MODULE 4

1. Consider a medical diagnosis problem in which there are two alternative hypotheses: 1. that the patient has a particular form of cancer (+) and 2. That the patient does not (-). A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer. Determine whether the patient has Cancer or not using MAP hypothesis.

Solution :-

* From the above situation, probabilities can be summarized as follow.

$P(cancer) = 0.008 \qquad P(\neg cancer) = 0.992$

$P(\oplus | cancer) = 0.98 \qquad P(\ominus | cancer) = 0.02$

$P(\oplus | \neg cancer) = 0.03 \qquad P(\ominus | \neg cancer) = 0.97$

Should we diagnose the patient as having cancer or not?

$P(\oplus | cancer) \, P(cancer) = (0.98) * (0.08) = 0.0078$

$P(\oplus | \neg cancer) \, P(\neg cancer) = 0.03 * 0.992 = 0.0298$

So, $h_{MAP} = \neg cancer$

* The exact posterior probabilities can also be determined by normalizing the above quantities so that they sum to 1

$$P(cancer | \oplus) = \frac{0.0078}{0.0078 + 0.0298} = 0.21$$

$$P(\neg cancer | \oplus) = \frac{0.0298}{0.0078 + 0.0298} = 0.79$$

2. Apply Naïve Bayes classifier for ***PlayTennis*** concept learning problem to classify the following novel instance

< Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong >

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Table: Training examples for the target concept *PlayTennis*

⇒ Solution

* Task is to predict the target value (yes or no) of the target concept PlayTennis for this new instance.

* $$V_{NB} = \underset{v_j \in \{yes, no\}}{\text{argmax}} \; P(v_j) \prod_i P(a_i | v_j) \longleftarrow \text{equ } 1$$

$$V_{NB} = \underset{v_j \in \{yes, no\}}{\text{argmax}} \; P(v_j) \cdot P(\text{Outlook} = \text{Sunny} | v_j) \cdot$$

$$P(\text{Temp} = \text{cool} | v_j) \cdot P(\text{Humidity} = \text{high} | v_j) \cdot P(\text{Wind} = \text{Strong} | v_j)$$

* First, the probabilities of the different target values is estimated based on their frequencies over 14 training examples

$$P(\text{playTennis} = \text{yes}) = \frac{9}{14} = 0.64$$

$$P(\text{playTennis} = \text{no}) = \frac{5}{14} = 0.36$$

* Next, estimate the conditional probabilities

$$P(\text{Outlook} = \text{sunny} \mid \text{playTennis} = \text{yes}) = \frac{2}{9} = 0.22$$

$$P(\text{Outlook} = \text{sunny} \mid \text{playTennis} = \text{no}) = \frac{3}{5} = 0.60$$

$$P(\text{Temp} = \text{cool} \mid \text{playTennis} = \text{yes}) = \frac{3}{9} = 0.33$$

$$P(\text{Temp} = \text{cool} \mid \text{playTennis} = \text{no}) = \frac{1}{5} = 0.20$$

$$P(\text{Humidity} = \text{high} \mid \text{playTennis} = \text{yes}) = 3/9 = 0.33$$

$$P(\text{Humidity} = \text{high} \mid \text{playTennis} = \text{no}) = 4/5 = 0.80$$

$$P(\text{wind} = \text{Strong} \mid \text{playTennis} = \text{yes}) = \frac{3}{9} = 0.33$$

$$P(\text{wind} = \text{Strong} \mid \text{playTennis} = \text{no}) = \frac{3}{5} = 0.60$$

* Next calculate V_{NB} according equation 1

$$= P(\text{yes}) \cdot P(\text{sunny} \mid \text{yes}) \cdot P(\text{cool} \mid \text{yes}) \cdot P(\text{high} \mid \text{yes}) \cdot P(\text{strong} \mid \text{yes})$$

$$= 0.64 * 0.22 * 0.33 * 0.33 * 0.33$$

$$= \underline{0.0053}$$

= P(no) . P(Sunny|no) . P(cool|no) . P(high|no) . P(Strong|no)

= 0.036 * 0.60 * 0.20 * 0.80 * 0.60

= 0.0206

Thus, the naive Bayes classifier assigns the target value "no" to the new instance.

ie., PlayTennis = no

Outlook	Temp	Humidity	Wind	PlayTennis
Sunny	Cool	high	Strong	no

* Normalizing the quantities to sum to 1, calculate the conditional probability of target values.

$$\text{Yes} - \frac{0.0053}{0.0053 + 0.0206} = 0.205$$

$$\text{No} - \frac{0.0206}{0.0053 + 0.0206} = 0.795$$