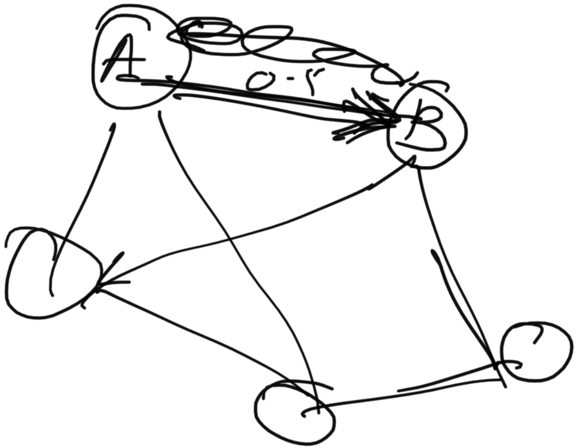
CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Graphical Models

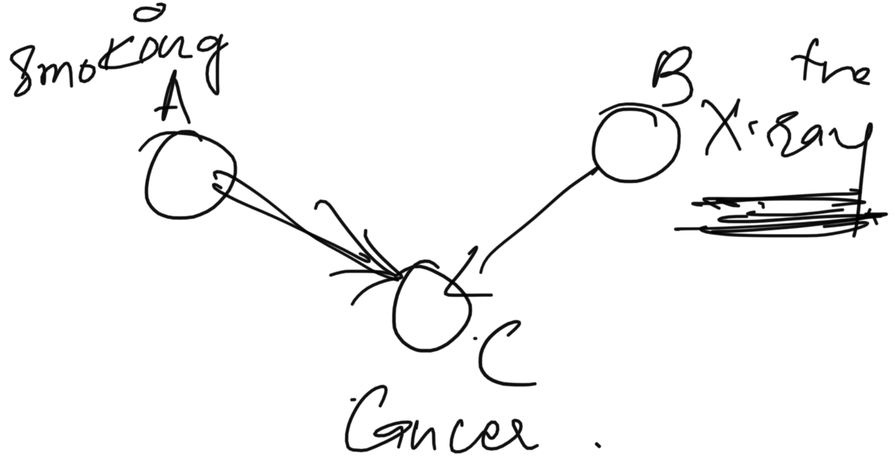
Arizona State University

Table of contents



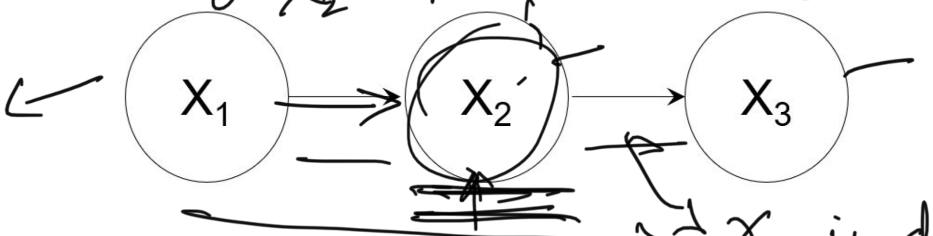
Bayesian Networks

- A probabilistic graphical model representing a set of variables and their conditional dependencies via a directed acyclic graph.



Conditional Independence

0 X_2 dependent on X_1 2



X_3 is dependent on X_2

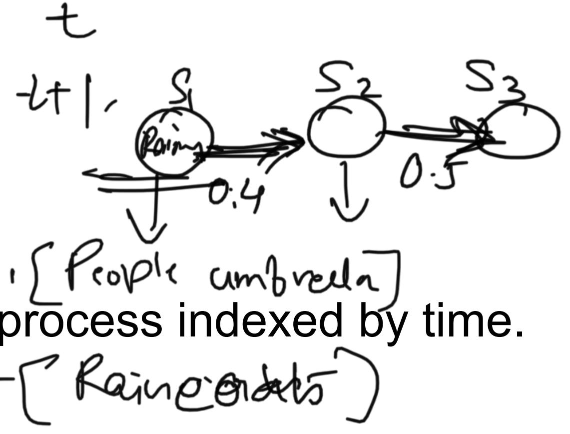
- Are X_1 and X_3 independent?

— No.

- Are they conditionally independent given X_2?

— Yes.

Hidden Markov Model (HMM)



- Dynamic Bayesian network - modeling the process indexed by time.

