CSC 665: Artificial Intelligence

Lecture: PGMs I

1 The need to model uncertainty

As we've seen before, the choice of representation (e.g. propositional vs. first-order logic) that we use to model a problem domain has a significant impact on the types of inference tasks we're able to perform on that representation, and even our ability to solve the problem in a meaningful way. In this unit of the course, we examine how explicitly integrating uncertainty into our models helps to produce higher-fidelity representations of the world, with important consequences for inference.

1.1 Problems with logic

Example: Suppose you have a flight out of SFO, and you are trying to determine when you should leave for the airport. Let A_t denote the action of leaving t minutes before your flight's departure time. Can propositional logic help us determine the best value of t? If we define the symbol CatchFlight to mean that we are on time for our flight, then we are interested in logical formulas of the form

$$A_{90} \implies \text{CatchFlight}.$$

What is the truth value of this formula? Even if the airport is close enough (say, a 10 minute drive away) such that leaving 90 minutes in advance usually gives you more than enough time to catch your flight, there is no guarantee that things will not go terribly wrong: the road to the airport may be closed, the TSA line may be extremely long, a meteor may strike the runway, etc. What this means is that for any value of t, the formula

$$A_t \implies \text{CatchFlight}$$

must be false. In other words, logic gives us no way to distinguish among the various actions we have available. This is unsatisfying — we want to be able to say that A_{90} is a better solution than, say, A_5 or A_{1440}.

This example shows that for many problems, making strict logical claims and assigning binary truth values to statements is not sufficient to produce a good model of the problem. In many cases, the rational decision depends on

1. the relative importance of the various goals we might have — we can model this with *utility theory*.
2. the likelihood those goals will be achieved if we take a given action — we can model this with *probability theory*.

Together, utility theory and probability theory constitute *decision theory*, which is an approach to making rational decisions (for a certain definition of rationality).

1.2 Uncertainty everywhere

Fundamentally, the reason why logic fails as a representational form for the flight example above is that uncertainty is baked into the problem: leaving earlier makes it more likely, but never guarantees, that we will catch our flight.

Example: Consider the problem of making a diagnosis as a dentist. Once again, propositional logic is ill suited as a representational form for this problem. The formula

$$\text{Toothache} \implies \text{Cavity}$$

is false, because there are things other than cavities that can cause toothaches. Writing down a true propositional formula would require us to enumerate a near endless list of possible causes:

$$\text{Toothache} \implies \text{Cavity} \vee \text{GumDisease} \vee \text{Abscess} \vee \ldots$$

Having seen two examples where uncertainty causes logic to fail, it's worth asking how uncertainty can arise in general. We might classify the sources of uncertainty as follows:

1. **Laziness:** It's too much work to list out all the antecedents or conclusions that would be required to write down a true logical implication (as in the dentist example above).

2. **Theoretical ignorance:** Dental medicine has no complete theory of diagnosis. There are gaps in our knowledge due to the current limits of science.

3. **Practical ignorance:** Even if we had a complete set of diagnostic rules, it may not be practical to collect all of the necessary diagnostic information (e.g. by running tests or ordering labs).

Many (most?) problem domains feature uncertainty in some combination of these types. As a result, we need a framework that allows us to reason quantitatively about uncertainty.

2 Probability: modeling

Probability theory will serve as the tool we use to quantitatively model uncertainty. In the rest of this lecture, we introduce the basic concepts of probability.

Definition: The sample space, denoted Ω, is the set of all possible realizable worlds for a given problem. These worlds must be mutually exclusive and exhaustive for the problem at hand.

Example: Suppose we roll two 6-sided dice. Then a natural choice for the sample space is

$$\Omega = \{(1,1), (1,2), \ldots, (2,1), (2,2), \ldots, (6,5), (6,6)\}.$$

In this case, the size (cardinality) of the sample space is $|\Omega| = 36$.

Definition: An outcome ω is simply an element of the sample space Ω.

Definition: The probability of an outcome ω is denoted $P(\omega)$. Giving a rigorous interpretation for the meaning of probabilities turns out to be philosophically difficult (and even contentious). For now, your natural intuition about probabilities should be mostly correct: the higher the probability of an outcome, the more we believe that the outcome is likely to happen.

Axiom: For every outcome $\omega \in \Omega$, we require

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

Axiom: We require that the probabilities of all possible outcomes sum to one:

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

Definition: An event A is a subset of the sample space Ω.

Example: If we let A denote the event that the sum of of the two dice is 11, then $A = \{(5,6),(6,5)\}$.

Definition: The probability of an event A is obtained by summing the probabilities of the outcomes in A:

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Example: For the event A of rolling an 11, we have

$$P(A) = P((5,6)) + P((6,5)).$$

If the die is fair, then $P((5,6)) = P((6,5)) = 1/36$, so $P(A) = 1/18$.

Definition: The conditional probability of an event A given an event B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Conditional probabilities allow us to modify our beliefs in the presence of new information. In the absence of any relevant data, our beliefs about the event A are captured by the unconditional probability $P(A)$. After observing that event B has taken place, we update our beliefs to the conditional probability $P(A \mid B)$. We do this by ignoring all outcomes that are not in B and renormalizing by dividing by $P(B)$.

Equivalent way to the definition of conditional probability is the *chain rule*:

$$\begin{aligned} P(A \cap B) &= P(A \mid B)P(B) \\ &= P(B \mid A)P(A) \end{aligned}$$

Example: Let Doubles be the event of rolling doubles and $\text{Die}_1 = 5$ be the event of rolling a 5 with the first die. Then

$$P(\text{Doubles} \mid \text{Die}_1 = 5) = \frac{P((5,5))}{P((5,1)) + P((5,2)) + P((5,3)) + P((5,4)) + P((5,5)) + P((5,6))}.$$

If the die is fair, then this probability is 1/6, as you should verify.

Definition: A random variable X is a variable that can take on a set of values according to a set of probabilities. The value of X depends on which outcome $\omega \in \Omega$ is realized.

Example: Let Die_1 be a random variable taking on values in the set $\{1, 2, \ldots, 6\}$. Then $\text{Die}_1 = 5$ for the outcomes $(5,1),(5,2),\ldots,(5,6) \in \Omega$.

Example: Let Sum be a random variable taking on values in the set $\{2, 3, \ldots, 12\}$. Then $\text{Sum} = 11$ for the outcomes $(5,6),(6,5) \in \Omega$.