

CS 480 – INTRODUCTION TO ARTIFICIAL INTELLIGENCE

TOPIC: UNCERTAINTY
CHAPTER: 12



Mustafa Bilgic



<http://www.cs.iit.edu/~mbilgic>



<https://twitter.com/bilgicm>

MEDICAL DIAGNOSIS - TOOTHACHE

- Toothache \Rightarrow Cavity
 - What's wrong?
 - Not all toothaches are due to cavity
- Toothache \Rightarrow Cavity \vee GumProblem \vee Abscess \vee ...
 - What's wrong?
 - We have to add almost an unlimited number of possible problems
- Cavity \Rightarrow Toothache
 - What's wrong?
 - Not all cavities cause toothache

LOGIC

○ Fails for three main reasons

1. **Laziness**

- Too much work to list all premises and conclusions
- Too hard to use such rules

2. **Theoretical ignorance**

- No complete theory for the domain

3. **Practical ignorance**

- Even if we knew all the rules, some information might be missing; e.g., lab tests for diagnosis

The butterfly effect

SOME SOURCES OF UNCERTAINTY

○ **Uncertainty in knowledge**

- E.g., We do not know all the causes of all the diseases

○ **Uncertainty in actions;** we cannot list all the pre-conditions of actions

- E.g., To be able to fly a plane from SFO to JFK, it must not be broken, the weather conditions have to be appropriate, the pilot must not be sick, you need to have enough fuel, ...

○ **Uncertainty in sensors**

- E.g., lightning conditions for a camera might not be enough

TASKS

1. Representation

- What is the formal and appropriate language to represent uncertainty?

2. Inference

Tell

- How can we infer uncertainty before or after we gather more information?

3. Decision making

Act

- How can a rational agent act in an uncertain world?

PROBABILITY MODEL

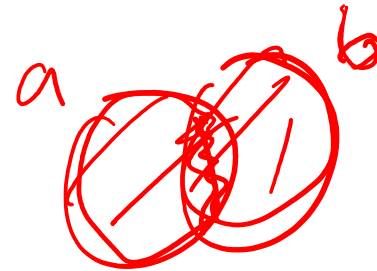
- It's all about the state the world is in, i.e., a possible world
- A possible world is an assignment of truth values to the predicates
 - Logical assertions rule out some of the possible worlds
 - E.g., $\text{cavity} \Rightarrow \text{toothache}$ rules out the worlds where $\text{cavity}=\text{true} \wedge \text{toothache}=\text{false}$
 - Probabilistic reasoning determines how probable the various worlds are
 - E.g., a world where $\text{cavity}=\text{true} \wedge \text{toothache}=\text{true}$ is more probable than the world where $\text{cavity}=\text{true} \wedge \text{toothache}=\text{false}$
- The set of all possible worlds is called the sample space
- The possible worlds are *mutually exclusive* and *exhaustive*
 - *Mutually exclusive*: the world can be only in one state
 - *Exhaustive*: the world has to be in one of the states

PROBABILITY MODEL

- A **probability model** associates a numerical probability $P(w)$ with each possible world w
 - $P(w)$ sums to 1 over all possible worlds
- An **event** is the set of possible worlds where a given predicate is true
 - Roll two dice; the possible worlds are (1,1), (1,2), ..., (6,6); 36 possible worlds
 - Predicate = two dice sum to 10
 - Event = {(4,6), (5,5), (6,4)}

AXIOMS OF PROBABILITY

1. The probability $P(a)$ of a proposition a is a real number between 0 and 1
2. $P(\text{true}) = 1$, $P(\text{false}) = 0$
3. $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



$P(\neg a)$

- $P(a \vee \neg a) = P(a) + P(\neg a) - P(a \wedge \neg a)$
- $P(\text{true}) = P(a) + P(\neg a) - P(\text{false})$
- $1 = P(a) + P(\neg a) - 0$
- $P(\neg a) = 1 - P(a)$
- Intuitive explanation:
 - The probability of all possible worlds is 1
 - Either a or $\neg a$ holds in one world
 - The worlds that a holds and the worlds that $\neg a$ holds are mutually exclusive and exhaustive

RANDOM VARIABLE

- Like CSP
 - A factored representation of the world; random variables
 - Each variable has a domain
 - Probabilities over the domain values of a variable sum to 1
 - The possible worlds where a random variable takes a certain value are mutually exclusive and exhaustive (from the viewpoint of that variable)
- E.g.
 - $D_1: \{1, 2, 3, 4, 5, 6\}$

JOINT DISTRIBUTION

- We have n random variables, V_1, V_2, \dots, V_n
- We are interested in the probability of a possible world, where
 - $V_1=v_1, V_2=v_2, \dots, V_n=v_n$
- $P(V_1, V_2, \dots, V_n)$ associates a probability for each possible world \equiv the **joint distribution**
- How many entries are there, if we assume the variables are all binary?
- How is this related to the truth tables in logic?

