Graphical Models
and
Bayesian Networks

Required reading:

• Ghahramani, section 2, "Learning Dynamic Bayesian Networks" (just 3.5 pages :-)

Optional reading:

• Mitchell, chapter 6.11 Bayesian Belief Networks

Machine Learning 10-701

Tom M. Mitchell
Center for Automated Learning and Discovery
Carnegie Mellon University

November 1, 2005

Graphical Models

- Key Idea:
  - Conditional independence assumptions useful
  - but Naïve Bayes is extreme!
  - Graphical models express sets of conditional independence assumptions via graph structure
  - Graph structure plus associated parameters define joint probability distribution over set of variables/nodes
- Two types of graphical models:
  - Directed graphs (aka Bayesian Networks) ← today
  - Undirected graphs (aka Markov Random Fields)

Graphical Models – Why Care?

- Among most important ML developments of the decade
- Graphical models allow combining:
  - Prior knowledge in form of dependencies/independencies
  - Observed data to estimate parameters
- Principled and ~general methods for
  - Probabilistic inference
  - Learning
- Useful in practice
  - Diagnosis, help systems, text analysis, time series models, ...

Marginal Independence

Definition: X is marginally independent of Y if

$$(\forall i, j) P(X = x_i, Y = y_j) = P(X = x_i) P(Y = y_j)$$

Equivalently, if

$$(\forall i, j) P(X = x_i | Y = y_j) = P(X = x_i)$$

Equivalently, if

$$(\forall i, j) P(Y = y_i | X = x_j) = P(Y = y_i)$$

Conditional Independence

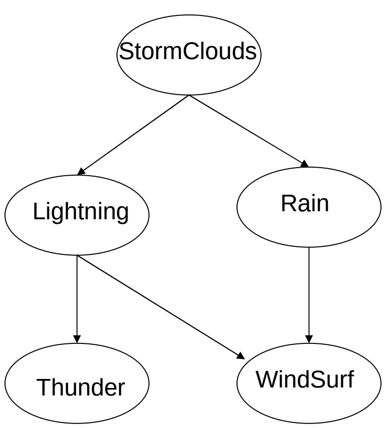
Definition: X is conditionally independent of Y given Z, if the probability distribution governing X is independent of the value of Y, given the value of Z

$$(\forall i, j, k) P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z = z_k)$$

Which we often write $P(X|Y, Z) = P(X|Z)$

E.g., $P(Thunder|Rain, Lightning) = P(Thunder|Lightning)$

Bayesian Network



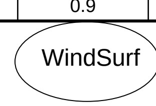
Bayes network: a directed acyclic graph defining a joint probability distribution over a set of variables

Each node denotes a random variable

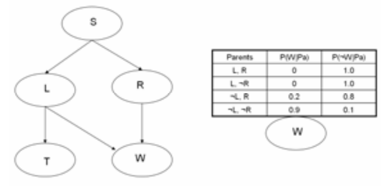
Each node is conditionally independent of its non-descendents, given its immediate parents.

A conditional probability distribution (CPD) is associated with each node N, defining P(N | Parents(N))

Parents	P(W|Pa)	P(¬W|Pa)
L, R	0	1.0
L, ¬R	0	1.0
¬L, R	0.2	0.8
¬L, ¬R	0.9	0.1



Bayesian Networks



- Each node denotes a variable
- Edges denote dependencies
- CPD for each node X_i describes $P(X_i \mid Pa(X_i))$
- Joint distribution given by

$$P(X_1 \ldots X_n) = \prod_i P(X_i | Pa(X_i))$$

- Node X_i is conditionally independent of its non-descendents, given its immediate parents

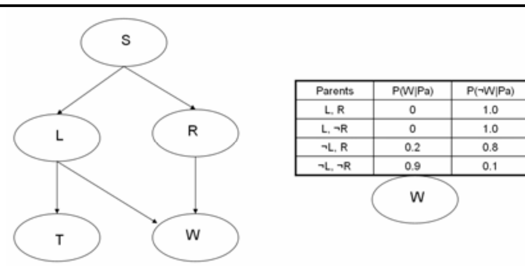
Parents = Pa(X) = immediate parents

Antecedents = parents, parents of parents, ...

Children = immediate children

Descendents = children, children of children, ...

Bayesian Networks



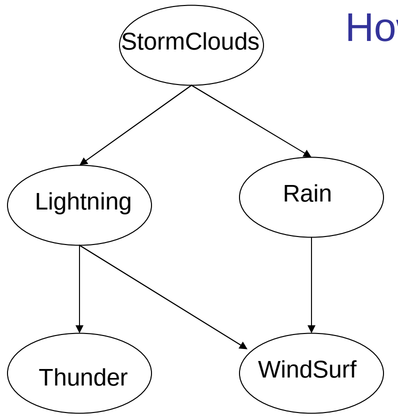
- CPD for each node X_i describes $P(X_i \mid Pa(X_i))$

- Chain rule of probability:

$$P(S, L, R, T, W) = P(S)P(L|S)P(R|S, L)P(T|S, L, R)P(W|S, L, R, T)$$

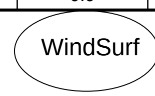
- But in Bayes net:

$$P(X_1 \ldots X_n) = \prod_i P(X_i | Pa(X_i))$$



How Many Parameters?

