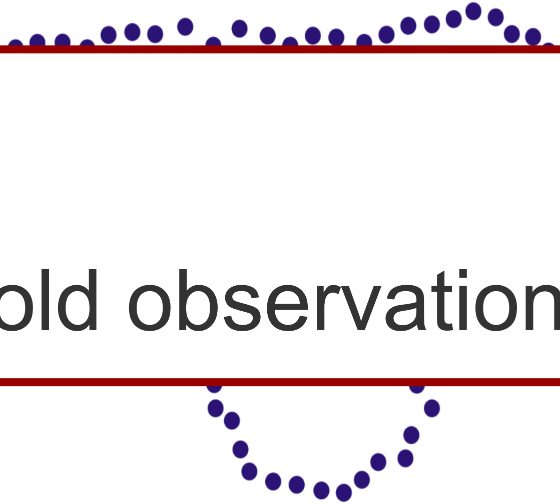
Hidden Markov Model

Motivation

Sequential data often arise from measurements of **time series**



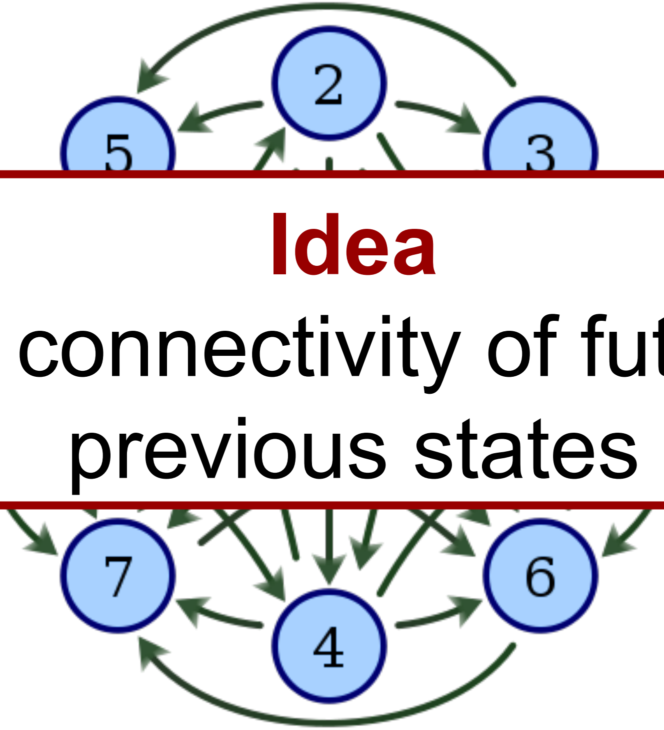
Goal

Predict new values given old observations

Weather Forecast

Person’s position over time

Markov Model



Idea

Restrict the connectivity of future states to previous states

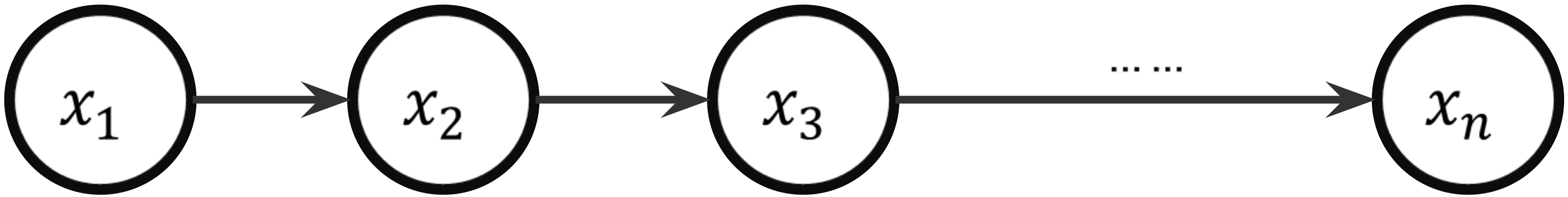
Outline

- Markov Model
- Hidden Markov Model
- Inference using HMM
- Case Study: Speech Recognition
- Conclusion

Markov Model

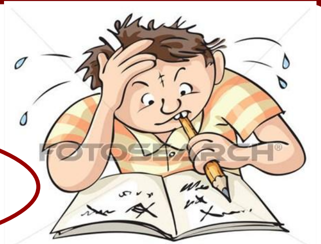
- **Markov Models:** Only assume a certain number of previous states to be relevant

- *First*-order Markov Models: Only the last state is relevant for the current state



Markov Model

- Recall: Chain Rule

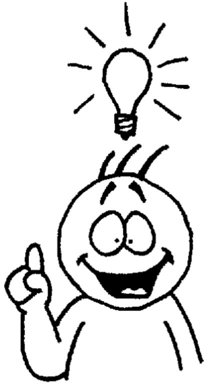


$$p(x_1, x_2, x_3, \ldots, x_n) = \prod_{k=1}^{N} p(x_k \mid x_1, \ldots, x_{k-1})$$

