CSC421 Intro to Artificial Intelligence

UNIT 22: Probabilistic Reasoning

Midterm Review

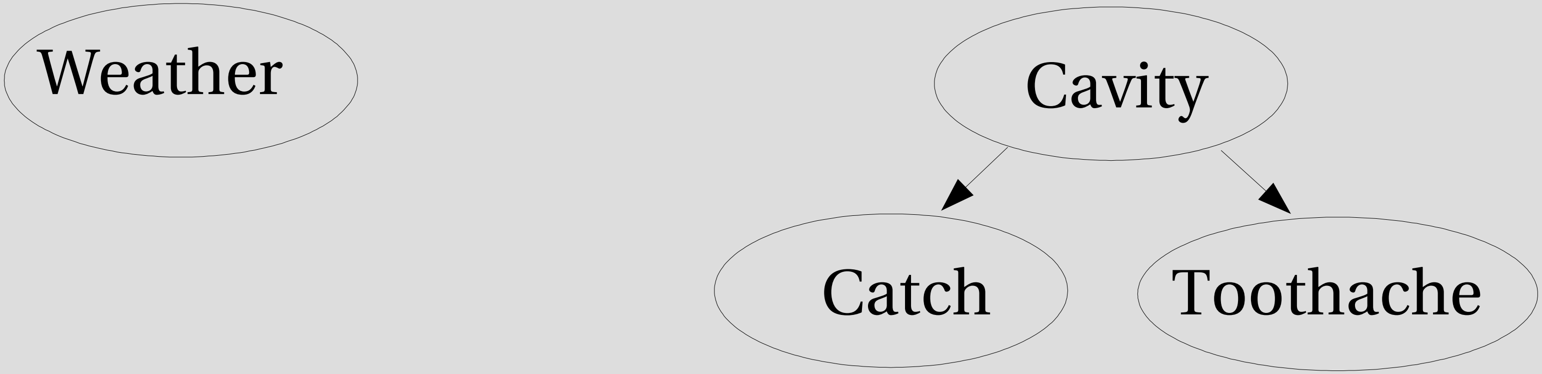
- Rooks, bishop, pawns
- Edit distance

Baysian Networks

- A simple, graphical notation for conditional independence assertions and for compact specification of full joint distributions
- Syntax:
  - Each node is a RV (continuous or discrete)
  - A directed acyclic graph (link “directly influences”)
  - A conditional distribution for each node given it's parents
    - $P(X_i \mid Parents(X_i)$
    - Simplest case (CPT) – distribution over X_i for every combination of parent values

