Probabilistic Relational Models

Often we want a random variable for each individual in a population

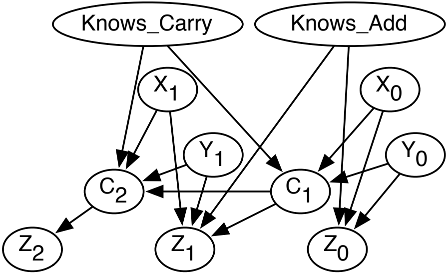
- build a probabilistic model before knowing the individuals
- learn the model for one set of individuals
- apply the model to new individuals
- allow complex relationships between individuals

Predicting students errors

$$
\begin{array}{rccc}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$

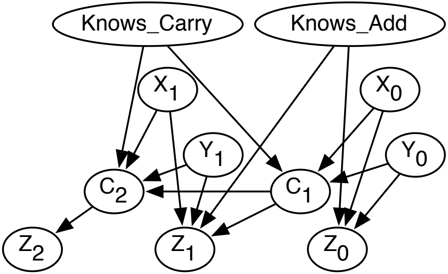
Predicting students errors

$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



Predicting students errors

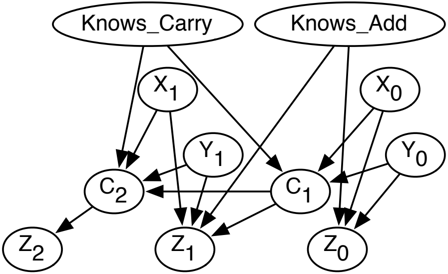
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits

Predicting students errors

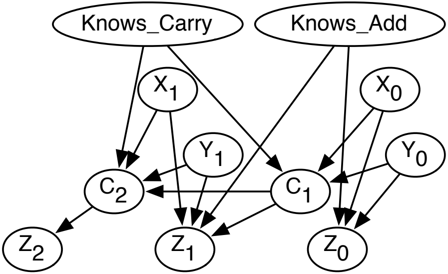
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems

Predicting students errors

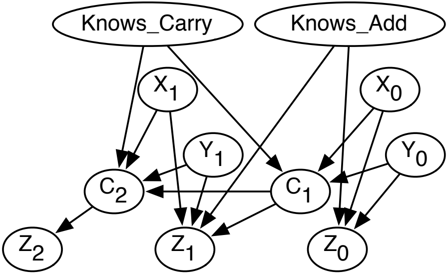
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems, students

Predicting students errors

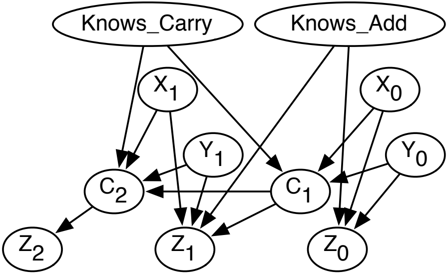
$$
\begin{array}{rccc}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$



What if there were multiple digits, problems, students, times?

Predicting students errors

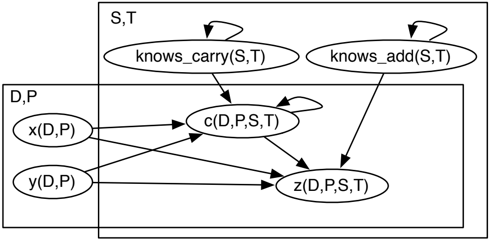
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems, students, times? How can we build a model before we know the individuals?

Multi-digit addition with parametrized BNs / plates

$$\begin{array}{rcccc} & x_{j_x} & \cdots & x_2 & x_1 \\ + & y_{j_z} & \cdots & y_2 & y_1 \\ \hline & z_{j_z} & \cdots & z_2 & z_1 \end{array}$$



Parametrized Random Variables: $x(D, P)$, $y(D, P)$, $knows\_carry(S, T)$, $knows\_add(S, T)$, $c(D, P, S, T)$, $z(D, P, S, T)$ for digit D, problem P, student S, time T.
There is a random variable for each assignment of a value to D and a value to P in $x(D, P)$....

Representing Conditional Probabilities

- $P(knows\_adn(X)|bright(X), taught\_adn(X))$ — parameter sharing — individuals share probability parameters.
- $P(happy(X)|friend(X, Y), mean(Y))$ — needs aggregation — $happy(a)$ depends on an unbounded number of parents.
- the carry of one digit depends on carry of the previous digit
- probability that two authors collaborate depends on whether they have a paper authored together

Independent Choice Logic

- A language for first-order probabilistic models.
- Idea : combine logic and probability, where all uncertainty in handled in terms of Bayesian decision theory, and a logic program specifies consequences of choices.
- Parametrized random variables are represented as logical atoms, and the plates correspond to logical variables.

