CS 486 — Lecture 10: Uncertainty and Probability

1 Intro to Uncertainty and Probability

- There is a lot of things that an agent may not be able to observe, or an action that did not have intended consequences.
- So, an agent may need to reason and make a decision based on uncertainty!
- Probability is the formal measure of uncertainty.
- There are two views of probability:
  - *Frequentists* see probability as something objective and observable; we can compute the probabilities by counting the frequencies of events.
  - *Bayesians* see probability as something subjective, and degrees of belief. We start with prior beliefs and update beliefs based on new evidence.
- In this course, we follow the Bayesian view.
- A *random variable* has a domain of possible values and an associated probability distribution which is a function from the domain of random variables to $[0, 1]$.
- Some stuff about probability (let A and B be boolean random variables):
  - $0 \leq P(A) \leq 1$
  - $P(A) = 1$ means it is always true, $P(A) = 0$ means it is always false (duh)
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- A probabilistic model contains a set of RV.
- An atomic event assigns a value to every RV in the model.
- A joint probability distribution assigns a probability to every atomic event.
- We define $P(X)$ as a prior/unconditional probability — the likelihood of X in the absence of any other information, based only on background info.
- We define $P(X|Y)$ as a posterior or conditional probability — the likelihood of X given Y, based on Y as evidence.
- Given a joint prob. distribution, how do we compute:
  - The probability over a subset of the variables?
  - A conditional probability?
- We can compute the probability over a subset of the variables using the *sum rule*.
- So, given a joint distribution over A, B, C, we can calculate $P(A \wedge B)$ by summing out C, as:

$$P(A \wedge B) = P(A \wedge B \wedge C) + P(A \wedge B \wedge \neg C)$$

- Basically fix all values we care about and add all combinations of values we don't care about.
- How about $P(A)$?

$$P(A \wedge B) = P(A \wedge B \wedge C) + P(A \wedge B \wedge \neg C) + P(A \wedge \neg B \wedge C) + P(A \wedge \neg B \wedge \neg C)$$

- Now, how do we calculate $P(A|B)$ given a joint distro. over A, B, C?
- Use:

$$\begin{aligned} P(A|B) &= \frac{P(A \wedge B)}{P(B)} \\ \implies P(A \wedge B) &= P(A|B)P(B) \end{aligned}$$

- Hence, the product rule.

2 Chain and Bayes' Rule

- The chain rule for two variables is $P(A \wedge B) = P(A|B) \times P(B)$.
- For three variables, $P(A \wedge B \wedge C) = P(A|B \wedge C) * P(B|C) * P(C)$.
- For n variables,

$$\begin{aligned} P(X_N \wedge X_{n-1} \cdots X_1) &= \prod_{i=1}^{n} P(X_i | X_{i-1} \wedge \cdots \wedge X_1) \\ &= P(X_n | Xn - 1 \wedge \cdots \wedge X_1) \times \cdots \times P(X_2|X_1) \times P(X_1) \end{aligned}$$

- How do we flip a conditional probability? For example, if you had $P(\text{symptom}|\text{disease})$, how do you get $P(\text{disease}|\text{symptom})$?
- We define the Bayes' rule as $P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)}$