Artificial Intelligence
15-381

Mar 22, 2007

Probability and Uncertainty 2: Probabilistic Reasoning

Review of concepts from last lecture

Making rational decisions when faced with uncertainty:

- *Probability*

  the precise representation of knowledge and uncertainty
- *Probability theory*

  how to optimally update your knowledge based on new information
- *Decision theory: probability theory + utility theory*

  how to use this information to achieve maximum expected utility

Basic concepts

- random variables
- probability distributions (discrete) and probability densities (continuous)
- rules of probability
- expectation and the computation of 1st and 2nd moments
- joint and multivariate probability distributions and densities
- covariance and principal components

Simple example: medical test results

- Test report for rare disease is positive, 90% accurate
- What's the probability that you have the disease?
- What if the test is repeated?

- This is the simplest example of reasoning by combining sources of information.

How do we model the problem?

- Which is the correct description of "Test is 90% accurate" ?

$$\begin{aligned} P(T = \text{true}) &= 0.9 \\ P(T = \text{true}|D = \text{true}) &= 0.9 \\ P(D = \text{true}|T = \text{true}) &= 0.9 \end{aligned}$$

- What do we want to know?

$$\begin{gathered} P(T = \text{true}) \\ P(T = \text{true}|D = \text{true}) \\ P(D = \text{true}|T = \text{true}) \end{gathered}$$

- More compact notation:

$$\begin{aligned} P(T = \text{true}|D = \text{true}) &\rightarrow P(T|D) \\ P(T = \text{false}|D = \text{false}) &\rightarrow P(\bar{T}|\bar{D}) \end{aligned}$$

Evaluating the posterior probability through Bayesian inference

- We want P(D|T) = "The probability of the having the disease given a positive test"
- Use Bayes rule to relate it to what we know: P(T|D)

$$\text{posterior}\quad P(D|T) = \frac{\overbrace{P(T|D)}^{\text{likelihood}}\overbrace{P(D)}^{\text{prior}}}{\underbrace{P(T)}_{\text{normalizing constant}}}$$

- What's the prior P(D)?
- Disease is rare, so let's assume

$$P(D) = 0.001$$

- What about P(T)?
- What's the interpretation of that?

Evaluating the normalizing constant

$$\text{posterior}\quad P(D|T) = \frac{\overbrace{P(T|D)}^{\text{likelihood}}\overbrace{P(D)}^{\text{prior}}}{\underbrace{P(T)}_{\text{normalizing constant}}}$$

- P(T) is the marginal probability of P(T,D) = P(T|D) P(D)
- So, compute with summation

$$P(T) = \sum_{\text{all values of D}} P(T|D)P(D)$$

- For true or false propositions:

$$P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D})$$

What are these?