Foundations of Artificial Intelligence

Unit-4

Probabilistic Models

* Types of variables - boolean random variables
  discrete random variable
  continuous random variable

* Where do probabilities come from?
  (i) Frequents
  (ii) Subjective
  (iii) Objectivist
  (iv) Prior probability
  (v) Posterior probability

* Axioms of probability -

(i) $0 \leq P(x) \leq 1$

(ii) $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

(iii) $P(a|b) = P(a \wedge b) | P(b)$ (product rule)

(iv) Theorem of total probability:

$$P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i)$$

$$= P(B|A_1) + P(B|A_2) + \cdots$$

(v) Bayes Theorem: $P(h|D) = \dfrac{P(D|h)P(h)}{P(D)}$

* Choosing hypotheses -

<u>maximum a posteriori</u> : h_{MAP}

$$h_{MAP} = \underset{h \in H}{\arg\max}\ P(h|D)$$

$$= \underset{h \in H}{\arg\max}\ \frac{P(D|h)\,P(h)}{P(D)}$$

$$= \arg\max\ P(D|h)\,P(h)$$

<u>maximum likelihood</u> $= \arg\max\ P(D|h)$

* Naive Bayesian Classification

Example : Make a naive bayes classifier as to whether to play tennis or not. The given table is as follows. and the sample is

X = <rain, hot, high, false>

Outlook	Temperature	Humidity	Windy	Class
sunny	H	H	F	N
sunny	H	H	T	N
overcast	H	H	F	P
rain	m	H	F	P
rain	C	N	F	P
rain	c	N	T	N
overcast	C	N	T	P
sunny	m	H	F	N
sunny	C	N	F	P.
rain	m	N	F	P
sunny	m	N	T	P
overcast	m	H	T	P
overcast	H	N	F	P
rain	m	H	T	N

* Bayes Theorem

Q Paper - CAT-2

Event A : A patient has liver disease = 10%.

Event B : Patient is an alcoholic = 5%.

In patients diagonosed with liver disease - 7% are alcoholics

If the patient is an alcoholic, find the probability of getting a liver disease in the future.

Ans: P(Liver) = 0.1

P(Alcoholic) = 0.05

P(Alcoholic | Liver) = 0.07

To Find : P(Liver | Alcoholic) $= \frac{0.07 \times 0.1}{0.05}$

$= 0.14$

* Bayesian Belief Network (BBN)

→ It is an acyclic directed graph where the nodes of the graph represent evidence of hypotheses, and the connection of 2 nodes represents the dependence between them.

$X \rightarrow Y$ = X is the parent of Y

Need for BBN - Joint probability distribution for n variables requires 2^n entries. ⇒ impractically

Step 1: Find $P(X \mid class)$

(i) $P(\text{Outlook} \mid class)$

$P(\text{sunny} \mid P) =$

$P(\text{sunny} \mid N) =$

$P(\text{rain} \mid P) =$

$P(\text{rain} \mid N)$

$P(\text{overcast} \mid P) =$

$P(\text{overcast} \mid N) =$

(ii) $P(\text{Temp} \mid class)$

$P(H \mid P) =$

$P(H \mid N) =$

$P(M \mid P) =$

$P(M \mid N) =$

$P(C \mid P) =$

$P(C \mid N) =$

(iii) $P(\text{Humidity} \mid class)$

$P(\text{H} \mid P) =$

$P(\text{Ham} \mid N) =$

$P(\text{Namal} \mid P) =$

$P(\text{Noimal} \mid N) =$

(iv) $P(\text{Windy} \mid class)$

$P(\text{windy} \mid P)$

$P(\text{windy} \mid N)$

$P(NW \mid P)$

$P(NW \mid N)$

$P(P) =$ $\quad$ $P(n) =$

For the given statement :

$X = \langle \text{rain}, \text{hot}, \text{high}, \text{false} \rangle$

$$P(X \mid P) = \frac{P(\text{rain} \mid P) \times P(\text{hot} \mid P) \times P(\text{high} \mid P) \times P(F \mid P) \times P(P)}{P(P)}$$

$$P(X \mid n) = P(\text{rain} \mid n) \times P(\text{hot} \mid n) \times P(\text{high} \mid n) \times P(F \mid n) \times P(n)$$

choose the larger value