Independent Choice Logic

- An alternative is a set of ground atomic formulas.
  $\mathcal{C}$, the choice space is a set of disjoint alternatives.
- $\mathcal{F}$, the facts is a logic program that gives consequences of choices.
- P_0 a probability distribution over alternatives:

$$\forall A \in \mathcal{C} \ \sum_{a \in A} P_0(a) = 1.$$

Meaningless Example

$$\mathcal{C} = \{\{c_1, c_2, c_3\}, \{b_1, b_2\}\}$$

$$\begin{aligned} \mathcal{F} = \{\ & f \leftarrow c_1 \wedge b_1, && f \leftarrow c_3 \wedge b_2, \\ & d \leftarrow c_1, && d \leftarrow \sim c_2 \wedge b_1, \\ & e \leftarrow f, && e \leftarrow \sim d\} \end{aligned}$$

$$\begin{array}{lll} P_0(c_1) = 0.5 & P_0(c_2) = 0.3 & P_0(c_3) = 0.2 \\ P_0(b_1) = 0.9 & P_0(b_2) = 0.1 & \end{array}$$

Semantics of ICL

- There is a possible world for each selection of one element from each alternative.
- The logic program together with the selected atoms specifies what is true in each possible world.
- The elements of different alternatives are independent.

Meaningless Example: Semantics

$$\begin{aligned}\mathcal{F} = \{\ & f \leftarrow c_1 \wedge b_1, & & f \leftarrow c_3 \wedge b_2, \\ & d \leftarrow c_1, & & d \leftarrow {\sim}c_2 \wedge b_1, \\ & e \leftarrow f, & & e \leftarrow {\sim}d\}\end{aligned}$$

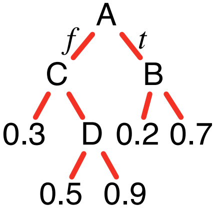
$P_0(c_1) = 0.5 \quad P_0(c_2) = 0.3 \quad P_0(c_3) = 0.2$
$P_0(b_1) = 0.9 \quad P_0(b_2) = 0.1$

$$\begin{array}{llllllll} & & \overbrace{\qquad\qquad}^{\text{selection}} & & \overbrace{\qquad\qquad\qquad}^{\text{logic program}} & & & \\ w_1 & \models & c_1 & b_1 & f & d & e & P(w_1) = 0.45 \\ w_2 & \models & c_2 & b_1 & {\sim}f & {\sim}d & e & P(w_2) = 0.27 \\ w_3 & \models & c_3 & b_1 & {\sim}f & d & {\sim}e & P(w_3) = 0.18 \\ w_4 & \models & c_1 & b_2 & {\sim}f & d & {\sim}e & P(w_4) = 0.05 \\ w_5 & \models & c_2 & b_2 & {\sim}f & {\sim}d & e & P(w_5) = 0.03 \\ w_6 & \models & c_3 & b_2 & f & {\sim}d & e & P(w_6) = 0.02 \end{array}$$

$P(e) = 0.45 + 0.27 + 0.03 + 0.02 = 0.77$

Belief Networks, Decision trees and ICL rules

- There is a local mapping from belief networks into ICL.
- Rules can represent decision tree representation of conditional probabilities:

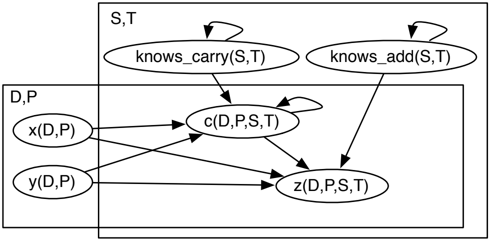


P(e|A,B,C,D)

$$
\begin{array}{ll}
e \leftarrow a \wedge b \wedge h_1. & P_0(h_1) = 0.7 \\
e \leftarrow a \wedge \sim b \wedge h_2. & P_0(h_2) = 0.2 \\
e \leftarrow \sim a \wedge c \wedge d \wedge h_3. & P_0(h_3) = 0.9 \\
e \leftarrow \sim a \wedge c \wedge \sim d \wedge h_4. & P_0(h_4) = 0.5 \\
e \leftarrow \sim a \wedge \sim c \wedge h_5. & P_0(h_5) = 0.3
\end{array}
$$

Example: Multi-digit addition

$$\begin{array}{cccccc} & x_{j_x} & \cdots & x_2 & x_1 \\ + & y_{j_z} & \cdots & y_2 & y_1 \\ \hline & z_{j_z} & \cdots & z_2 & z_1 \end{array}$$



ICL rules for multi-digit addition

$$
\begin{array}{l}
z(D,P,S,T) = V \leftarrow \\
\quad x(D,P) = Vx \land \\
\quad y(D,P) = Vy \land \\
\quad c(D,P,S,T) = Vc \land \\
\quad knows\_add(S,T) \land \\
\quad \neg mistake(D,P,S,T) \land \\
\quad V \text{ is } (Vx + Vy + Vc) \text{ div } 10.
\end{array}
$$

$$
\begin{array}{l}
z(D,P,S,T) = V \leftarrow \\
\quad knows\_add(S,T) \land \\
\quad mistake(D,P,S,T) \land \\
\quad selectDig(D,P,S,T) = V. \\
z(D,P,S,T) = V \leftarrow \\
\quad \neg knows\_add(S,T) \land \\
\quad selectDig(D,P,S,T) = V.
\end{array}
$$

Alternatives:

$\forall DPST \{noMistake(D,P,S,T), mistake(D,P,S,T)\}$

$\forall DPST \{selectDig(D,P,S,T) = V \mid V \in \{0..9\}\}$