Summary of the Bayes Net Formalism

David Danks
Institute for Human & Machine Cognition

Bayesian Networks

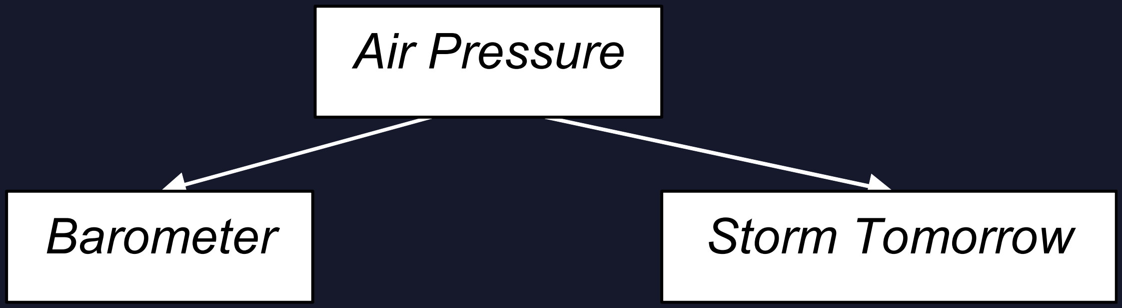
Two components:

1. Directed Acyclic Graph (DAG)
   - ***G***: There is a node for every variable
   - ***D***: Some nodes have arrows btw. them ($X \rightarrow Y$)
   - ***A***: Can't get from one node back to itself by following arrows

2. Joint Probability Distribution
   - For all $\{X, Y, \ldots, Z\}$, $P(X=x, Y=y, \ldots, Z=z)$ is defined

Example of a Bayes Net

- Directed Acyclic Graph:



- Joint Probability Distribution:
  - *P*(*AP*=*High*, *B*=*High*, *ST*=*Yes*) = 0.2
  - *P*(*AP*=*High*, *B*=*Low*, *ST*=*Yes*) = 0.005
  - …

Connecting the Graph & JPD

- Markov assumption:
  "*X* is (probabilistically) independent of its (graphical) non-descendants conditional on its (graphical) parents."
  - "No edge $\Rightarrow$ Conditional independence"

Connecting the Graph & JPD

- Markov assumption:
  "X is (probabilistically) independent of its (graphical) non-descendants conditional on its (graphical) parents."
  - "No edge $\Rightarrow$ Conditional independence"

- Example: $X \rightarrow Y \rightarrow Z \;\Rightarrow\; X \perp\!\!\!\perp Z \mid Y$

Connecting the Graph & JPD

- The Markov assumption implies a factorization of the JPD based on the graph:

  $P(X{=}x, Y{=}y, \ldots, Z{=}z) = \prod P(v \mid parents(v))$

- The Markov assumption allows us to move from graph to probability distribution

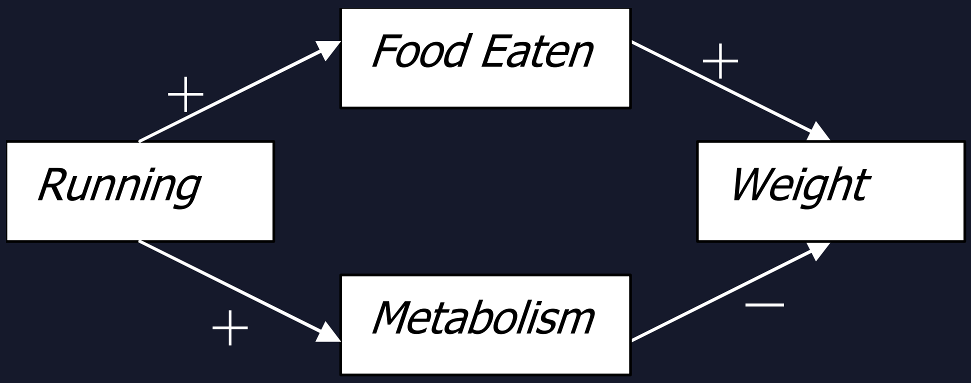
Connecting the Graph & JPD

- Faithfulness assumption:
  "The (probabilistic) effects of (graphical) paths never exactly offset."
  - "Conditional independence $\Rightarrow$ No edge"

