Announcements

Midterm 2

- Mon, 11/9, during lecture
- See Piazza for details

Schedule change

- Lecture on Friday instead of recitation
- Pre-recorded
- Suggested that you watch during your recitation timeslot
- TAs will be available during the Zoom session
- Any polls will be open all day

Plan

Last Time

- Generative models $\text{argmax}_\theta\, p(\,\boldsymbol{x} \mid y, \theta\,)\, p(y \mid \theta)$
- Naïve Bayes $\text{argmax}_\theta \prod_{m=1}^{M} p(\,x_m \mid y, \theta\,)\, p(y \mid \theta)$

Today

- Wrap up generative models and naïve Bayes
- Probability primer
- Bayes nets
- Markov chains

Wrap Up Generative Models and Naïve Bayes

Generative models and naïve Bayes slides…

Plan

Last Time

- Generative models $p(\boldsymbol{x}, y) = p(\,\boldsymbol{x} \mid y\,)\, p(y)$
- Naïve Bayes $p(\boldsymbol{x}, y) = \prod_{m=1}^{M} p(\,x_m \mid y\,)\, p(y)$

Today

- Bayes nets $p(z_1, z_2, z_3, z_4, z_5) = \prod_i p(\,z_i \mid parents(z_i)\,)$
- Markov chains $p(y_1, y_2, y_3, \ldots) = p(y_1)\, p(\,y_2 \mid y_1\,)\, p(\,y_3 \mid y_2\,) \ldots$

Next Time

- Hidden Markov models

Introduction to Machine Learning

Bayes Nets & Markov Chains

Instructor: Pat Virtue

Outline

1. Probability primer
2. Generative stories and Bayes nets
   - Bayes nets definition
   - Naïve Bayes
   - Markov chains

Probability Tools Summary

Our toolbox

- Definition of conditional probability

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

- Product Rule

$$P(A,B) = P(A \mid B)P(B)$$

- Bayes' theorem

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Discrete Probability Tables

Random variables, outcomes, and discrete distributions

- Capital letters/words are random variables and represent all possible discrete outcome
- Lowercase letters/words are specific outcomes of a random variable
- Example: Random variable $Weather$ (W) with three outcomes, $sun, rain, snow$

Discrete probability tables

- The probability distribution for discrete random variables can be represented as a table of parameters for each outcome, i.e. a Categorial distribution

W	$P(W)$
sun	0.5
$rain$	0.4
$snow$	0.1

Discrete Probability Tables

Joint distribution tables

- Tables contain entries of all possible combinations of outcomes for the multiple random variables
- Joint tables should sum to one
- Example: Random variables $Weather\ (W)$ and $Traffic\ (T)$

W	T	$P(W,T)$
sun	$light$	0.40
$rain$	$light$	0.12
$snow$	$light$	0.01
sun	$heavy$	0.10
$rain$	$heavy$	0.28
$snow$	$heavy$	0.09

