For Wednesday

- Read chapter 14, sections 1-3
- Wumpus program due

Program 2

- Any questions?

Uncertainty

- Everyday reasoning and decision making is based on uncertain evidence and inferences.
- Classical logic only allows conclusions to be strictly true or strictly false
- We need to account for this uncertainty and the need to weigh and combine conflicting evidence.

Coping with Uncertainty

- Straightforward application of probability theory is impractical since the large number of conditional probabilities required are rarely, if ever, available.
- Therefore, early expert systems employed fairly ad hoc methods for reasoning under uncertainty and for combining evidence.
- Recently, methods more rigorously founded in probability theory that attempt to decrease the amount of conditional probabilities required have flourished.

Probability

- Probabilities are real numbers 0-1 representing the a priori likelihood that a proposition is true.

  $P(Cold) = 0.1$

  $P(\neg Cold) = 0.9$

- Probabilities can also be assigned to all values of a random variable (continuous or discrete) with a specific range of values (domain), e.g. low, normal, high.

  $P(temperature=normal)=0.99$

  $P(temperature=98.6) = 0.99$

Probability Vectors

- The vector form gives probabilities for all values of a discrete variable, or its probability distribution.

  P(temperature) = <0.002, 0.99, 0.008>

- This indicates the prior probability, in which no information is known.

Conditional Probability

- Conditional probability specifies the probability given that the values of some other random variables are known.

  $P(Sneeze \mid Cold) = 0.8$

  $P(Cold \mid Sneeze) = 0.6$

- The probability of a sneeze given a cold is 80%.
- The probability of a cold given a sneeze is 60%.

Cond. Probability cont.

- Assumes that the given information is all that is known, so all known information must be given.

  $P(\text{Sneeze} \mid \text{Cold} \wedge \text{Allergy}) = 0.95$

- Also allows for conditional distributions

$P(X \mid Y)$ gives 2-D array of values for all $P(X=x_i \mid Y=y_j)$

- Defined as

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

Axioms of Probability Theory

- All probabilities are between 0 and 1.

$$0 \leq P(A) \leq 1$$

- Necessarily true propositions have probability 1, necessarily false have probability 0.

$$P(true) = 1 \qquad\qquad P(false) = 0$$

- The probability of a disjunction is given by

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

Joint Probability Distribution

- The joint probability distribution for a set of random variables $X_1 \ldots X_n$ gives the probability of every combination of values (an n-dimensional array with vn values if each variable has v values)

  $P(X_1,...,X_n)$

	Sneeze	¬Sneeze
Cold	0.08	0.01
¬Cold	0.01	0.9

- The probability of all possible cases (assignments of values to some subset of variables) can be calculated by summing the appropriate subset of values from the joint distribution.
- All conditional probabilities can therefore also be calculated

Bayes Theorem

$$P(H \mid e) = \frac{P(e \mid H)\, P(H)}{P(e)}$$

- Follows from definition of conditional probability:

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

Other Basic Theorems

- If events A and B are independent then:

$$P(A \wedge B) = P(A)P(B)$$

- If events A and B are incompatible then:

$$P(A \vee B) = P(A) + P(B)$$

Simple Bayesian Reasoning

- If we assume there are n possible disjoint diagnoses, $d_1 \ldots d_n$

$$P(d_i \mid e) = \frac{P(e \mid d_i)\, P(d_i)}{P(e)}$$

- P(e) may not be known but the total probability of all diagnoses must always be 1, so all must sum to 1
- Thus, we can determine the most probable without knowing P(e).

Efficiency

- This method requires that for each disease the probability it will cause any possible combination of symptoms and the number of possible symptom sets, e, is exponential in the number of basic symptoms.
- This huge amount of data is usually not available.

Bayesian Reasoning with Independence ("Naïve" Bayes)

- If we assume that each piece of evidence (symptom) is independent given the diagnosis (conditional independence), then given evidence e as a sequence $\{e_1, e_2, \ldots, e_d\}$ of observations, $P(e \mid d_i)$ is the product of the probabilities of the observations given d_i.
- The conditional probability of each individual symptom for each possible diagnosis can then be computed from a set of data or estimated by the expert.
- However, symptoms are usually not independent and frequently correlate, in which case the assumptions of this simple model are violated and it is not guaranteed to give reasonable results.

Bayes Independence Example

- Imagine there are diagnoses ALLERGY, COLD, and WELL and symptoms SNEEZE, COUGH, and FEVER

Prob	Well	Cold	Allergy
P(d)	0.9	0.05	0.05
P(sneeze|d)	0.1	0.9	0.9
P(cough | d)	0.1	0.8	0.7
P(fever | d)	0.01	0.7	0.4

- If symptoms sneeze & cough & no fever:

P(well | e) = (0.9)(0.1)(0.1)(0.99)/P(e) = 0.0089/P(e)

P(cold | e) = (.05)(0.9)(0.8)(0.3)/P(e) = 0.01/P(e)

P(allergy | e) = (.05)(0.9)(0.7)(0.6)/P(e) = 0.019/P(e)

- Diagnosis: allergy

P(e) = .0089 + .01 + .019 = .0379

P(well | e) = .23

P(cold | e) = .26

P(allergy | e) = .50

Problems with Probabilistic Reasoning

- If no assumptions of independence are made, then an exponential number of parameters is needed for sound probabilistic reasoning.
- There is almost never enough data or patience to reliably estimate so many very specific parameters.
- If a blanket assumption of conditional independence is made, efficient probabilistic reasoning is possible, but such a strong assumption is rarely warranted.

Practical Naïve Bayes

- We’re going to assume independence, so what numbers do we need?
- Where do the numbers come from?