Mid-semester feedback

1. What’s been most useful to you (and why)?
   - 12: projects
   - 2: book reading assignments
   - open forum
   - lectures better than book
   - class discussions

2. What could be going better / be more useful (and why)?
   - 2: grade code correctness not output / don’t like autograder
   - 2: more detail on slides
   - 2: skip math/tech details during reading
   - Give more math details
   - more competition
   - less reading
   - book is challenging
   - help to have book solutions, essay questions
   - Better grades
Mid-semester feedback

What could students do to improve the class?
• 3: Do projects earlier,
• 2: Let Matt answer questions and then let students answer/respond
• talking during lecture is distracting
• only have 1 person answer at once
• give Matt more feedback
• be more involved in discussions
• not turn a nap into a full-blow 8 hr sleep

What could Matt do to improve the class?
• 5: More examples/problems/in-class activities
• 2: Fewer reading responses
• 2: go over response ?’s in class
• give a easy question for each reading
• keep moodle up to date
• Cover Neural networks
• more lol cats/memes
• 2: more food
## Mid-semester feedback

How many hours per week do you spend on the course *outside* of class?

<table>
<thead>
<tr>
<th>Hours</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>5.5</td>
<td>9.2 +/- 7</td>
</tr>
<tr>
<td>6</td>
<td>7.2 +/- 1.7</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
• Can you clarify what a hybrid Bayesian network is?

• What are linear Gaussian and condition Gaussian distributions? What is the difference between them? What do the equations mean? Why are they important to know?

• What are probit and logit distributions? What is the difference between them? Are they related to Gaussian distributions in any way? Why are they important?

• Question: Fig. 14.7 a) looks like a regular Normal distribution, not a probit one!

• Question: Could you please explain a Markov Blanket.

• Can you please explain Elimination-ask algorithm

• I'd write something about that Inference by enumeration thing, but I don't understand it. I hope if you want us to know this you go over this well in class, because the book is losing me.
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

- Sensors are noisy, but we know $P(\text{Color} \mid \text{Distance})$

<table>
<thead>
<tr>
<th>$P(\text{red} \mid 3)$</th>
<th>$P(\text{orange} \mid 3)$</th>
<th>$P(\text{yellow} \mid 3)$</th>
<th>$P(\text{green} \mid 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.15</td>
<td>0.5</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Uncertainty

• General situation:
  – **Evidence**: Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  – **Unobserved 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
Events

• An event is a set E of outcomes

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

• 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?

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

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

\[
P(T, W)
\]

\[
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\text{hot} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
\]

\[
P(T) = \sum_{s} P(t, s)
\]

\[
P(T)
\]

\[
\begin{array}{|c|c|}
\hline
T & P \\
\hline
\text{hot} & 0.5 \\
\text{cold} & 0.5 \\
\hline
\end{array}
\]

\[
P(W)
\]

\[
\begin{array}{|c|c|}
\hline
W & P \\
\hline
\text{sun} & 0.6 \\
\text{rain} & 0.4 \\
\hline
\end{array}
\]

\[
P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)
\]
Conditional Probabilities

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

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

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[ P(W = r|T = c) = ??? \]
Conditional Distributions

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

**Conditional Distributions**

\[
P(W|T = \text{hot})
\]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\[
P(W|T = \text{cold})
\]

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>rain</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Joint Distribution**

\[
P(T, W)
\]

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>
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)

\[ P(T, W) \]

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[ P(T, r) \]

Select

<table>
<thead>
<tr>
<th>T</th>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Normalize

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.25</td>
</tr>
<tr>
<td>cold</td>
<td>0.75</td>
</tr>
</tbody>
</table>

- Why does this work? Sum of selection is \( P(\text{evidence}) \)\! (\( P(r) \), here)

\[
P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}
\]
Probabilistic Inference

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

• We generally compute conditional probabilities
  – \( P(\text{on time} \mid \text{no reported accidents}) = 0.90 \)
  – These represent the agent’s beliefs given the evidence

• Probabilities change with new evidence:
  – \( P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95 \)
  – \( P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80 \)
  – Observing new evidence causes beliefs to be updated
Inference by Enumeration

• $P(\text{sun})$?

• $P(\text{sun} \mid \text{winter})$?

• $P(\text{sun} \mid \text{winter, hot})$?

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Inference by Enumeration

• General case:
  – Evidence variables: \( E_1 \ldots E_k = e_1 \ldots e_k \)
  – Query* variable: \( Q \)
  – Hidden variables: \( H_1 \ldots H_r \)

\[ \begin{align*}
\{ \quad & X_1, X_2, \ldots X_n \\
\text{All variables} & \end{align*} \]

• We want: \( P(Q|e_1 \ldots e_k) \)

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

\[
P(Q, e_1 \ldots e_k) = \alpha \sum_{h_1 \ldots h_r} P(Q, h_1 \ldots h_r, e_1 \ldots e_k) \quad X_1, X_2, \ldots X_n
\]

• 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

\[ P(x|y) = \frac{P(x, y)}{P(y)} \quad \quad P(x, y) = P(x|y)P(y) \]

- Example:

<table>
<thead>
<tr>
<th>P(W)</th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td></td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td></td>
<td></td>
<td>0.2</td>
</tr>
</tbody>
</table>

| P(D|W) | D | W | P |
|-------|---|---|---|
|       | wet | sun | 0.1 |
|       | dry | sun | 0.9 |
|       | wet | rain | 0.7 |
|       | dry | rain | 0.3 |

<table>
<thead>
<tr>
<th>P(D, W)</th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wet</td>
<td>sun</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>dry</td>
<td>sun</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>wet</td>
<td>rain</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>dry</td>
<td>rain</td>
<td>0.96</td>
</tr>
</tbody>
</table>
The Chain Rule

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

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]

\[ P(x_1, x_2, \ldots x_n) = \prod_{i} P(x_i|x_1 \ldots x_{i-1}) \]
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:

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

• Example:
  – m is meningitis, s is stiff neck

\[
\begin{align*}
P(s|m) &= 0.8 \\
P(m) &= 0.0001 \\
P(s) &= 0.1
\end{align*}
\]

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

– Note: posterior probability of meningitis still very small
– Note: you should still get stiff necks checked out! Why?
Let’s say we have two distributions:

- Prior distribution over ghost location: \( P(G) \)
  - Let’s say this is uniform
- Sensor reading model: \( P(R \mid G) \)
  - Given: we know what our sensors do
  - \( R = \) reading color measured at \((1,1)\)
  - E.g. \( P(R = \text{yellow} \mid G=(1,1)) = 0.1 \)

We can calculate the posterior distribution \( P(G \mid r) \) over ghost locations given a reading using Bayes’ rule:

\[
P(g \mid r) \propto P(r \mid g)P(g)
\]