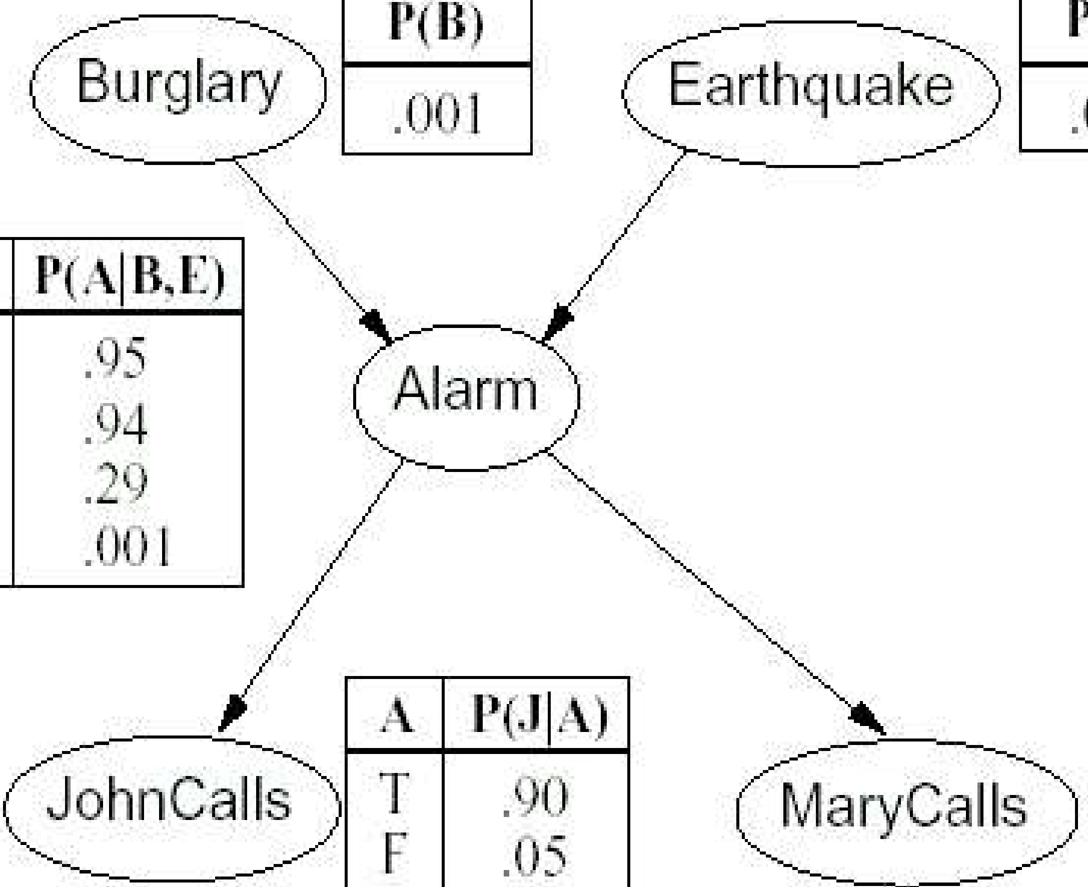
Topology

Topology of network encodes conditional independence assertions



Weather is independent of the other variables
Toothache and Catch are conditionally independent given cavity

Example



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

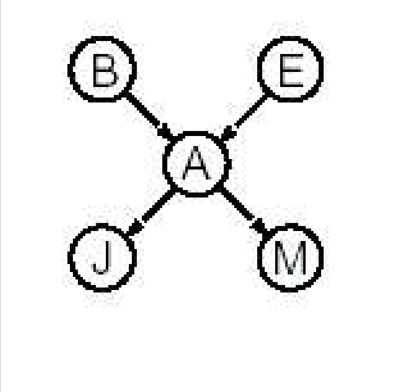
A	P(M|A)
T	.70
F	.01

Compactness

- A CPT for boolean X_i with k boolean parents has 2^k rows for the combinations of parent values
- Each row requires one number p for X_i = true (1-p for false)
- If each variables has no more than k parents then complete network $O(n * 2^k)$ numbers
- Full joint $O(2^n)$

Global Semantics

- Global semantics defines the joint distribution as the product of the local continuous distributions
- $P(x_1, \ldots x_N) = \Pi_{i=1}^{N} P(x_i \mid \text{Parent}(X_i))$
- e.g $P(j, m, a, \neg b, \neg a) = ?$

