# Naive Bayes and Bayesian Networks

Frederik Mallmann-Trenn
6CCS3AIN

# Bayes' Rule & conditional independence

- In naive Bayes models, one assumes that

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i | Cause)$$

- Conditional independence is an example of naive Bayes

# Naive Bayes

- Assuming conditional independent effects, reduces the model the problem

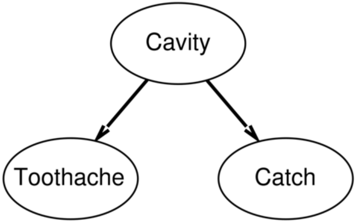
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i | Cause)$$

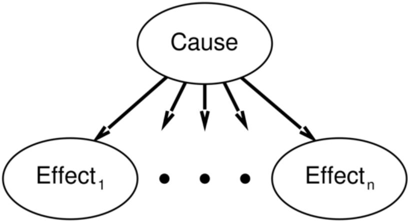
- Total number of parameters is linear in the number of conditionally independent effects $n$.
- It is called 'naive', because it is oversimplifying: in many cases the 'effect' variables aren't actually conditionally independent given the cause variable. Example:
  - $Cause$: it rained yesterday
  - $Effect_1$: the streets are wet this morning
  - $Effect_2$: I'm late for my class
  - If the streets were still wet, then an accident was more likely to happen and the caused traffic jam could be the reason for being late

# Naive Bayes

- $$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i | Cause)$$

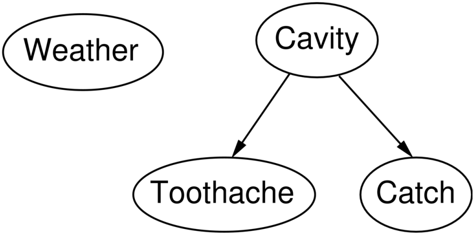
- Visualise as:



# Bayesian networks

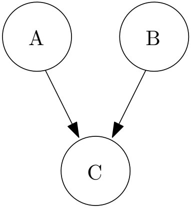
- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions
- Topology of network encodes conditional independence assertions:



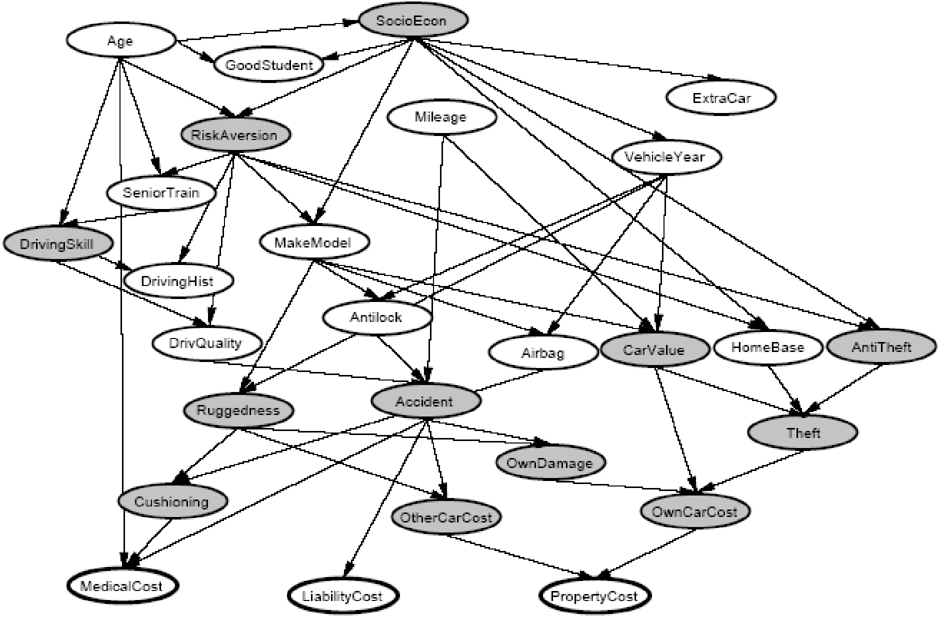

- $Weather$ is independent of the other variables
- $Toothache$ and $Catch$ are conditionally independent given $Cavity$

# Bayesian Networks

- Bayesian networks are a way to represent these dependencies:
  - Each node corresponds to a random variable (which may be discrete or continuous)
  - A directed edge (also called link or arrow) from node $u$ to node $v$ means that $u$ is the parent of $v$.
  - Likewise, $v$ is a child of $u$
  - The graph has no directed cycles (and hence is a directed acyclic graph, or DAG).
  - Each node $u$ has a conditional probability $P(u \mid Parents(u))$ that quantifies the effect of the parent nodes
- Example: $C$ depends on $A$ and $B$, and $A$ and $B$ are independent.



# Bayesian networks




*(http://www.igi.tugraz.at)*

# Bayesian networks

- How can we represent the knowledge about the probabilities?
- Conditional distribution represented as a conditional probability table (CPT) giving the distribution over $u$ for each combination of parent values

| A | B | $P(C \mid A, B)$ |
|---|---|---|
| T | T | 0.2 |
| T | F | 0.123 |
| F | T | 0.9 |
| F | F | 0.51 |

# Bayesian networks

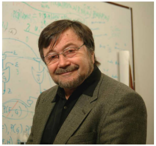
- Bayesian networks $\neq$ Naive Bayes
- These are somewhat orthogonal. Naive Bayes might be used in Bayesian networks.
- Also don't confuse them with Bayes' rule

# Bayesian networks

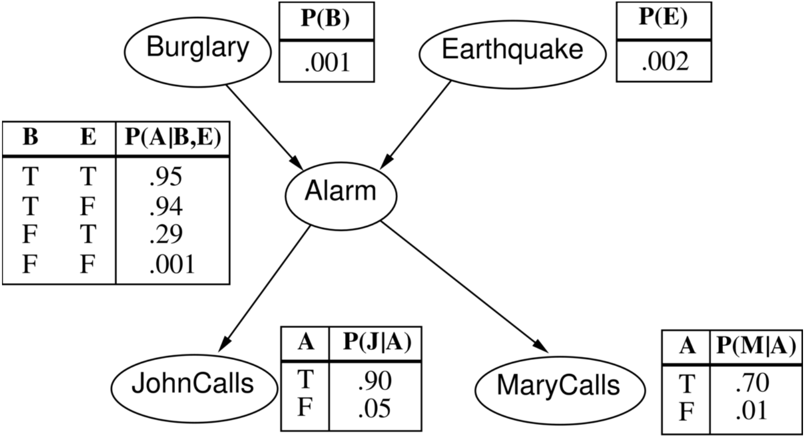
- An example (from California):

  *I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?*

- Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

# Bayesian networks



# A note on CPTs

- The CPTs in the previous slide appear to be missing some values:

  | $A$ | $P(J\|A)$ |
  |---|---|
  | $T$ | $0.90$ |
  | $F$ | $0.05$ |

  has two values rather then the four which would completely specify the relation between $J$ and $A$.
- The table tells us that:

  $$P(J = T|A = T) = 0.9$$

  which means:

  $$P(J = F|A = T) = 0.1$$

  because $P(J = T|A = T) + P(J = F|A = T) = 1$

# A note on CPTs

- Or, writing the values of $J$ and $A$ the other way:

$$P(j|a) = 0.9$$
$$P(\neg j|a) = 0.1$$

because $P(j|a) + P(\neg j|a) = 1$

# Applications