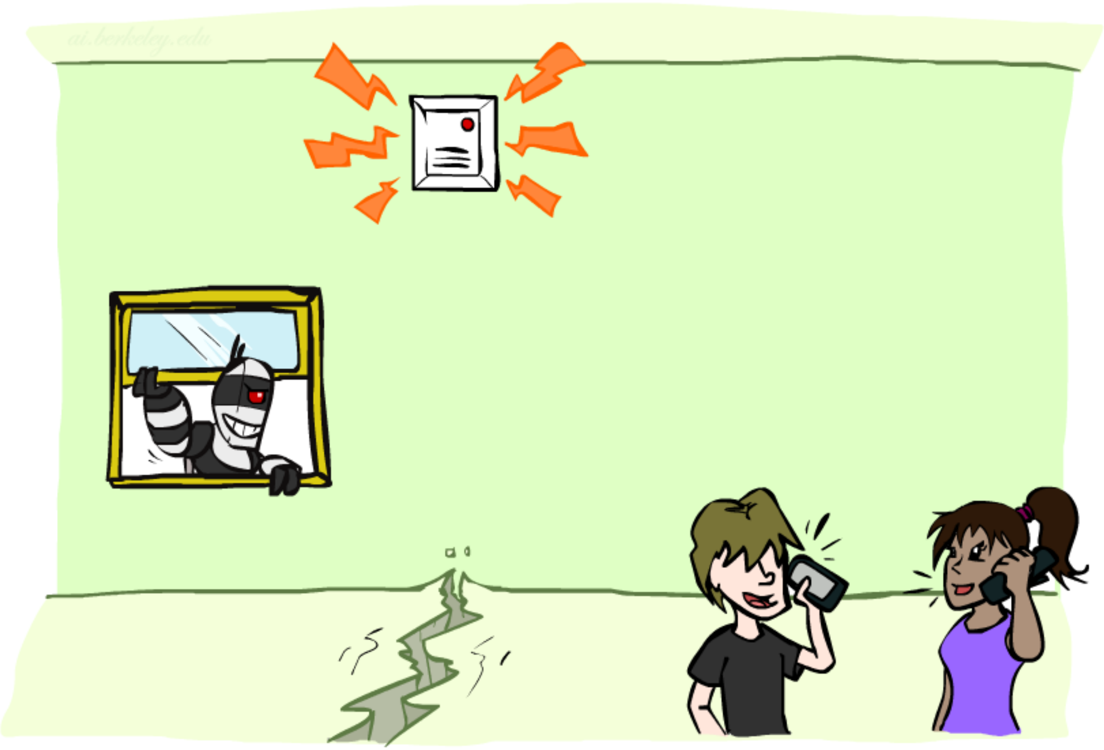
Discrete Probability Tables

Conditional probability tables (CPT)

- Tables can contain entries of all possible combinations of outcomes for the multiple random variables
- Joint tables won't necessarily sum to one. Why not?
- Example: Random variables $Weather\ (W)$ and $Traffic\ (T)$

W	T	$P(T \mid W)$
sun	$light$	0.8
$rain$	$light$	0.3
$snow$	$light$	0.1
sun	$heavy$	0.2
$rain$	$heavy$	0.7
$snow$	$heavy$	0.9

Piazza Poll 2



Variables, all binary

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

How many parameters are in the table $P(B, A, M, J, E)$?

