CSE 250A - Lecture 1 01-OCT-2019

Motivation for Probabilistic Reasoning

* Modeling of uncertainty
  - Inherent randomness (eg: radioactive decay)
  - Gross statistical dependecies of complex deterministic world (eg: coin toss)
* Probability as guardian of common sense reasoning
* Many emporical successes:
  - Natural language, speech, robotics, bio informatics.

Review

* Discrete random variable X

  Domain of possible values $\{x_1, x_2 \ldots x_m\}$

  Ex: month M, $\{m_1 = \text{Jan}, m_2 = \text{Feb}, \ldots m_{12} = \text{Dec}\}$
* Unconditional (or prior) probability $\Rightarrow P(X=x)$

* Basic axioms

  (i) $P(X=x) \geq 0$

  (ii) $\sum_i P(X=x_i) = 1$

  (iii) $P(X=x_i \text{ or } X=x_j) = P(X=x_i) + P(X=x_j)$ if $x_i \neq x_j$

  Probs add for union of mutually exclusive events

* Conditional probabilities

  $P(X=x_i \mid Y=y_j)$ - Probability that $X=x_i$ given $Y=y_j$

In general : $P(X=x_i \mid Y=y_j) \neq P(X=x_i)$

Ex: Dependent random variables

Weather W $\{w_1 = \text{sunny}, w_2 = \text{rainy}\}$

$P(W=\text{sunny}) = 0.9$

$P(W=\text{sunny} \mid M=\text{Aug}) = 0.97$

$P(W=\text{sunny} \mid M=\text{Jan}) = 0.83$

Ex: Independent random variables (Marginally independent)

Day of week D $\{d_1 = \text{sun}, d_2 = \text{mon}, \ldots \; d_7 = \text{sat}\}$

~~P(W = sunny) =~~

$P(W=\text{sunny}) = 0.9$

$P(W=\text{sunny} \mid D=\text{sunday}) = 0.9$

$P(W=\text{sunny} \mid D=\text{saturday}) = 0.9$

Ex: Conditionally independent random variables

Binary $\{0,1\}$ random variable

R = is it raining?

W = is the sidewalk wet?

U = people using umbrellas?

$P(W=1) < P(W=1 \mid U=1)$ $\Rightarrow$ W and U are not (marginally) independent

$P(W=1 \mid R=1) = P(W=1 \mid R=1, U=1)$ W and U are conditionally independent given R.

↳ assuming above is true for all possible values of W, R and U

Ex: Conditionally dependent random variables

End of lecture

* Also true

(i) $P(X=x_i \mid Y=y_j) \geq 0$

(ii) $\sum_i P(X=x_i \mid Y=y_j) = 1$ **Note:** sum over i, not j

* Joint probabilities

$P(X=x_i, Y=y_j)$ - Probability that $X=x_i$ and $Y=y_j$

* Product rule

For all values of i, j

$$P(X=x_i, Y=y_j) = P(X=x_i) \cdot P(Y=y_j \mid X=x_i)$$
$$= P(Y=y_j) \cdot P(X=x_i \mid Y=y_j)$$

* Marginalization

$$P(X=x_i) = \sum_j P(X=x_i, Y=y_j)$$

$$P(X=x_i, Y=y_j) = \sum_k P(X=x_i, Y=y_j, Z=z_k)$$

* Note: In general, easier to ~~ele~~ elicit conditional probability from experts than joint probabilities

* shorthand notation:

(i) Implied universality

$$P(X,Y) = P(X) \cdot P(Y \mid X) = P(Y) \cdot P(X \mid Y)$$

↳ prod rule in short hand notation

(2) Implied assignment

$$P(x,y,z) = P(X=x, Y=y, Z=z)$$

* Generalized Product Rule

$$P(A,B,C,D,\ldots) = P(A)\cdot P(B|A)\cdot P(C|A,B)\cdot P(D|A,B,C)\ldots$$

* Bayes Rule

$$P(X|Y) = \frac{P(Y|X)\cdot P(X)}{P(Y)}$$

* Bayes rule - Conditioned on additional event Z

$$P(X|Y,Z) = \frac{P(Y|X,Z)\cdot P(X|Z)}{P(Y|Z)}$$

$\boxed{\text{Alarm example}}$

* Binary random variables
  - B - is there a burglary
  - E - is there an earthquake
  - A - does the alarm go off.

