Naive Bayes ML Algorithm — classification

A probabilistic classification algo. based on Bayes theorem, which assumes that features are independent of each other

① Independent Event

Rolling a dice

{1, 2, 3, 4, 5, 6}

$Pr(1) = \frac{1}{6}$, $Pr(2) = \frac{1}{6}$

$Pr(3) = \frac{1}{6}$

mean, two event never impact eachother is called Independent.

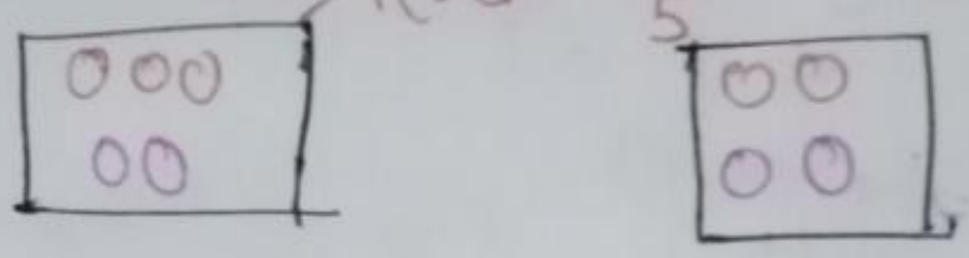
② Tossing a Coin

$P(H) = \frac{1}{2} = P(T) = \frac{1}{2}$

② Dependent Event

$P(R) = \frac{2}{4}$

$P(B) = \frac{3}{5}$



one event impact other event

$Pr(R \text{ and } B) = P(R) * P(B/R)$

(Conditional Probability)

$Pr(A \text{ and } B) = P(A) * P(B/A)$

$(Pr(A \text{ and } B) = Pr(B \text{ and } A))$

$P(A) * P(B/A) = Pr(B) * P(A/B)$

$$P(B/A) = \frac{Pr(B) * Pr(A/B)}{P(A)}$$ — Bayes theorem.

A			B
x_1	x_2	x_3	y
–	–	–	–
–	–	–	–
–	–	–	–

$$P(y/x_1, x_2, x_3) = \frac{Pr(y) * Pr(x_1, x_2, x_3/y)}{Pr(x_1, x_2, x_3)}$$

$$P(B/A) = \frac{P(B) * P(A/B)}{P(A)}$$

Ex- supervised classification

	x_1	x_2	x_3	x_4	Y
	—	—	—	—	O/P
	—	—	—	—	No
	—	—	—	—	Yes

$$P(Y/x_1 x_2 x_3 x_4 \text{---} x_n) = \frac{P(y) * P(x_1, x_2, x_3, x_4 \text{---} x_n/y)}{P(x_1, x_2, x_3 \text{---} x_4)}$$

$$= P(y) * P(x_1/y) * P(x_2/y) * P(x_3/y) \text{---} P(x_n/y)$$

Constant — $P(x) \quad P(x_2) \quad P(x_3) \text{---} P(x_4)$

$$P(N/x_1 x_2 x_3 \text{---} x_4) = \frac{P(N) * P(x_1/y) * P(x_2/y) * \text{---} P(x_n)}{\text{Constant}}$$

Constant — $P(x_1) * P(x_2) * P(x_3) * P(x_4) \text{---} P(x_n)$

Ex:-

DAY	Outlook	Temp	Humi°	Wind	Play Tennis
D1	Sunny	HOT	High	weak	No
D2	Sunny	HOT	High	Strong	No
D3	Overcast	HOT	High	weak	Yes
4	Rain	Mild	High	W	Y
5	Rain	Cold	Nor	W	Y
6	Rain	cold	Nor	S	N
7	Overcast	Cold	Nor	S	Y
8	Sunny	Mild	High	W	N
9	Sunny	Cool	Nor	W	Y
10	Rain	Mild	Nor	W	Y
11	Sunny	Mild	Nor	S	Y
12	Overcast	Mild	High	S	Y
13	Overcast	Hot	Nor	W	Y
14	Rain	Mild	High	S	N

Outlook

	P(Sunny/Yes)		P(Y)	P(N
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	4/5
	9	5	=14	

Temp.

	Yes	No	P(Y)	P(N)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cold	3	1	3/9	1/5
	9	5		

Play

		P(Y	P(N
Yes	9	9/14	5/14.
No	5		

New. P(Yes/Sunny, HOT) ⇒ ?

$$\begin{cases} P(Yes/Sunny\,HOT) = P(Yes) * P(Sunny/Yes) * P(HOT/Yes) \\ P(No/SunnyHOT) = P(No) * P(Sunny/No) * P(HOT/No) \end{cases}$$

$$\Rightarrow P(Yes/Sunny) = 9/14 * 2/9 * 2/3$$

$$\frac{2}{63} = 0.031 \rightarrow \frac{0.031}{0.031+0.08} \approx 27\%$$

$$P(No/Sunny/HOT) = 5/14 * 3/5 * 2/5$$

$$\frac{5}{85} \qquad 0.085 \qquad \frac{0.085}{0.031+0.085} \approx 73\%$$