A. 1

B. 5

C. 10

D. *25*

E. 2^5

F. 5!

Image: http://ai.berkeley.edu/

Probability Tools Summary

Our toolbox

- Product Rule

$$P(X_1, X_2) = P(X_1 \mid X_2)P(X_2)$$

- Chain Rule

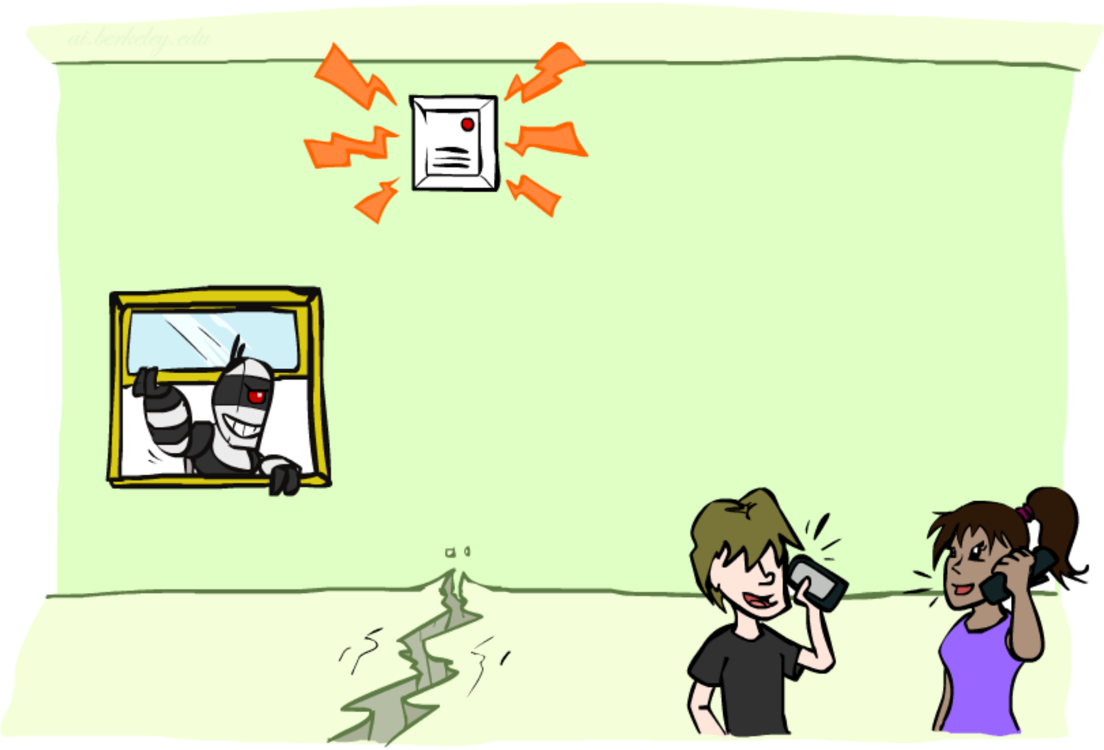
$$P(X_1, X_2, X_3) = P(X_1 \mid X_2, X_3)P(X_2, X_3)$$
$$= P(X_1 \mid X_2, X_3)P(X_2 \mid X_3)P(X_3)$$

$$P(X_1, \ldots, X_N) = \prod_{n=1}^{N} P(X_n \mid X_1, \ldots, X_{n-1})$$

Piazza Poll 3

Variables

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



How many different ways can we write the chain rule for $P(B,A,M,J,E)$?

A. 1

B. 5

C. 5 $choose$ 5

D. 5!

E. 5^5

Image: http://ai.berkeley.edu/

Probability Tools Summary

- Marginalization

$$P(A) = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} P(A, b, c)$$

- Normalization

$$P(B \mid a) = \frac{P(a, B)}{P(a)}$$

$$P(B \mid a) \propto P(a, B)$$

$$P(B \mid a) = \frac{1}{z} P(a, B)$$

$$z = P(a) = \sum_{b} P(a, b)$$

Probability Tools Summary

- Independence

  If A and B are independent, then:

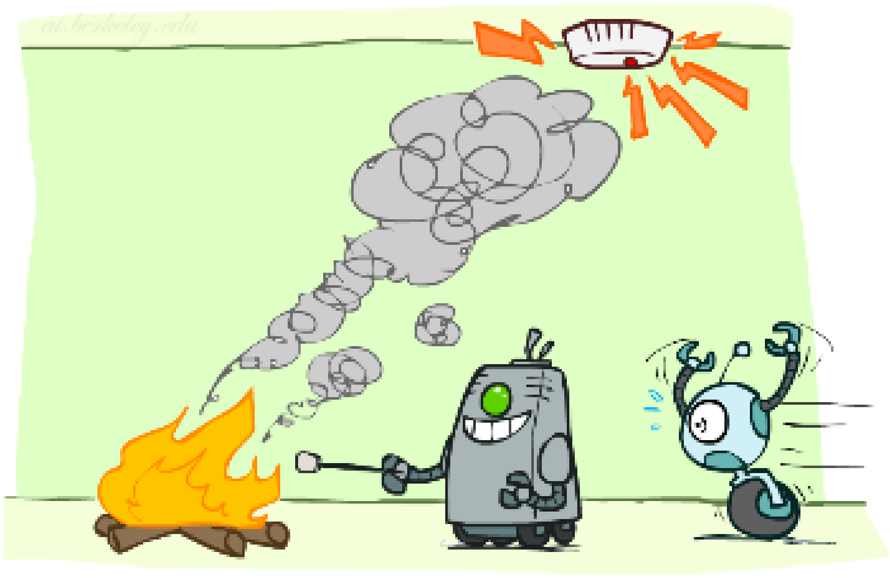
$$P(A,B) = P(A)P(B)$$
$$P(A \mid B) = P(A)$$
$$P(B \mid A) = P(B)$$