- The Faithfulness assumption allows us to move from probability distribution to graph

Bayesian Network Example

1. Running causes you to eat more
2. Eating more causes you to gain weight
3. Running increases your metabolism
4. Increased metabolism leads to weight loss



- Note: Faithfulness rules out "Running $\perp\!\!\!\perp$ Weight"

Learning Bayes Nets

- Given some data from the world, why would we want to learn a Bayes net?

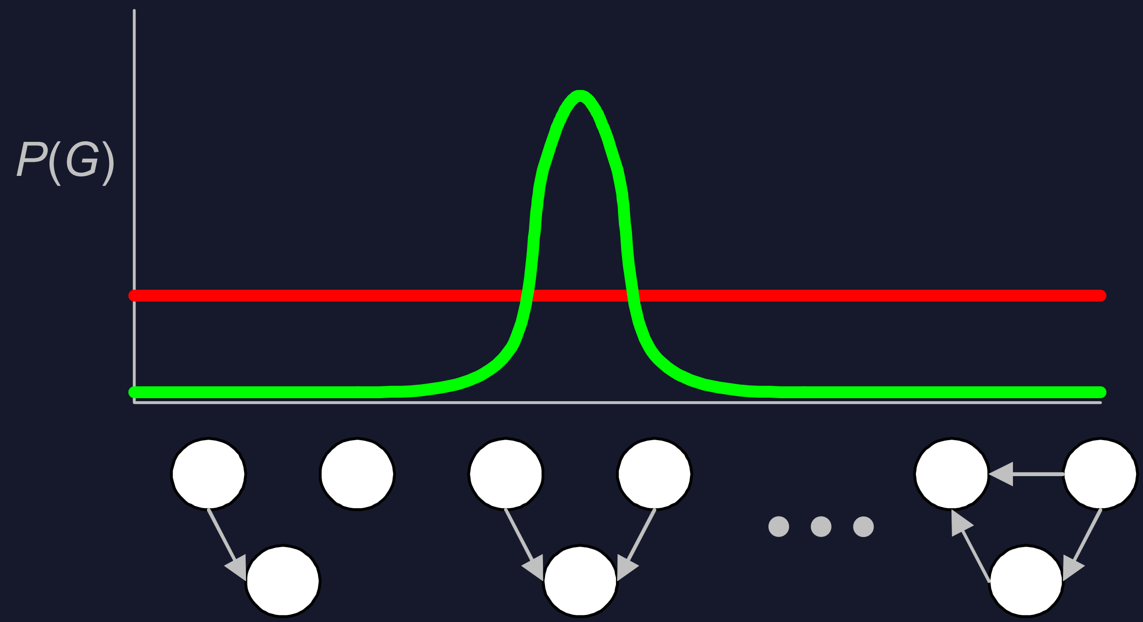
1. Compact representation of the data
   - There are fast algorithms for prediction/inference given *observations* of the environment
2. Causal knowledge
   - There are fast algorithms for prediction/inference given *interventions* in the environment

Bayesian Updating

- Start with a probability distribution over possible graphs

Bayesian Updating

- Start with a probability distribution over possible graphs, then figure out which graph is most likely given the observed data



Features of Bayesian Updating

- Advantages:
  1. Output is fine-grained probabilistic information
  2. "Rational" basis for the learning algorithm
  3. Robust to data errors
- Disadvantages:
  1. Number of possible graphs is super-exponential, & likelihood functions not always solvable analytically $\Rightarrow$ almost always use heuristics
  2. Cannot easily incorporate unobserved variables