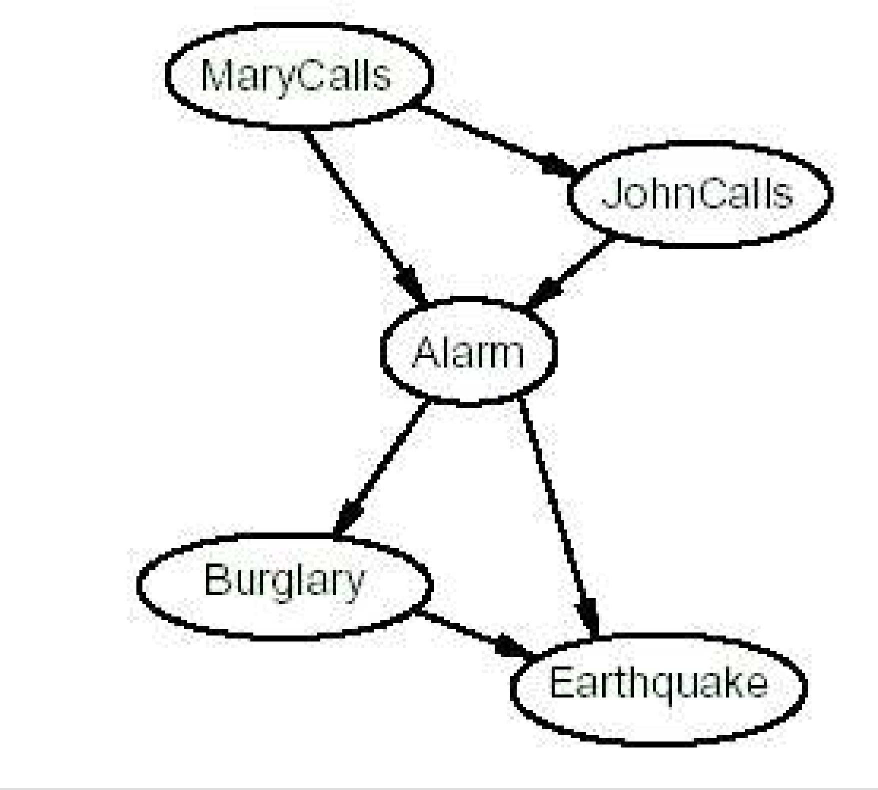


Constructing BNs

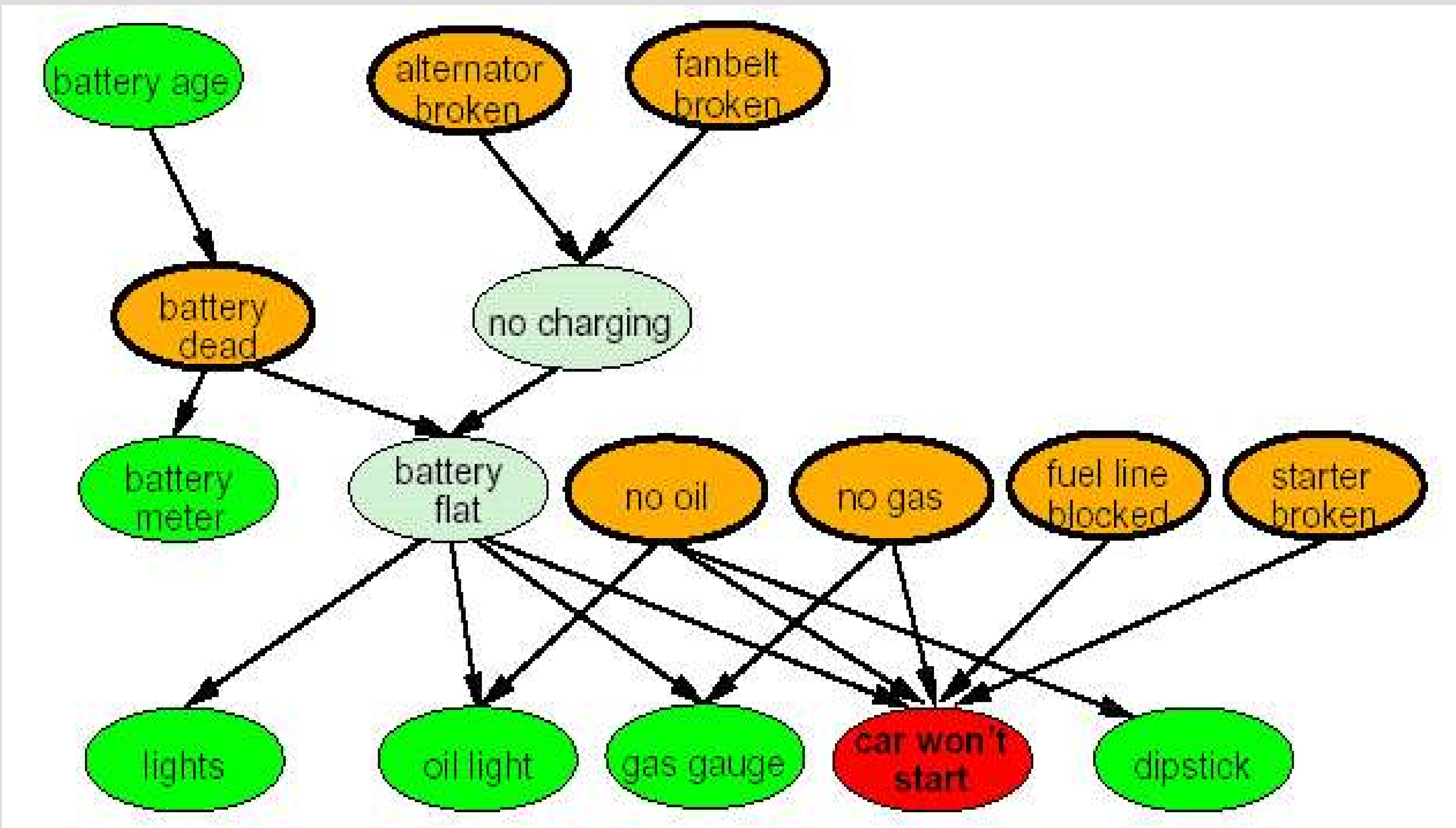
- Need a method such that a series of locally testable assertions conditional independence guarantee global semantics
- Intuitively:
  - Start from root causes and expand effects
  - (follow causality)
- Details:
  - Read textbook

