Outline

• Syllabus
• Probabilistic Reasoning
• Structure in Probabilistic Models
• Incorporating Evidence
• Belief Networks
• Theano: A quick look
• Homework
<table>
<thead>
<tr>
<th>Week</th>
<th>Syllabus</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intro</td>
<td>Belief networks I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 1, 3</td>
</tr>
<tr>
<td>2</td>
<td>Belief Networks II</td>
<td>Graphical Models I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 3, 4</td>
</tr>
<tr>
<td>3</td>
<td>Graphical Models II</td>
<td>Exact Inference (Trees) I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 4, 5</td>
</tr>
<tr>
<td>4</td>
<td>Exact Inference (Trees) II</td>
<td>Exact Inference (Junction Tree)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 5, 6</td>
</tr>
<tr>
<td>5</td>
<td>Approximate Inf (Sampling)</td>
<td>Approximate Inference (Samp)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 27</td>
</tr>
<tr>
<td>6</td>
<td>Approximate Inf (Variational)</td>
<td>Approximate Inference (Variational)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 28</td>
</tr>
<tr>
<td>7</td>
<td>Learning as inference</td>
<td>Learning with hidden variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ch 9, 11</td>
</tr>
<tr>
<td>8</td>
<td>Single Neurons</td>
<td>Feedforward Neural Networks</td>
</tr>
<tr>
<td>9</td>
<td>Autoencoders</td>
<td>Stacked Denoising Autoencoders</td>
</tr>
<tr>
<td>10</td>
<td>Hopfield Networks / Boltzmann</td>
<td>Restricted Boltzmann Machines</td>
</tr>
<tr>
<td>11</td>
<td>Deep Boltzmann Machines</td>
<td>Deep Belief Networks</td>
</tr>
<tr>
<td>12</td>
<td>Convolutional Neural Networks</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Applications</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Applications</td>
<td></td>
</tr>
</tbody>
</table>
Probabilistic Reasoning

• Characterise the problem space by identifying:

  • Identify variables and their domains (e.g. X in \{0,1\})
  • The joint distribution (e.g. p(X,Y,Z))
  • Perform inference on joint distribution
At work, Bob’s boss is angry because he’s underperforming. Bob has a headache. Possibly Bob is just demotivated, possibly he went to a party last night and is exhausted. What is the probability that he went to a party last night?
Identify Variables and Domains

- \( H \) in \( \{t, f\} \) : Bob has a headache
- \( P \) in \( \{t, f\} \) : Bob went to a party
- \( D \) in \( \{t, f\} \) : Bob is demotivated
- \( U \) in \( \{t, f\} \) : Bob is underperforming
- \( A \) in \( \{t, f\} \) : The boss is angry
Define Joint Distribution

- \( P(H,P,A,U,D) \) fully characterizes the problem
- To answer our question, calculate \( p(P=t \mid H=t, A=t) \)

\[
p(P = t \mid H = t, A = t) = \frac{p(P = t, H = t, A = t)}{p(H = t, A = t)}
\]

\[
= \frac{\sum_{U,D} p(P = t, H = t, A = t, U, D)}{\sum_{P,U,D} p(P, H = t, A = t, U, D)}
\]
Inference

All the ways in which Bob has a headache, the Boss is angry AND Bob went to a party

\[ \frac{\sum_{U,D} p(P = t, H = t, A = t, U, D)}{\sum_{P,U,D} p(P, H = t, A = t, U, D)} \]

All the ways in which Bob has a headache and the Boss is angry.
There's just one problem...

- How do we define the joint distribution?

- We could explicitly enumerate each case:
  
  - $p(H=f,P=f,A=f,U=f,D=f) = \ldots$
  
  - $P(H=f,P=f,A=f,U=f,D=t) = \ldots$

- Difficult to put down a sensible number

- Exponential growth in number of variables
Structured Models

• Influence between variables often indirect

• For example:
  • The boss is angry because Bob is underperforming, not because Bob might have gone to a party.