T_o C_a
→ →
...

$P(T_o, C_a)$

2^n

$m(KB)$

TOOTHACHE EXAMPLE

| T_0 Toothache | C_a Cavity | P(T,C) |
|--------------------|-----------------|--------|
| toothache | cavity | 0.15 |
| toothache | \neg cavity | 0.10 |
| \neg toothache | cavity | 0.05 |
| \neg toothache | \neg cavity | 0.70 |

These probabilities are different from what is given in the textbook

NOTATION

- An upper-case letter A_1 represents a variable and its all possible values
- A lower-case letter a_1 represents a particular value of the variable A_1
- $P(A_1)$ represents a table/function specifying a probability for each possible value of A_1
- $P(A_1=a_1)$ represents a scalar value specifying the probability of $A_1=a_1$
- We often abbreviate $P(A_1=a_1)$ as $P(a_1)$

PRIOR AND POSTERIOR

- Prior probability
 - Probability of a proposition in the absence of any other information
 - E.g., $P(V_1, V_3, V_5)$
- Conditional/posterior probability
 - Probability of a proposition given another piece of information
 - E.g., $P(V_2, V_3 \mid V_5 = T, V_7 = F)$
 - $P(A \mid B) = P(A \wedge B) / P(B)$

NUMBER OF PARAMETERS

- Assuming everything is binary
- $P(V_1)$ requires
 - 1 independent parameter
- $P(V_1, V_2, \dots, V_n)$ requires
 - $2^n - 1$ independent parameters
- $P(V_1 | V_2)$ requires
 - 2 independent parameters
- $P(V_1, V_2, \dots, V_n | V_{n+1}, V_{n+2}, \dots, V_{n+m})$ requires
 - $2^m \times (2^n - 1)$ independent parameters

MARGINALIZATION

- Given $P(V_1, V_2, \dots, V_n \mid V_{n+1}, V_{n+2}, \dots, V_{n+m})$, where $n > 0$ and $m \geq 0$, we can find, for example
 - $P(V_i, V_j, V_k \mid V_{n+1}, V_{n+2}, \dots, V_{n+m})$ where $i, j, k < n$ by summing out all the irrelevant variables
- Examples

LET'S ANSWER A FEW QUERIES

| Toothache | Cavity | P(T,C) |
|------------------|---------------|--------|
| toothache | cavity | 0.15 |
| toothache | \neg cavity | 0.10 |
| \neg toothache | cavity | 0.05 |
| \neg toothache | \neg cavity | 0.70 |

- $P(\text{cavity}) = ?$
- $P(\neg \text{cavity}) = ?$
- $P(\text{toothache}) = ?$
- $P(\neg \text{toothache}) = ?$

LET'S ANSWER A FEW QUERIES

| Toothache | Cavity | P(T,C) |
|------------------|---------------|--------|
| toothache | cavity | 0.15 |
| toothache | \neg cavity | 0.10 |
| \neg toothache | cavity | 0.05 |
| \neg toothache | \neg cavity | 0.70 |

- $P(\text{cavity} \mid \text{toothache}) = ?$
- $P(\text{cavity} \mid \neg \text{toothache}) = ?$
- $P(\neg \text{cavity} \mid \text{toothache}) = ?$
- $P(\neg \text{cavity} \mid \neg \text{toothache}) = ?$
- $P(\text{toothache} \mid \text{cavity}) = ?$
- $P(\neg \text{toothache} \mid \text{cavity}) = ?$
- $P(\text{toothache} \mid \neg \text{cavity}) = ?$
- $P(\neg \text{toothache} \mid \neg \text{cavity}) = ?$

BAYES' RULE

- $P(B | A) = P(A | B) * P(B) / P(A)$
- Example use
 - $P(\text{cause} | \text{effect}) = P(\text{effect} | \text{cause}) * P(\text{cause}) / P(\text{effect})$
- Why is this useful?
 - Because in practice it is easier to get probabilities for $P(\text{effect} | \text{cause})$ and $P(\text{cause})$ than for $P(\text{cause} | \text{effect})$
 - E.g., $P(\text{disease} | \text{symptoms}) = P(\text{symptoms} | \text{disease}) * P(\text{disease}) / P(\text{symptoms})$
 - It is easier to know what symptoms diseases cause. It is harder to diagnose a disease given symptoms

