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(application-accept) = 0.2
- Propositions include random variables X
  - Each random variable X has domain of values: *{red, blue, …green}*
  - P(*X=Red*) means the probability of X to be Red

Unconditional Probability

- If application-accept is binary random variable -> values = {true,false}
  - P(application-accept) same as P(app-accept = True)
  - P(~app-accept) same as P(app-accept = False)
- *If* Status-of-application domain: *{reject, accept, wait-list}*
  - We are allowed to make statements such as:
    P*(status-of-application = reject)* = 0.2
    P*(status-of-application = accept)* = 0.3
    P*(status-of-application = wait-list)* = 0.5

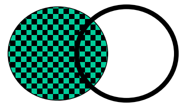
Conditional Probability

- 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=accept|Weaker application was accepted)*
  - If we know B and also know C, then P(A| B $\wedge$ C)

Product rule

– $P(A \wedge B) = P(A|B)*P(B)$
– $P(A \wedge B) = P(B|A)*P(A)$

– $P(A|B) = P(A \wedge B) / P(B)$
– $P(B|A) = P(A \wedge B) / P(A)$



Probability Distribution

- Probability of all the possible values of X Denote by **P**(X)
  - Note that P is in bold
  - In our example:
    X = Status-of-application
    $X_i \in$ *{reject, accept, wait-list}*
    P(X) = *<0.2, 0.3, 0.5>*
- Σ P(X=xi) = 1

Joint Probability Distribution

- Joint probability distribution is a table
  - *Assigns probabilities to all possible assignment of values for combinations of variables*
- $\mathbf{P}(X_1, X_2, ..X_n)$ assigns probabilities to all possible assignment of values to variables X_1, X_2,..X_n

Joint Probability Distribution

- X_1 = Status of your application
- X_2 = Status of your friend's application
- Then **P**(X1,X2)

		X1 Reject	X1 Accept	X1 Wait-list
X2	Reject	0.15	0.3	0.02
	Accept	0.3	0.02	0.09
	Wait-list	0.02	0.09	0.01

Bayes' Rule

- *Given that*
  - $P(A \wedge B) = P(A|B)*P(B)$
  - $P(A \wedge B) = P(B|A)*P(A)$

  $\rightarrow P(B|A) = \frac{P(A|B)*P(B)}{P(A)}$
- Determine P(B|A) given P(A|B), P(B) and P(A)
- Generalize to some background evidence e

  $\mathbf{P}(Y \mid X, e) = \frac{\mathbf{P}(X \mid Y, e) * \mathbf{P}(Y \mid e)}{\mathbf{P}(X \mid e)}$

127

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:
  - $P(S \mid M) = 0.5$
  - $P(M) = 1/50{,}000$
  - $P(S) = 1/20$
  - $P(M|S) = P(S|M) * P(M) / P(S) = 0.0002$
- If a patient complains about stiff-neck, P(meningitis) only 0.0002

128

Bayes' Rule

- How can it help us?
  - P(A|B) may be causal knowedge, P(B|A) diagnostic knowledge
  - *E.g., A is symptom, B is disease*
- Diagnostic knowledge may vary:
  - *Robustness by allowing P(B | A) to be computed from others*

129

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

130

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))

131

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

132

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…

133

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

134

Simple Example

- Suppose two identical boxes
- Box_1:
  - colored red from inside
  - has 1/3 black balls, 2/3 red balls
- **Box_2**:
  - colored black from inside
  - has 1/3 red balls, 2/3 black balls
- We select one Box at random; cant tell how it is colored inside.
- What is the probability that Box is red inside?

135

Applying Bayes' Rule

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

$$P(\text{Red-box} \mid \text{Red-ball}) = \frac{P(\text{Red-ball} \mid \text{Red-box}) * P(\text{Red-box})}{P(\text{Red-ball})}$$
$$= 2/3 * 0.5 / P(\text{Red-ball})$$

How to calculate P(Red-ball)?

$$P(\textbf{Black}\text{-box} \mid \text{Red-ball}) = \frac{P(\text{Red-ball} \mid \textbf{Black}\text{-box}) * P(\textbf{Black}\text{-box})}{P(\text{Red-ball})}$$
$$= 1/3 * 0.5 / P(\text{Red-ball})$$

Thus, by our approach #2: $2/3 * 0.5 / P(\text{Red-ball}) + 1/3 * 0.5 / P(\text{Red-ball}) = 1$

Thus, P(Red-ball) = 0.5, and P(Red-box | Red-ball) = 2/3

136

Absolute and Conditional Independence

- Absolute: $\mathbf{P}(X|Y) = \mathbf{P}(X)$ or $\mathbf{P}(X \wedge Y) = \mathbf{P}(X)\mathbf{P}(Y)$
- **Conditional: $P(A \wedge B \mid C) = P(A \mid C)\, P(B \mid C)$**
- **$P(A \mid B \wedge C)$**
  - *If A and B are conditionally independent given C Then, probability of A is not dependent on B*
  - **$P(A \mid B \wedge C) = P(A \mid C)$**
- E.g. Two independent sensors S1 and S2 and a jammer J1
  - *P(Si) = Probability Si can read without jamming*
  - *$P(S1 \mid J1 \wedge S2) = P(S1 \mid J1)$*

137

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?

138

Combining Evidence

- $P(C| A \wedge B) = P(C \wedge A \wedge B) / P (A \wedge B)$

- Numerator:
  - $P(C \wedge A \wedge B) = P(B | A \wedge C) * P(A \wedge C)$
    $= P(B | C) * P(A \wedge C)$
    $= P(B | C) * P(A | C) * P (C)$

- Going back to our example:

$$P(M | S \wedge H) = \frac{P(S| M) * P(H| M) * P(M)}{P(S \wedge H)}$$

139