Course 440 : Introduction To Artificial Intelligence
Lecture 7

Probabilistic Reasoning

Abdeslam Boularias

Wednesday, October 26, 2016

RUTGERS

THE STATE UNIVERSITY
OF NEW JERSEY

Outline

We show how to reason and act under uncertainty.

1. Decision-making under uncertainty
2. Basic probability notations
3. Inference using joint probability distributions
4. Independence
5. Bayes' rule

Why do we need to reason with probabilities ?

Planning a trip to the airport

- An automated taxi has the goal of delivering a passenger to the airport on time. The airport is only 10 miles away.
- Suppose the taxi can choose between two plans, A_{90} and A_{180}
  - A_{90} : Leave home 90 minutes before the flight departs.
  - A_{180} : Leave home 180 minutes before the flight departs.
- Which plan should the agent choose ?
- Plan A_{90} will get the passenger at the airport in time, but there is a significant risk of missing the flight due to traffic jam, accidents, car problems, etc.
- Plan A_{180} has a better chance of succeeding, but will result in a longer waiting time.

Why do we need to reason with probabilities ?

Planning a trip to the airport

- To solve this search problem, we need to define an objective function : a high cost for missing the flight (such as -1000), and a low cost for waiting (such as -1 per minute).
- We also need, in advance, to know all what will happen during the trip : accidents, car problems, etc.

Why do we need to reason with probabilities ?

Planning a trip to the airport

- To solve this search problem, we need to define an objective function : a high cost for missing the flight (such as -1000), and a low cost for waiting (such as -1 per minute).
- We also need, in advance, to know all what will happen during the trip : accidents, car problems, etc.
- It is almost impossible to include all these details for two main reasons :
  - **Complexity** : We may be able to predict the state of traffic, but we will need to track every car on the road. We could also predict car problems, but we will need to include the state of each piece of the car in our search problem. In general, precise answers require a lot of details about the world and increase the complexity of the problem.
  - **Ignorance** : predicting accidents requires mental models of all the drivers on the road, which is simply something we do not have. Therefore, we cannot know for sure if there will be an accident.

Why do we need to reason with probabilities ?

Planning a trip to the airport

- To solve this type of problems, we summarize information using probabilities.
- Instead of tracking cars to predict the state of traffic, we look at how was traffic in each particular segment of the road at the same time of the day in the previous days. We know that it should roughly be the same.
- We can also use historical data on accidents to know how often accidents happen.
- Degrees of beliefs are represented by probabilities.
- Philosophical question : uncertainty is the result of our limited perception and reasoning capabilities, are there phenomena that are inherently uncertain (e.g. quantum mechanics) ?

Interpreting probabilities

- Example : tossing a fair coin comes up heads with probability 0.5
- **Frequentist interpretation** : the 0.5 probability is a physical property of the coin and tossing operation, it comes from the fact that if we toss the coin n times, it comes up with heads $\frac{n}{2}$ times when $n \to \infty$
- **Bayesian interpretation** : the 0.5 is our prior belief on how often the coin will come up heads. It is our subjective understanding based on the evidence we have seen so far. The 0.5 may change based on new evidence.

The frequentist interpretation has the advantage of being objective (detached from prior beliefs) and experimentally driven. However, there are certain probabilistic statements that do not have a frequentist interpretation (e.g. probability of an event that never occurred before).

Basic probability notations

- Probabilities are assertions about possible worlds.
- The set of all possible worlds is called the sample space.
- The possible worlds are mutually exclusive and exhaustive.
- A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world ω, such that :

$$\forall \omega \in \Omega : 0 \leq P(\omega) \leq 1 \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1.$$

Basic probability notations

- Probabilistic queries are not usually about particular possible worlds, but about sets of them.
- Example : We throw two dice and we wonder what the probability of getting the sum of the dice equal to 11.

$$P(Total = 11) = P((5,6)) + P((6,5)) = 1/36 + 1/36 = 1/18$$

- In general, if ϕ is a proposition, then

$$P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Inference using joint probability distributions

Probabilistic inference : computation of posterior probabilities for query propositions given observed evidence.

Simple example : a domain consisting of just the three Boolean variables Toothache, Cavity, and Catch.

