343H: Honors AI

Lecture 9: Bayes nets, part 1

2/13/2014

Kristen Grauman

UT Austin

Slides courtesy of Dan Klein, UC Berkeley

Unless otherwise noted

Outline

- Last time: Probability
  - Random Variables
  - Joint and Marginal Distributions
  - Conditional Distribution
  - Product Rule, Chain Rule, Bayes' Rule
  - Inference

- Today:
  - Independence
  - Intro to Bayesian Networks

Quiz: Bayes' Rule

- What is P(W | dry) ?

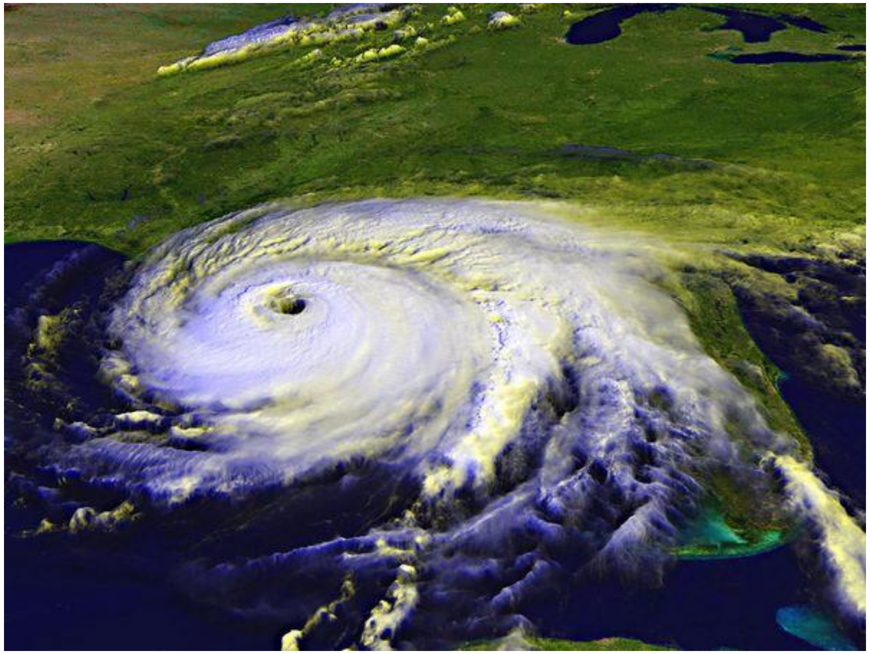
$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

Models and simplifications



Probabilistic Models

- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - "All models are wrong; but some are useful."
    – George E. P. Box

- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables

$$P(X_1, X_2, \ldots X_n)$$

- Given a joint distribution, we can reason about unobserved variables given observations (evidence)
- General form of a query:

$$P(X_q | x_{e_1}, \ldots x_{e_k})$$

Stuff you care about → X_q

Stuff you already know → $x_{e_1}, \ldots x_{e_k}$

- This kind of posterior distribution is also called the belief function of an agent which uses this model

Independence

- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = P(x)$$

  - We write: $X \perp\!\!\!\perp Y$

- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence?

$P(T)$

T	P
warm	0.5
cold	0.5

$P_1(T, W)$

T	W	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W)$

T	W	P
warm	sun	
warm	rain	
cold	sun	
cold	rain	

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

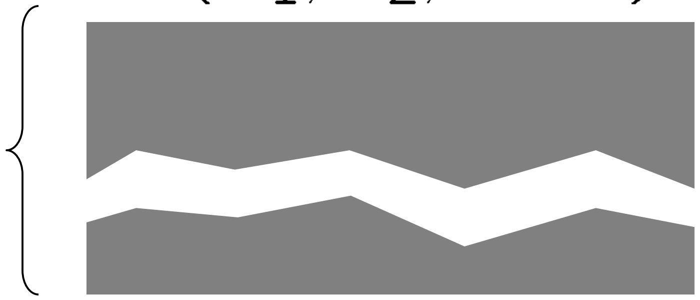
$\dots$

$P(X_n)$

h	0.5
t	0.5

$P(X_1, X_2, \dots X_n)$

2^n



Conditional Independence

- P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- The same independence holds if I don’t have a cavity:
  - P(+catch | +toothache, $\neg$cavity) = P(+catch| $\neg$cavity)

- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)

- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)

Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z iff: $X \perp\!\!\!\perp Y | Z$

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

- Or, equivalently, iff:

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

Conditional independence

- What about this domain?
  - Traffic
  - Umbrella
  - Raining


Conditional independence

- What about this domain?
  - Fire
  - Smoke
  - Alarm

Cond indep and the Chain Rule

$$P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$$

- Trivial decomposition:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- With assumption of conditional independence:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- Bayes' nets / graphical models help us express conditional independence assumptions

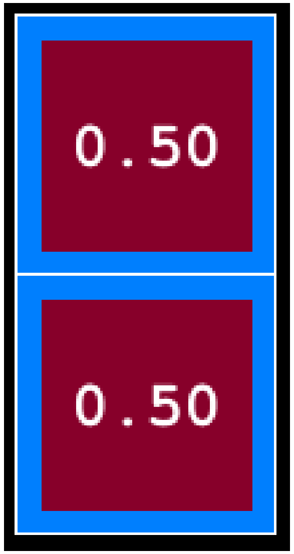
Ghostbusters Chain Rule

- Each sensor depends only on where the ghost is
- That means, the two sensors are conditionally independent, given the ghost position
- T: Top square is red
  B: Bottom square is red
  G: Ghost is in the top
- Givens:

