MARKOV MODELS 1

ARTIFICIAL INTELLIGENCE | COMP 131

TODAY ON AI

- Probability theory recap
- Reasoning over time or space
- Markov models
- Hidden Markov models
- Fundamental challenges in HMMs
- Questions?

- **Conditional probability**: $P(X|Y) = \frac{P(X,Y)}{P(Y)}$

- **Product rule**: $P(X,Y) = P(X|Y)P(Y)$

- **Chain rule**: $P(X_1, \ldots, X_n) = \prod_i P(X_i|X_1, \ldots, X_{i-1})$

- **X and Y are independent iff** $P(X,Y) = P(X)P(Y)$

- **X and Y are conditionally independent given Z iff**
  $X \perp Y|Z$: $P(X,Y|Z) = P(X|Z)P(Y|Z)$

- **Bayes rule**: $P(Cause|Effect) = \frac{P(Effect|Cause)\,P(Cause)}{P(Effect)}$

Often, we want a model able to handle a sequence of observations:

- Speech recognition
- Robot localization
- Classification
- Filtering / Smoothing of sensor data
- State estimate of a system

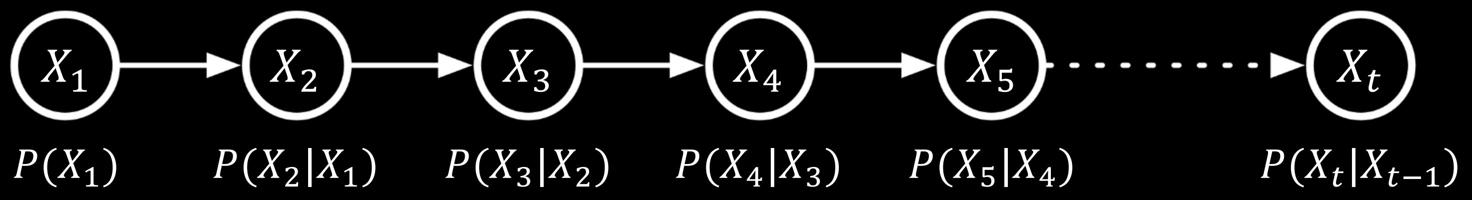
How do we introduce **time** (or **space**) in our models?

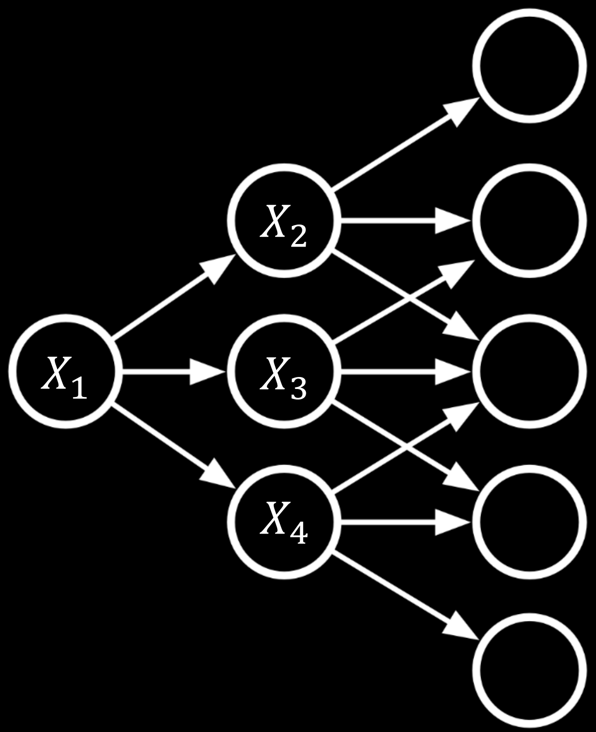
SECTION 01

Markov models

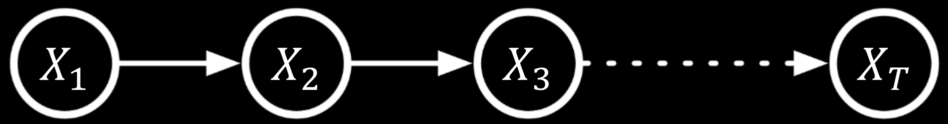
Markov models are Bayes networks with the following assumptions:

- The future is always independent from the past, given the present
- **First-order Markov property**: each time step only depends on the previous
- Value of X at a given time is called **state** of the Markov model





- Primary components of a Markov model are:
  - An **initial prior** specifies where the model starts from. Sometimes is not known
  - A **transition probability** or model dynamics specifies how the state evolves over time



$P(X_1)$ $P(X_2|X_1)$ $P(X_3|X_2)$ $P(X_T|X_{T-1})$

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2)$$

$$P(X_1, \dots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2) \dots P(X_T|X_{T-1})$$
$$= \mathrm{P}(\mathrm{X}_1) \prod_{t=2}^{T} P(X_t|X_{t-1})$$

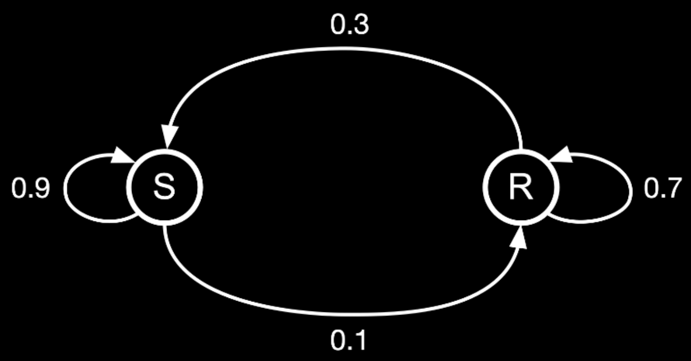
- The **stationarity assumption** states that transition probabilities do not change with time

The transition probability can be represented in four ways:

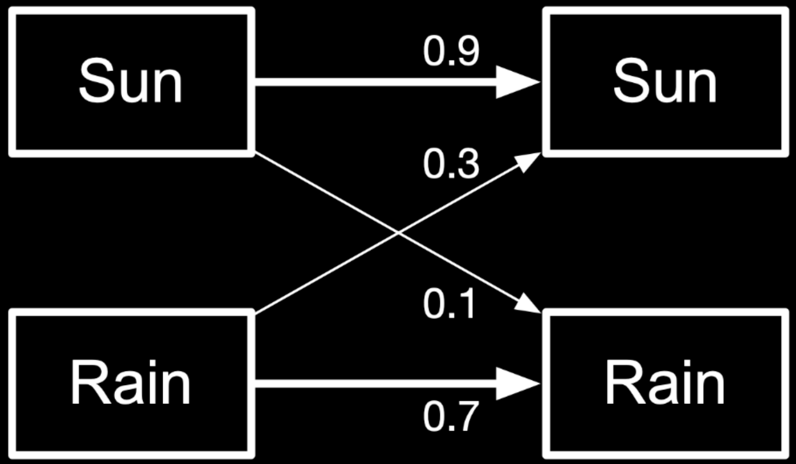
- A transition probability table (CPT):

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

- A **finite state machine** with edges that specify the probability of transition:



- A widely used representation is the **Trellis diagram** through time:



- A **matrix representation**:

$$\begin{array}{cc} sun_{t-1} & \multirow{2}{*}{$\begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$} \\ rain_{t-1} & \\ & sun_t \; rain_t \end{array}$$