Constraint-Based Learning

1. Determine the (un)conditional associations and independencies in the data
2. Determine the set of Bayes nets that could have produced that data

Constraint-Based Learning

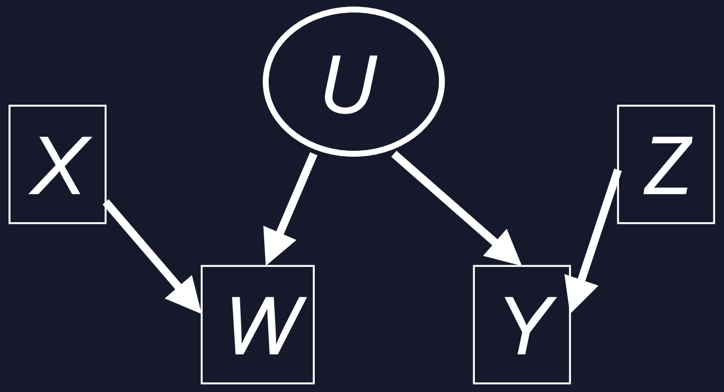
1. Determine the (un)conditional associations and independencies in the data
2. Determine the set of Bayes nets that could have produced that data

$$\left\{ \begin{array}{l} X \not\perp\!\!\!\perp W \ (\mathit{all}) \\ Y \not\perp\!\!\!\perp Z \ (\mathit{all}) \\ X \not\perp\!\!\!\perp Y, Z \mid W \\ Z \not\perp\!\!\!\perp X, W \mid Y \end{array} \right\}$$

Constraint-Based Learning

1. Determine the (un)conditional associations and independencies in the data
2. Determine the set of Bayes nets that could have produced that data

$$\left\{\begin{array}{l} X \cancel{\perp\!\!\!\perp} W \text{ (all)} \\ Y \cancel{\perp\!\!\!\perp} Z \text{ (all)} \\ X \cancel{\perp\!\!\!\perp} Y,Z \mid W \\ Z \cancel{\perp\!\!\!\perp} X,W \mid Y \end{array}\right\} \Rightarrow$$

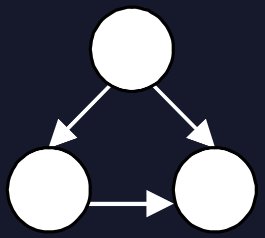


Features of C-B Learning

- Advantages:
  1. Feasible, asymptotically correct algorithms (though worst case requires exponential comp.)
  2. Easily incorporates unobserved variables
  3. Gives exactly the information in the data
- Disadvantages:
  1. Susceptible to mistaken independence judgments (big problem for small datasets)
  2. Cannot use fine-grained prior knowledge

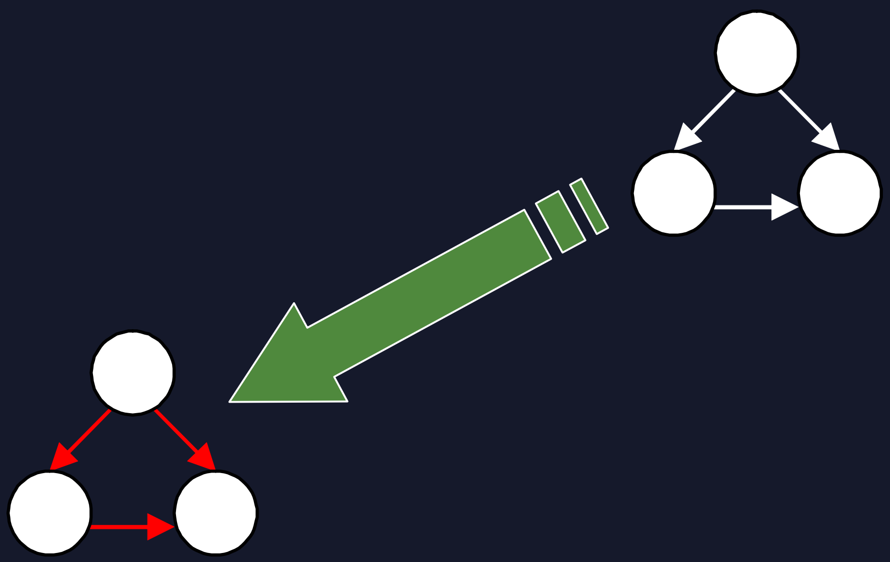
Layers of “Bayes Nets”

