Directed Acyclic Graphical Models

Given $X_1, X_2, \ldots, X_n$ (assume discrete, but equally applicable to cont.)

If you know the joint pf : $P(X_1, X_2, \ldots, X_n)$, then you can answer any probabilities queries on any combinations of $X$'s

$$P(X_1, X_2) = \sum_{X_3} \sum_{X_4} \cdots \sum_{X_n} P(X_1, X_2, \ldots, X_n)$$

$$P(X_1 | X_2, X_3) = \frac{P(X_1, X_2, X_3)}{P(X_2, X_3)}$$

ex $X$'s are binary, 40 random variables
How big is $P(X_1, X_2, \ldots, X_{40})$? $2^{40}$ entries. $\approx 16$ TB

Motivations: Graphical Models incorporate "human knowledge" about the random variables so as to produce a more compact representation of the probability function.

Independent Variables

\[ \begin{align*}
&X_1 \\
&X_2 \\
\end{align*} \]

\[ \begin{align*}
&X_3 \\
&X_4 \\
\end{align*} \]

Independence $\Rightarrow$ no edges at all.

$f_1(X_1)$, $f_2(X_2)$, $f_3(X_3)$, $f_4(X_4)$ are arbitrary functions with the following properties:

1. $f_i(X_i = x_i) > 0 \ \forall X_i \ \forall i$
2. $\sum_{\forall x_i} f_i(X_i = x_i) = 1 \ \forall i$
Define \( g(X_1, X_2, X_3, X_4) = f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \)

3 Claims:

1. \( g(X_1, X_2, X_3, X_4) \) is a probability mass function
2. \( f_i(x_i) = P(X_i) \) given \( g(\cdot) \) is the joint
3. \( X_i \)'s are independent given \( g(\cdot) \) is the joint

Proof:

1. \( g(\cdot) \geq 0 \) obvious

\[
\sum_{x_1, x_2, x_3, x_4} g(x_1, x_2, x_3, x_4) = \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \\
= \sum_{x_1} \sum_{x_2} \sum_{x_3} f_1(x_1) f_2(x_2) f_3(x_3) \sum_{x_4} f_4(x_4) \\
= 1 \quad \text{apply the same trick 3 more times}
\]

2. \( P(X_1) = \sum_{x_2} \sum_{x_3} \sum_{x_4} P(X_1, X_2, X_3, X_4) \)

\[
= \sum_{x_2} \sum_{x_3} \sum_{x_4} g(x_1, x_2, x_3, x_4) \\
= \sum_{x_2} \sum_{x_3} \sum_{x_4} f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \\
= f_1(x_1)
\]

(3) is "free".

Economy of Representation: \( n \) R.V.'s

Full joint pdf: \( 2^n \) rows (assume binary r.v.'s)

Independent R.V.'s: Store \( f_1, \ldots, f_n \) separately, each has 2 rows

\( \Rightarrow 2 \cdot n \) rows
Somewhere between full joint pf and total independence: graphical models

Directed Acyclic Graph (DAG)

Directed: X → Y  

parent  child

Acyclic: No directed cycle

X → Y  

NO!

X → Z  

OKAR!

Ex/ 

X₁ → X₂ → X₃ → X₄ → X₅ → X₆

No directed cycle ⇒ you can always find the "oldest" node in the graph (nobody is pointing at)

"Topological Order"

"oldest" next  "youngest"
<table>
<thead>
<tr>
<th>Parameterization</th>
<th>R. V.</th>
<th>Parents</th>
<th>&quot;Arbitrary $f_{ij}$&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td></td>
<td></td>
<td>$f_1(X_1)$</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$X_1$</td>
<td></td>
<td>$f_2(X_2, X_1)$</td>
</tr>
<tr>
<td>$X_3$</td>
<td>$X_1$</td>
<td></td>
<td>$f_3(X_3, X_1)$</td>
</tr>
<tr>
<td>$X_4$</td>
<td>$X_2$</td>
<td></td>
<td>$f_4(X_4, X_2)$</td>
</tr>
<tr>
<td>$X_5$</td>
<td>$X_3$</td>
<td></td>
<td>$f_5(X_5, X_3)$</td>
</tr>
<tr>
<td>$X_6$</td>
<td>$X_4, X_5$</td>
<td></td>
<td>$f_6(X_6, X_4, X_5)$</td>
</tr>
</tbody>
</table>

$$f_i(X_i, X_{\pi_i})$$

↑ the indices of all the parent

Basic properties that $f_i$ must satisfy:

1. $f_i(X_i, X_{\pi_i}) \geq 0$
2. $\sum_{X_i} f_i(X_i, X_{\pi_i}) = 1$ for all possible configurations of $X_{\pi_i}$

3 Claims

1. $g(X_1, \ldots, X_n) = \prod_{i=1}^{n} f_i(X_i, X_{\pi_i})$ is a pdf of $X_1, \ldots, X_n$
2. If $g(\cdot)$ is the joint, then $f_i(X_i, X_{\pi_i}) = P(X_i | X_{\pi_i})$
3. If $g(\cdot)$ is the joint, $X_i \perp X_{\text{non-descendants of } i | X_{\pi_i}}$