Some BNs are better than others

- Deciding conditional independence in non-causal directions is hard
- Assessing conditional probabilities in non-causal directions is hard
- Network is less compact

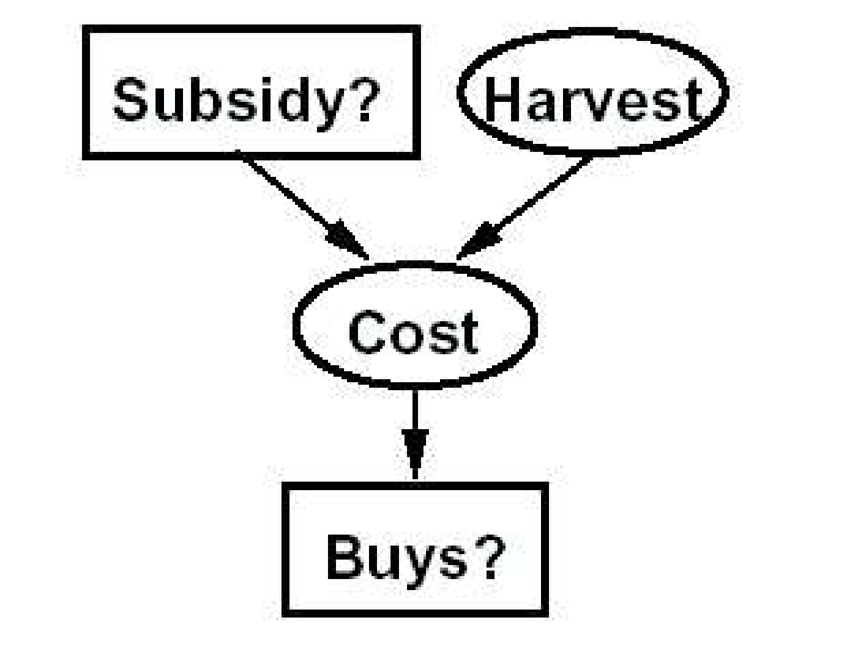


Example: Car Diagnosis



Hybrid (discrete + continuous) networks

Discrete (Subsidy?, Buys?)
Continuous (Harvest, Cost)



Option 1: Discretization - possibly large erros, large CPTs
Option 2: Finitely parametrized canonical families