- *First-order* Markov Models simplify this to

$$p(x_1, \ldots, x_K) = p(x_1) \prod_{k=2}^{K} p(x_k \mid x_{k-1})$$

by only taking
the **last state** into account



Example

R_{k-1}	$p(R_k)$	$p(\neg R_k)$
true	0.7	0.3
false	0.3	0.7

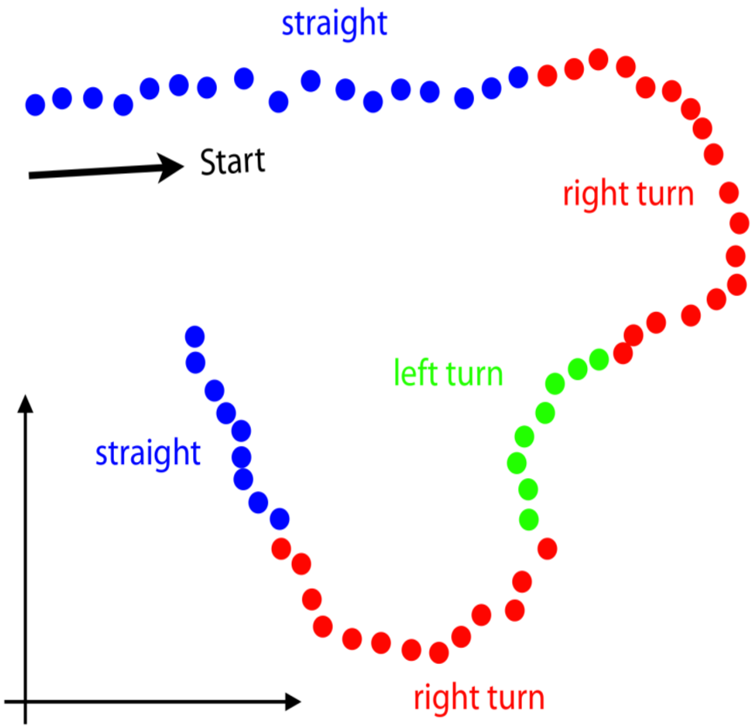
Transition model

State Space Model

Input: Human trajectory in 2d

Output: Classification
- Walking straight
- Right turn
- Left turn

Problem:
State space is not observation space



Hidden Markov Model


Is it raining outside??

Weather (Rain/ No rain) → Hidden

Umbrella (Yes/ No) → Observation

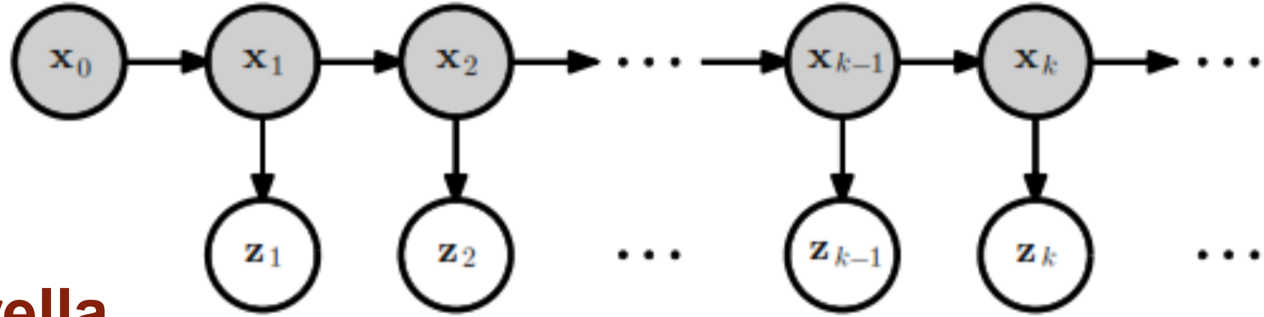
The rain **causes** the umbrella to appear !

Hidden Markov Model

Weather

Latent (Hidden) State Variable

Discrete



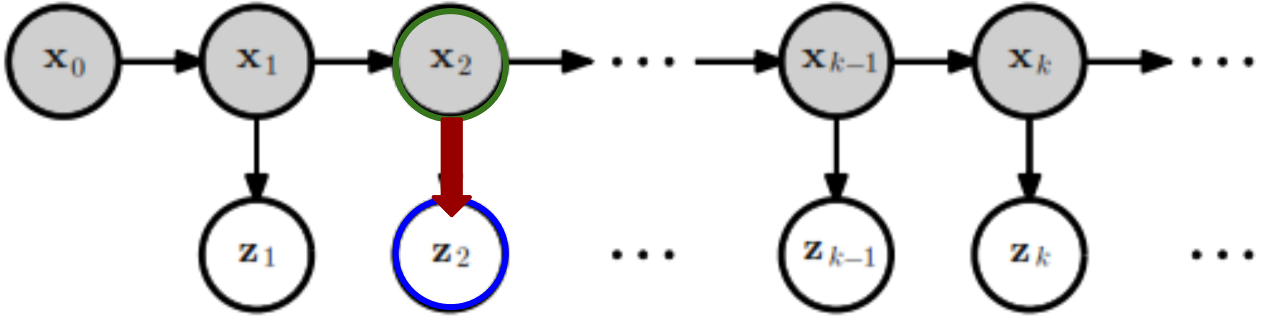
Umbrella

Observation Variable

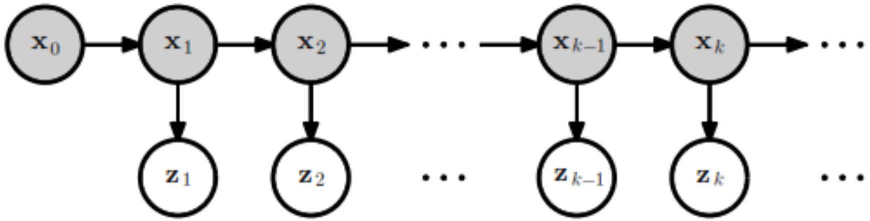
Discrete or Continuous

State Space Model

An **observation** at time k is only dependent on the **state** of time k and independent of all states from the beginning to k-1