• Captured using conditional independence:

\[ p(A|P, H, U, D) = P(A|U) \]

• Given we know that Bob is underperforming or not, we know the probability that the boss will be angry.
However ...

- $p(H,P,A,U,D)$ isn’t conditioned on anything :(.
- But we know $p(X,Y) = p(X|Y)p(Y)$, so


Note: This is just one possible order!
Independence Assumptions

\[ p(A|H,P,U,D) = p(A|U) \]

Bob’s boss is only angry because Bob is underperforming, not for any other reason.
Independence Assumptions

\[ p(A|U) \ p(U|H,P,D) \ p(H|P,D) \ p(P|D) \ p(D) \]

Bob’s underperforming is either because he went to the party and is tired, or that he is demotivated.

Alternative choice: The headache might also play a role.

\[ p(U|H,P,D) = p(U|P,D) \]
Independence Assumptions

\[ p(A|U) \ p(U|P,D) \ p(H|P,D) \ p(P|D) \ p(D) \]

Bob’s headache is only dependent on whether or not he went to the party.

\[ p(H|P,D) = p(H|P) \]
Independence Assumptions

\[ p(A|U) \quad p(U|P,D) \quad p(H|P) \quad p(P|D) \quad p(D) \]

Bob’s headache is only dependent on whether or not he went to the party.

\[ p(P|D) = p(P) \]

A wild prior appears!
Independence Assumptions

\[ p(A|U) \ p(U|P,D) \ p(H|P) \ p(P) \ p(D) \]

Whether or not Bob is demotivated is prior to this situation. Nothing to do, since \( p(D) \) is already a prior.
Belief Networks

\[ p(A|H,P,U,D) \quad p(U|H,P,D) \quad p(H|P,D) \quad p(P|D) \quad p(D) \]

Directed Acyclic Graph (DAG)
Each edge is directed, no path following the arrows crosses the same node twice
Belief Networks

\[ p(A|H,P,U,D) \] \[ p(U|H,P,D) \] \[ p(H|P,D) \] \[ p(P|D) \] \[ p(D) \]

\[ p(A|H,P,U,D) = p(A|U) \]
Belief Networks

\[ p(A|U) \ p(U|H,P,D) \ p(H|P,D) \ p(P|D) \ p(D) \]

\[ p(U|H,P,D) = p(U|P,D) \]
Belief Networks

\[ p(A|U) \quad p(U|P,D) \quad p(H|P,D) \quad p(P|D) \quad p(D) \]

\[ p(H|P,D) = p(H|P) \]
Belief Networks

\[
p(A|U) \quad p(U|P,D) \quad p(H|P) \quad p(P|D) \quad p(D)
\]

\[
p(P|D) = p(P)
\]
Belief Networks

$p(A|U) \ p(U|P,D) \ p(H|P) \ p(P) \ p(D)$
Belief Networks

\[ p(A|U) \quad p(U|P,D) \quad p(H|P) \quad p(P) \quad p(D) \]
Belief Networks

\[ p(A|U) \quad p(U|P,D) \quad p(H|P) \quad p(P) \quad p(D) \]
## BN Specification

Conditional Probability Tables (CPDs)

| \( P(A|U) \) | \( U=F \) | \( U=T \) |
|--------------|----------|----------|
| A=F          |          |          |
| A=T          |          |          |

\( p(A|U) \) \( p(U|P,D) \) \( p(H|P) \) \( p(P) \) \( p(D) \)
**BN Specification**

**Conditional Probability Tables (CPDs)**

| P(A|U) | U=F | U=T |
|-------|-----|-----|
| A=F   |     |     |
| A=T   | 0.5 |     |

Grouchy!
## BN Specification

### Conditional Probability Tables (CPDs)

| P(A|U) | U=F | U=T |
|--------|-----|-----|
| A=F    |     |     |
| A=T    | 0.5 | 0.95|

Rage!
### BN Specification

Conditional Probability Tables (CPDs)

| P(A|U) | U=F | U=T |
|--------|-----|-----|
| A=F    | 0.5 | 0.05|
| A=T    | 0.5 | 0.95|

Redundant…
## BN Specification

### Conditional Probability Tables (CPDs)

| $P(U|P,D)$ | $P=F$ | $P=T$ | $P=F$ | $P=T$ |
|------------|-------|-------|-------|-------|
| $D=F$      |       |       |       |       |
| $D=T$      |       |       |       |       |