- **STATE**

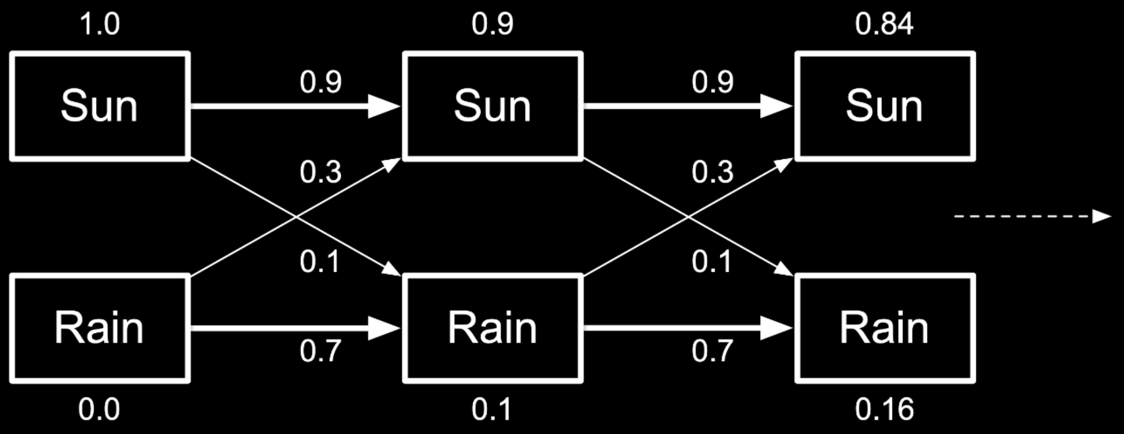
$X = \{rain, sun\}$

- **INITIAL PRIOR**

$P(X_1 = sun) = 1.0$

- **TRANSITION PROBABILITY**

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



$$P(X_2 = sun) = P(X_2 = sun|X_1 = sun)P(X_1 = sun) + P(X_2 = sun|X_1 = rain)P(X_1 = rain)$$

$$P(X_2 = sun) = \mathbf{0.9}$$

What about after time t?

$$P(X_1) = \textit{known prior}$$

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

Any Markov models converge to a **stationary distribution P_∞**:

- Influence of the initial distribution gets less and less over time
- P_∞ does not depend on the initial distribution
- It is defined as follow: $P_\infty(X) = P_{\infty+1}(X) = \sum_{X_{t-1}} P(X_t|X_{t-1})P_\infty(X_{t-1})$

- For the example:

$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9\ P_\infty(sun) + 0.3\ P_\infty(rain)$$
$$P_\infty(rain) = 0.1\ P_\infty(sun) + 0.7\ P_\infty(rain)$$

$$\begin{cases} P_\infty(sun) = 3\ P_\infty(rain) \\ P_\infty(rain) = \frac{1}{3} P_\infty(sun) \\ P_\infty(sun) + P_\infty(rain) = 1 \end{cases}$$

$$P_\infty(sun) = \frac{3}{4}$$
$$P_\infty(rain) = \frac{1}{4}$$

SECTION 02

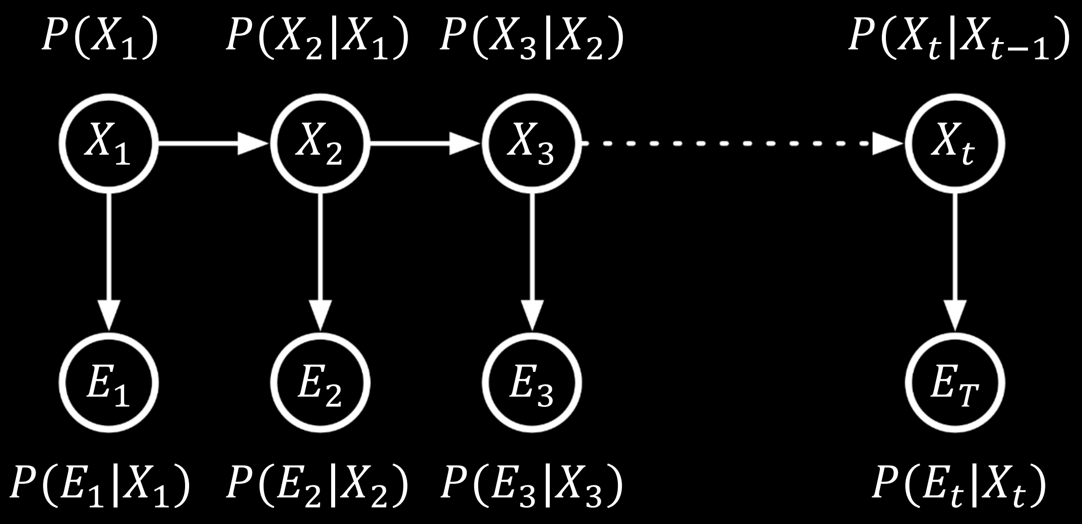
Hidden Markov models

- Speech recognition HMMs:
  - Observations are acoustic signals (continuous valued)
  - States are specific positions in specific words (so, tens of thousands)

- Classification:
  - Observations are sensor readings
  - States are the classes of the samples

- Robot tracking:
  - Observations are range readings (continuous)
  - States are positions on a map (continuous)

