BAYESIAN NETWORKS

Vibhav Gogate

Today – Bayesian networks

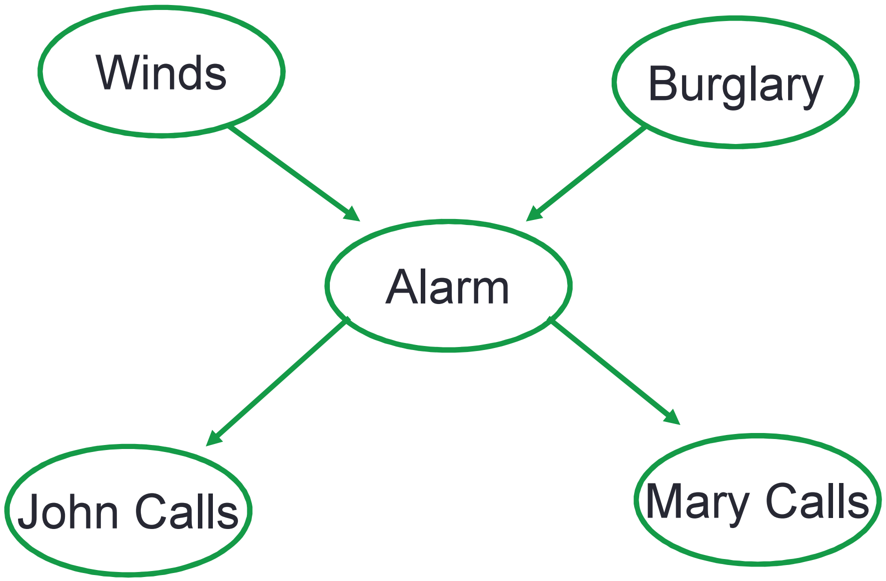
- One of the most exciting advancements in statistical AI and machine learning in the last 10-15 years
- Generalizes naïve Bayes and logistic regression classifiers
- Compact representations of exponentially-large probability distributions
- Exploit conditional independences

Judea Pearl: Turing award in 2011 for his contributions to Bayesian networks

CAUSAL STRUCTURE

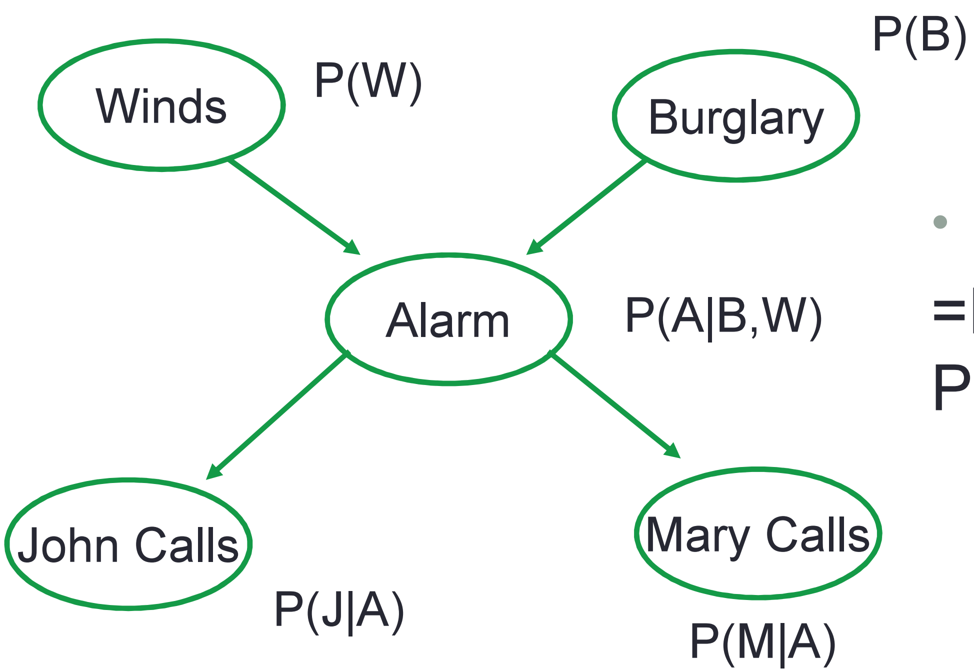
- Draw a directed acyclic (causal) graph for the following
  - Direct arrows from cause to effect
- Story:
  - There is a Burglar alarm that rings when we have Burglary
  - However, sometimes it may ring because of winds that exceed 60mph
  - When the alarm rings your neighbor Mary Calls
  - When the alarm rings your neighbor John Calls