* Joint distribution.

$$P(B,E,A) = P(B)\cdot P(E|B)\cdot P(A|B,E)$$

* Prior knowledge about world

Burglaries are rare $P(B=1) = 0.001$

Earthquakes are rare $P(E=1|B=0) = 0.002$

$P(E=1|B=1) = 0.002$ → Implies B and E are independent

Assumption: Burgularies & earthquakes are independent

$P(E|B) = P(E)$ (or) $P(B|E) = P(B)$

B	E	P(A|B,E)
0	0	0.001
1	0	0.94
0	1	0.29
1	1	0.95

Probs of complimentary events are easy to compute: $P(B=0) = 1 - P(B=1)$

$P(A=0|B,E) = 1 - P(A=1|B,E)$

Inference

Do rules of probability capture common sense reasoning?

Let's compare the following:

① $P(B=1) = 0.001$

② $P(B=1|A=1) = ?$

③ $P(B=1|A=1, E=1) = ?$

② $$P(B=1|A=1) = \frac{P(A=1|B=1) \cdot \overbrace{P(B=1)}^{0.001}}{P(A=1)}$$

Numerator: $P(A=1|B=1) = \sum_e P(A=1, E=e | B=1)$ conditional marginalization

$$= \sum_e P(E=e | B=1) \cdot P(A=1 | E=e, B=1)$$

Prod rule

$$= \sum_{e} P(E=e) \cdot P(A=1 \mid E=e, B=1)$$ marginally independent

$$P(A=1 \mid B=1) = P(E=0) \cdot P(A=1 \mid E=0, B=1) + P(E=1) \cdot P(A=1 \mid E=1, B=1)$$

substituting values

$$\boxed{P(A=1 \mid B=1) = 0.94002}$$

<u>Denominator:</u>

$$P(A=1) = \sum_{b,e} P(B=b, E=e, A=1)$$ marginalization.

$$= \sum_{b,e} P(B=b) \cdot P(E=e \mid B=b) \cdot P(A=1 \mid E=e, B=b)$$ Prod rule

$$= \sum_{b,e} P(B=b) \cdot P(E=e) \cdot P(A=1 \mid E=e, B=b)$$ marginally independent

substituting values and sum over 4 terms

$$P(A=1) = 0.00252$$

$$P(B=1 \mid A=1) = \frac{P(A=1 \mid B=1) \cdot P(B=1)}{P(A=1)}$$

$$= \frac{0.94002 \times 0.001}{0.00252} = 0.37$$

We have

$$P(B=1) = 0.001$$
$$P(B=1 \mid A=1) = 0.37$$

agrees with common sense

3 Last term:

$$P(B=1 \mid A=1, E=1) = \frac{\overset{0.95}{P(A=1 \mid B=1, E=1)} \cdot \overset{0.001}{P(B=1 \mid E=1)}}{P(A=1 \mid E=1)}$$ ↳ conditional Bayes rule

$$\underbrace{P(B=1 \mid E=1) = P(B=1)}_{\text{independent}} = 0.001$$

Denominator:

Similar to calculating $P(A=1 \mid B=1)$

$$P(A=1 \mid E=1) = \sum_b P(A=1, B=b \mid E=1)$$ marginalization

$$= \sum_b P(B=b \mid E=1) \cdot P(A=1 \mid B=b, E=1)$$ prod rule

$$= \sum_b P(B=b) \cdot P(A=1 \mid B=b, E=1)$$ marginal independence

Substitute and sum

$$P(A=1 \mid E=1) = 0.29066$$

$$P(B=1 \mid A=1, E=1) = \frac{0.95 \times 0.001}{0.29066} = 0.0033$$

Compare:

$$\left.\begin{array}{l} P(B=1) = 0.001 \\ P(B=1 \mid A=1) = 0.37 \uparrow \\ P(B=1 \mid A=1, E=1) = 0.0033 \downarrow \end{array}\right\}$$ Example of non-monotonic reasoning

↙

Pattern of reasoning known as "Explaining away"
Earthquake explains away the alarm,
thereby diminishing our belief in burglary.

Ex: Conditional dependence

$P(B) = P(B|E)$ → marginal independence
(B is independent of E)

$P(B|A) \neq P(B|E,A)$
↳ conditional dependence
B and E are conditionally dependent given A.