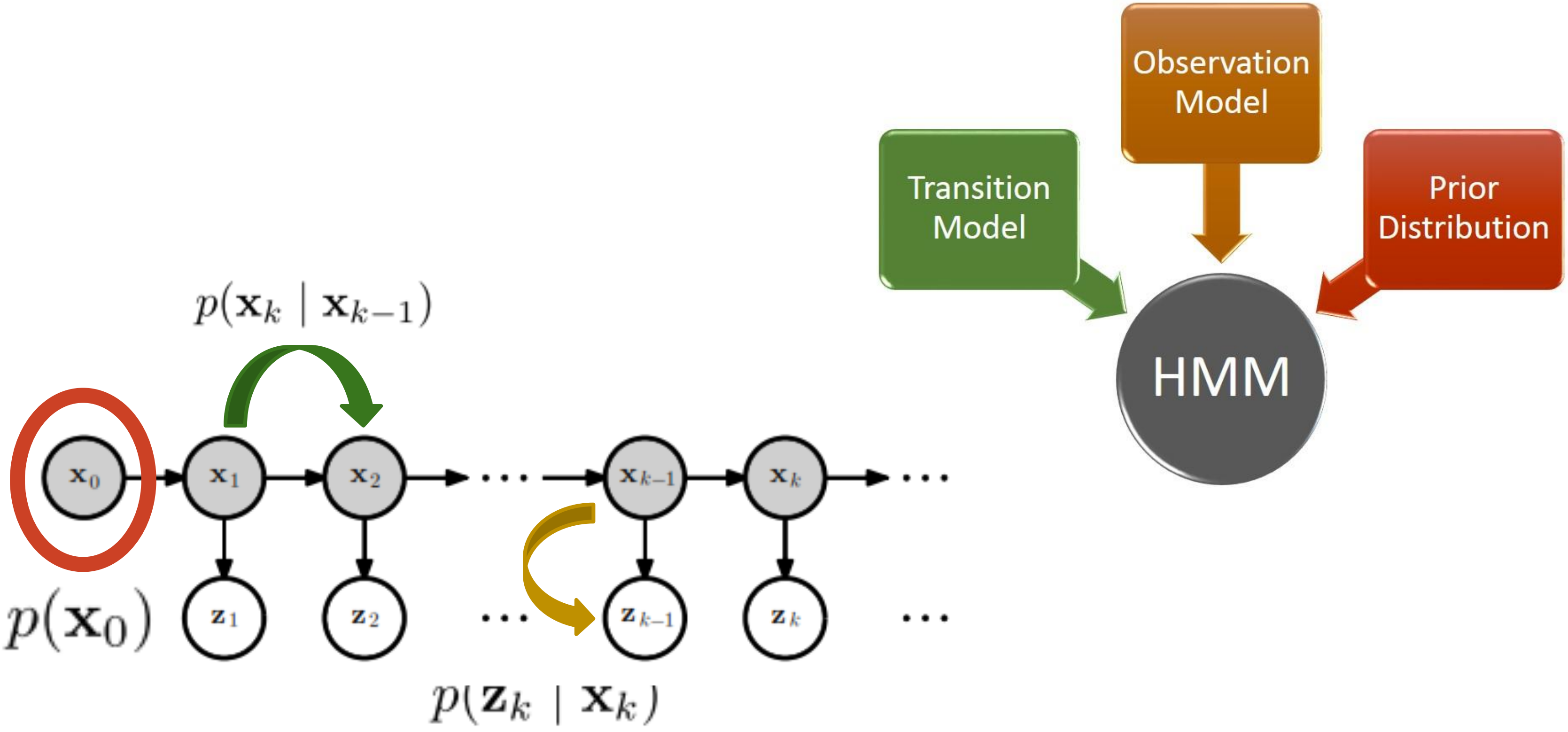


Hidden Markov Model



$$p(\mathbf{x}_1, \ldots, \mathbf{x}_K, \mathbf{z}_1, \ldots, \mathbf{z}_K)$$

Hidden Markov Model



Transition Model

$$p(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$$

j (next)

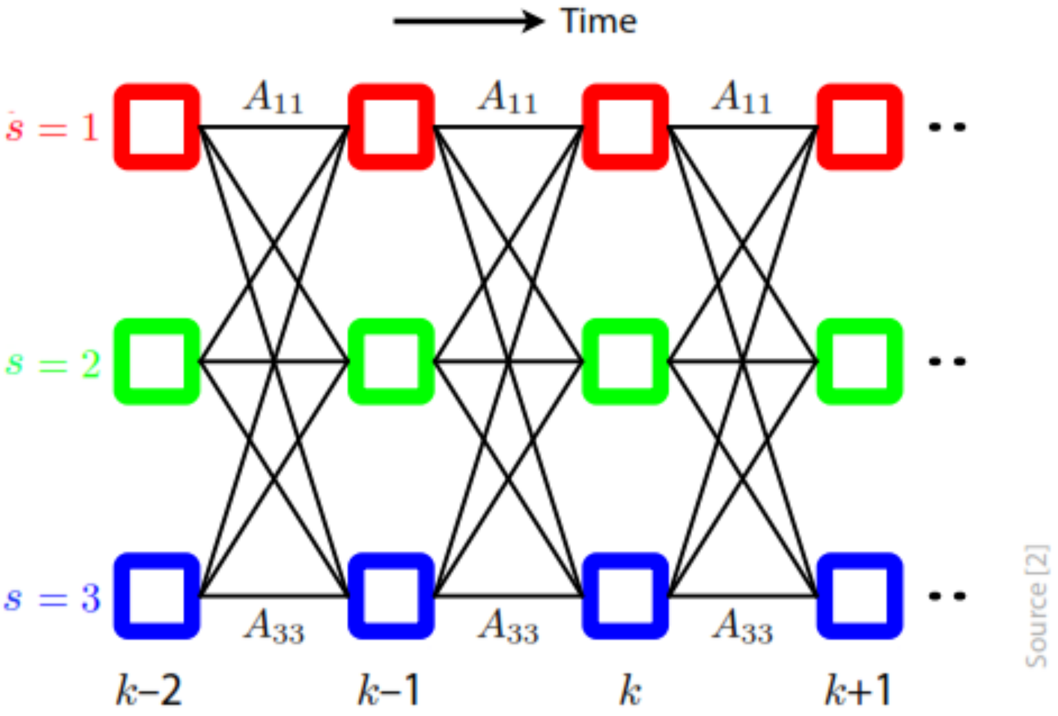
i (current)