$P(+g) = 0.5$ } $P(G)$

$P(+t \mid +g) = 0.8$
$P(+t \mid \neg g) = 0.4$ } $P(T|G)$

$P(+b \mid +g) = 0.4$
$P(+b \mid \neg g) = 0.8$ } $P(B|G)$



$$P(T,B,G) = P(G)\ P(T|G)\ P(B|T,G)$$
$$= P(G)\ P(T|G)\ P(B|G)$$

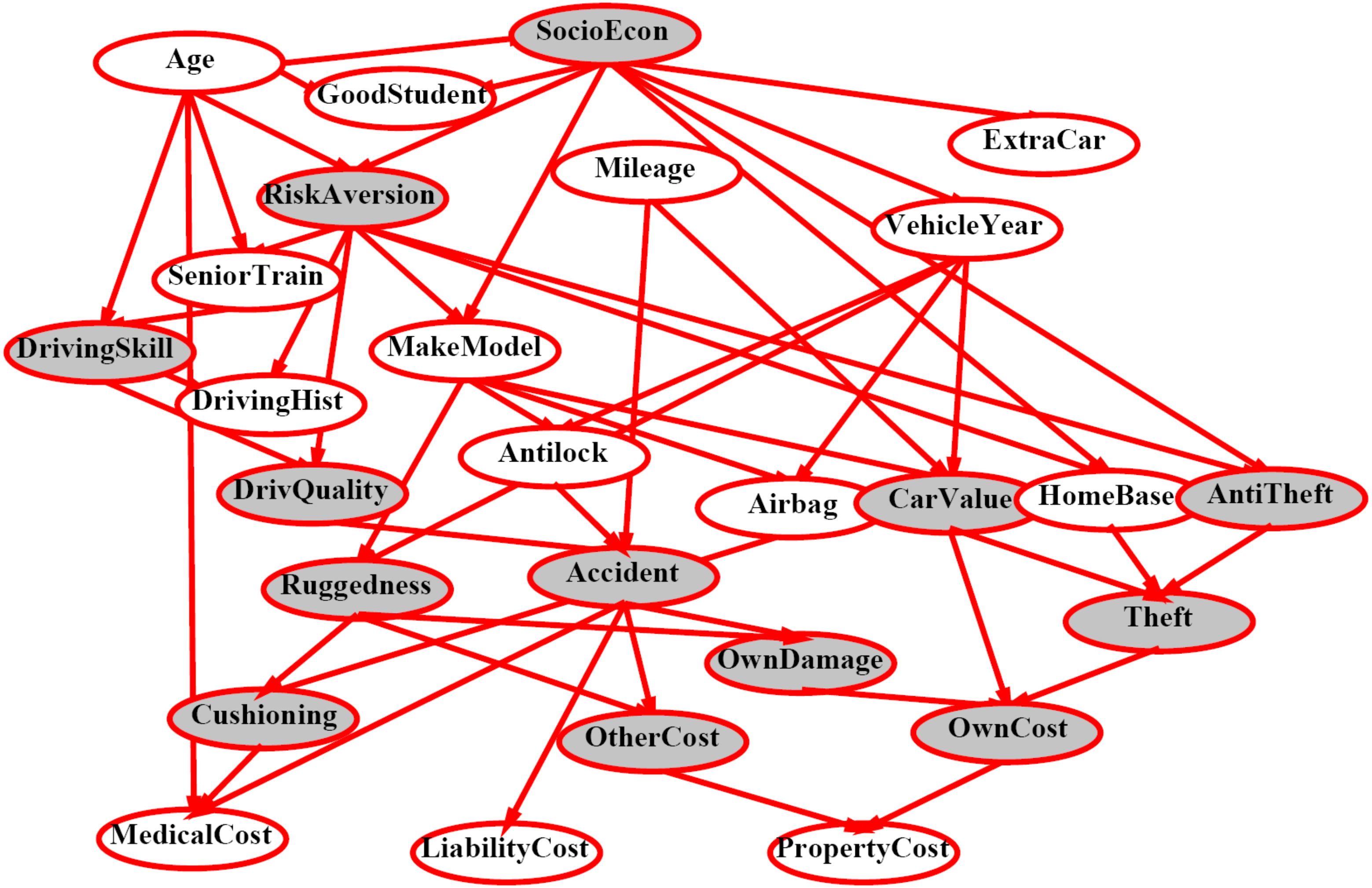
T	B	G	P(T,B,G)
+t	+b	+g	0.16
+t	+b	¬g	0.16
+t	¬b	+g	0.24
+t	¬b	¬g	0.04
¬t	+b	+g	0.04
¬t	+b	¬g	0.24
¬t	¬b	+g	0.06
¬t	¬b	¬g	0.06