- Start with simple association graph



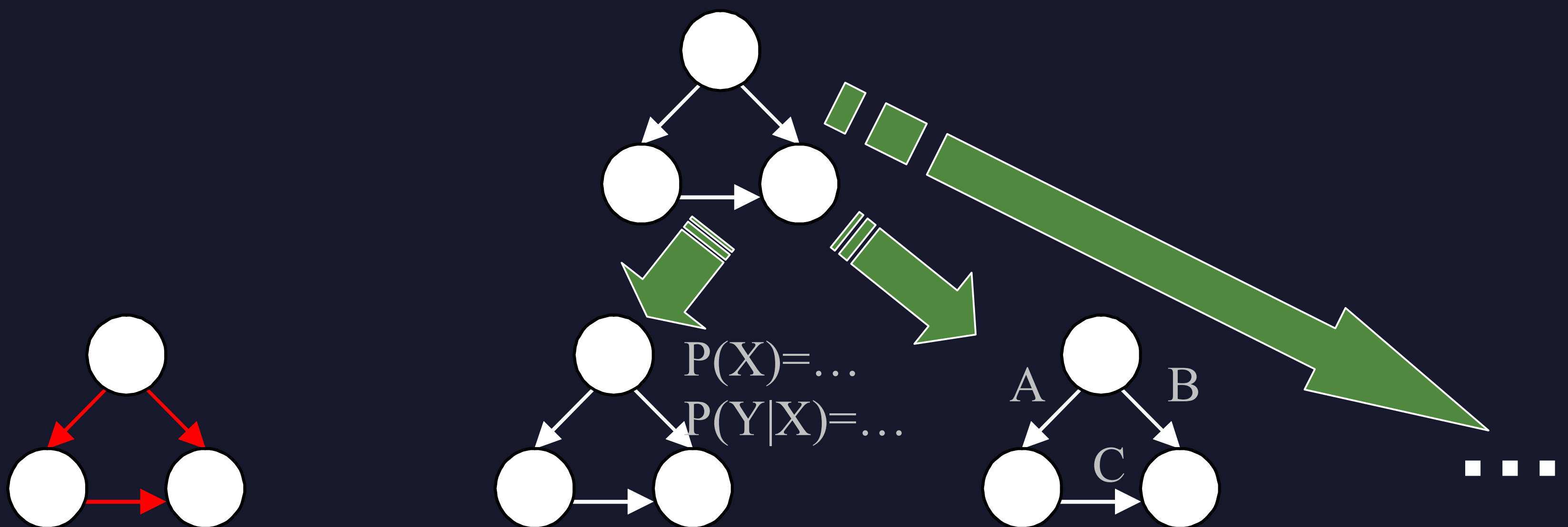
Layers of “Bayes Nets”

- Then we can add a causal interpretation:



