CS 561 Artificial Intelligence
Lecture # 2 Probability Review

Rashmi Dutta Baruah

Department of Computer Science & Engineering

IIT Guwahati

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी

Indian Institute of Technology Guwahati

Guwahati - 781039, INDIA

Basic Probability Terminologies

- **Probability :** a measure of belief (as opposed to being a frequency) – Bayesian Probability or Subjective Probability

- Example: Suppose there are three agents, Alice, Bob and Chris, and one die has been tossed.
  - Alice observes that the outcome is a "6" and tells Bob that the outcome is even but Chris knows nothing about the outcome
  - Alice has probability (degree of belief) about the outcome of the toss to be "6" is "1", Bob has probability of "1/3" (Bob believes Alice and considers that all events outcome are equally likely), Chris has probability "1/6".

Basic Probability Terminologies

- Axioms of Probability
  - Suppose P is a function from propositions (or events) into real numbers that satisfies the following three axioms of probability:
  - Axiom 1: $0 \leq P(\alpha)$ for any proposition α i.e. the belief in any proposition cannot be negative.
  - Axiom 2: $P(\tau) = 1$ if τ is a tautology i.e. if τ is true in all possible worlds, its probability is 1.
  - Axiom 3: $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$ if α and β are contradictory propositions i.e. if $\sim(\alpha \wedge \beta)$ is a tautology. If two propositions can not both be true (they are mutually exclusive) then the probability of their disjunction is the sum of their probabilities.

Basic Probability Terminologies

- Conditional Probability
- We do not only want to know the prior probability of some proposition, but we want to know how this belief is updated when an agent observes new evidence.
- The **unconditional or prior probability** refers to the degree of belief in proposition in the absence of any other information (or when agent has not observed anything).
- The measure of belief in proposition α based on proposition β is called **conditional (or posterior)** probability of α given β and written as $P(\alpha|\beta)$. β is also referred to as **evidence**.
- For an agent the conjunction of all his observations of the world is evidence.
- When taking decision an agent has to condition on **all** the evidence it has observed.

Basic Probability Terminologies

- Conditional Probability
- Conditional probabilities are defined in terms of unconditional probabilities
  - $P(\alpha|\beta) = \frac{P(\alpha \wedge \beta)}{P(\beta)}$ which holds whenever $P(\beta) > 0$
  - It can also be written in the form of **product rule**
  - $P(\alpha \wedge \beta) = P(\alpha|\beta)\ P(\beta)$ for α and β to be true , we need β to be true and also need α to be true given β.

Basic Probability Terminologies

- Joint Probability Distribution
  - Random variable (neither random nor variable) : function from some discrete domain (in our case) → [0,1]
  - Domain of random variable is the set of all possible values that it can take on.
  - Example: random variable **Weather** and its domain is { **sunny, rain, cloudy, snow**}
  - $Weather(sunny) = 0.2$, this is usually written as
    - $P(Weather = sunny) = 0.2$ or $P(sunny) = 0.2$
  - **Probability Distribution:** probability assignment of all possible values for a random variable

Basic Probability Terminologies

- Joint Probability Distribution
  - **Probability Distribution:** probability assignment of all possible values for a random variable

    $P(Weather = sunny) = 0.6$

    $P(Weather = rain) = 0.1$

    $P(Weather = cloudy) = 0.29$

    $P(Weather = snow) = 0.01$

    $\mathbf{P}(Weather) = <0.6, 0.1, 0.29, 0.01>$, **P** indicates that the result is a vector of numbers and defines a probability distribution
  - For multiple random variables, we use the term **Joint Probability Distribution:** probability assignment to all combinations of the values of the random variables

Basic Probability Terminologies

- Joint Probability Distribution
  - Example: **Dentistry domain** with two propositional random variables : ***Cavity*** and ***Toothache***
  - $Cavity = \{cavity, \sim cavity\}$: Does the patient have a cavity or not?
  - $Toothache = \{toothache, \sim toothache\}$: Does the patient have toothache?

	toothache	*~toothache*
cavity	0.04	0.06
~cavity	0.01	0.89

- Inference with Joint Probability Distribution: a simple method of probabilistic inference
  - Joint probability distribution table can be used as a "knowledge base" for answering any query
  - Example: What is the probability of cavities?
  - Answer: The probability of cavities is 0.1 (add elements of *cavity* row i.e. cavities with and without toothache) obtained by Marginalization (or summing out)
  - $\boldsymbol{P(Y) = \sum_{z \in Z} P(Y, z)}$

Basic Probability Terminologies

	toothache	*~toothache*
cavity	0.04	0.06
~cavity	0.01	0.89

- Inference with Joint Probability Distribution
  - Example: What is the probability that if a patient comes with a toothache has a cavity?
  - Answer: $P(cavity \mid toothache) = P(cavity \wedge toothache) / P(toothache) = 0.04/0.05 = 0.8$
  - So, the probability that someone has cavity is 0.1 but if we know that he/she has toothache then the probability increases to 0.8.