Bayes' Nets: Big Picture

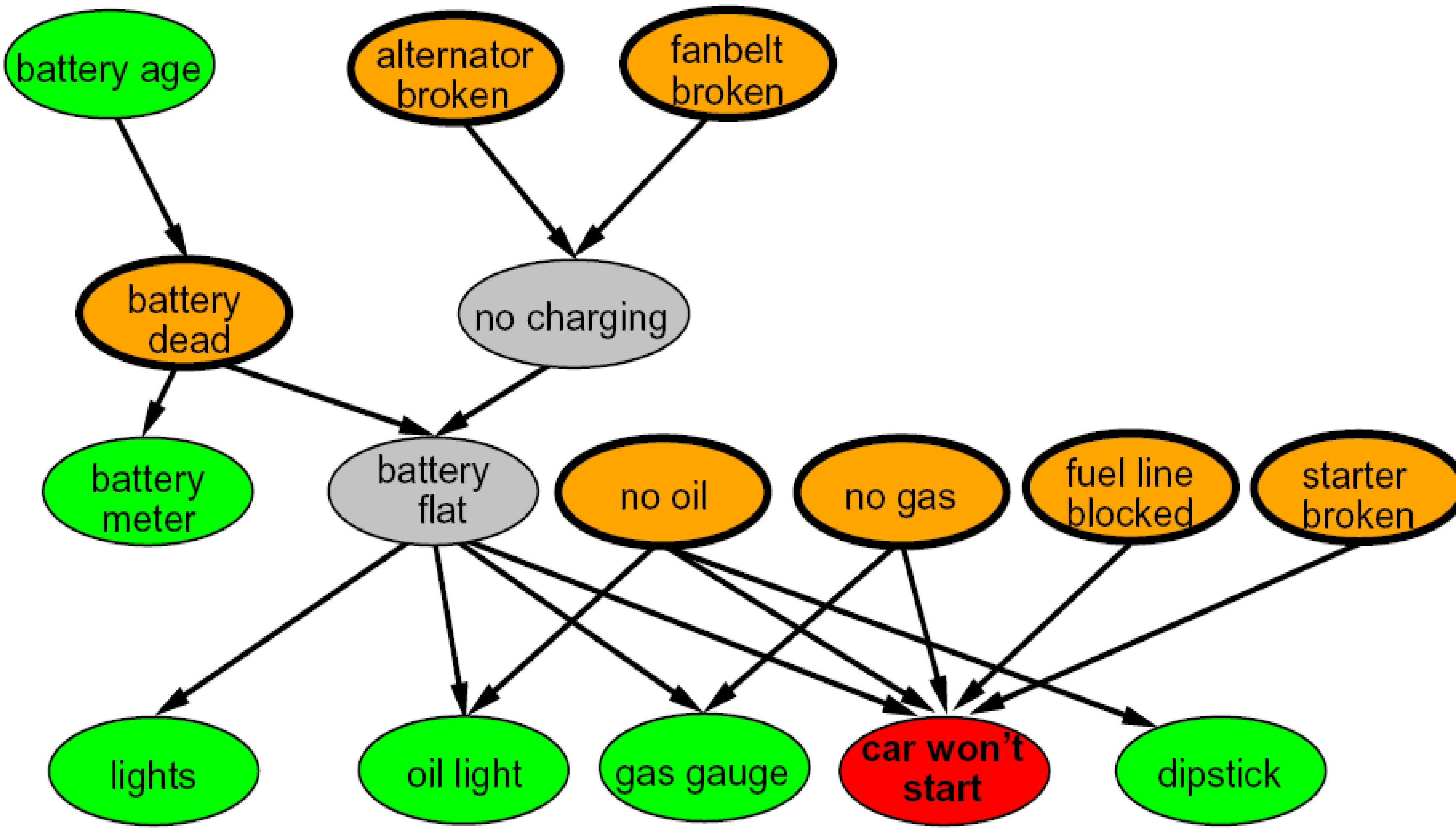
- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables *locally* interact
  - Local interactions chain together to give global, indirect interactions
  - For now, we'll be vague about how these interactions are specified

Example Bayes' Net: Insurance

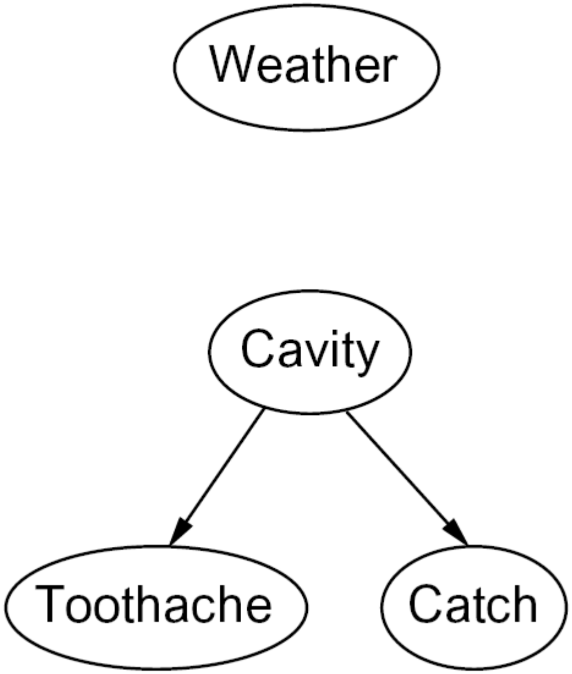


Example Bayes' Net: Car



Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don’t!)



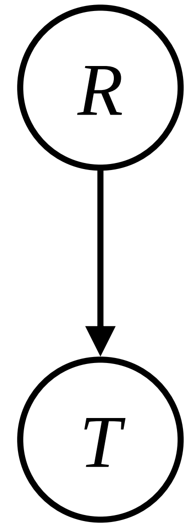
Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic



- Model 1: independence

- Model 2: rain causes traffic

- Why is an agent using model 2 better?

Example: Traffic II

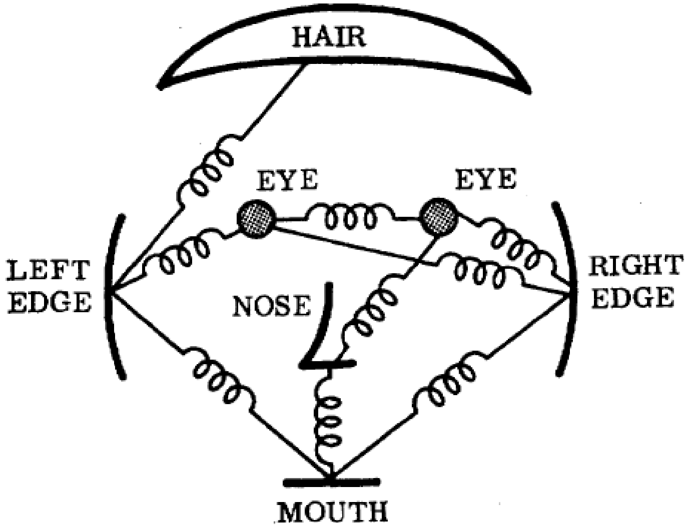
- Let’s build a causal graphical model

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!