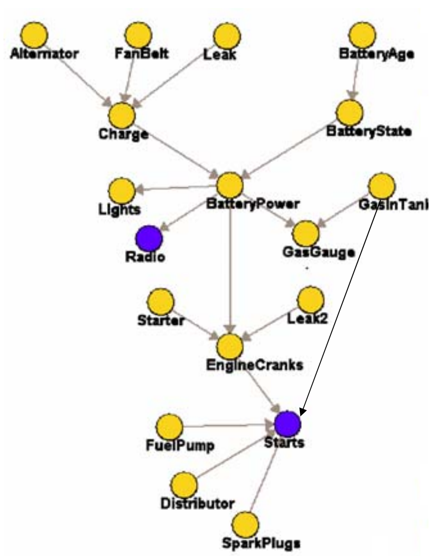
Parents	P(W|Pa)	P(¬W|Pa)
L, R	0	1.0
L, ¬R	0	1.0
¬L, R	0.2	0.8
¬L, ¬R	0.9	0.1



In full joint distribution?

Given this Bayes Net?

Bayes Net



Inference:

P(BattPower=t | Radio=t, Starts=f)

Most probable explanation:

What is most likely value of Leak, BatteryPower given Starts=f?

Active data collection:

What is most useful variable to observe next, to improve our knowledge of node X?

Algorithm for Constructing Bayes Network

- Choose an ordering over variables, e.g., X_1, X_2, ... X_n
- For i=1 to n
  - Add X_i to the network
  - Select parents $Pa(X_i)$ as minimal subset of X_1 ... X_{i-1} such that

$$P(X_i|Pa(X_i)) = P(X_i|X_1, \ldots, X_{i-1})$$

Notice this choice of parents assures

$$P(X_1 \ldots X_n) = \prod_i P(X_i|X_1 \ldots X_{i-1}) \quad \text{(by chain rule)}$$

$$= \prod_i P(X_i|Pa(X_i)) \quad \text{(by construction)}$$

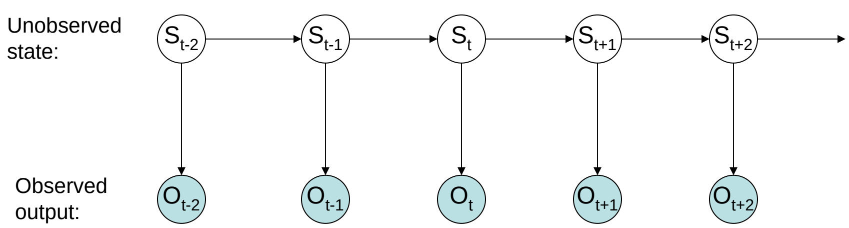
Example

- Bird flu and Allegies both cause Nasal problems
- Nasal problems cause Sneezes and Headaches

What is the Bayes Network for Naïve Bayes?

Bayes Network for a Hidden Markov Model

Assume the future is conditionally independent of the past, given the present

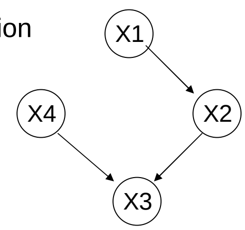


$$P(S_{t-2}, O_{t-2}, S_{t-1}, \ldots, O_{t+2}) =$$

Conditional Independence, Revisited

- We said:
  - Each node is conditionally independent of its non-descendents, given its immediate parents.

- Does this rule give us all of the conditional independence relations implied by the Bayes network?
  - No!
  - E.g., X1 and X4 are conditionally indep given {X2, X3}
  - But X1 and X4 not conditionally indep given X3
  - For this, we need to understand D-separation



Explaining Away

X and Y are conditionally independent given Z, iff X and Y are D-separated by Z.

D-connection:
If G is a directed graph in which X, Y and Z are disjoint sets of vertices, then X and Y are d-connected by Z in G if and only if there exists an undirected path U between some vertex in X and some vertex in Y such that (1) for every collider C on U, either C or a descendent of C is in Z, and (2) no non-collider on U is in Z.

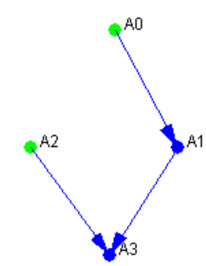
X and Y are **D-separated** by Z in G if and only if they are not D-connected by Z in G.

See d-Separation tutorial

http://www.andrew.cmu.edu/user/scheines/tutor/d-sep.html

See d-Separation Applet

http://www.andrew.cmu.edu/user/wimberly/dsep/dSep.html



X, Y	Cond Set	D-Separated	D-Connected
A0 A2	A1 A3	***	
A0 A2	A3		***
A0 A2	A1	***	
A0 A2		***	

A0 and A2 conditionally indep. given {A1, A3}

Inference in Bayes Nets

- In general, intractable (NP-complete)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
  - Or for singly connected graphs (ie., no undirected loops)
    - Belief propagation
- For multiply connected graphs (no directed loops)
    - Junction tree
- Sometimes use Monte Carlo methods
  - Generate a sample according to known distribution
- Variational methods for tractable approximate solutions

Learning in Bayes Nets

- Four categories of learning problems
  - Graph structure may be known/unknown
  - Variables may be observed/unobserved
- Easy case: learn parameters for known graph structure, using fully observed data
- Gruesome case: learn graph and parameters, from partly unobserved data
- More on these in next lectures

Java Bayes Net Applet

http://www.pmr.poli.usp.br/ltd/Software/javabayes/Home/applet.html

What You Should Know

- Bayes nets are convenient representation for encoding dependencies / conditional independence
- BN = Graph plus parameters of CPD's
  - Defines joint distribution over variables
  - Can calculate everything else from that
  - Though inference may be intractable
- Reading conditional independence relations from the graph
  - N cond indep of non-descendents, given parents
  - D-separation
  - 'Explaining away'