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

$\sum_j A_{1j} = 1$

Probability of a transition from state 3 to state 2

Transition Model



Observation Model

$$p(\mathbf{z}_k \mid \mathbf{x}_{0:k}, \mathbf{z}_{0:k-1}) = p(\mathbf{z}_k \mid \mathbf{x}_k)$$

j (observation symbol)

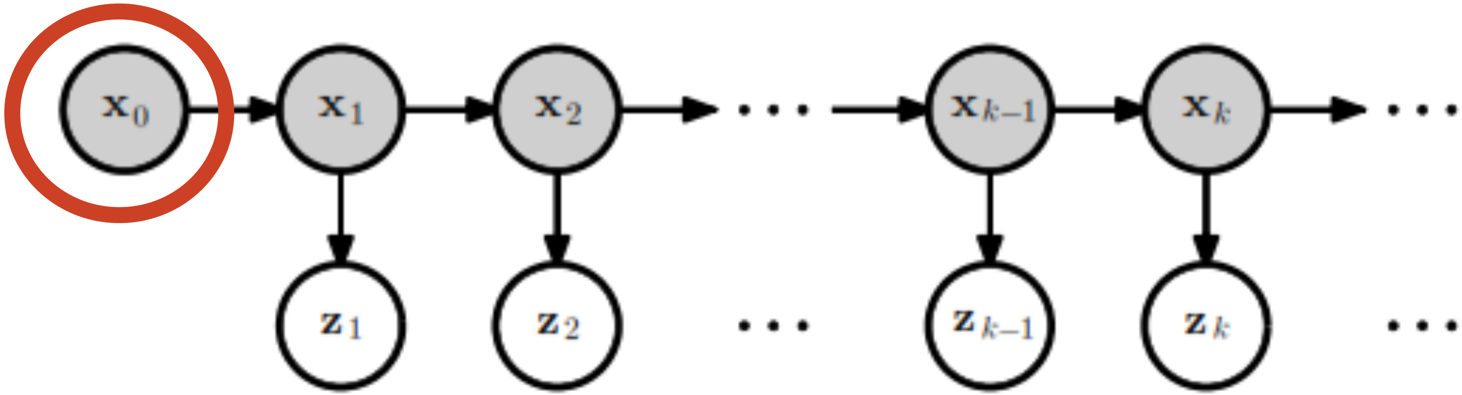
i (states)

$$E = \begin{pmatrix} E_{11} & E_{12} & E_{13} & E_{14} & E_{15} \\ E_{21} & E_{22} & E_{23} & E_{24} & E_{25} \\ E_{31} & E_{32} & E_{33} & E_{34} & E_{35} \end{pmatrix}$$