Example: Part-based object models



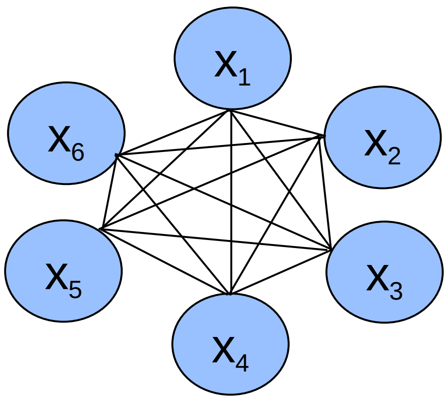
[Fischler and Elschlager, 1973]

Kristen Grauman

Example: Part-based object models

One possible graphical model:

Fully connected constellation model

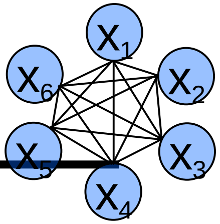


- **e.g. Constellation Model**
- **Parts fully connected**

N image features, P parts in the model

Slide credit: Rob Fergus

Probabilistic constellation model



$$P(image \mid object) = P(appearance, shape \mid object)$$

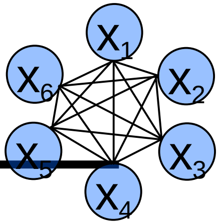
Part descriptors → appearance

Part locations → shape

Candidate parts

Source: Lana Lazebnik

Probabilistic constellation model



$$P(image \mid object) = P(appearance, shape \mid object)$$

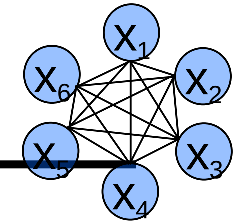
Part 1

Part 3

Part 2

Source: Lana Lazebnik

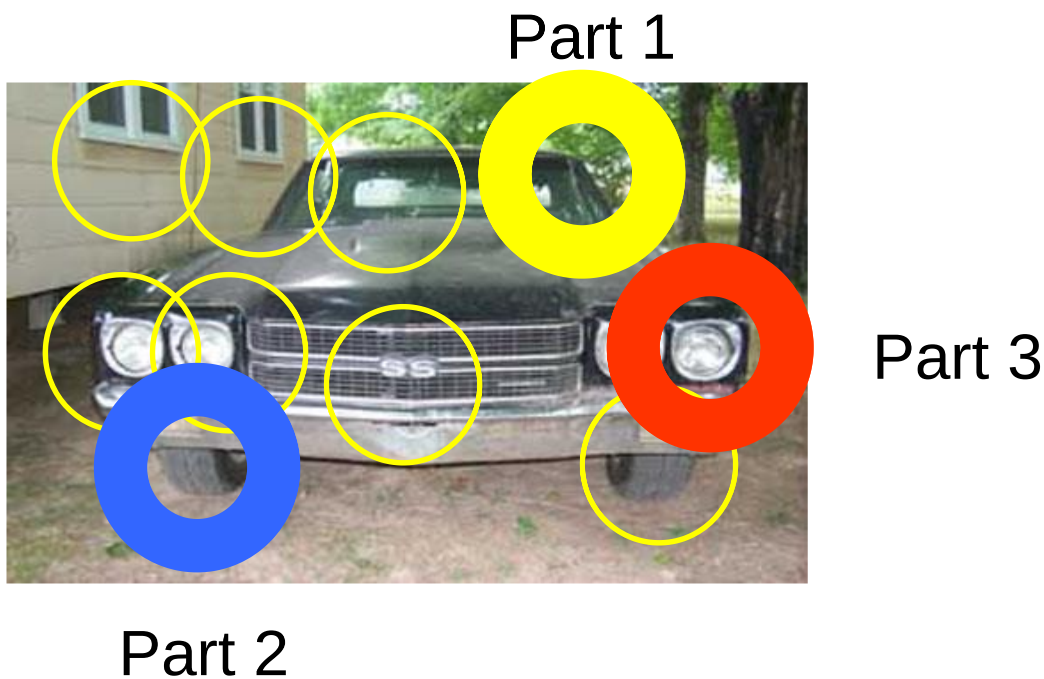
Probabilistic constellation model



$$P(image \mid object) = P(appearance, shape \mid object)$$

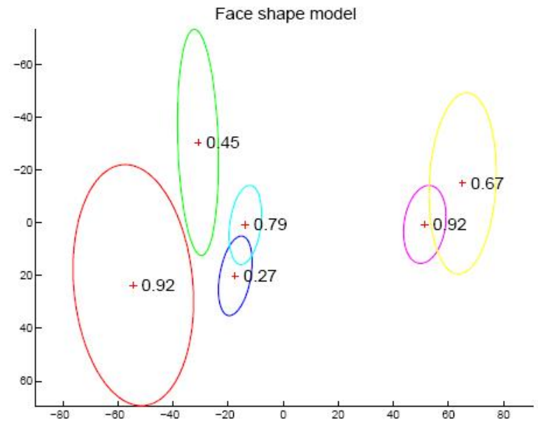
$$= \max_h P(appearance \mid h, object)\, p(shape \mid h, object)\, p(h \mid object)$$