<table>
<thead>
<tr>
<th>$U=F$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$U=T$</th>
</tr>
</thead>
</table>
### BN Specification

Conditional Probability Tables (CPDs)

| P(U|P,D) | P=F | P=T | P=F | P=T |
|---------|-----|-----|-----|-----|
| D=F     | D=F | D=F | D=T | D=T |

| U=F | P(U|P,D) | P=F | P=T |
|-----|---------|-----|-----|
| U=T | 0.01    |     |     |
### BN Specification

**Conditional Probability Tables (CPDs)**

| P(U|P,D) | P=F | P=T | P=F | P=T |
|---|---|---|---|---|
| D=F | | | | |
| D=F | | | | |
| D=T | | | | |
| D=T | | | | |

| U=F | | | | |
| U=T | 0.01 | 0.9 | | |

... he hasn’t partied too hard ...
BN Specification

Conditional Probability Tables (CPDs)

| P(U|P,D)  | P=F | P=T | P=F | P=T |
|----------|-----|-----|-----|-----|
| D=F      |     |     |     |     |
| U=F      |     |     |     |     |
| U=T      | 0.01| 0.9 | 0.9 |     |
| D=T      |     |     |     |     |

... and he hasn’t become demotivated.
### BN Specification

#### Conditional Probability Tables (CPDs)

| \( P(U|P,D) \) | \( P=F \) | \( P=T \) | \( P=F \) | \( P=T \) |
|----------------|---------|---------|---------|---------|
| \( D=F \)     |         |         |         |         |
| \( D=F \)     |         |         |         |         |
| \( U=F \)     |         |         |         |         |
| \( U=T \)     | 0.01    | 0.9     | 0.9     | 0.999   |

Let alone both!
BN Specification

Conditional Probability Tables (CPDs)

| P(U|P,D) | P=F | P=T | P=F | P=T |
|---------|-----|-----|-----|-----|
| U=F     | 0.99 | 0.1 | 0.1 | 0.001 |
| U=T     | 0.01 | 0.9 | 0.9 | 0.999 |

Redundant...
### BN Specification

Conditional Probability Tables (CPDs)

| P(H|P) | P=F | P=T |
|-------|-----|-----|
| H=F   |     |     |
| H=T   |     |     |

p(A|U) p(U|P,D) \( p(H|P) \) p(P) p(D)
BN Specification

Conditional Probability Tables (CPDs)

| P(H|P) | P=F | P=T |
|-------|-----|-----|
| H=F   |     |     |
| H=T   |     | 0.2 |

No parties, no problems ... kind of.
The party lingers …
BN Specification

Conditional Probability Tables (CPDs)

| $P(H|P)$ | $P=F$ | $P=T$ |
|----------|-------|-------|
| $H=F$    | 0.8   | 0.1   |
| $H=T$    | 0.2   | 0.9   |

Redundant...
### BN Specification

Conditional Probability Tables (CPDs)

| P(P) | p(A|U) | p(U|P,D) | p(H|P) | P(P) | P(D) |
|------|--------|----------|--------|------|------|
| P=F  |        |          |        |      |      |
| P=T  |        |          |        |      |      |
On a given school night, 20% chance of party.

