So far …

- Part I: Search and Planning!
  - We’ve seen how AI methods can solve problems in:
    - Search
    - Constraint Satisfaction Problems
    - Games
    - Markov Decision Problems
    - Reinforcement Learning

Next up …

- (Probability Review)
- Part II: Machine Learning
  - naive bayes
  - perceptron
  - kernels and clustering
  - neural networks/deep learning
- Part III: Probabilistic Reasoning
  - HMMs
  - particle filtering
  - bayes nets (not covered)
- Advanced Applications (Vision, NLP)

CS 188: Artificial Intelligence

Instructor: Marco Alvarez — University of Rhode Island

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.]

Probability

Today

- Probability
  - Random Variables
  - Joint and Marginal Distributions
  - Conditional Distribution
  - Product Rule, Chain Rule, Bayes’ Rule
  - Inference
  - Independence

You’ll need all this stuff A LOT for the next few weeks, so make sure you go over it now!

Inference in Ghostbusters

- A ghost is in the grid somewhere
- Sensor readings tell how close a square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green

P(red | 3)
P(orange | 3)
P(yellow | 3)
P(green | 3)

0.05
0.15
0.5
0.3

Sensors are noisy, but we know P(Color | Distance)

Video of Demo Ghostbuster – No probability

Probability Distributions

- Associate a probability with each value
  - Temperature:
  - Weather:
Unobserved random variables have distributions

\[ P(T), P(W) \]

- Must have: \( P(W = \text{rain}) = 0.1 \)
- Must obey: \( \sum_x P(X = x) = 1 \)

A distribution is a table of probabilities of values

A probability (lower case value) is a single number

Typically, the events we care about are partial assignments, like \( P(T=\text{hot}) \)

\[ P(E) = \sum_{x_1, \ldots, x_n} P(x_1, \ldots, x_n) \]

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

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

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

\[ P(X_1, X_2, \ldots, X_n) \]

Joint distributions: say whether assignments are likely

Marginal distributions are sub-tables which eliminate variables

Constraint satisfaction problems:

- Variables with domains
- Constraints: state whether assignments are allowed
- Ideally: only certain variables directly interact

Marginal distributions are sub-tables which eliminate variables

Constraint satisfaction problems:

- Variables with domains
- Constraints: state whether assignments are allowed
- Ideally: only certain variables directly interact
### Conditional Distributions
- Conditional distributions are probability distributions over some variables given fixed values of others.

<table>
<thead>
<tr>
<th>P(X = hot)</th>
<th>P(X = cold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>P</td>
</tr>
<tr>
<td>hot</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
</tr>
</tbody>
</table>

### Normalization Trick
- The joint probability distribution is:
  \[ P(T, W) \]

- The conditional probability distribution is:
  \[ P(W | T) \]

- The marginal probability distribution is:
  \[ P(T) \]

### Quiz: Normalization Trick
- \[ P(X = y | Y = y) ? \]

- \[ P(X = y) \]

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

- We generally compute conditional probabilities from other known probabilities (e.g., \[ P(T = c) \]).

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

- We generally compute conditional probabilities from other known probabilities (e.g., \[ P(T = c) \]).

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

- We generally compute conditional probabilities from other known probabilities (e.g., \[ P(T = c) \]).
Obvious problems:
- Worst-case time complexity $O(d^n)$
- Space complexity $O(d^n)$ to store the joint distribution

Sometimes have conditional distributions but want the joint:
$$P(y)P(x|y) = P(x, y)$$

Example:
$$P(D)P(W|D) = P(W,D)$$

More generally, can always write any joint distribution as an incremental product of conditional distributions:
$$P(x_1, x_2, x_3) = P(x_3| x_2)P(x_2|x_1, x_3)$$
$$P(x_1, x_2, ..., x_n) = \prod_i P(x_i|x_1, ..., x_{i-1})$$

Why is this always true?

The Chain Rule

Bayes Rule

Two ways to factor a joint distribution over two variables:
$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

Dividing, we get:
$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

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:
$$P(M|w) = \frac{P(w|M)P(M)}{P(w)}$$

Examples:
- M: meningitis, S: stiff neck
- $P(M)$ - prior
- $P(S|M)$ - likelihood
- $P(S)$ - marginal probability

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

Quiz: Bayes’ Rule

- Given:
  - $P(W)$
  - $P(D|W)$

- What is $P(W|dry)$?

Ghostbusters, Revisited

Let’s say we have two distributions:
- Prior distribution over ghost location: $P(G)$
- Sensor reading model: $P(R|G)$

Given: we know what our sensors do
- $R = \text{reading color measured at (1,1)}$
- E.g. $P(R = \text{yellow} | G=(1,1)) = 0.1$

We can calculate the posterior distribution $P(G|r)$ over ghost locations given a reading using Bayes’ rule:
$$P(g|r) \propto P(r|g)P(g)$$
Video of Demo Ghostbusters with Probability

Ghostbusters, Revisited

- Let's say we have two distributions:
  - A normal distribution with mean 0 and standard deviation 1
  - A Cauchy distribution with location 1 and scale 1
- Plot these two distributions
- Compare the two distributions
- Discuss the implications of using a Cauchy distribution
- Conclusion