Problem: the size of the table

Bayes' Rule and its Use

- Product rule:
  - $P(a \wedge b) = P(a|b)\ P(b)$, it can also be written as
  - $P(a \wedge b) = P(b|a)\ P(a)$
- Equating the two RHS and dividing by $P(a)$, we get the equation known as Bayes' Rule:
  - $P(b|a) = P(a|b)P(b)\ /\ P(a)$
- Bayes' Rule is useful for assessing diagnostic probability from causal probability

  cause effect

  - $P(disease|symptom) = P(symptom|disease)\ P\ (disease)\ /\ P(symptom)$
  - Example: *disease* = measles , *symptom* = high fever
    - $P(disease|symptom)$ may be different in India vs US : diagnostic direction
    - $P(symptom\ |\ disease)$ should be same : causal direction
    - So, it is more useful to learn relationships in causal direction and use it to compute the diagnostic probabilities.

Bayes' Rule and its Use

- Conditioning (can be used to determine *P(symptom)*)
  - $P(a) = P(a \wedge b) + P(a \wedge {\sim}b) = P(a \mid b)\, P(b) + P(a \mid {\sim}b) P({\sim}b)$
- Independence
  - $P(a \wedge b) = P(a).P(b)$
  - $P(a \mid b) = P(a)$: knowing that b is true does not give us any more information about the truth of a
  - $P(b \mid a) = P(b)$
- Independence is helpful in efficient probabilistic reasoning.
- Conditional Independence
  - a and b are conditionally independent given c
  - $P(a \mid b, c) = P(a \mid c)$
  - $P(b \mid a,c) = P(b \mid c)$
  - $P(a \wedge b \mid c) = P(a \mid c).\, P(b \mid c)$

Using Bayes' Rule: combining evidence

- Example: Dentistry domain
  - *Toothache*
  - *Cavity*
  - *Xray-Spot*
- given variables are not independent but *toothache* and *xrayspot* are conditionally independent given *cavity*
- **Combining evidence**
  - $\mathbf{P}(Cavity \mid toothache, xrayspot) = \dfrac{\mathbf{P}(toothache, xrayspot \mid Cavity)\ \mathbf{P}(Cavity)}{P(toothache, xrayspot)}$
  - Assuming that *xrayspot* and *toothache* are conditionally independent
  - $\mathbf{P}(Cavity \mid toothache, xrayspot) =$
    $$\frac{\mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(xrayspot \mid Cavity)\ \mathbf{P}(Cavity)}{P(toothache, xrayspot)}$$

Using Bayes' Rule: combining evidence

- **Combining evidence**
  - $\mathbf{P}(Cavity \mid toothache , xrayspot) = \dfrac{\mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(xrayspot \mid Cavity)\ \ \mathbf{P}(Cavity)}{P(toothache,\ xrayspot)}$
  - We started with prior probability of someone having cavity and as the evidence arrives (xrayspot) we multiplied a factor to and so on as new evidences arrive.
- Normalizing constant ($P(toothache, xrayspot)$)

$$P(cavity \mid toothache , xrayspot) + P(\sim cavity \mid toothache , xrayspot) = 1$$

$$\frac{P(toothache \mid cavity)\ P(xrayspot \mid cavity)\ \ P(cavity)}{P(toothache,\ xrayspot)} + \frac{P(toothache \mid \sim cavity)\ P(xrayspot \mid \sim cavity)\ \ P(\sim cavity)}{P(toothache,\ xrayspot)} = 1$$

$$P(toothache \mid cavity)\ P(xrayspot \mid cavity)\ \ P(cavity) + P(toothache \mid \sim cavity)\ P(xrayspot \mid \sim cavity)\ \ P(\sim cavity) = P(toothache,\ xrayspot)$$

Conditional Independence

- Given a cause (disease) that influences a number of effects (symptoms), which are conditionally independent, the full joint distribution can be written as

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\ \Pi_i\ \mathbf{P}(Effect_i \mid Cause)$$

  - this is called the *naïve Bayes model*
    - it makes the simplifying assumption that *all* effects are conditionally independent
    - it is naïve in that it is applied to many problems although the effect variables are not precisely conditionally independent given the cause variable
    - nevertheless, such systems often work well in practice