Lec - 21 25|02|21

Naive Bayes classifier

→ Probability is the science of data mining.

→ A Naive Bayes classifier is a simple probability based model.

→ It uses Baye's theorem and it performs well in complex real world situation.

→ It is fast building model for both binary and multi-class classifications, relatively for low volumes of data.

→ The algo. makes predictions using Baye's Thm, which incorporate evidence or prior knowledge in it's predictions (Hand et al., 2001)

Baye's Theorem relates the conditional and marginal probability of stochastic events C and A, which is mathematically stated as:

$$P(C|A) = \frac{P(A|C)\,P(C)}{P(A)}$$

→ prior

posterior prob.

likelihood $\boxed{L(A|C)}$

where P stands for prob. of variable within paranthesis

$L(A|C)$ is the likelihood of A given C.

$P(C)$ is the prior prob. or marginal prob. of C
It is prior because it has not yet accounted for the information available in A

$P(C|A)$ is the conditional prob. of C given A. It is called posterior prob. because it has already incorporated the out come of event A

$P(A|C)$ is the condn. prob. of A given C

$P(A)$ is the prior or marginal prob. of A, which is normally the evidence.

Baye's Theorem as a ML model :

Let a classifier is conditional upon the attributes $A_1, A_2 \cdots A_n$ i.e

$P(C|A_1 A_2 \cdots A_n)$ where n is the no. of attributes

Using Baye's Theorem,

$$\underbrace{P(C|A_1 A_2 \cdots A_n)}_{\text{posterior}} = \frac{\overbrace{P(A_1 \cdots A_n | C)}^{\text{Likelihood}} \times \overbrace{P(C)}^{\text{prior}}}{P(A_1 \cdots A_n)}$$

Numerical :

EID	Age	Income	student	CR (credit rating)	buys-comp (Yes/NO)
1	Youth	H	NO	fair	NO
2	Y	H	NO	Ex	NO
3	Mid. (M)	H	NO	f	Y
4	senior (S)	m	NO	f	Y
5	S	L	Y	f	Y
6	S	L	Y	Ex	NO
7	m	L	Y	Ex	Yes
8	Y	M	N	f	NO
9	Y	L	Y	f	Yes
10	S	M	Y	f	Yes
11	Y	M	Y	Ex	Yes

12	m	M	N	Ex	Yes
13	m	H	Y	fair	Yes
14	s	M	N	ex	NO

$C_1 \rightarrow$ buys - comp (Yes)

$C_2 \rightarrow$ buys - comp (NO)

X = < age = Youth, income = medium, student = Yes, CR = fair > ??

C1 / C2 ??

$P(C_1 | X) = P(X | C_1) \cdot P(C_1)$

$P(C_2 | X) = P(X | C_2) \cdot P(C_2)$

$P(C_1) = P(\text{buys} = \text{Yes}) = \left(\frac{9}{14}\right) = 0.643$

$P(C_2) = P(\text{buys} = \text{NO}) = (5/14) = 0.357$

Compute $P(X | C_i)$

$P(\text{age} = \text{Youth} \mid \overset{\text{buys}}{C_1} = \text{Yes}) = \frac{2}{9} = 0.22$

$P(\text{age} = \text{Youth} \mid \text{buys} = \text{NO} \ (C_2)) = \frac{3}{5} = 0.6$ (Δ)

$P(\text{income} = \text{medium} \mid C_1) = 4/9 = 0.44$

$P(\text{income} = \text{medium} \mid C_2) = 2/5 = 0.4$ (Δ)

$P(\text{stud} = \text{Yes} \mid C_1) = 6/9 = 0.667$

$P(\text{stud} = \text{Yes} \mid C_2) = 1/5 = 0.2$ (Δ)

$P(CR = \text{fair} \mid C_1) = 6/9 = 0.667$

$P(CR = \text{fair} \mid C_2) = 2/5 = 0.4$ (Δ)

$$P(X \mid buys = yes\ (C_1)) = 0.22 \times 0.44 \times 0.667 \times 0.667$$
$$= 0.044$$

$$P(X \mid buys = NO\ (C_2)) = 0.6 \times 0.4 \times 0.2 \times 0.4$$
$$= 0.019$$

To find class c_i that maximizes $(P(X \mid c_i) \cdot P_i)$

$$\Rightarrow P(X \mid c_i) \cdot P_i$$

$$P(C_1 \mid X) = P(X \mid C_1) \cdot P(C_1) = 0.044 \times 0.643 = \boxed{0.028}$$

$$P(C_2 \mid X) = P(X \mid C_2) \cdot P(C_2) = 0.019 \times 0.357 = \underline{0.007}$$

$\therefore$ The Naive Bayes classifier classifies the unknown tuple X to class C1 (having maximum posterior probability).