- Conditional independence

  If A and B are conditionally independent given C, then:

$$P(A,B \mid C) = P(A \mid C)P(B \mid C)$$
$$P(A \mid B,C) = P(A \mid C)$$
$$P(B \mid A,C) = P(B \mid C)$$

Generative Stories and Bayes Nets



Fire, Smoke, Alarm

- Generative story and Bayes net

- Assumptions

- Joint distribution

Image: http://ai.berkeley.edu/

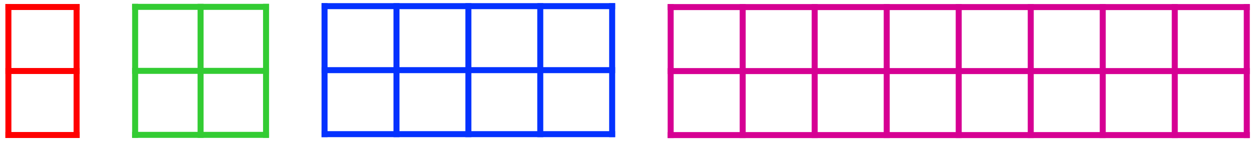
Bayesian Networks

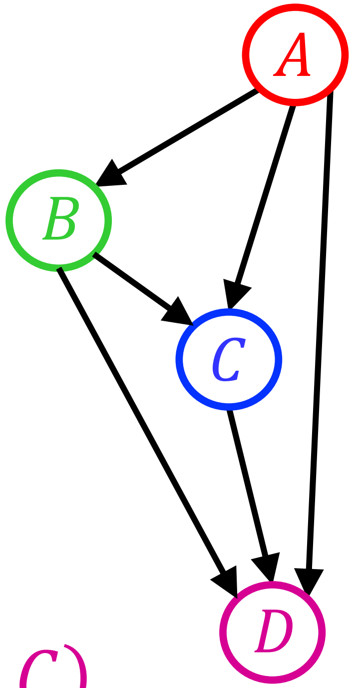
Bayes net

One node per random variable

Directed-Acyclic-Graph

One CPT per node: P(node | *Parents*(node))





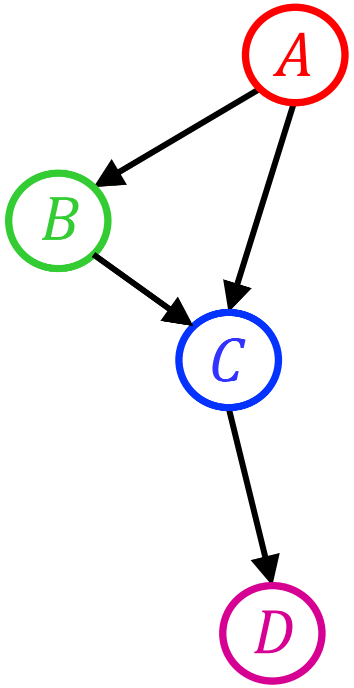
$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|A,B,C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1,\ldots,X_N) = \prod_i P(X_i|\,Parents(X_i))$$

Bayesian Networks

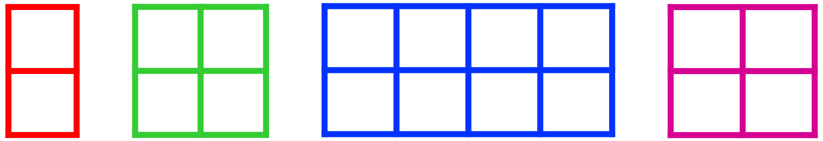
Bayes net



One node per random variable

Directed-Acyclic-Graph

One CPT per node: P(node | *Parents*(node))



$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1,\ldots,X_N) = \prod_i P(X_i|\,Parents(X_i))$$