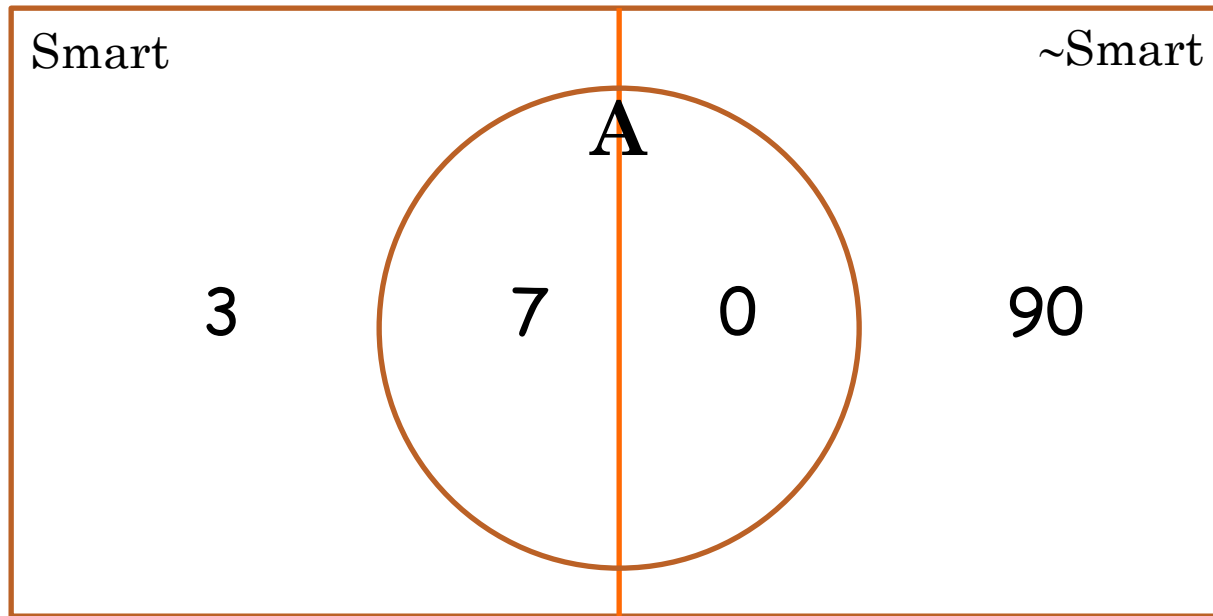
BAYES RULE

- Can we compute $P(\alpha|\beta)$ from $P(\beta|\alpha)$?
- E.g.,
 - In a class, 70% of the smart students got an A.
 - $P(a \mid \text{smart}) = 0.7$
 - John got an A. What is the probability of John being smart given he got an A?
 - $P(\text{smart} \mid a) = ?$

CLASS EXAMPLE

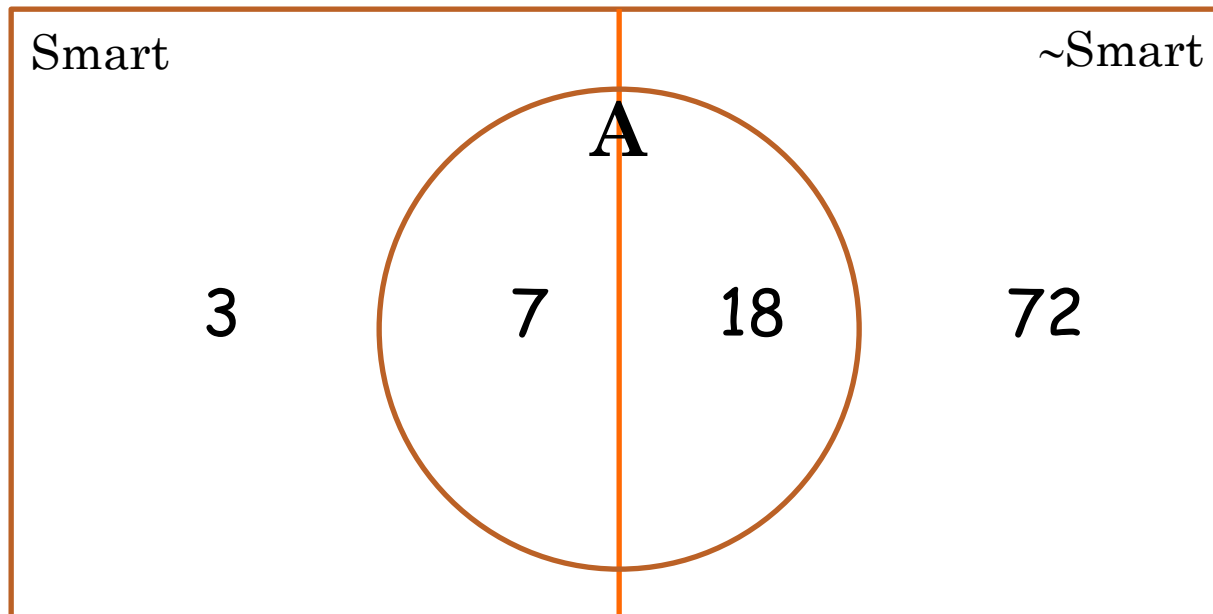
- Let's say there are 100 students in the class
- Let's say 10 of them are Smart, 90 are \sim Smart
- Probability of a randomly picked student being smart
 - $P(s) = 0.1$
- We know that 70% of the smart students got an A.
 - $P(a | s) = 0.7$
 - 7 smart students got an A; 3 did not get an A.
- What is $P(s|a) = ?$
 - Depends on $P(a)$