$\sum_j E_{1j} = 1$

Emission probability of symbol 4 from state 3

Prior Probability



Hidden Markov Model

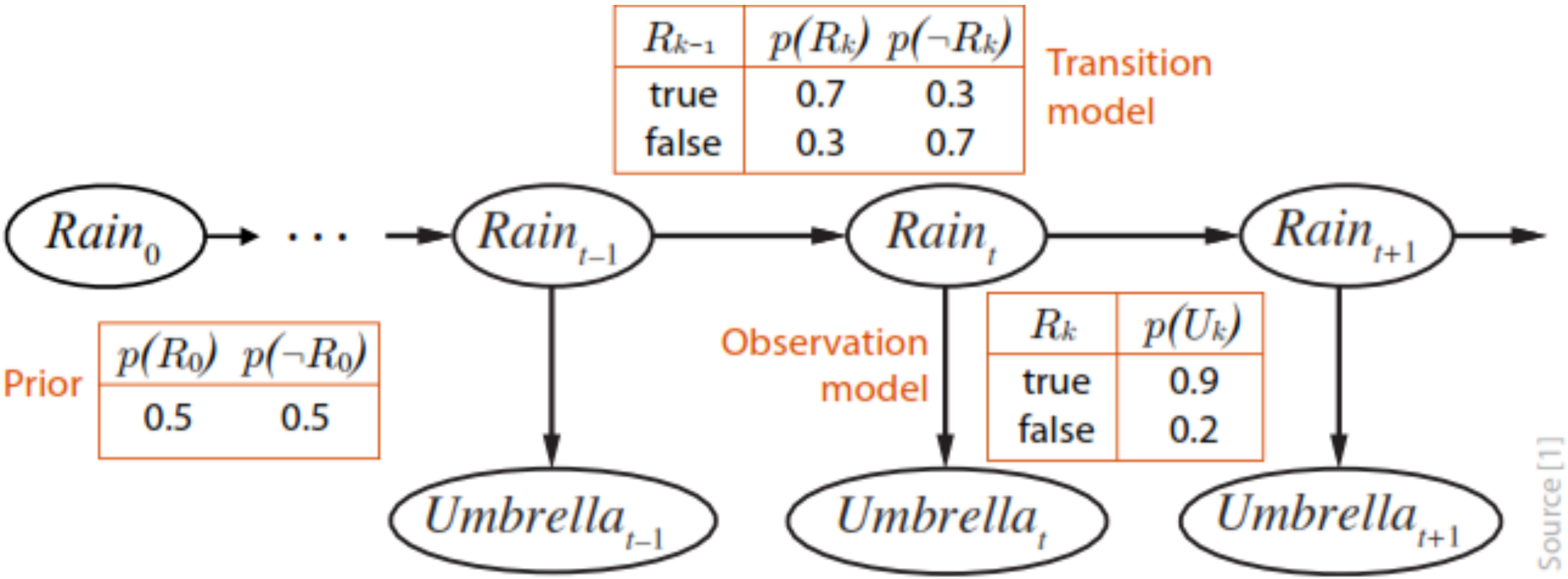
$$p(\mathbf{x}_{0:K}, \mathbf{z}_{1:K}) = p(\mathbf{x}_0) \prod_{k=1}^{K} p(\mathbf{x}_k | \mathbf{x}_{k-1})\, p(\mathbf{z}_k | \mathbf{x}_k)$$

Prior

Transition model

Observation model

Hidden Markov Model



R_{k-1}	$p(R_k)$	$p(\neg R_k)$
true	0.7	0.3
false	0.3	0.7

$p(R_0)$	$p(\neg R_0)$
0.5	0.5

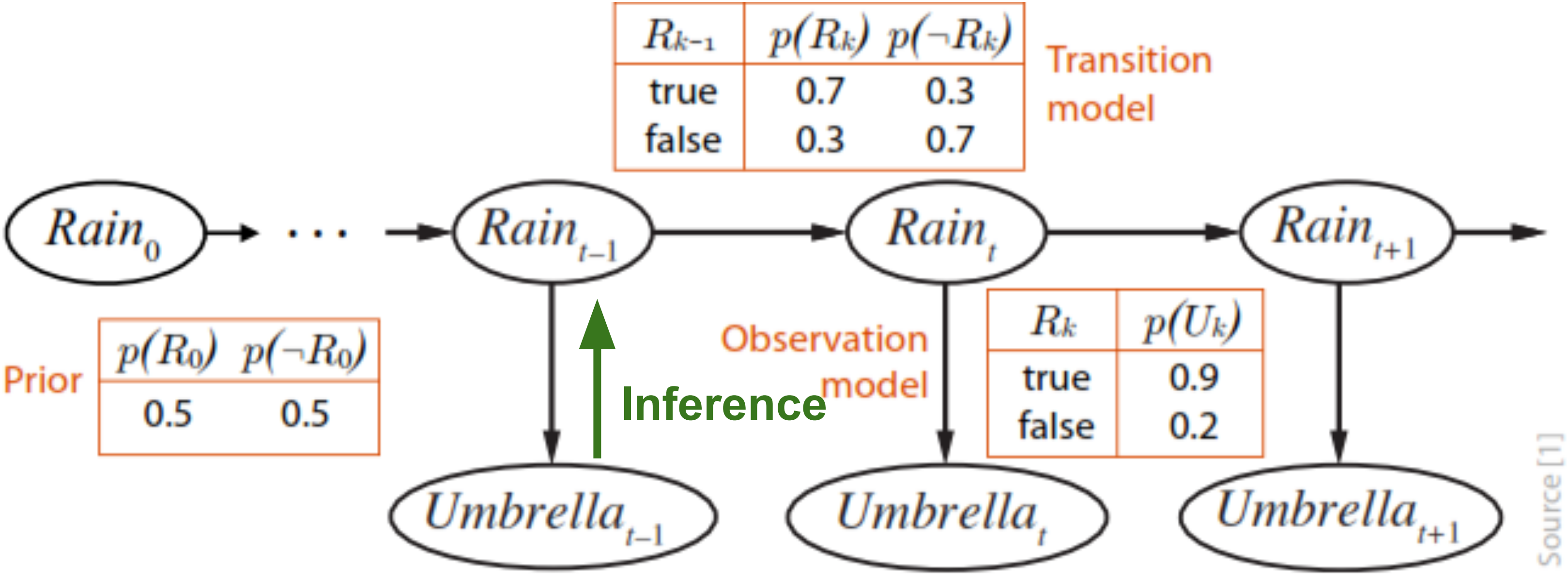
R_k	$p(U_k)$
true	0.9
false	0.2

Important Note

- First-order Markov Models are often inaccurate, since you throw away any information about the past
- Increasing the accuracy by more general models
  - Increasing the order of the Markov model
  - Adding more state variables

Inference

Hidden Markov Model



Inference Tasks

1. Filtering

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$$

Present

Inference Tasks

1. Filtering $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$
2. Smoothing

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:k}) \quad 0 \leq t < k$$

Past

Inference Tasks

1. Filtering $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$
2. Smoothing $p(\mathbf{x}_t \mid \mathbf{z}_{1:k}) \quad 0 \leq t < k$
3. Prediction

$$p(\mathbf{x}_{k+t} \mid \mathbf{z}_{1:k}) \quad t > 0$$

Future

Inference Tasks

1. **Filtering** $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$
2. **Smoothing** $p(\mathbf{x}_t \mid \mathbf{z}_{1:k}) \quad 0 \leq t < k$
3. **Prediction** $p(\mathbf{x}_{k+t} \mid \mathbf{z}_{1:k}) \quad t > 0$
4. **Most Likely Sequence**

$$\arg\max_{\mathbf{x}_{1:k}} p(\mathbf{x}_{1:k} \mid \mathbf{z}_{1:k})$$

Inference??

