G53KRR handout on Bayesian networks.

Bayesian approach (subjective probability) The basic idea is that we assign degrees of belief (subjective probabilities) to statements like "Tweety can fly" or "Patient a has disease b". Both sentences are either true or false in the real world, so standard objective statistical probabilities don't apply. However we can base our degree of belief on statistical information: namely, if 95% of birds can fly we may believe with degree 95% that a particular bird Tweety can fly. This will be an a priori degree of belief, before we know anything else about Tweety. Once we discover other facts about Tweety, our belief will be based on the conditional probability that Tweety flies given other facts (that it is a penguin for example).

Probability axioms and useful rules. Given a universal set U of all possible event occurrences (here we assume it is finite), an event a is a subset of U. For example, if U is a large set of medical histories, a is a subset of them where the patient has been diagnosed with flu. A probability function Pr is a function from events to numbers in $[0, 1]$ satisfying the folowing postulates:

1. $Pr(U) = 1$
2. If $a_1, \ldots, a_n$ are disjoint events, $Pr(a_1 \cup \ldots \cup a_n) = Pr(a_1) + \ldots + Pr(a_n)$.

Some consequences:

- $Pr(\bar{a}) = 1 - Pr(a)$ ($\bar{a}$ is the complement of a)
- $Pr(\emptyset) = 0$
- $Pr(a \cup b) = Pr(a) + Pr(b) - Pr(a \cap b)$

Conditional probability:

$$Pr(a \mid b) = \frac{Pr(a \wedge b)}{Pr(b)}$$

Conditionaly independent events: $Pr(a \mid b) = Pr(a)$.
a and b are conditionally independent given c: $Pr(a \mid b \cap c) = Pr(a \mid c)$.
Conditional version of the negation rule $Pr(\bar{a} \mid b) = 1 - Pr(a \mid b)$.

Bayes' rule

$$Pr(a \mid b) = \frac{Pr(a) \cdot Pr(b \mid a)}{Pr(b)}$$

If a is a disease and b a symptom, and we want to know the probability that someone has the disease given they have the symptom; it is easier to find the a priori probability of the disease and what is the probability that a patient who has the disease will display the symptom, and the a priori probability of the symptom.

Probabilities of sentences (basic Bayesian approach) Suppose there are n propositional variables of interest: $p_1, \ldots, p_n$ (corresponding to sentences like "Tweety flies" or "John has flu"). There are 2^n possible states of the world (truth assignments to those variables). J is a *joint probability distribution* if for every assignment I, $J(I)$ is a number between 0 and 1 and $\Sigma\ J(I) = 1$ (the probability that one of the assignments corresponds to reality is 1). The probability of a sentence α is the sum of probabilities of the worlds where α is true:

$$Pr(\alpha) = \Sigma_{I \models \alpha} J(I)$$

We can now find the probability of any sentence. Unfortunately, this requires us to keep 2^n numbers (probability of each assignment).

Let us represent an assignment to $\{p_1, \ldots, p_n\}$ as $\langle P_1, \ldots, P_n \rangle$ where P_i is p_i if p_i is assigned true and $\neg p_i$ otherwise. For example, an assignment which assigns true to p_1 and false to p_2 can be represented as $\langle p_1, \neg p_2 \rangle$.

$$J(\langle P_1, \ldots, P_n \rangle) = Pr(P_1 \wedge \ldots \wedge P_n)$$

By the chain rule (which follows from the definition of conditional probability),

$$Pr(P_1 \wedge \ldots \wedge P_n) = Pr(P_1) \cdot Pr(P_2|P_1) \cdots Pr(P_n|P_1 \wedge \ldots \wedge P_{n-1})$$

If all the variables were conditionally independent of each other:

$$Pr(P_i|P_1 \wedge \ldots \wedge P_{i-1}) = Pr(P_i)$$

then we could compute probabilities of each interpretation from n numbers. But normally we cannot assume that all variables are conditionally independent.

Belief networks The idea is to represent explicitly which variables *are* conditionally dependent on each other. The nodes in the network are variables p_i and there is an arc from p_i to p_j if p_j is conditionally dependent on p_i (its probability given p_i is different from its prior probability).

If there is an arc from p_i to p_j we call p_i a *parent* of p_j in the network.

Each propositional variable in the belief network is conditionally independent from non-parent variables given its parent variables:

$$Pr(P_i \mid P_1 \wedge \ldots \wedge P_{i-1}) = Pr(P_i \mid parents(P_i))$$

where $parents(P_i)$ is the conjunction of literals which correspond to parents of p_i in the network.

Exercise Do exercise 2 after Chapter 12:

Consider the following example: *Metastatic cancer is a possible cause of a brain tumor and is also an explanation for an increased total serum calcium. In turn, either of these could cause a patient to fall into occasional coma. Severe headache could also be explained by a brain tumor.*

(a) Represent these causal links in a belief network. Let a stand for 'metastatic cancer', b for 'increased total serum calcium', c for 'brain tumor', d for 'occasional coma', and e for 'severe headaches'.

(b) Give an example of an independence assumption that is implicit in this network.

(c) Suppose the following probabilities are given: $Pr(a) = 0.2, Pr(b|a) = 0.8, Pr(b|\neg a) = 0.2, Pr(c|a) = 0.2, Pr(c|\neg a) = 0.05, Pr(e|c) = 0.8, Pr(e|\neg c) = 0.6, Pr(d|b \wedge c) = 0.8, Pr(d|b \wedge \neg c) = 0.8, Pr(d|\neg b \wedge c) = 0.8, Pr(d|\neg b \wedge \neg c) = 0.05$ and assume that it is also given that some patient is suffering from severe headaches but has not fallen into a coma. Calculate joint probabilities for the eight remaining possibilities (that is, according to whether a, b, and c are true or false).

(d) According to the numbers given, the a priori probability that the patient has metastatic cancer is 0.2. Given that the patient is suffering from severe headaches but has not fallen into a coma, are we now more or less inclined to believe that the patient has cancer? Explain.