CS3243 **AY20/21 Sem1**

Lecture 10: October 28

Lecturer: MEEL, Kuldeep S. *Scribes: Song Qifeng*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

10.1 Recap: Probability

10.1.1 Axioms

1.$0 \leq P(a) \leq 1$

2.$P(a \bigcup b) = P(a) + P(b) - P(a \bigcap b)$

3.$P(True) = 1, P(False) = 0$

10.1.2 Bayes' theorem

$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$

10.1.3 Independence

If $P(a|b) = P(a)$, then a is independent of b.

A special case of independence is if $P(a|b, c) = P(a|b)$, then a, given b, is **conditionally independent of** c.

10.2 Bayesian Networks

10.2.1 What

When drawing a probability table for a event c, if it is dependent on event a and b, then the table will look like the following example:

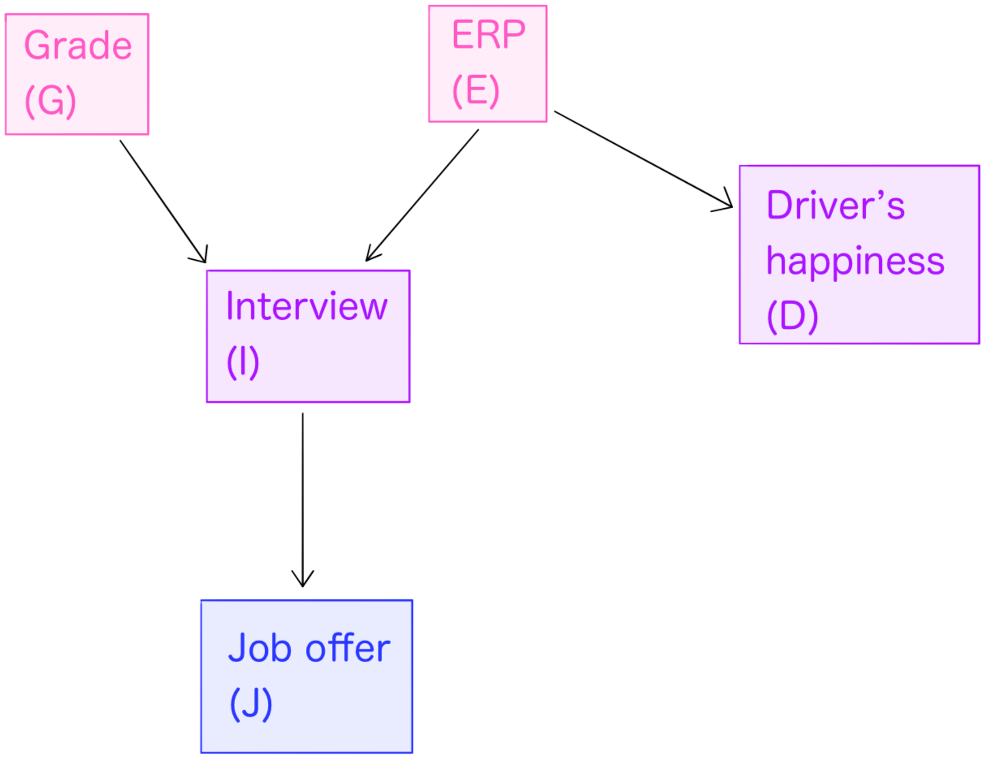
a	b	c
T	T	P(c|a,b)
T	F	P(c|a,!b)
F	T	P(c|!a,b)
F	F	P(c|!a,!b)

Since c is dependent on 2 events, therefore its probability table has 4 rows. For an event that is dependent on k other events, its table will have 2^k rows.

In real world, there are way too many events. When drawing a probability table for a single event, it is impossible to involve all other events because it will result in a table of 2^n where n is a huge number.

Bayesian Networks abstract causal relationships from the real world and form a **directed acyclic graph**, where each directed edge represents a *'ancestor-descendent'* or *'depended-depends'* relationship.

The example mentioned in lecture is attached below:



10.2.2 Axiom

Each node in a Bayesian Network is independent of its non-descendants, given its parents.

10.2.3 How

Based on the axiom, $P(I|G, E, D) = P(I|G, E)$ because D is a non-descendant of I.

Suppose we want to know $P(G, E, I, D, J)$, we can use the axiom smartly to reduce the complexity:

$$\begin{aligned} P(G, E, I, D, J) &= P(J, I, D, G, E) \\ &= P(J|I, D, G, E)P(I|D, G, E)P(D|G, E)P(G|E)P(E) \\ &= P(J|I)P(I|G, E)P(D|E)P(G)P(E) \end{aligned} \tag{10.1}$$

Now, we only need probability tables of 10 rows.

Note that to best exploit the axiom, we need to arrange the events in **topological order**, this can be easily done by recursively picking the leaf node.