Naive Baye's Algorithm (Classification)

1. Probability
2. Baye's Theorem

Independent Events

Rolling a Dice $\{1,2,3,4,5,6\}$

$Pr(1) = \frac{1}{6}$ $Pr(2) = \frac{1}{6}$ $Pr(3) = \frac{1}{6}.$

Dependent Events

1. What is the probability of removing a orange marble and then a yellow marble

O O O
O O

$\rightarrow P(O) = \frac{3}{5}$ → 1^{st} Event → Orange marble has been Removed

O O
O O

$\rightarrow P(Y/O) = \frac{2}{4} = \frac{1}{2}$ ⇒ 2^{nd} Event ⇒ Removed the Yellow Marble

$$P(O \text{ and } Y) = \underset{1^{st}}{P(O)} * \underset{II^{nd}}{\boxed{P(Y/O)}}$$ → Conditional Probability

$$= 3/5 * \frac{1}{2} = \boxed{\frac{3}{10}}$$

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

Bayes Theorem

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

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

$$\boxed{P(A/B) = \frac{P(A) * P(B/A)}{P(B)}} \Rightarrow \text{Bayes Theorem.}$$

$P(A/B)$ = Probability of Event A given B has occured

$P(A)$ = Probability of Event A

$P(B)$ = Probability of Event B

$P(B/A)$ = Probability of Event B given A has occured.

$$\boxed{P(A/B) = \frac{P(A) * P(B/A)}{P(B)}} \Rightarrow \text{Bayes Theorem.}$$

$\underbrace{x_1 \quad x_2 \quad x_3}_{\text{I/P features}}$			y (Dependent)
x_1	x_2	x_3	y
—	—	—	Yes
—	—	—	No
—	—	—	Yes
—	—	—	No

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

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

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

$\underbrace{x_1 \quad x_2 \quad x_3}_{}$ I/P features			↙ Dependent y
x_1	x_2	x_3	y
—	—	—	Yes
—	—	—	No
—	—	—	Yes
—	—	—	No

$[\boxed{0.60}, 0.40]$

⇓

Yes

New test data

$$Pr(Yes/(x_1,x_2,x_3)) = \frac{P(Yes) * P(x_1/Yes) * P(x_2/Yes) * P(x_3/Yes)}{\cancel{P(x_1) * Pr(x_2) * Pr(x_3)}} = \boxed{0.60}$$

$\cancel{P(x_1) * Pr(x_2) * Pr(x_3)} \Rightarrow$ Constant

⇓

Yes

$$Pr(No/(x_1,x_2,x_3)) = \frac{Pr(No) * Pr(x_1/No) * Pr(x_2/No) * Pr(x_3/No)}{\cancel{P(x_1) * Pr(x_2) * Pr(x_3)}} = 0.40$$

$\cancel{P(x_1) * Pr(x_2) * Pr(x_3)} \Rightarrow$ Constant

Lets Solve this Problem

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Outlook

	Yes	No	P(E/Yes)	Pr(E/No)
Sunny	2	3	$\boxed{2/9}$	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5

Temperature

→ Test (Sunny, Hot) → O/P PLAY (Y/N)

	Yes	No	P(E/Yes)	P(E/No)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	4/5
Cool	3	1	3/9	3/5

		P(Yes)	P(No)
Yes	9	9/14	5/14
No	5		

$$P(Yes/Sunny, Hot) = \frac{P(Yes) * Pr(Sunny/Yes) * Pr(Hot/Yes)}{\cancel{Pr(Sunny) * Pr(Hot)}}$$

$$= \frac{\cancel{9}}{\cancel{14}_{7}} * \frac{\cancel{2}^{1}}{\cancel{9}} * \frac{2}{9}$$

$$= \frac{2}{63} = 0.031$$

$$P(No \mid (Sunny, Hot)) = Pr(No) * Pr(Sunny/No) * Pr(Hot/No)$$

$$= \frac{\cancel{5}}{\cancel{14}_{7}} * \frac{3}{\cancel{5}} * \frac{\cancel{2}^{1}}{5}$$

$$= \frac{3}{35} = 0.085$$

$$Pr(Yes/(Sunny, hot)) = \frac{0.031}{(0.031 + 0.085)} = 0.27 = 27\%$$

$$Pr(No \mid Sunny, hot) = \frac{0.085}{(0.031 + 0.085)} = 0.73 = 73\%$$

New Test ⇒

Outlook	Temperature	O/P
Sunny	Hot	73% ⇒ They will not play Tennis 27% ⇒ They will play Tennis.

⇓

0 ⇒ Person is not going to playing Tennis