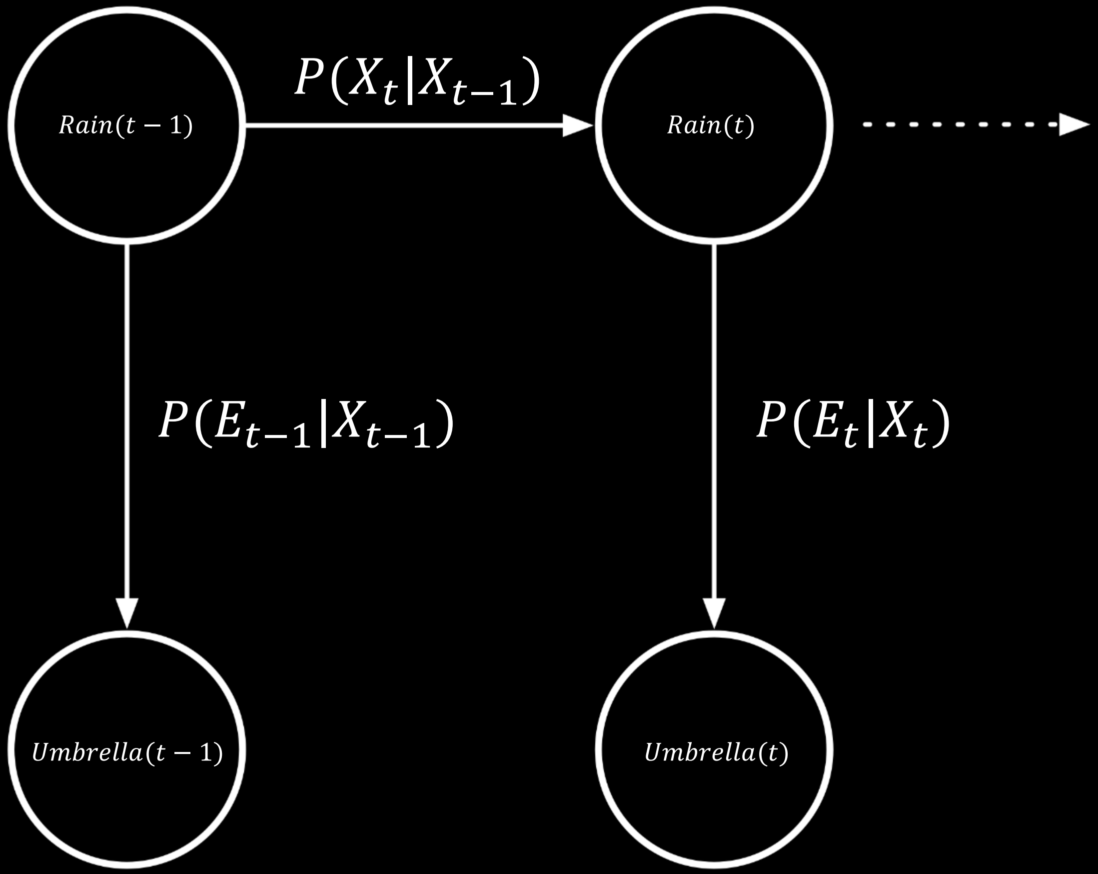
Hidden Markov models (HMMs) describe an underlying hidden state conditionally dependent to some observed evidence:



$$X_t \perp X_1, \dots, X_{t-2} \mid X_{t-1}$$

$$P(X_1, E_1, \dots, X_t, E_t) = P(X_1)P(E_1|X_1)\prod_{i=2}^{t} P(X_i|X_{i-1})P(E_i|X_i)$$

- In addition to the definitions for MMs, there is an **Emission CPT**
- Evidence variables are not independent because correlate via the hidden states



TRANSITION CPT

R_{t-1}	R_t	$P(R_t \mid R_{t-1})$
$\neg rain$	$\neg rain$	0.7
$\neg rain$	$rain$	0.3
$rain$	$\neg rain$	0.3
$rain$	$rain$	0.7

EMISSION CPT

R_t	U_t	$P(U_t \mid R_t)$
$rain$	$umbrella$	0.9
$rain$	$\neg umbrella$	0.1
$\neg rain$	$umbrella$	0.2
$\neg rain$	$\neg umbrella$	0.8

READINGS

Chapter 15, 16, and 17

QUESTIONS ?

Tufts UNIVERSITY

ARTIFICIAL INTELLIGENCE

COMP 131

FABRIZIO SANTINI

VERSION 3.0