CAUSAL STRUCTURE



- There is a Burglar alarm that rings when we have Burglary
- However, sometimes it may ring because of winds that exceed 60mph
- When the alarm rings your neighbor Mary Calls
- When the alarm rings your neighbor John Calls

Representation of Joint Distribution



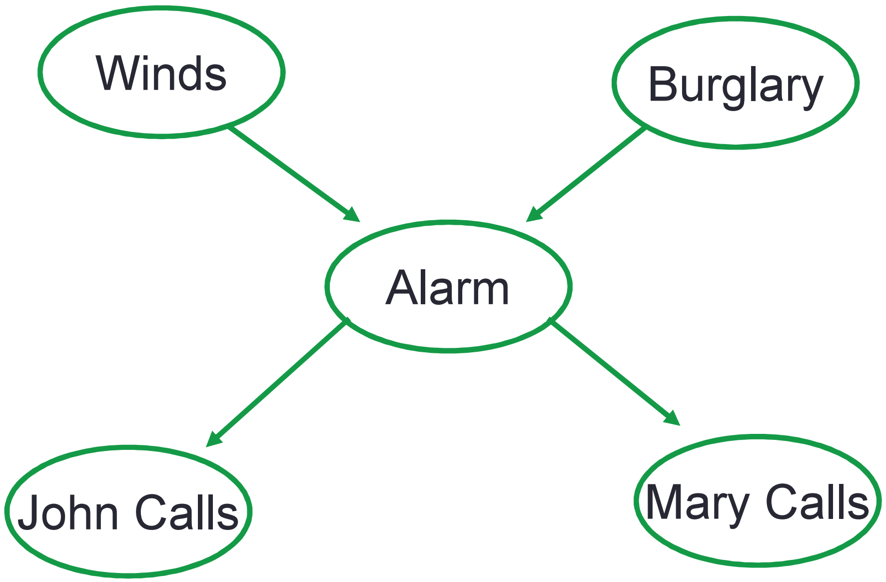
- P(W,B,A,J,M)

=P(W)P(B)P(A|B,W)P(J|A) P(M|A)

In general:

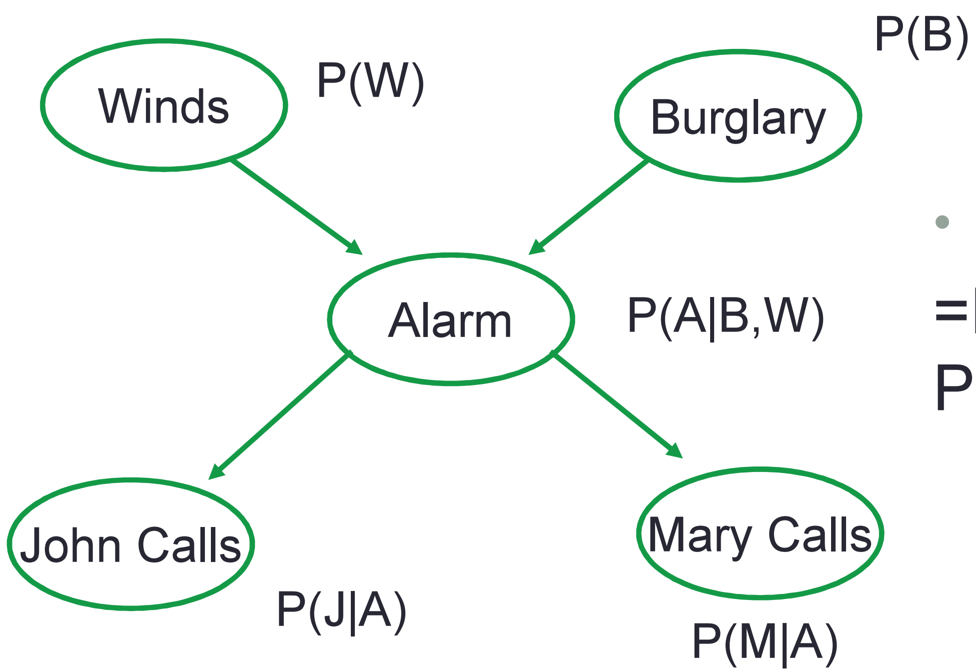
$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i \mid pa(x_i))$$

Possible Queries



- Inference
  - P(W=?|J=True)
- Most probable explanation
  - Assignment of values to all other variables that has the highest probability given that J=True and M=False
- Maximum Aposteriori Hypothesis.