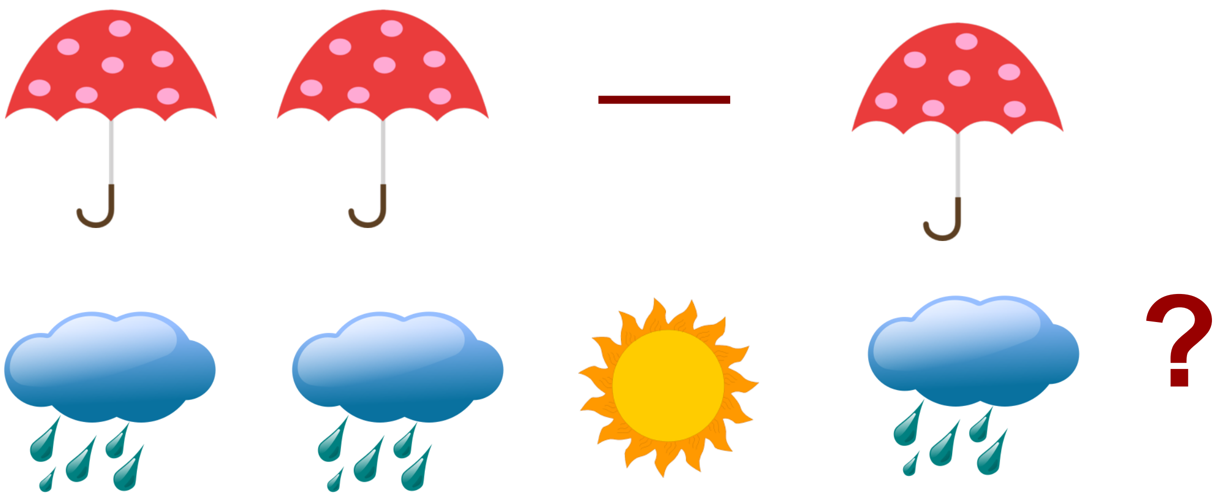
Prediction??

Inference Example

	Day 1	Day 2	Day 3	Day 4

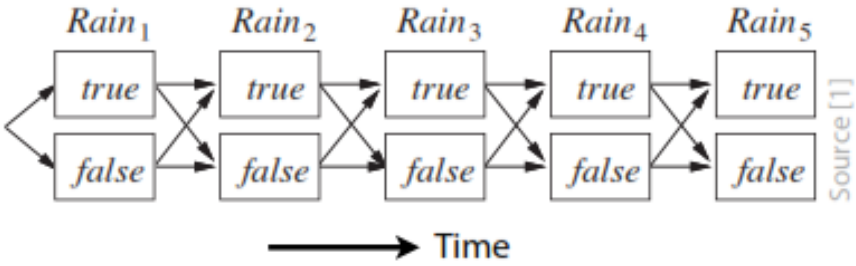
Observation:

Weather:



What is the **most likely weather sequence**??

