Dublin R Workshop on Probabilistic Graphical Models

Mick Cooney
michael.cooney@applied.ai

Nov 19, 2015

`https://bitbucket.org/kaybenleroll/dublin_r_workshops`.

Code is available in the `wspg201511` directory.

Content in this workshop is based on the book Graphical Models with R by Søren Højsgaard.

Also look at the vignettes for the packages `gRain` and `gRbase`

1. Introduction

A graph is a mathematical object that can be defined as a pair $\mathcal{G} = (V, E)$, where V is a set of *vertices* or *nodes*, and E is a set of *edges* that joins two vertices. Edges in general may be directed, undirected or bidirected. They are typically visualised by using shapes or points for the nodes and lines for the edges.

The concept of *conditional independence* is related to that of *statistical independence*. Suppose we have three random variables A, B and C, then A and B are *conditionally independent* given C, writen $A \perp B|C$, iff, for every given value c in C, A and B are independent in the conditional distribution given $C = c$.

Another way of saying this is that for some f a generic density or probability mass function, then one characteristic of $A \perp B|C$ is that

$$f(a, b|c) = f(a|c)f(b|c).$$

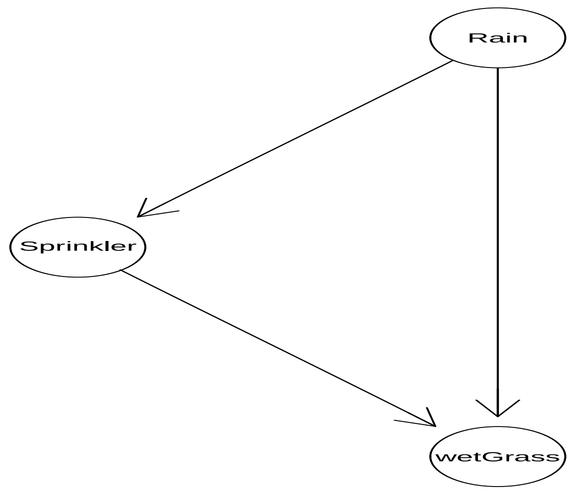
An equivalent characterisation is that the joint density of A, B and C factorises as

$$f(a, b, c) = g(a, c)h(b, c).$$

Finally, we will also make heavy use of Bayes' Rule, the standard formula for relating conditional probabilities:

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}.$$

2. The Sprinkler Network



Two events can cause grass to be wet: Either the sprinkler is on or it is raining. Rain has a direct effect on the use of the sprinkler (namely that when it rains, the sprinkler is usually not turned on).

This can be modeled with a Bayesian network. The variables (R)ain, (S)prinkler, Wet(G)rass have two possible values: (y)es and (n)o.

We can factorise the joint probability mass function as

$$p_{GSR}(g, s, r) = p_{G|SR}(g|s, r)p_{S|R}(s|r)p_R(r)$$

or overloading the notation a little:

$$P(G, S, R) = P(G|S, R)P(S, R) = P(G|S, R)P(S|R)P(R)$$

This means we can construct the joint probability table by starting with the *conditional probability tables* (CPTs).

Exercise 2.1 Create the 3 CPTs using the `parray` function and the following conditional probabilities:

$$\begin{aligned}
P(R) &= 0.2 \\
P(S|R) &= 0.01 & P(S|\neg R) &= 0.4 \\
P(G|S, R) &= 0.99 & P(G|S, \neg R) &= 0.9 & P(G|\neg S, R) &= 0.8 & P(G|\neg S, \neg R) &= 0
\end{aligned}$$

Exercise 2.2 Calculate the full joint probability function $P(G, S, R)$. *HINT:* The function `tabListMult` might be of use.

Exercise 2.3 Calculate the probability that it is raining given that the grass is wet. *HINT:* The functions `tabMarg` and `tabDiv` may be of use.

3. Genetic Inheritance