VERY HARD CLASS



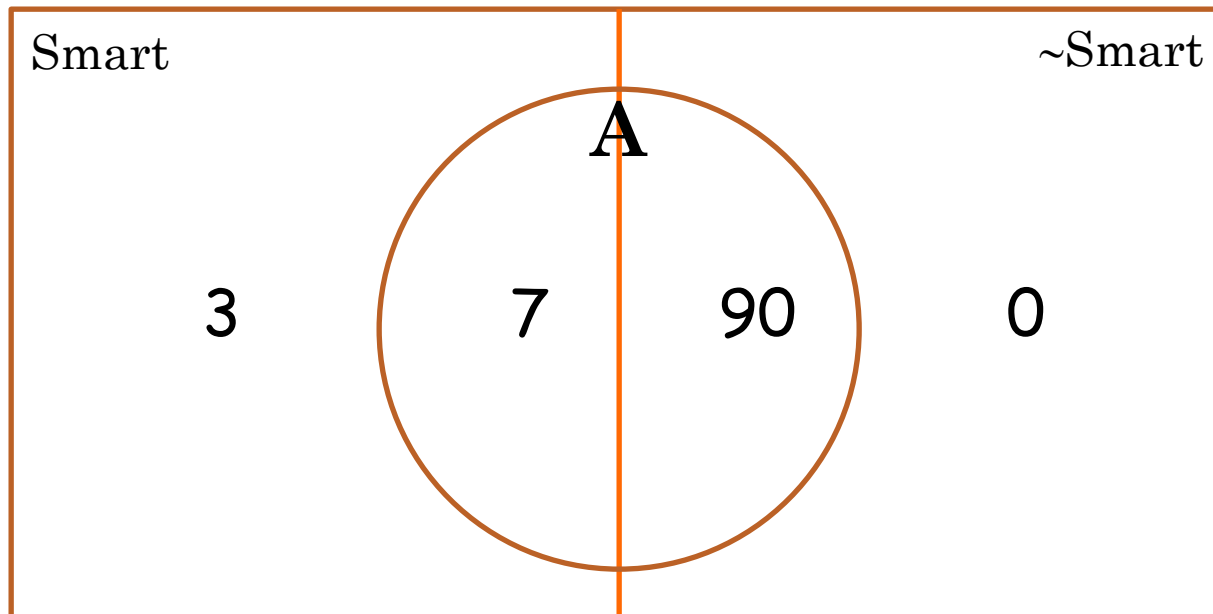
$$P(s | a) = ?$$

MEDIUM HARD CLASS



$$P(s | a) = ?$$

WEIRD CLASS



$$P(s | a) = ?$$

TUBERCULOSIS TEST

○ Tuberculosis test

- The test is 90% accurate
 - If you have TB, the test is positive with 90% probability
 - If you don't have TB, the test is negative with 90% probability
- John takes the test and the result is positive
- What is the probability that John has TB?

○ Formally

- $P(+ | TB) = 0.9$
- $P(- | \neg TB) = 0.9$
- $P(TB | +) = ?$

CHAIN RULE

- $P(X_1, X_2, X_3, \dots, X_k) =$
 - $P(X_1) P(X_2 | X_1) P(X_3 | X_1, X_2) \dots P(X_k | X_1, X_2, X_3, \dots, X_{k-1})$
 - or
 - $P(X_2) P(X_1 | X_2) P(X_3 | X_1, X_2) \dots P(X_k | X_1, X_2, X_3, \dots, X_{k-1})$
 - or
 - $P(X_2) P(X_3 | X_2) P(X_1 | X_3, X_2) \dots P(X_k | X_1, X_2, X_3, \dots, X_{k-1})$
 - or
 - Pick an order, then
 - $P(\text{first})P(\text{second} | \text{first})P(\text{third} | \text{first}, \text{second}) \dots P(\text{last} | \text{all_previous})$