<table>
<thead>
<tr>
<th>P(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P=F</td>
</tr>
<tr>
<td>P=T</td>
</tr>
</tbody>
</table>
### BN Specification

#### Conditional Probability Tables (CPDs)

<table>
<thead>
<tr>
<th>P(D)</th>
<th>D=F</th>
<th>D=T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $p(A|U)$
- $p(U|P,D)$
- $p(H|P)$
- $p(P)$
- $p(D)$
BN Specification
Conditional Probability Tables (CPDs)

<table>
<thead>
<tr>
<th>P(D)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D=F</td>
<td>0.6</td>
</tr>
<tr>
<td>D=T</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Not Bob’s dream job ...
Incorporating Evidence

At work, Bob’s boss is angry because he’s underperforming. Bob has a headache. Possibly Bob is just demotivated, possibly he went to a party last night and is exhausted. What is the probability that he went to a party last night?
Incorporating Evidence

\[ p(H, P, A, U, D) = p(A|U) \cdot p(U|P, D) \cdot p(H|P) \cdot p(P) \cdot p(D) \]
Incorporating Evidence

\[ p(H=t, P, A=t, U, D) = p(A=t | U) \ p(U | P, D) \ p(H=t | P) \ p(P) \ p(D) \]
Inference

At work, Bob’s boss is angry because he’s underperforming. Bob has a headache. Possibly Bob is just demotivated, possibly he went to a party last night and is exhausted. What is the probability that he went to a party last night?

\[ p(P=t \mid H=t, A=t) \]
Inference

\[ p(P = t | H = t, A = t) = \frac{p(P = t, H = t, A = t)}{p(H = t, A = t)} \]

\[ = \frac{\sum_{U,D} p(P = t, H = t, A = t, U, D)}{\sum_{P,U,D} p(P, H = t, A = t, U, D)} \]

Using the CPDs, we can answer this now!

\[ p(H, P, A, U, D) = p(A | U) \cdot p(U | P, D) \cdot p(H | P) \cdot p(P) \cdot p(D) \]
\[
\frac{\sum_{U,D} p(P = t, H = t, A = t, U, D)}{\sum_{P,U,D} p(P, H = t, A = t, U, D)}
\]

\[
\frac{\sum_{U,D} p(A = t | U)p(U | P = t, D)p(H = t | P = t)p(P = t)p(D)}{\sum_{P,U,D} p(A = t | U)p(U | P, D)p(H = t | P)p(P)p(D)}
\]

Profit! Or, at least, the probability \( p(P=t | H=t, A=t) \)
Look at the numerator as an example, and imagine H, P and A were actually not given...

\[ \sum_{U,D} p(A|U)p(U|P,D)p(H|P)p(P)p(D) \]

Two-dimensional summation  Five-dimensional table

This is fine for now, but what if we had LOTS more variables?
Look carefully at which variables are in which CPDs

Rearrange CPDs ...

\[
\sum_{U,D} p(A|U)p(U|P,D)p(H|P)p(P)p(D)
\]

\[
\sum_{U,D} p(H|P)p(P)p(A|U)p(U|P,D)p(D)
\]
Sneak preview

\[ \sum_{U,D} p(H|P)p(P)p(A|U)p(U|P,D)p(D) \]

Rearrange summations …

\[ p(H|P)p(P) \sum_{U} p(A|U) \sum_{D} p(U|P,D)p(D) \]

1D Summation

Three dimensional table
Sneak preview

1D Summation

\[ p(H|P)p(P) \sum_U p(A|U) \sum_D p(U|P, D)p(D) \]

After inner summation ...

1D Summation

\[ p(H|P)p(P) \sum_U p(A|U)\sigma(U, P) \]

Three dimensional table (D is gone!)
Belief networks enable …

- ... representation of complex joint probability distributions
- ... intuitive definition of conditional independence relationships
- ... managing complexity by decomposing the model
- ... possibly lowering of computational complexity
Homework (Optional)

• Exercise 3.1 (Party animal), 3.4, 3.6, 3.8.1

• For at least one of these: Implement in Python (Numpy) Naively (always possible), and with rearranged summations (may not be possible for each case).

• Have a quick look at Theano:

http://deeplearning.net/software/theano/

Port your Numpy code to run in Theano

My office A510, or e-mail, if you have any questions.