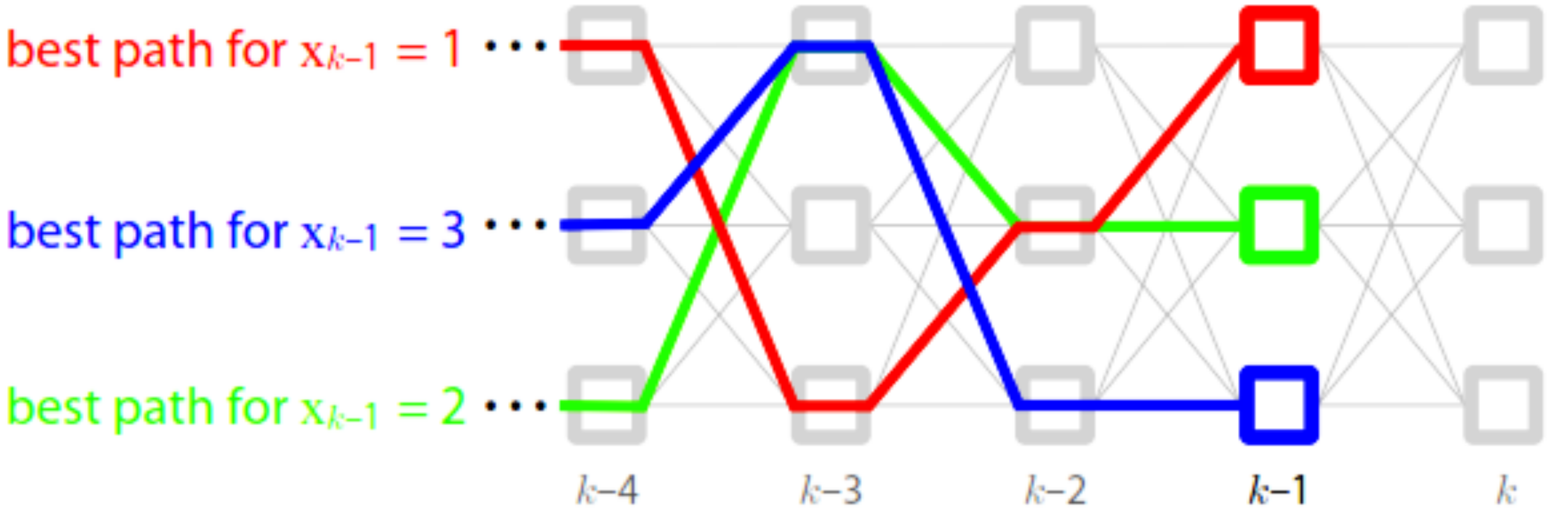
Most Likely Sequence



- Recall, we want to find the state sequence $\mathbf{x}^*$ that **maximizes the probability along its path**, i.e. $\mathbf{x}^* = \arg\max_{\mathbf{x}_{1:k}} p(\mathbf{x}_{1:k} \mid \mathbf{z}_{1:k})$

Viterbi Algorithm

- Maximizing path for **k-1**

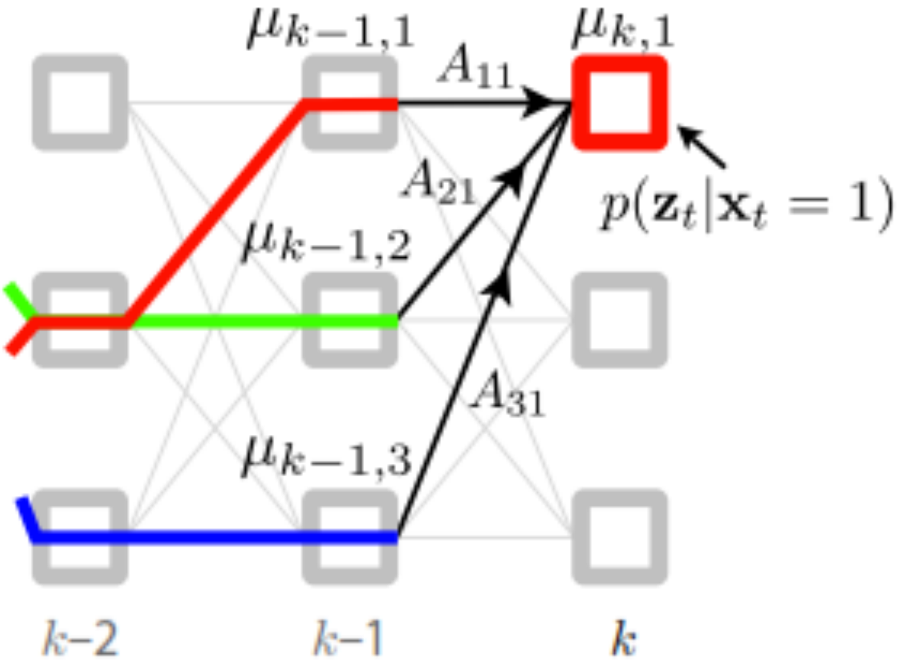


Viterbi Algorithm

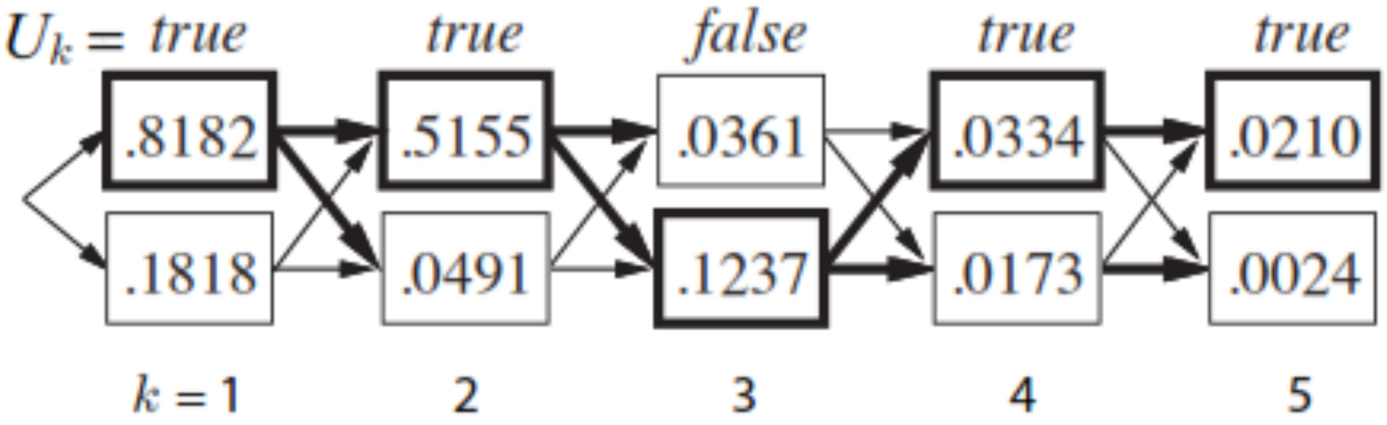
The algorithm computes

$\mu_k = \max_{\mathbf{x}_{1:k-1}} p(\mathbf{x}_{1:k}, \mathbf{z}_{1:k})$ as

$$\mu_k = p(\mathbf{z}_k \mid \mathbf{x}_k) \max_{\mathbf{x}_{k-1}} \left(p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \, \mu_{k-1} \right)$$