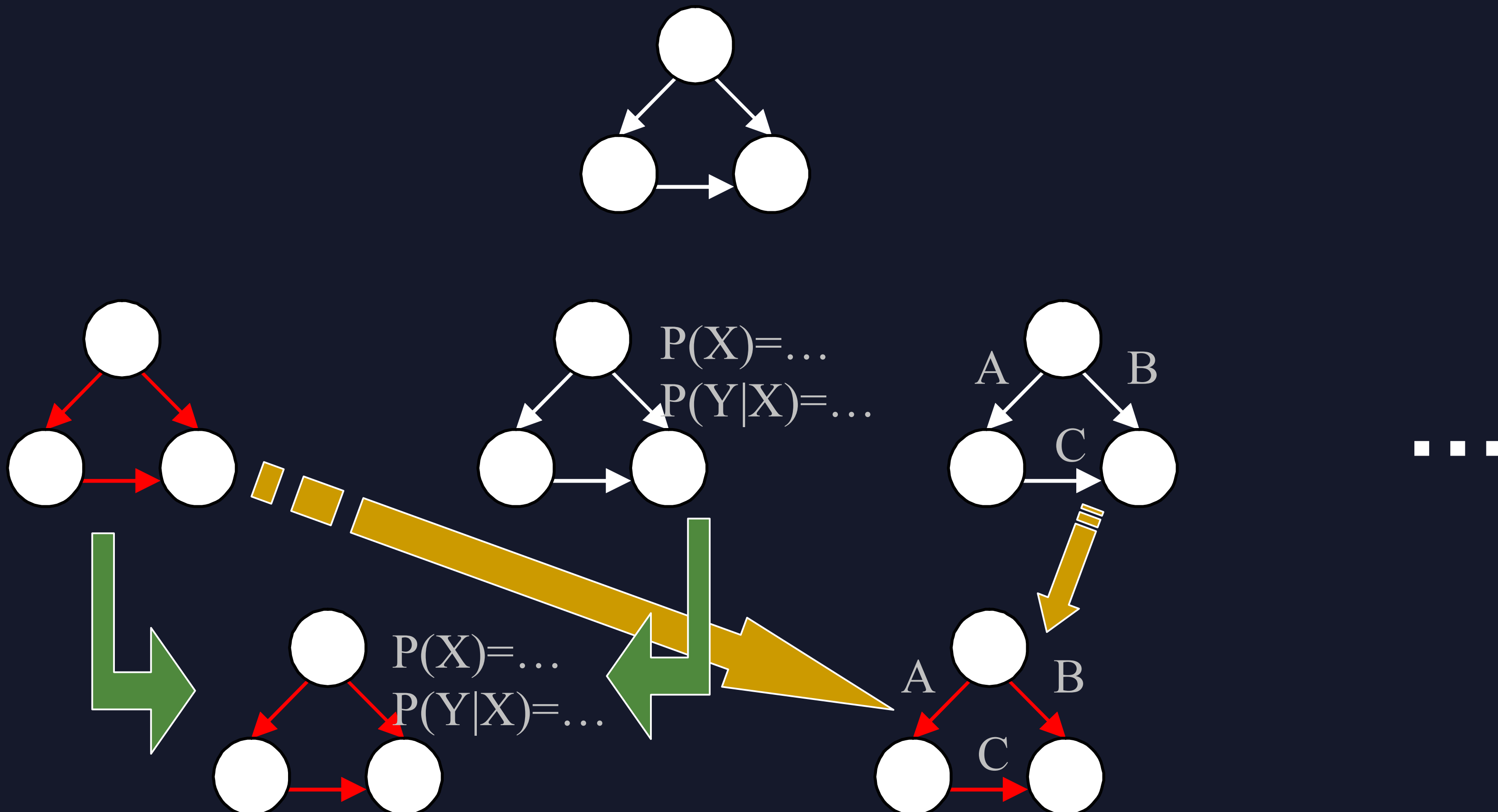
Layers of “Bayes Nets”

- Or we can provide parameter information:



Layers of “Bayes Nets”

- Or we can combine causation & parameters



Useful Websites

Tutorials:

http://www.cs.berkeley.edu/~murphyk/Bayes/bayes.html

Constraint-based learning software:

http://www.phil.cmu.edu/tetrad/index.html (free)

Bayesian learning software:

http://www.hugin.com (commercial)

http://research.microsoft.com/~dmax/winmine/tooldoc.htm (free)