Number of Parameters



- P(W,B,A,J,M) =P(W)P(B)P(A|B,W)P(J|A) P(M|A)

- Assume Binary variables
- How many parameters with each CPT $P(x_i|x_1, \ldots, x_j)$?
  - (Domain-size of x_i-1)*(All possible combinations of $x_1,\ldots,x_j$).

Can any distribution be represented as a Bayesian network

Let $\mathbf{X} = \{X_1, \ldots, X_n\}$ be a set of discrete random variables such that each variable X_i takes values from a finite domain $D(X_i)$. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an assignment to all variables $\mathbf{X}$ such that each variable X_i is assigned the value x_i from its domain $D(X_i)$. Let $(X_1, \ldots, X_n)$ be any ordering of the variables, then:

$$\Pr(\mathbf{x}) = P(x_1) \prod_{i=2}^{n} P(x_i | x_1, \ldots, x_{i-1})$$

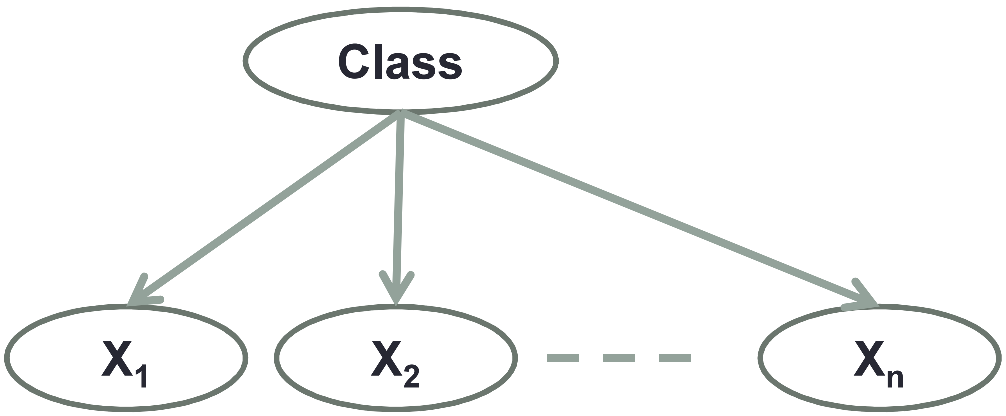
Let $pa(X_i)$ be a subset of variables of the set $\{X_1, \ldots, X_{i-1}\}$ such that X_i is conditionally independent of all other variables ordered before it given $pa(X_i)$. Then, we can rewrite the joint probability distribution as:

$$\Pr(\mathbf{x}) = \prod_{i=1}^{n} P(x_i | \pi(\mathbf{x}, pa(X_i))) \qquad (1)$$

where $\pi(\mathbf{x}, pa(X_i))$ is the projection of $\mathbf{x}$ on the parents of X_i. (For instance, the projection of the assignment $(X_1 = 0, X_2 = 0, X_3 = 1)$ on $\{X_1, X_3\}$ is the assignment $(X_1 = 0, X_3 = 1)$).

Relationship to Naïve Bayes

- Distribution?



Number of parameters?

Real Bayesian networks applications

- Diagnosis of lymph node disease
- Speech recognition
- Microsoft office and Windows
  - http://www.research.microsoft.com/research/dtg/
- Study Human genome
- Robot mapping
- Robots to identify meteorites to study
- Modeling fMRI data
- Anomaly detection
- Fault dianosis
- Modeling sensor network data

Now What?

- Given a Bayesian network, answer queries
  - Inference
- Learning Bayesian networks from Data
  - Structure and **weight learning**
  - **Partially and fully observed data**
  - **MLE** and Bayesian approach
  - Requires Inference, i.e., computing P(X|evidence) where evidence is an assignment to a subset of variables
- Unfortunately, Inference is NP-hard.
  - Tractable classes based on treewidth
  - Approximate Inference approaches