Probability 101++

Many slides courtesy of Dan Klein, Stuart Russell, or Andrew Moore

CS 5300 / CS 6300
Artificial Intelligence
Spring 2009

Hal Daumé III
hal@cs.utah.edu

www.cs.utah.edu/~hal/courses/2009S_AI

Announcements

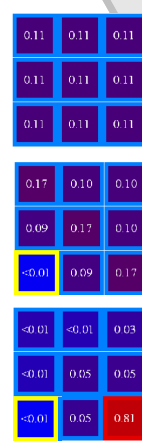
- All homework/project solutions online now
  - (Except HW5 which will be up this afternoon)
- We're making a few minor changes to the contest setup to make it more fun (should be finalized next week)
  - Note: max team size is 6, not unlimited :)

Today

- Probability
  - Random Variables
  - Joint and Conditional Distributions
  - Inference, Bayes' Rule
  - Independence
- You'll need all this stuff for the next few weeks, so make sure you go over it!

Uncertainty

- General situation:
  - Evidence: Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - Hidden variables: Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
  - Model: Agent knows something about how the known variables relate to the unknown variables
- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge



Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - D = How long will it take to drive to work?
  - L = Where am I?
- We denote random variables with capital letters
- Like in a CSP, each random variable has a domain
  - R in {true, false} (often write as {r, ¬r})
  - D in [0, ∞)
  - L in possible locations

Probabilities

- We generally calculate 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 reported accidents, 5 a.m.) = 0.95
  - P(on time | no reported accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

Probabilistic Models

- CSPs:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact
- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact

T	W	P
hot	sun	T
hot	rain	F
cold	sun	F
cold	rain	T

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

Joint Distributions

- A *joint distribution* over a set of random variat[redacted] specifies a real number for each assignment (or *outcome*):

[redacted]

[redacted]

  - Size of distribution if n variables with domain sizes d?
  - Must obey: [redacted]

[redacted]

- For all but the smallest distributions, impractical to write out

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

Events

- An *event* is a set E of outcomes

[redacted]

- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny?
  - Probability that it's hot?
  - Probability that it's hot OR sunny?
- Typically, the events we care about are *partial assignments*, like P(T=h)

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

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

[redacted]

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

[redacted]

[redacted]

T	P
hot	0.5
cold	0.5

[redacted]

W	P
sun	0.6
rain	0.4

[redacted]

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

[redacted]

[redacted]

W	P
sun	0.8
rain	0.2

[redacted]

W	P
sun	0.4
rain	0.6

Joint Distribution

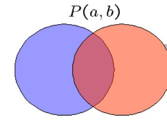
[redacted]

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

Conditional Distributions

- A simple relation between joint and conditional probabilities
  - In fact, this is taken as the *definition* of a conditional probability

$$P(a|b) = \frac{P(a,b)}{P(b)}$$



$P(T,W)$

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

$$P(W = r|T = c) = ???$$

Normalization Trick

- A trick to get a whole conditional distribution at once:
  - Select the joint probabilities matching the evidence
  - Normalize the selection (make it sum to one)

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

Select →

[redacted]

T	P
hot	0.1
cold	0.3

Normalize →

[redacted]

T	P
hot	0.25
cold	0.75

  - Why does this work? Because sum of selection is P(evidence)!

[redacted]

The Product Rule

- Sometimes have a joint distribution but want a conditional
- Sometimes the reverse

[redacted] ⟷ [redacted]

- Example:

[redacted]

R	P
sun	0.8
rain	0.2

[redacted]

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

⟷

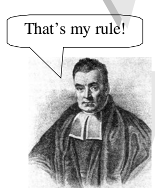
[redacted]

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

Bayes' Rule

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

[redacted]



- Dividing, we get:

[redacted]

- 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!

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

[redacted]

- Example:
  - m is meningitis, s is stiff neck [redacted] } Example givens

[redacted]

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

Ghostbusters

- Let's say we have two distributions:
  - Prior distribution over ghost locations: P(L)
    - Say this is uniform (for now)
  - Sensor reading model: P(R | L)
    - Given by some known black box process
    - E.g. P(R = yellow | L=(1,1)) = 0.1
    - For now, assume the reading is always for the lower left corner

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

- We can calculate the posterior distribution over ghost locations using Bayes' rule:

[redacted]

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

Inference by Enumeration

- P(sun)?
- P(sun | winter)?
- P(sun | winter, warm)?

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

Independence

- Two variables are *independent* in a joint distribution if:

[redacted]

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Usually variable aren't independent!
- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity}?
- Independence is like something from CSPs: what?

Example: Independence

- N fair, independent coin flips:

[redacted]

H	0.5
T	0.5

[redacted]

H	0.5
T	0.5

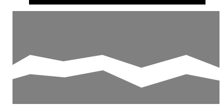
[redacted]

[redacted]

H	0.5
T	0.5

[redacted]

[redacted]



Example: Independence?

- Arbitrary joint distributions can be poorly modeled by independent factors

[redacted]

T	P
warm	0.5
cold	0.5

[redacted]

W	P
sun	0.6
rain	0.4

[redacted]

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

[redacted]

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

Conditional Independence

- Warning: we're going to use domain knowledge, not laws of probability, here to simplify a model!
- 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)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)

Conditional Independence

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

[redacted]

- What about this domain:
  - Traffic
  - Umbrella
  - Raining
- What about fire, smoke, alarm?

The Chain Rule II

- Can *always* write any joint distribution as an incremental product of conditional distributions

[redacted]

- Why?
- This actually claims nothing…
- What are the sizes of the tables we supply?

The Chain Rule III

- Trivial decomposition:

[redacted]

- With conditional independence:

[redacted]

- Conditional independence is our most basic and robust form of knowledge about uncertain environments
- Graphical models (next class) will help us work with and think about conditional independence