CS 188: Artificial Intelligence
Spring 2010

Lecture 14: Bayes' Nets

3/4/2010

Pieter Abbeel – UC Berkeley

Many slides throughout the course adapted from Dan Klein, Stuart Russell, Andrew Moore

Announcements

- Assignments
  - P3 due tonight
  - W4 going out tonight
- Midterm
  - 3/18, 6-9pm, 0010 Evans
  - No lecture on 3/18

2

Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

3

Inference by Enumeration

- P(sun)? $= \sum_{s,t} P(s,t,sun)$
  $= 0.3 + 0.1 + 0.1 + 0.15$
  $= .65$

- P(sun | winter)? $\dfrac{\sum_t P(sun, winter, t)}{\sum_{w,t} P(w, winter, t)}$
  $= \dfrac{0.1 + 0.15}{0.1 + 0.05 + 0.15 + 0.20} = \dfrac{.25}{.50} = .5$

- P(sun | winter, warm)?
  $= \dfrac{P(sun, winter, warm)}{P(winter, warm)} = \dfrac{P(sun, winter, warm)}{P(winter, warm, sun) + P(winter, warm, rain)}$
  $= \dfrac{0.1}{0.1 + 0.05} = \dfrac{2}{3}$

Joint probability

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

4

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

  $\Bigg\}$ $X_1, X_2, \dots X_n$ *All variables*

- We want: $P(Q|e_1 \dots e_k)$

$$P(Q=q|e_1 \dots e_k) = \frac{P(Q=q, e_1, e_2, \dots, e_k)}{\sum_q P(Q=q, e_1, e_2, \dots, e_k)}$$

- First, select the entries consistent with the evidence
- Second, sum out H to get joint of Query and evidence:

$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Finally, normalize the remaining entries to conditionalize

- Obvious problems:
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

* *Works fine with multiple query variables, too*

The Product Rule

- Sometimes have conditional distributions but want the joint

defn. cond. prob.

$$P(x|y) = \frac{P(x,y)}{P(y)} \quad \Longleftrightarrow \quad \boxed{P(x,y) = P(x|y)P(y)}$$

- Example:

$P(W)$

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

$P(D,W)$

D	W	P
wet	sun	0.08
dry	sun	0.72
wet	rain	0.14
dry	rain	0.06

P(D,W)

D	W	P
w	s	.8*.1=.08
d	s	.8*.9=.72
w	r	.2*.7=.14
d	r	.2*.3=.06

6

The Chain Rule

Product rule
$P(x_1, x_2) = P(x_1) . P(x_2 | x_1)$

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

$$P(x_1, x_2, x_3) = \overbrace{P(x_1)P(x_2|x_1)}^{P(x_1, x_2)}P(x_3|x_1, x_2)$$

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

$y = (x_1, x_2)$

$P(y, x_3) = P(y) . P(x_3 | y)$

$P(x, y, z) = P(x) P(y|x) P(z|x, y)$
$= P(y) P(x|y) P(z|x, y) = P(z) P(y|z) . P(x|y, z)$

- Why is this always true?

7

Bayes' Rule

- Two ways to factor a joint distribution over two variables:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

That's my rule!

- Dividing, we get:

$$P(x|y) = \frac{P(y|x)}{P(y)}\underbrace{P(x)}_{\text{prior}}$$

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple
  - Foundation of many systems we'll see later (e.g. ASR, MT)
- In the running for most important AI equation!

8

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

$$P(\text{Cause}|\text{Effect}) = \frac{P(\text{Effect}|\text{Cause})P(\text{Cause})}{P(\text{Effect})}$$

- Example:
  - m is meningitis, s is stiff neck

$$P(s|\neg m) = 0.1$$

$$\left.\begin{array}{l} P(s|m) = 0.8 \\ P(m) = 0.0001 \\ \cancel{P(s) = 0.1} \end{array}\right\} \text{Example givens}$$

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

$$\frac{P(s|m)\cdot P(m)}{P(s)} = \frac{P(s|m)P(m)}{P(s,m)+P(s,\neg m)} = \frac{P(s|m)P(m)}{P(m)P(s|m)+P(\neg m)P(s|\neg m)}$$

$$= \frac{0.8 * 0.0001}{0.8 * 0.0001 + 0.9999 * 0.1} \approx 0.0008$$

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?

