Lecture 2: Uncertainty

Deductive Logic-based Systems

- Automated Theorem Provers
- Rule-based Expert Systems
- Planners and Schedulers
- Robots
- Constraint Satisfaction Systems
- Natural Language Processing

Assumptions Inherent in Deductive Logic-based Systems

- All the assertions we wish to make and use are **universally true**.
- Observations of the world (percepts) are **complete** and **error-free**.
- All conclusions consistent with our knowledge are **equally viable**.
- All the desirable inference rules are **truth-preserving**.

Completely Accurate Assertions

- Initial intuition: if an assertion is not completely accurate, replace it by several more specific assertions.
- **Qualification Problem:** would have to add too many preconditions (or might forget to add some).
- Example: Strep Throat if: sore throat and positive culture *and lab didn't mistakenly switch samples and swab was sterile and* ….

Complete and Error-Free Perception

- Errors are common: biggest problem in use of Pathfinder for diagnosis of lymph system disorder is human error in feature detection.
- Some tests are impossible, too costly, or dangerous. "We could determine if your hip pain is really due to a lower back problem if we cut these nerve connections."

Consistent Conclusions are Equal

- A diagnosis of either early smallpox or cowpox is consistent with our knowledge and observations.
- But cowpox is more likely (e.g., if the sores are on your cow-milking hand).

Truth-Preserving Inference

- Might want to use abductive or inductive inference.
- Even if our inference rules are truth-preserving, if there’s a slight probability of error in our assertions or observations, during chaining (e.g., resolution) these probabilities can compound quickly, *and we are not estimating them*.

Solution: Reason Explicitly About Probabilities

- Full joint distributions.
- Certainty factors attached to rules.
- Dempster-Shafer Theory.
- Qualitative probability and non-monotonic reasoning.
- Possibility theory (within fuzzy logic, which itself does not deal with probability).
- Bayesian Networks.

Start with the Terminology of Most Rule-based Systems

- Atomic proposition: assignment of a value to a variable (parameter).
- Domain of possible values: variables may be Boolean, Discrete (finite domain), or Continuous (floating point in practice).
- Compound assertions can be built with standard logical connectives.

Rule-based Systems (continued)

- State of the world (model, interpretation): a complete setting of all the variables.
- States are mutually exclusive (at most one is actually the case) and collectively exhaustive (at least one must be the case).
- A proposition is equivalent to the set of all states in which it is true; standard compositional semantics of logic applies.

To Add Probability

- Replace variables with **random variables**.
- State of the world (setting of all the random variables) will be called an **atomic event**.
- Apply probabilities or *degrees of belief* to propositions: P(*Weather*=*sunny*) = 0.7.

Prior Probability

- The **unconditional** or **prior probability** of a proposition is the degree of belief accorded to it *in the absence of any other information*.
- Example: P(*Cavity* = *true*) or P(*cavity*)
- Example: P(*Weather* = *sunny*)
- Example: P(*cavity* $\wedge$ (*Weather* = *sunny*))

Distribution of Probability

- Start with an individual random variable.
- Given a random variable X with a set $\mathcal{C}$ of possible values, a *probability set function P* tells how the probability is distributed over various subsets C of $\mathcal{C}$ (of particular type). For any C, P(C) is 0 or greater; P($\mathcal{C}$) is 1; for disjoint subsets $C_1, C_2, \ldots, C_n$ of $\mathcal{C}$: $P(C_1 \cup C_2 \cup C_n) = P(C_1)+P(C_2)+\ldots+P(C_n)$.

Simpler Methods

- For discrete random variables, defining a probability distribution is easier, as here:
  P(*Weather*=*sunny*)=0.7
  P(*Weather*=*rain*)=0.2
  P(*Weather*=*cloudy*)=0.08
  P(*Weather*=*snow*)=0.02
- For the continuous case, a probability density function (p.d.f.) often can be used.

Conditional (Posterior) Probability

- $P(a|b)$: the probability of *a* given that all we know is *b*.
- P(*cavity*|*toothache*) = 0.8: if a patient has a toothache, and no other information is available, the probability that the patient has a cavity is 0.8.
- To be precise: $P = \dfrac{P(a \wedge b)}{P(b)}$

Product Rule

- Equivalent to the previous equation is the following, known as the *product rule*.

$$P(a \wedge b) = P(a|b)\,P(b)$$

$$P(a \wedge b) = P(b|a)\,P(a)$$

The Axioms of Probability

- For any proposition *a*, $0 \leq P(a) \leq 1$
- P(*true*) = 1 and P(*false*) = 0
- The probability of a disjunction is given by

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

Boldface Notation

- Sometimes we want to write an equation that holds for a vector (ordered set) of random variables. We will denote such a set by **boldface** font. So *Y* denotes a random variable, but **Y** denotes a set of random variables.
- *Y* = *y* denotes a setting for *Y*, but **Y=y** denotes a setting for all variables in **Y**.

P Notation

- P(X = x) denotes the probability that the random variable X takes the value x; **P**(X) denotes a probability distribution over X.
- For example, if we want to say that observing a value for Y does not change the probability distribution over X, we can write **P**(X|Y) = **P**(X) rather than repeating for every combination of settings for X and Y.

Marginalization & Conditioning

- Marginalization (summing out): for any sets of variables Y and Z: $\mathbf{P}(\mathrm{Y}) = \sum_{\mathrm{z} \in \mathrm{Z}} \mathbf{P}(\mathrm{Y}, \mathrm{z})$
- Conditioning(variant of marginalization):

$$\mathbf{P}(\mathrm{Y}) = \sum_{\mathrm{z} \in \mathrm{Z}} \mathbf{P}(\mathrm{Y}|\mathrm{z})\, \mathbf{P}(\mathrm{z})$$

Often want to do this for $\mathbf{P}(Y|X)$ instead of $\mathbf{P}(\mathrm{Y})$.

Recall $\mathbf{P}(Y|X) = \dfrac{\mathbf{P}(X \wedge Y)}{\mathbf{P}(X)}$

Example

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Related Example

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

- In the two preceding examples the denominator (P(*toothache*)) was the same, and we looked at all possible values for the variable *Cavity* given *toothache*.
- The denominator can be viewed as a normalization constant α.
- We don't have to compute the denominator -- just normalize 0.12 and 0.08 to sum to 1.

General Inference Procedure

- Let *X* be a random variable about which we want to know its probabilities, given some evidence (values **e** for a set **E** of other variables). Let the remaining (unobserved) variables be **Y**. The query is **P**(X|**e**), and it can be answered by

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X,\mathbf{e}) = \alpha\sum_{\mathrm{y}}\mathbf{P}(X,\mathbf{e},\mathbf{y})$$