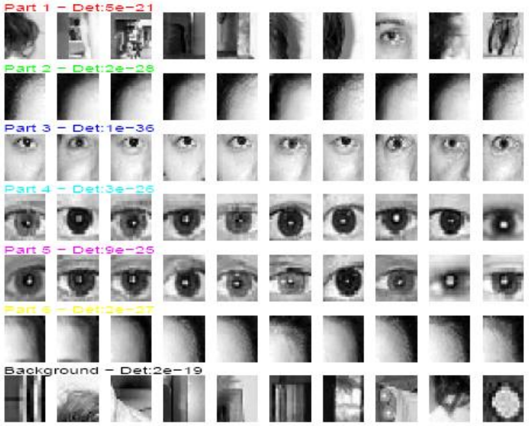
h: assignment of features to parts



Source: Lana Lazebnik

Face model

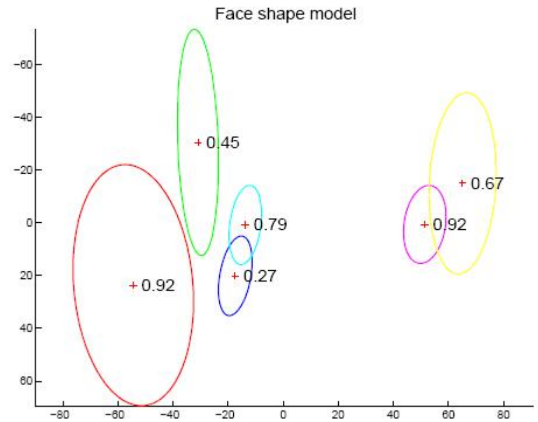


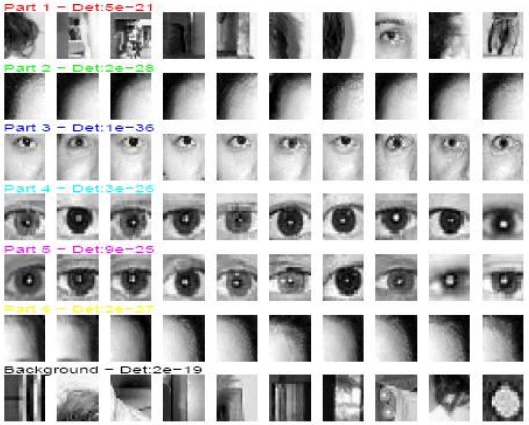


Appearance: 10 patches closest to mean for each part

Fergus et al. CVPR 2003

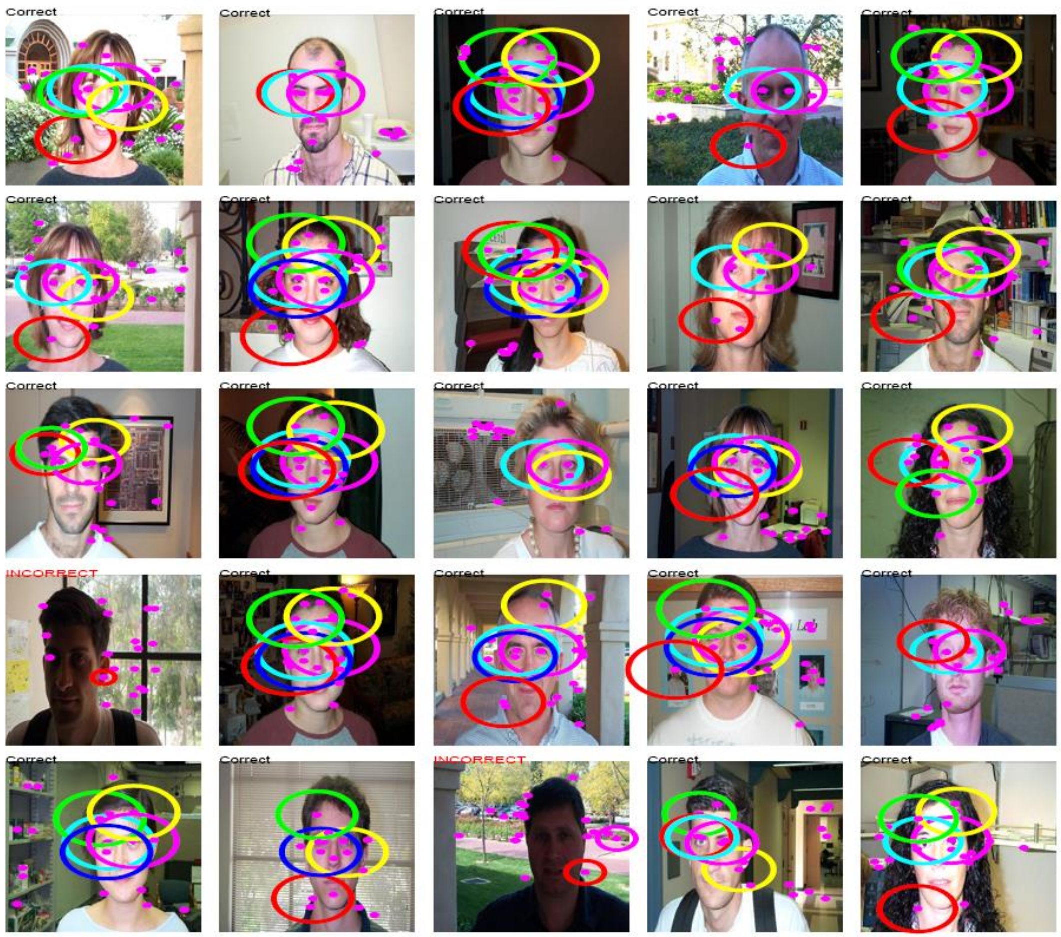
Face model





Appearance: 10 patches closest to mean for each part

Recognition results

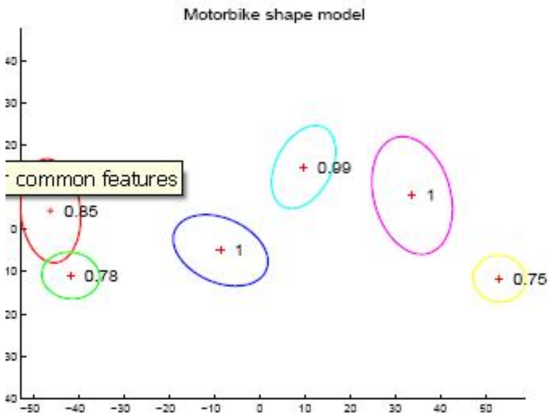


Test images: size of circles indicates score of hypothesis

Kristen Grauman

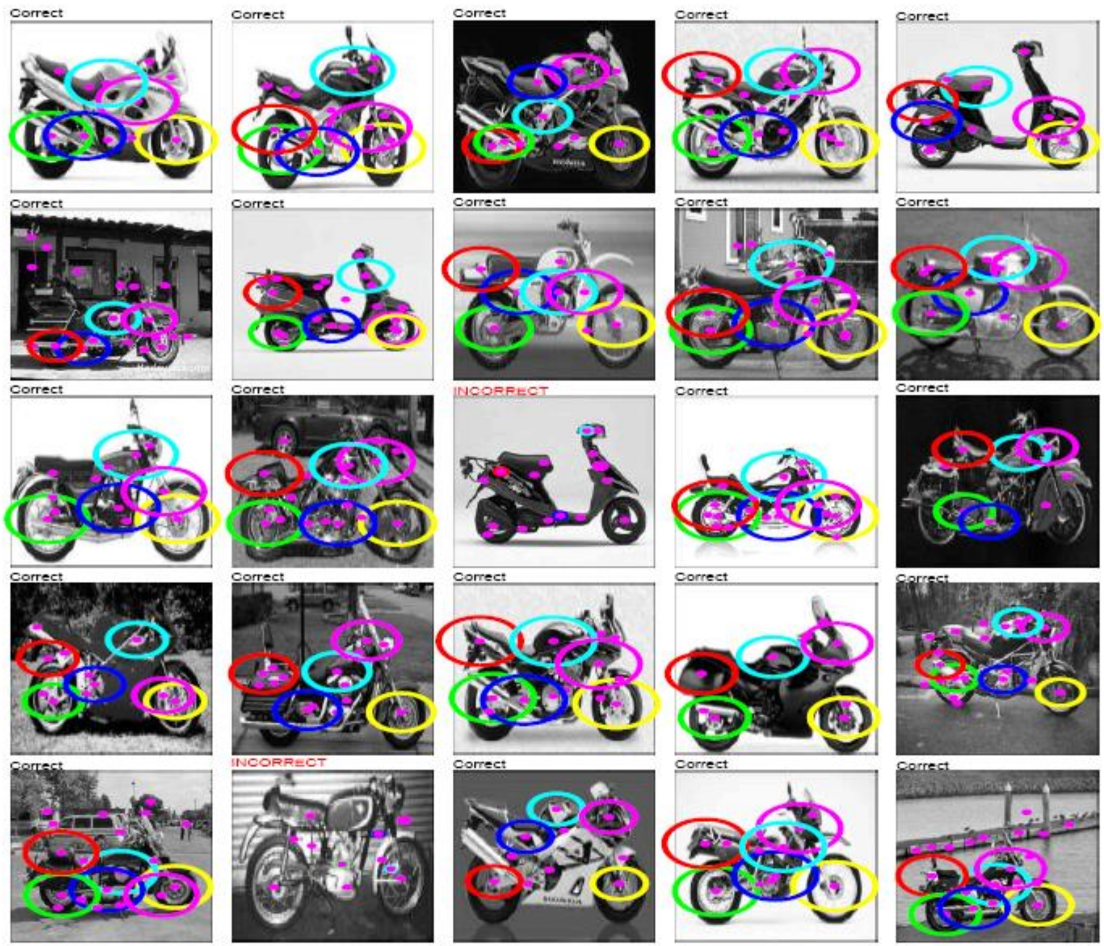
Fergus et al. CVPR 2003

Motorbike model



Appearance: 10 patches closest to mean for each part

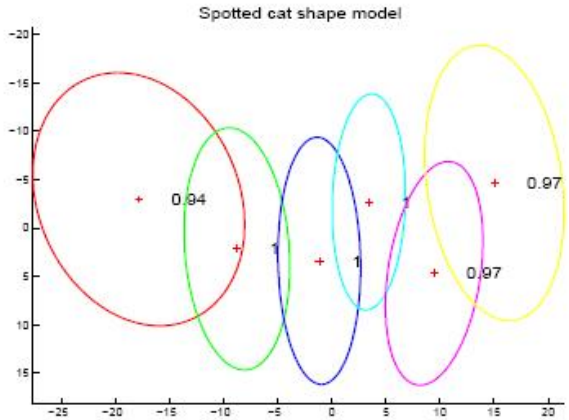
Recognition results

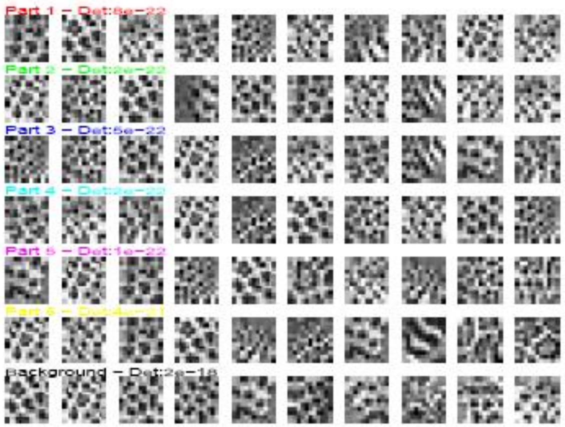


Kristen Grauman

Fergus et al. CVPR 2003