9

Ghostbusters, Revisited

- Let's say we have two distributions:
  - Prior distribution over ghost location: P(G)
    - Let's say this is uniform
  - Sensor reading model: P(R | G)
    - Given: we know what our sensors do
    - R = reading color measured at (1,1)
    - E.g. P(R = yellow | G=(1,1)) = 0.1

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

- We can calculate the posterior distribution P(G|r) over ghost locations given a reading using Bayes' rule:

$$P(g|r) \propto P(r|g)P(g)$$

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

10

Ghostbusters, Revisited

- P(G): Prior distribution over ghost location
- Sensor reading model: P(R | G)

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

reading = yellow

P(R | 0)

R	P
r	.8
o	.18
y	.02
g	0

P(R | 1)

R	P
r	.2
o	.5
y	.27
g	.03

P(R | 2)

R	P
r	.05
o	.15
y	.5
g	.3

P(R | 3)

R	P
r	0
o	.1
y	.3
g	.6

$P(g, R=y)$

- Bayes' rule: $P(g|r) \propto P(r|g)P(g)$

$P(y\|2)\cdot P(1,3) = .5 * .11$	$P(y\|3)\cdot P(2,3) = .3 * .11$	$P(y\|3+)\cdot P(3,3) = .3 * .11$
$P(y\|1)\cdot P(1,2) = .27 * .11$	$P(y\|2)\cdot P(2,2) = .5 * .11$	$P(y\|3+)\cdot P(3,2) = .3 * .11$
$P(y\|0)\cdot P(1,1) = 0.02 * 0.11$	$P(y\|1)\ P(2,1) = .27 * .11$	$P(y\|2)\cdot P(3,1) = .5 * .11$

11

Independence

- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y) \quad (1)$$

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

$$\forall x, y : P(x|y) = P(x) \quad (2)$$

$\Leftarrow$ (1) $\Rightarrow$ (2)

(2) $P(x|y) \overset{\text{always}}{=} \frac{P(x,y)}{P(y)} \overset{\text{independence}}{=} \frac{P(x)\cdot \cancel{P(y)}}{P(y)}$

  - We write: $\boxed{X \perp\!\!\!\perp Y}$

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

13

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

$P_2(T, W)$ $P(T) \cdot P(W)$

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

$P(W)$

W	P
sun	0.6
rain	0.4

14

Example: Independence

- N fair, independent coin flips:

$2 \cdot n$

$P(X_1)$

H	0.5
T	0.5

$P(X_2)$

H	0.5
T	0.5

$\cdots$

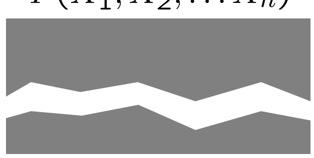
$P(X_n)$

H	0.5
T	0.5

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

2^n

2^{n-1}



15

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

16

Probabilistic Models

- For n variables with domain sizes d, joint distribution table with $d^n - 1$ free parameters [recall probabilities sum to one]
- Size of representation if we use the chain rule

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

Concretely, counting the number of free parameters accounting for that we know probabilities sum to one:

$P(X_1)$ $P(X_2|X_1)$ $P(X_3|X_1,X_2) \rightarrow P(X_3|X_2)$

$1+q+\cdots q^{n-1} = \frac{q^n - 1}{q-1}$

$(d-1) + d(d-1) + d^2(d-1) + \ldots + d^{n-1}(d-1)$

$= (d^n-1)/(d-1)\ (d-1)$

$= d^n - 1$

conditional independence $P(X_3 = x_3 | X_1 = x_1, X_2 = x_2) = P(X_3 = x_3 | X_2 = x_2)$

[why do both representations have the same number of free parameters?]

17

$P(x|y) = P(x)$ (1)
$P(x,y) = P(x)\,P(y)$ (2)

Independence $P(x|y) = P(x)$

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, ¬cavity) = P(+catch| ¬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) (1)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity) (2)
  - One can be derived from the other easily

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

20

Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments:

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

$$\boxed{X \perp\!\!\!\perp Y | Z}$$

$Y \perp X | Z$

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

$\boxed{U \perp\!\!\!\perp T \mid R}$

$R \to U$, $R \to T$

$\cancel{Z \perp\!\!\!\perp X | Y}$

- What about fire, smoke, alarm?

$F \perp\!\!\!\perp A \mid S$ (assumes A triggered by smoke)

$F \to S \to A$

21