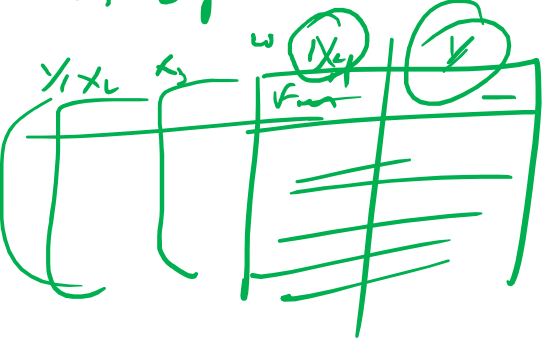
→ Naive Bayes classification

→ Probabilistic ML Model

→ used for classification

→ Supervised ML



Bayes Theorem

$P(B/A)$ = formulae ???

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

$$P(Y/X) = \frac{P(X/Y) * P(Y)}{P(X)}$$

$$P(Y/X_1, X_2, X_3 \ldots X_n) = \frac{P(X_1/Y) * P(X_2/Y) * \ldots * P(X_n/Y) * P(Y)}{P(X_1) * P(X_2) * P(X_3) \ldots P(X_n)}$$

	X_1	X_2	X_3	Y
S.No.	Long	Sweet	Yellow	Fruit
1	1	1	0	B
2	0	0	1	Or
3	0	1	0	Oth
⋮	⋮	⋮	⋮	⋮
1000				

1000 fruit
=
500 Banana
300 Orange
200 other

long	Sweet	Yellow	
0	1	1	B

(1) Summary → Category-wise

Type	L (a=1)	L' (a=0)	S	S'	Y	Y'	Total
B	400	100	350	150	450	50	500
Or	0 +1	300	150	150	300	0 +1	300
Oth	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

(2) → Calc the prob.

→ $P(B) = \frac{500}{1000} = 0.5$ | $P(or) = \frac{300}{1000} = 0.3$ | $P(om) = \frac{200}{1001} = 0.2$

→ $P(L) = \frac{500}{1000} = 0.5$ | $P(S) = \frac{650}{1000} = 0.65$ | $P(Y) = \frac{800}{1000} = 0.8$

→

$P(L/B) = \frac{400}{500} = 0.8$ | $P(L/or) = \frac{1}{300} = 0.003$ | $P(L/om) = \frac{100}{200} = 0.5$

$P(S/B) = \frac{350}{500} = 0.7$ | $P(S/or) = \frac{150}{300} = 0.5$ | $P(S/om) = \frac{150}{200} = 0.75$

$P(Y/B) = \frac{450}{500} = 0.9$ | $P(Y/or) = \frac{300}{300} = 1$ | $P(Y/om) = \frac{50}{200} = 0.25$

Training completed

Prediction Process

1	0	1

→

L	S	Y
1	1	1

Bayes

① Calculate using Bayes Theorem (Category wise)

$$P(B/L,S',Y) \Rightarrow \frac{P(L/B) * P(S'/B) * P(Y/B) * P(B)}{P(L) * P(S') * P(Y)} \Rightarrow \frac{0.8 * 0.4 * 0.9 * 0.5}{0.5 * 0.65 * 0.8} = 0.69$$

$$P(Or/L,S',Y) \Rightarrow \frac{P(L/Or) * P(S'/Or) * P(Y/Or) * P(Or)}{P(L) * P(S') * P(Y)} \Rightarrow \frac{0.003 * 0.5 * 2 * 0.3}{0.5 * 0.65 * 0.8} = 0$$

$$P(Om/L,S',Y) \Rightarrow \frac{P(L/Om) * P(S'/Om) * P(Y/Om) * P(Om)}{P(L) * P(S') * P(Y)} \Rightarrow \frac{0.5 * 0.45 * 0.25 * 0.2}{0.5 * 0.65 * 0.8} = 0.072$$

→ Laplace Correction → Add (+1) to zero-prob so that the overall prob does not become zero

→ Bayes th

Classifier of NB Algo

* Gaussian NB :- Most features are in Normal Dist

+ Multinomial NB :- Most features are discrete

+ Bernoulli NB :- Most features are Binary