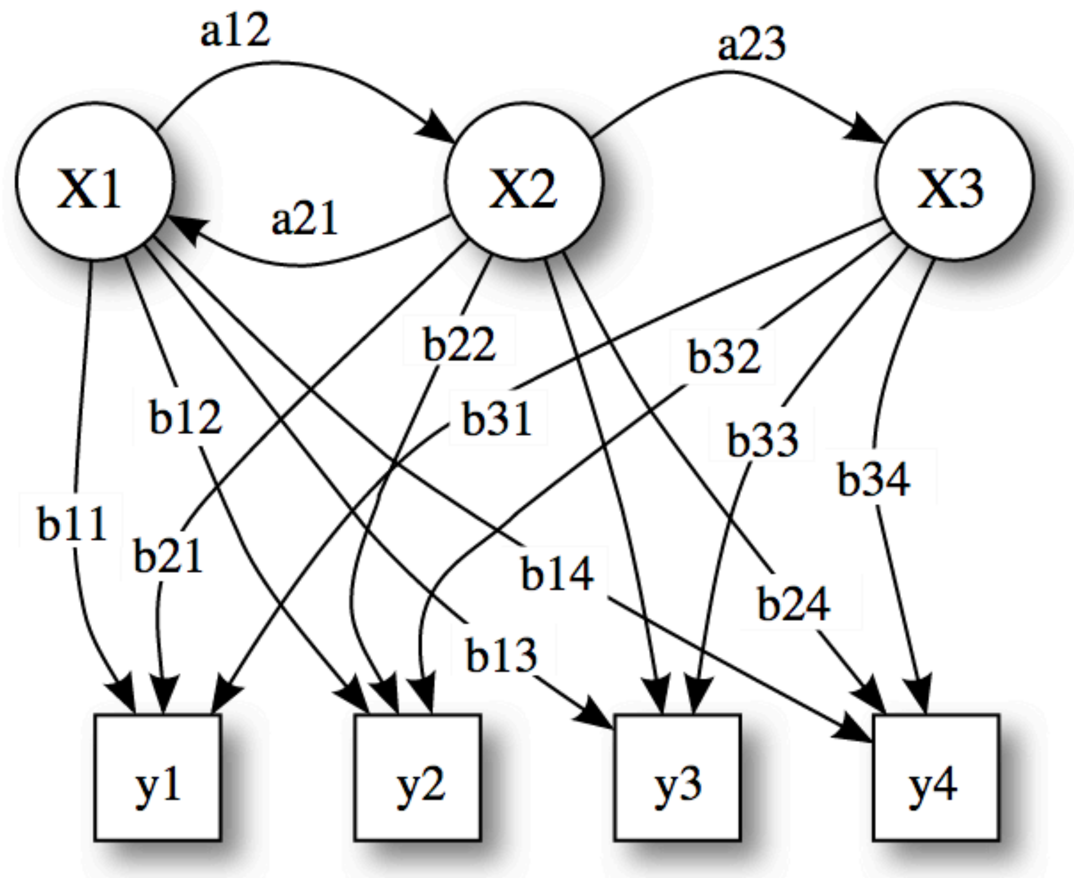
HMM Q2

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

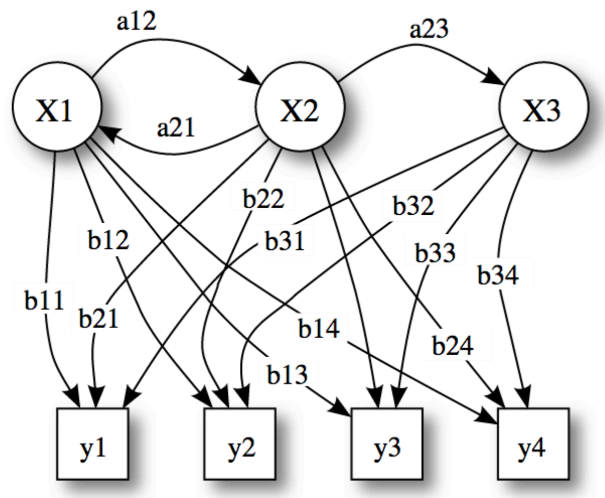
$$
\begin{aligned}
&= \frac{P(u_3=F|w_3=F)P(u_2=T|w_2=F)P(w_3=F|w_2=F)P(w_2=F|w_1=S)}{P(u_3=F)P(u_2=T)} + \\
&\quad \frac{P(u_3=F|w_3=F)P(u_2=T|w_2=R)P(w_3=F|w_2=R)P(w_2=R|w_1=S)}{P(u_3=F)P(u_2=T)} + \\
&\quad \frac{P(u_3=F|w_3=F)P(u_2=T|w_2=S)P(w_3=F|w_2=S)P(w_2=S|w_1=S)}{P(u_3=F)P(u_2=T)} \\
&= \frac{(0.7)(0.3)(0.5)(0.15)}{(0.5)(0.5)} + \\
&\quad \frac{(0.7)(0.8)(0.2)(0.05)}{(0.5)(0.5)} + \\
&\quad \frac{(0.7)(0.1)(0.15)(0.8)}{(0.5)(0.5)} \\
&= 0.119
\end{aligned}
$$

Example



a12
a23
X1
X2
X3
a21
b22
b32
b31
b12
b33
b34
b11
b21
b14
b24
b13
y1
y2
y3
y4

How many parameters?



- The hidden state space is assumed to consist of one of N possible values, modeled as a categorical distribution.

- This means that for each of the N possible states that a hidden variable at time t can be in, there is a transition probability from this state to each of the N possible states of the hidden variable at time t+1, for a total of N^2 transition probabilities.

How many parameters?

- The set of transition probabilities for transitions from any given state must sum to 1. Because any one transition probability can be determined once the others are known, there are a total of N(N-1) transition parameters.

- In addition, for each of the N possible states, there is a set of emission probabilities. The size of this set depends on the nature of the observed variable.

- For example, if the observed variable is discrete with M possible values, governed by a categorical distribution, there will be M-1 separate parameters, for a total of N(M-1) emission parameters over all hidden states.

- If the observed variable is an M-dimensional vector distributed according to an arbitrary multivariate Gaussian distribution, there will be M parameters controlling the means and M(M+1)/2 parameters controlling the covariance matrix, for a total of

$$N\left(M + \frac{M(M+1)}{2}\right) = NM(M+3)/2 = O(NM^2)$$