CSC384
Introduction to Artificial Intelligence: Uncertainty

November 6, 2014

November 6th

- This material will cover chapters 13 and 14 of the textbook
- The remainder of the course will cover chapters 16 and 17 of the textbook
- Today we review probability theory (Chapter 13) and work into new material (Chapter 14)

Uncertainty

Probablility Review

Until now we have primarily dealt with deterministic worlds

- Many problems feature nondeterminism, unknowns in knowledge or outcomes
- In probabilistic outcomes we do not what will happen, but we know how likely the outcomes are
- By basing are actions on the likelihood of outcomes we can “gamble” in an optimal way

We assign probabilities to events in a universe U of events.

- $P(U) = 1$
- $\forall a \in U, P(a) \in [0, 1]$
- For a set of events F
$$P(F) = \sum_{a \in F} P(a)$$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Uncertainty

Feature Vectors

If we have variables and domains:

- Set of variables $V_1, ..., V_n$
- Domains of the variables $D_1, ..., D_n$

Events can be interpretted as a vector of values

$$< v_1, ..., v_n >$$

With $v_i \in D_i \forall i$.

The universe U is the collection of all such vectors.

Uncertainty

Feature Vectors

Problem: if we have n different variables, how many different feature vectors can we have?

Can we feasibly track this many events?

Uncertainty

Conditional Probability

To reduce computation, we can use conditional probability and independence.

$P(A|B)$: The probability that events A occur given that events B have occurred

$p(A \cap B) = P(B|A)P(A)$

Uncertainty

Chain Rule

If $P(A) \neq 0$, $p(A \cap B) = P(B|A)P(A)$

For notational convience we can write $P(A \cap B)$ as $P(AB)$

Using what we have so far we can derive the chain rule:

$$P(A_1 A_2 ... A_n) = P(A_1|A_2...A_n) * P(A_2|A_3...A_n) * \\ P(A_3|A_4...A_n) * ... * P(A_{n-1}|A_n) * P(A_n)$$

Uncertainty

Independence

We say that events A and B are independent if $P(A|B) = P(A)$

If all of our variables are independent, how many values must we know in order to compute the value of any given feature vector? (Consider the chain rule)

In practice all of the variables will not be independent

Uncertainty

Bayes Theorem

Given:

$$P(AB) = P(B|A)P(A)$$
$$P(AB) = P(A|B)P(B)$$

We can derive the following:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

This is known as Baye's Theorem.

Uncertainty

Terminology

Some terminology:

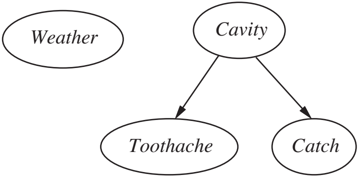
- $P(V_1)$: the marginal distribution over V_1
- $P(V_1, V_2)$: a joint probability distribution
- $P(V_3 = 4)$: a *prior* or unconditional probability
- $P(V_3 = 2 | V_6 = 1)$: a *posterior* or conditional probability
- $P(V_1, V_2)$: a joint probability distribution

A Bayesian network consists of the following:

- A directed acyclic graph in which the nodes are variables and the edges indicate conditional **dependence**
- A conditional probability table for each variable

Uncertainty

Bayesian Networks



Uncertainty

Bayesian Networks