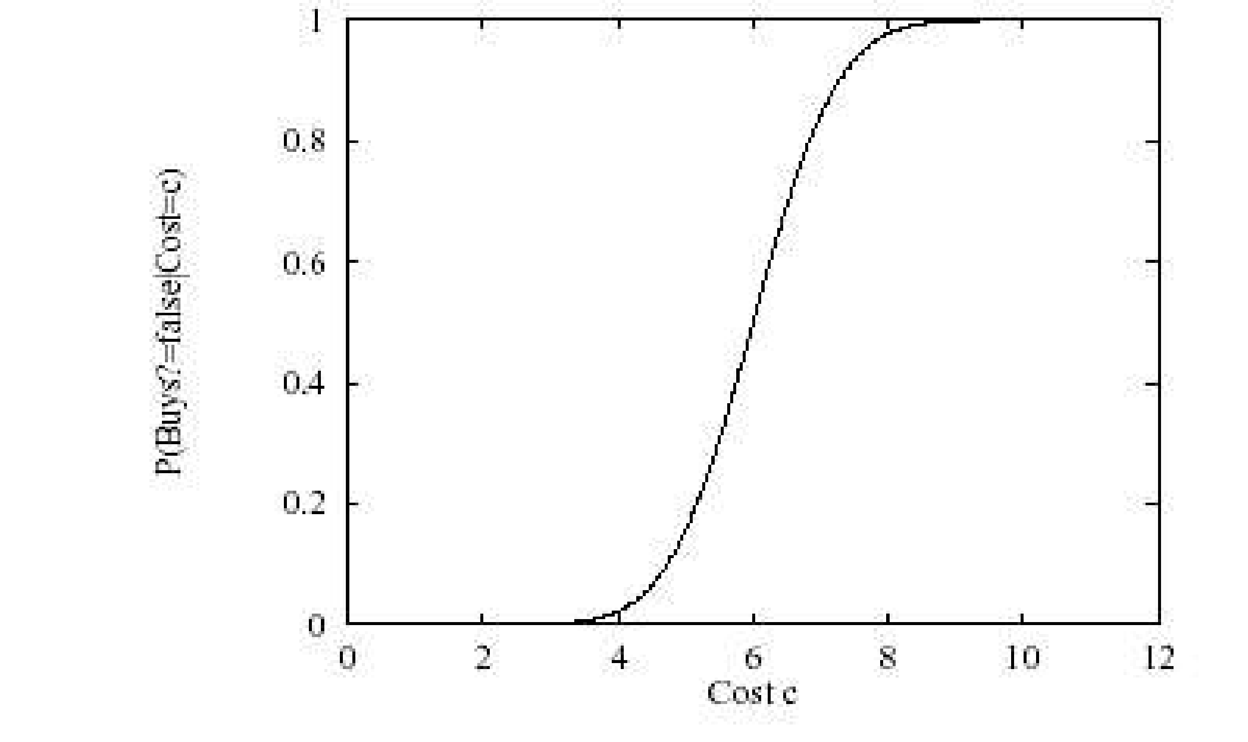
1) Continuous variable (discrete & continuous parents) (e.g. Cost)
2) Discrete variable (continuous parents) (e.g Buys?)

Continuous Child Variables

- Need one conditional density function for child variable given continuous parents for each possible assignment to discrete parents
- Most common linear Gaussian model:
  - P(Cost = c | Harvest =h, Subsidy = true) = $N(a_t h + b_{t,}, \sigma_t)(c)$
- All continuous network with LGs then joint distribution is multivariate Gaussian
- Discrete + continuous = conditional Gaussian (multivariate Gaussian for every setting of all discrete variables)