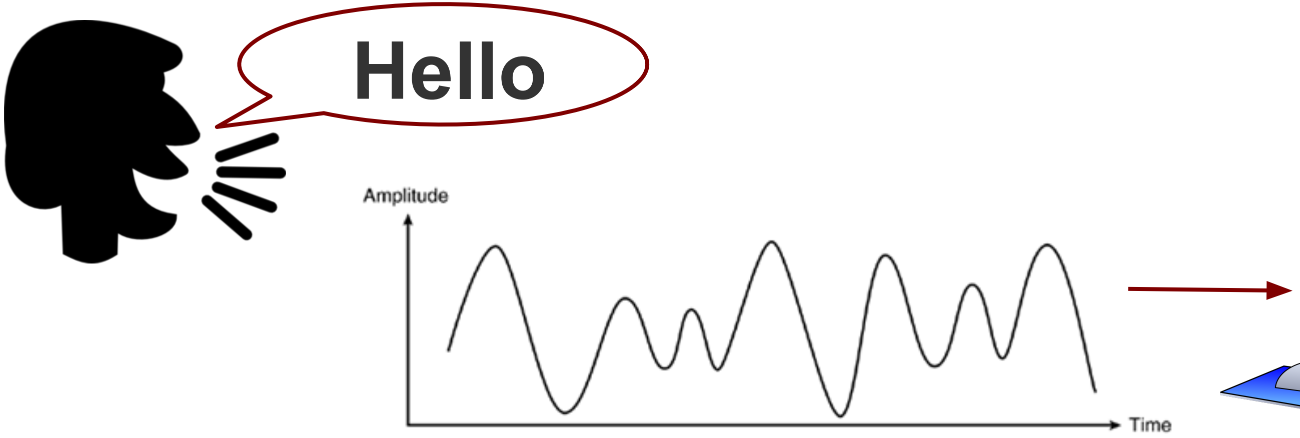


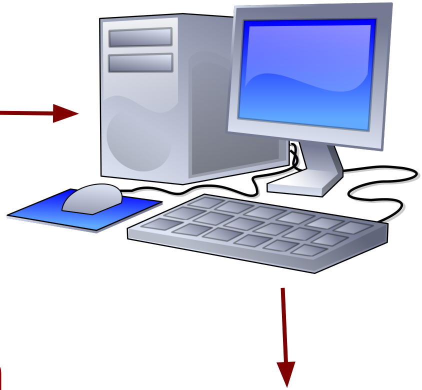
Viterbi Algorithm



Source [1]

Speech Recognition





States?

Observations?

Transition Model?

"Hello"

Observation Model?

Prior?

Strengths & Weaknesses of HMM

HMM **provides better results** than MM

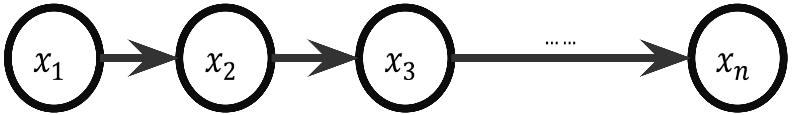
Principle of HMM can be adapted to **many problems**, e.g. finding alignment

Computationally **expensive**, both in memory and time

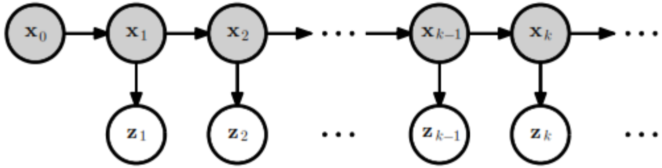
Need **more training** than classical Markov Models

Summary

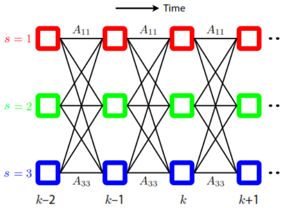
Markov Model



Hidden MM



Transition Model



Observation Model

j (observation symbol)

$$E = \begin{pmatrix} E_{11} & E_{12} & E_{13} & E_{14} & E_{15} \\ E_{21} & E_{22} & E_{23} & E_{24} & E_{25} \\ E_{31} & E_{32} & E_{33} & E_{34} & E_{35} \end{pmatrix}$$

i (states)

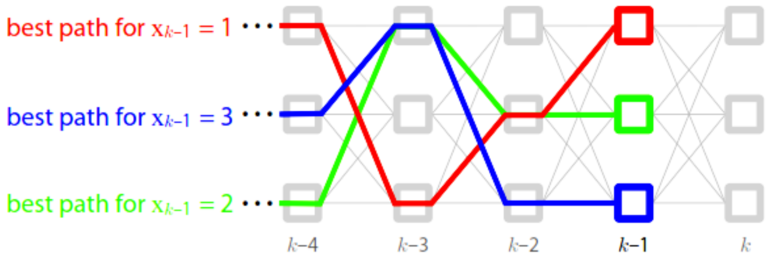
Emission probability of

Prior

Summary

Inference using HMM

- Filtering
- Smoothing
- Prediction
- Most likely Sequence
  - Viterbi Algorithm



Thank you

Sources

Most important graphics from this presentation are taken from

http://srl.informatik.uni-freiburg.de/teachingdir/ws13/slides/09-TemporalReasoning-1.pdf