Fresh our memory with PROBABILITY

Lesson 9

Unconditional Probability

• **Unconditional** or prior probability that a proposition A is true: \( P(A) \)
  
  – In the absence of any other information, the probability to event A is \( P(A) \).
  
  – Probability of application accepted: \( P(\text{application-accept}) = 0.2 \)

• Propositions include random variables \( X \)
  
  – Each random variable \( X \) has domain of values: \{red, blue, … green\}
  
  – \( P(X=\text{Red}) \) means the probability of \( X \) to be Red

Unconditional Probability

• If application-accept is binary random variable -> values = \{true,false\}
  
  – \( P(\text{application-accept}) \) same as \( P(\text{app-accept} = \text{True}) \)
  
  – \( P(\neg\text{app-accept}) \) same as \( P(\text{app-accept} = \text{False}) \)

• If Status-of-application domain: \{reject, accept, wait-list\}
  
  – We are allowed to make statements such as:
    
    \[
    \begin{align*}
    P(\text{status-of-application} = \text{reject}) &= 0.2 \\
    P(\text{status-of-application} = \text{accept}) &= 0.3 \\
    P(\text{status-of-application} = \text{wait-list}) &= 0.5
    \end{align*}
    \]

Conditional Probability

• What if agent has some evidence?
  
  – E.g. agent has a friend who has applied with a much weaker qualification, and that application was accepted?

• **Posterior** or conditional probability
  
  \( P(A|B) \) probability of \( A \) given all we know is \( B \)
  
  – \( P(X=\text{accept}|\text{Weaker application was accepted}) \)
  
  – If we know \( B \) and also know \( C \), then \( P(A|B \land C) \)
**Product rule**

\[ P(A \land B) = P(A|B)*P(B) \]
\[ P(A \land B) = P(B|A)*P(A) \]
\[ P(A|B) = \frac{P(A \land B)}{P(B)} \]
\[ P(B|A) = \frac{P(A \land B)}{P(A)} \]
Bayes’ Rule

Given that
- \( P(A \land B) = P(A|B) \cdot P(B) \)
- \( P(A \land B) = P(B|A) \cdot P(A) \)

\[ P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \]

- Determine \( P(B|A) \) given \( P(A|B) \), \( P(B) \) and \( P(A) \)
- Generalize to some background evidence \( e \)

\[ P(Y | X, e) = \frac{P(X | Y, e) \cdot P(Y | e)}{P(X | e)} \]

Bayes’ Rule Example

- \( S \): Proposition that patient has stiff neck
- \( M \): Proposition that patient has meningitis
- Meningitis causes stiff-neck, 50% of the time

- Given:
  - \( P(S | M) = 0.5 \)
  - \( P(M) = 1/50,000 \)
  - \( P(S) = 1/20 \)
  - \( P(M|S) = P(S| M) \cdot P(M) / P(S) = 0.0002 \)

- If a patient complains about stiff-neck, \( P(\text{meningitis}) \) only 0.0002

Bayes’ Rule Use

- \( P(S | M) \) is causal knowledge, does not change
  - It is "model based"
  - It reflects the way meningitis works

- \( P(M | S) \) is diagnostic; tells us likelihood of \( M \) given symptom \( S \)
  - Diagnostic knowledge may change with circumstance, thus helpful to derive it
  - If there is an epidemic, probability of Meningitis goes up; rather than again observing \( P(M | S) \), we can compute it
Computing the denominator: P(S)

We wish to avoid computing the denominator in the Bayes’ rule
− May be hard to obtain
− Introduce 2 different techniques to compute (or avoid computing P(S))

Computing the denominator:

#1 approach - compute relative likelihoods:
• If M (meningitis) and W(whiplash) are two possible explanations:
  − P(M|S) = P(S|M) * P(M) / P(S)
  − P(W|S) = P(S|W) * P(W)/ P(S)
• P(M|S)/P(W|S) = P(S|M) * P(M) / P(S|W) * P(W)
• Disadvantages:
  − Not always enough
  − Possibility of many explanations

Computing the denominator:

#2 approach - Using M & ~M:
• Checking the probability of M, ~M when S
  − P(M|S) = P(S|M) * P(M) / P(S)
  − P(~M|S) = P(S|~M) * P(~M)/ P(S)
• P(M|S) + P(~M | S) = 1 (must sum to 1)
  − [P(S|M)*P(M)/ P(S)] + [P(S|~M) * P(~M)/P(S)] = 1
  − P(S|M) * P(M) + P(S|~M) * P(~M) = P(S)
• Calculate P(S) in this way…

Computing the denominator:
The #2 approach is actually - normalization:
• 1/P(S) is a normalization constant
  − Must ensure that the computed probability values sum to 1
  − For instance: P(M|S)+P(~M|S) must sum to 1
• Compute:
  − (a) P(S|~M) * P(~M)
  − (b) P(S | M) * P (M)
  − (a) and (b) are numerators, and give us "un-normalized values"
  − We could compute those values and then scale them so that they sum to 1
Simple Example

- Suppose two identical boxes
  - **Box 1:**
    - colored red from inside
    - has $\frac{1}{3}$ black balls, $\frac{2}{3}$ red balls
  - **Box 2:**
    - colored black from inside
    - has $\frac{1}{3}$ red balls, $\frac{2}{3}$ black balls
- We select one Box at random; can’t tell how it is colored inside.
- What is the probability that Box is red inside?

Applying Bayes’ Rule

What if we were to select a ball at random from Box, and it is red, does that change the probability?

$$P(\text{Red-box} | \text{Red-ball}) = \frac{P(\text{Red-ball} | \text{Red-box}) \cdot P(\text{Red-box})}{P(\text{Red-ball})}$$

How to calculate $P(\text{Red-ball})$?

Thus, by our approach #2:

$$2/3 \cdot 0.5 / P(\text{Red-ball}) + 1/3 \cdot 0.5 / P(\text{Red-ball}) = 1$$

Thus, $P(\text{Red-ball}) = 0.5$, and $P(\text{Red-box} | \text{Red-ball}) = 2/3$

Absolute and Conditional Independence

- Absolute: $P(X|Y) = P(X)$ or $P(X \land Y) = P(X)P(Y)$
- Conditional: $P(A \land B | C) = P(A | C)P(B | C)$
- $P(A| B \land C)$
  - If $A$ and $B$ are conditionally independent given $C$ then, probability of $A$ is not dependent on $B$
  - $P(A| B \land C) = P(A| C)$
- E.g. Two independent sensors $S_1$ and $S_2$ and a jammer $J_1$
  - $P(S_i) =$ Probability Si can read without jamming
  - $P(S_1 | J_1 \land S_2) = P(S_1 | J_1)$

Combining Evidence

- Example:
  - S: Proposition that patient has stiff neck
  - H: Proposition that patient has severe headache
  - M: Proposition that patient has meningitis
  - Meningitis causes stiff-neck, 50% of the time
  - Meningitis causes head-ache, 70% of the time
- Probability of Meningitis should go up, if both symptoms reported
- How to combine such symptoms?
Combining Evidence

• \[ P(C \mid A \land B) = \frac{P(C \land A \land B)}{P(A \land B)} \]

• Numerator:
  \[ P(C \land A \land B) = P(B \mid A \land C) \cdot P(A \land C) \]
  \[ = P(B \mid C) \cdot P(A \mid C) \cdot P(C) \]

• Going back to our example:
  \[ P(M \mid S \land H) = \frac{P(S \mid M) \cdot P(H \mid M) \cdot P(M)}{P(S \land H)} \]