Discrete Variable with continuous parents

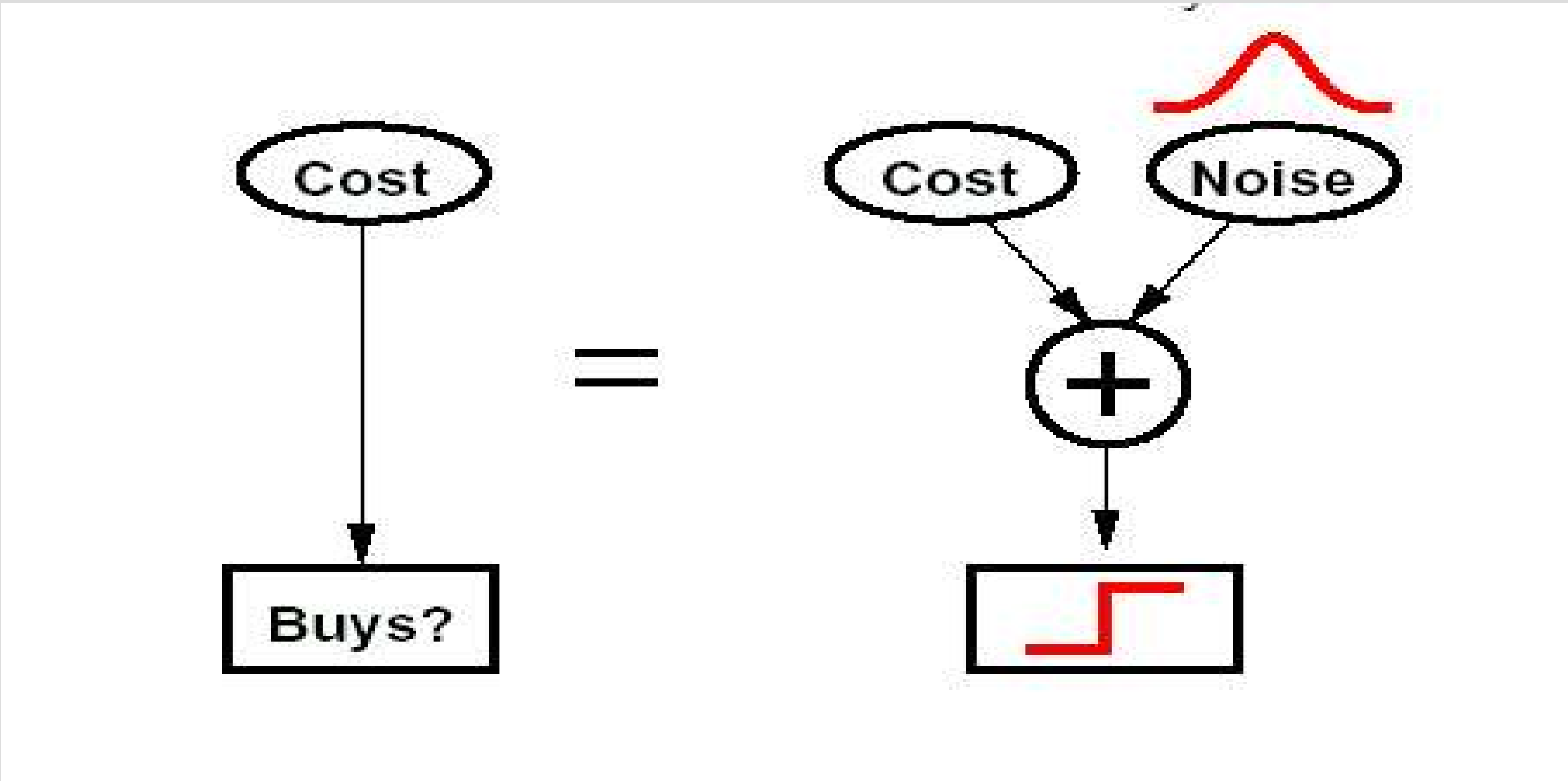
- Probability of Buys ? given Cost should be a soft threshold



Probit distribution = integral of Gaussian
Sigmoid (or Logit) distribution = also used in Neural Networks

Why probit ?

- It's sort of the right shape
- Can be viewed as a hard threshold whose location is subject to noise



Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPT = compact reprsentation of joint distribution
- Generally easy for (non) experts to construct
- Continuous variables => parametrized distributions (e.g linear Gaussian)
- Extra:
  - Canonical distributions (noisy-OR)