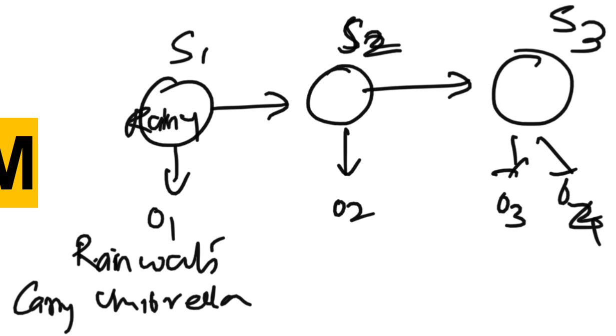
- Two assumptions -

  - First-order Markov chain - $P(s^t=S_j \mid s^{t-1}=S_i)$

  - $a_{ij} = P(s^t=S_j \mid s^{t-1}=S_i)$, $1 \le i,j \le N$, for any t — Stationary.

$P(S_3 \mid S_2) = 0.5$

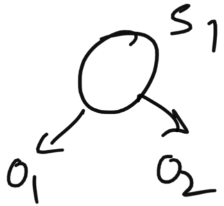
Specifying HMM



- Θ - set of hidden states $= \{S_1, S_2, S_3\}$
- **Ω** - set of output observations $= \{O_1, O_2, O_3, O_4\}$
- π - initial state distribution
  - $\pi = \{\pi_i\}$, $\pi_i = P(s^1 = S_i)$ $= S_2$

$\{0, 1, 0\}$ $P(S^1 = S_2) = 1$

Specifying HMM



$P(o_1 | s_1) = 0.5$

$P(o_2 | s_1) = 0.5$

- State transition matrix - A — Shape of A = $n \times n$
  - $A=\{a_{ij}\}$ where $a_{ij} = P(s^t=S_j \mid s^{t-1}=S_i)$ for $1 \le i,j \le N$
- Observation/emission probability matrix B
  - $B = \{b_{jk}\} = P(o^t=v_k \text{ at } t \mid s^t=S_j)$ where v_k is the k^{th} symbol in Ω

$n \rightarrow$ states K – observations

$$\begin{array}{cc} & \begin{array}{ccc} o_1 & o_2 & K \end{array} \\ \begin{array}{c} S_1 \\ n \\ \\ \end{array} & \begin{bmatrix} 0.5 & 0.5 & \\ & & \\ & & \end{bmatrix} \end{array}$$

Problems in HMM

state transition matrix

Emission prob. matrix

- For a given HMM $\Lambda=\{\Theta,\Omega,A,B,\pi\}$
    - Estimation of model parameters — A, B
    - Given an observation sequence **O**= $\{o^1,o^2, \dots ,o^k\}$, what is the most likely state sequence $S = \{s^1, s^2, \dots ,s^k\}$ that has produced O? —Decoding.
    - How likely is an observation O? —likelihood.

HMM Parameter Estimation

$S_1(O_1) - S_2(O_2)$

- Given labeled data - state and observation

$\rightarrow S_1(O_1) \rightarrow S_2(O_2) \rightarrow S_3(O_1)$ —

$$a(S_i|S_j) = \frac{\text{number of } (s^t = S_i, s^{t-1} = S_j)}{\text{number of } S_j}$$

transition probability

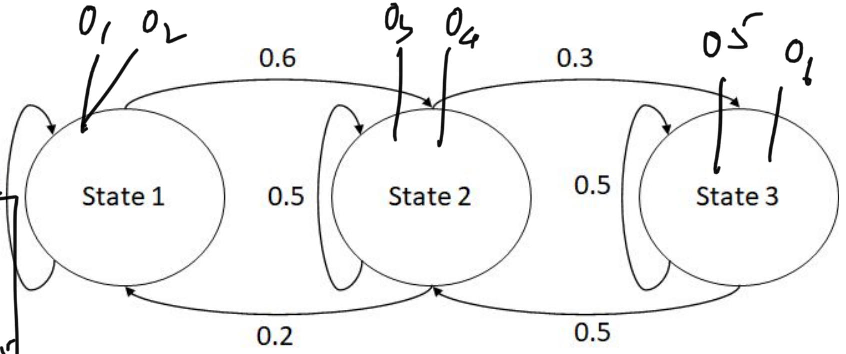
$$b(o_r|S_j) = \frac{\text{number of } (o^t = o_r, s^t = S_j)}{\text{number of } S_j}$$

emission probability

~~Sta~~ . Given data - sequence of observations.

Forward-Backward Algorithm.

HMM Parameter Estimation Example



$$
\begin{array}{c|cccccc}
 & o_1 & o_2 & o_3 & o_4 & o_5 & o_6 \\
\hline
s_1 & 0.5 & 0.5 & & & & \\
s_2 & & & 0.5 & 0.5 & & \\
s_3 & & & & & 0.5 & 0.5
\end{array}
$$

$$
A = \begin{array}{c|ccc}
 & s_1 & s_2 & s_3 \\
\hline
s_1 & 0.4 & 0.6 & 0 \\
s_2 & 0.2 & 0.5 & 0.3 \\
s_3 & 0 & 0.5 & 0.5
\end{array} = 1
$$

Questions?