Spotted cat model





Appearance: 10 patches closest to mean for each part

Recognition results

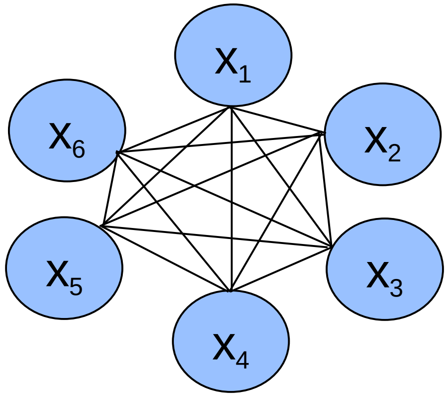
Kristen Grauman

Fergus et al. CVPR 2003

Example: Part-based object models

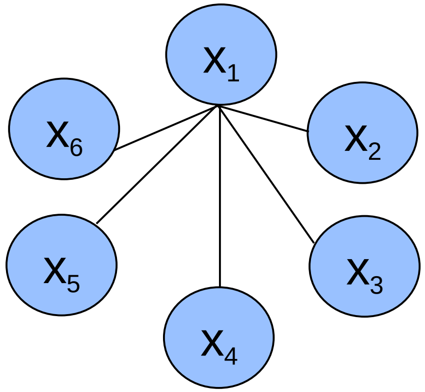
Two possible graphical models:

Fully connected constellation model



- **e.g. Constellation Model**
- **Parts fully connected**
- **Recognition complexity: $O(N^P)$**

"Star" shape model

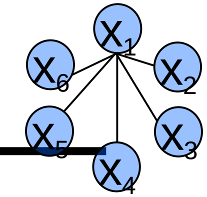


- **e.g. implicit shape model**
- **Parts mutually independent**
- **Recognition complexity: O(NP)**

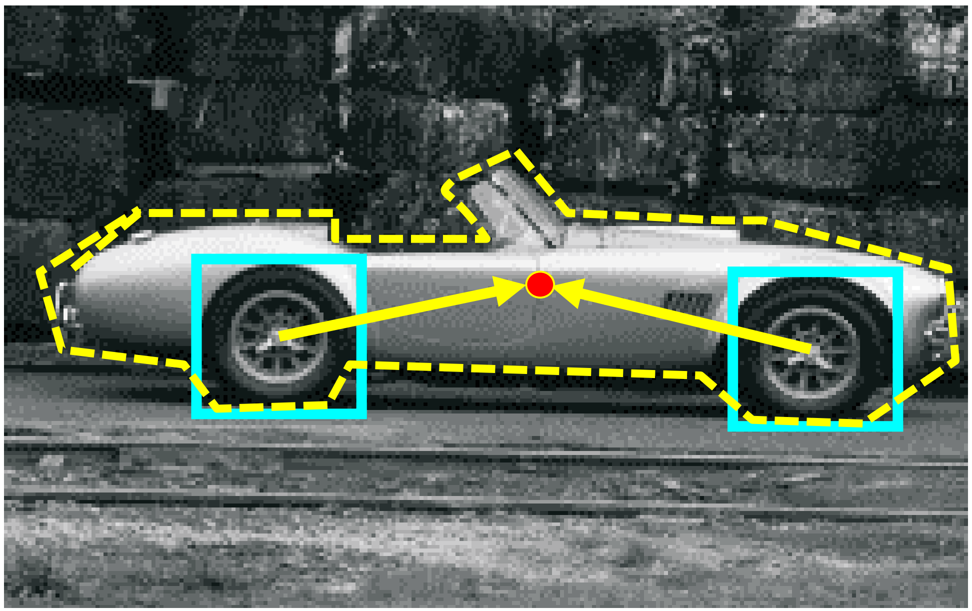
N image features, P parts in the model

Slide credit: Rob Fergus

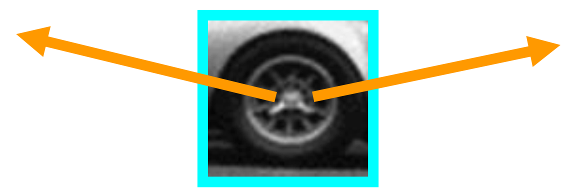
Star-shaped graphical model



- Discrete set of part appearances are used to index votes for object position



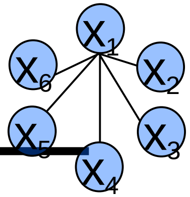
training image annotated with object localization info



Part with displacement vectors

B. Leibe, A. Leonardis, and B. Schiele, Combined Object Categorization and Segmentation with an Implicit Shape Model, ECCV Workshop on Statistical Learning in Computer Vision 2004

Star-shaped graphical model

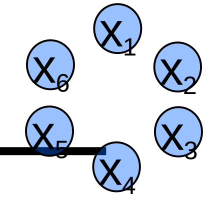


- Discrete set of part appearances are used to index votes for object position

test image

B. Leibe, A. Leonardis, and B. Schiele, [Combined Object Categorization and Segmentation with an Implicit Shape Model](), ECCV Workshop on Statistical Learning in Computer Vision 2004

Naïve Bayes model of parts



$$c^* = \arg\max_c \; p(c \mid w) \propto p(c)\, p(w \mid c) = p(c) \prod_{n=1}^{N} p(w_n \mid c)$$