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

A full joint distribution for the Toothache, Cavity, Catch world

- From the joint distribution, we can calculate the probability of any proposition.
- Example : P(*cavity* ∨ *toothache*) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28

Inference using joint probability distributions

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

A full joint distribution for the Toothache, Cavity, Catch world

- From the joint distribution, we can calculate the probability of any proposition.
- Example : P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2 .
- This process is called **marginalization**, or **summing out**, it eliminates hidden variables
- In general,

$$P(X) = \sum_{Y \in \mathcal{Y}} P(X, Y),$$

where $\sum_{Y \in \mathcal{Y}}$ means to sum over all the possible combinations of values of the set of variables Y.

Inference using joint probability distributions

Marginalization

- Example :

$$P(Cavity) = \sum_{Y \in \{(Catch, Toothache)\}} P(Cavity, Y),$$

- Another variant of marginalization is called **conditioning** :

$$P(X) = \sum_{Y \in \mathcal{Y}} P(X|Y)P(Y).$$

Independence

- Example : Let us add a variable "*Weather*" to the previous example. We consider four states of weather.

- The full joint distribution then becomes P(*Toothache,Catch,Cavity,Weather*), which has now $2 \times 2 \times 2 \times 4 = 32$ entries.

- We have P(toothache, catch, cavity, cloudy) = P(cloudy | toothache, catch, cavity) P(toothache, catch, cavity) .

- We know that toothaches, cavities, and dentistry have nothing to do with the weather.

- Therefore, P(cloudy | toothache, catch, cavity) = P(cloudy).

- We can write,
  P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity) .

- Therefore, we only need $8 + 4 = 12$ entries to represent the full distribution.

Independence

- We say that variables X and Y are independent if :
  $P(X \mid Y) = P(X)$ or $P(Y \mid X) = P(Y)$ or $P(X \wedge Y) = P(X)P(Y)$.
- Independence assertions are usually based on knowledge of the domain.
- Full joint distribution can be factored into separate joint distributions on independent subsets.
- For example, the full joint distribution on the outcome of n independent coin flips, $P(C_1, ..., C_n)$, has 2^n entries, but it can be represented as the product of n single-variable distributions $P(C_i)$.
- Independence is always good news in probabilistic inference, it leads to more efficient computation.
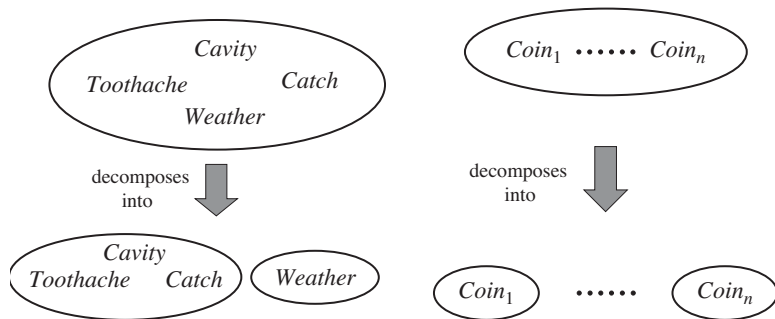
Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. Weather and dental problems are independent. Coin flips are independent.

Bayes' rule

- From the product rule, we have $P(X \wedge Y) = P(X \mid Y)P(Y)$ and $P(X \wedge Y) = P(Y \mid X)P(X)$, then

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}.$$

- This simple equation, known as Bayes' rule, underlies most modern AI systems for probabilistic inference.
- Bayes' rule is used to update an agent's belief on the basis of new evidence

$$P(\text{Hypothesis} \mid \text{Evidence}) = P(\text{Hypothesis})\frac{P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})}$$

Bayes' rule

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}.$$

- $P(X)$ is called the prior probability, $P(X \mid Y)$ is the posterior probability, $P(Y \mid X)$ is the likelihood, and $P(Y)$ is the marginal probability.
- Notice that $P(Y) = \sum_{X_i} P(Y \mid X_i)P(X_i)$ (summing out X).
- Therefore,

$$\begin{aligned} P(X \mid Y) &= \frac{P(Y \mid X)P(X)}{P(Y)} = \frac{P(Y \mid X)P(X)}{\sum_{X_i} P(Y \mid X_i)P(X_i)} \\ &= \alpha P(Y \mid X)P(X) \end{aligned}$$

where α is known as the normalization constant.
- We can also write $P(X \mid Y) \propto P(Y \mid X)P(X)$