Object class decision

Prior prob. of the object classes

Image likelihood given the class

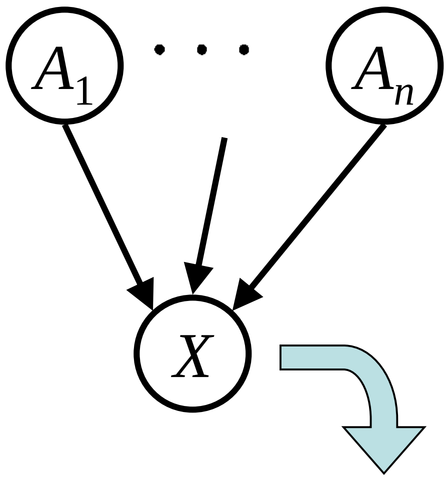
N patches

Bayes' Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

$$P(X|a_1 \ldots a_n)$$

  - CPT: conditional probability table
  - Description of a noisy "causal" process



$$P(X|A_1 \ldots A_n)$$

A Bayes net = Topology (graph) +

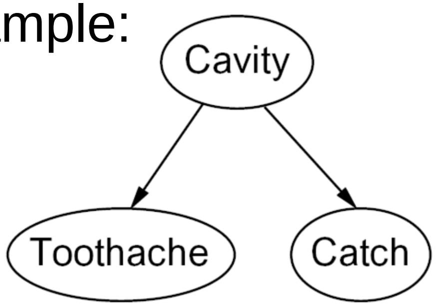
Local Conditional Probabilities

Probabilities in BNs

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Example:



P(Cavity) * P(Ache | Cavity) * P(Catch | Cavity)

$$P(+cavity, +catch, \neg toothache)$$

Probabilities in BNs

- Why are we guaranteed that setting

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \mathit{parents}(X_i))$$

results in a proper distribution?

Recall: The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

- Why is this always true?

$$P(x_1, x_2, x_3) = P(x_1, x_2)P(x_3|x_1, x_2)$$

$$= P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

Probabilities in BNs

- Why are we guaranteed that setting

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \mathit{parents}(X_i))$$

results in a proper distribution?

- Chain rule (valid for all distributions):

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i | x_1 \ldots x_{i-1})$$

- Due to <u>assumed</u> conditional independences:

$$P(x_i | x_1 \ldots x_{i-1}) = P(x_i | \mathit{parents}(X_i))$$

- Consequence:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \mathit{parents}(X_i))$$

Example: Coin Flips

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$\cdots$

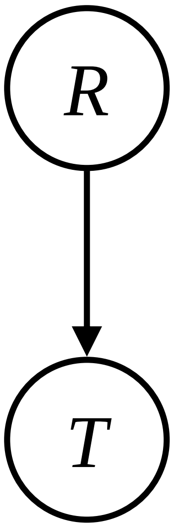
$P(X_n)$

h	0.5
t	0.5

$P(h, h, t, h) =$?

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic



$P(R)$

+r	1/4
¬r	3/4

$P(T|R)$

+r →

+t	3/4
¬t	1/4

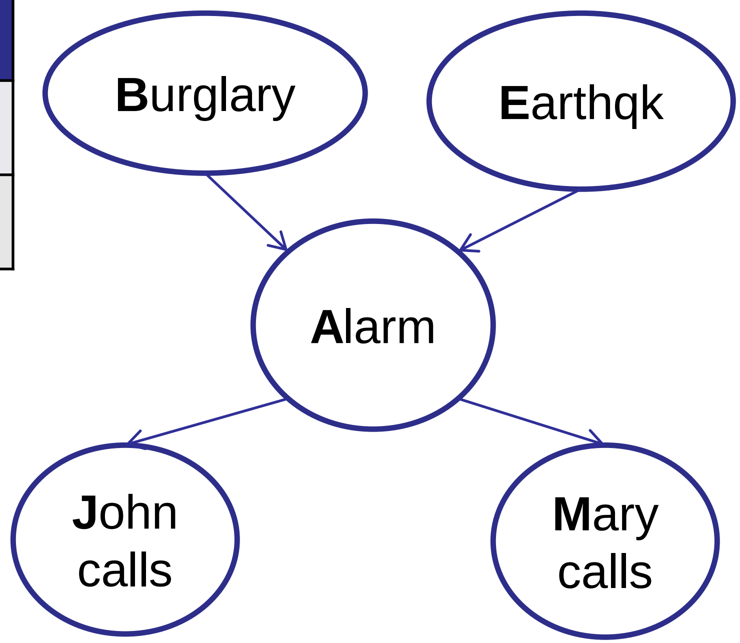
¬r →

+t	1/2
¬t	1/2

$$P(+r, \neg t) =$$

Example: Alarm Network

B	P(B)
+b	0.001
¬b	0.999



E	P(E)
+e	0.002
¬e	0.998

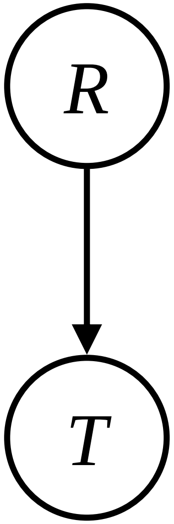
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	¬a	0.05
+b	¬e	+a	0.94
+b	¬e	¬a	0.06
¬b	+e	+a	0.29
¬b	+e	¬a	0.71
¬b	¬e	+a	0.001
¬b	¬e	¬a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	¬j	0.1
¬a	+j	0.05
¬a	¬j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	¬m	0.3
¬a	+m	0.01
¬a	¬m	0.99

Example: Traffic

- Causal direction



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

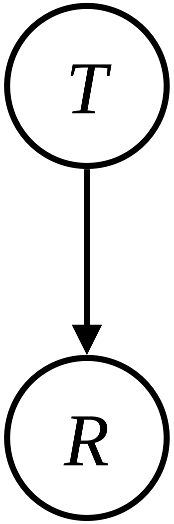
r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(T, R)$

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Example: Reverse Traffic

- Reverse causality?



$P(T)$

t	9/16
¬t	7/16

$P(R|T)$

t	r	1/3
	¬r	2/3
¬t	r	1/7
	¬r	6/7

$P(T,R)$

r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect **correlation**, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence

Summary: Bayes' Nets

- So far: how a Bayes' net encodes a joint distribution
- Next: how to answer queries about that distribution
  - Key idea: conditional independence
  - Today: assembled BNs using an intuitive notion of conditional independence as causality
  - Next: formalize these ideas
  - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)

Next week

- Making complex decisions