Naive Bayes (only for classification)

* Prob : how likely something is to happen

$$p = \frac{\text{No of fav outcomes}}{\text{Total no. of outcomes}}$$

$$p(H) = 1/2$$
$$p(T) = 1/2$$

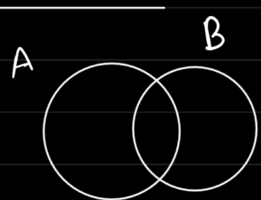
* Independent Events and dependent events.

Independent Events ↓ outcome of one event is not affected by another Events.

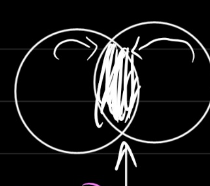
dependent events ⇓ outcome of an event is dependent on another event.

$$\boxed{P(A/B) = P(A) * P(B/A)}$$ → conditional prob.

* Baye's theorem



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



$$P(A|B) \cdot P(B) = P(B/A) \cdot P(A)$$

$$\boxed{P(A|B) = \frac{P(A) \cdot P(B/A)}{P(B)}}$$

← Bayes' theorem

$P(A|B)$ = Prob of event A given B has occurred

$P(A)$ = Prob of event A

$P(B)$ = " " event B.

$P(B/A)$ = Prob of event B given A has occurred.

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

x_1	x_2	x_3	y
—	—	—	—
—	—	—	—
—	—	—	—
—	—	—	—

$$P(\underset{A}{y}|\underset{B}{x_1 x_2 x_3}) = \frac{P(y) * P(x_1 x_2 x_3/y)}{P(x_1 x_2 x_3)}$$

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

Naive Bayes

y has two possibility → yes, NO

$$P(yes|x_1 x_2 x_3) = \frac{P(yes) * P(x_1/yes) * P(x_2/yes) * P(x_3/yes)}{\cancel{P(x_1) P(x_2) P(x_3)}} \text{ - Constant}$$

0.7

$$P(NO/(x_1 x_2 x_3)) = \frac{P(NO) * P(x_1/NO) * P(x_2/NO) * P(x_3/NO)}{\cancel{P(x_1) \cdot P(x_2) \cdot P(x_3)}} \text{ - Constant}$$

0.30

Predicted class = yes

* for multiclass classification

$$P(C_k|X) = \frac{P(X/C_k)\cdot P(C_k)}{P(X)}$$

$$P(C_1|x_1x_2x_3) = \frac{P(C_1) * P(x_1/C_1) * P(x_2|C_1) * P(x_3|C_1)}{\cancel{P(x_1)}\, P(x_2)\, P(x_3)}$$

$$P(C_2|x_1x_2x_3) = \frac{P(C_2) * P(x_1/C_2) * P(x_2|C_2) * P(x_3|C_2)}{\cancel{P(x_1)\, P(x_2)\, P(x_3)}}$$

$$P(C_3|x_1x_2x_3) = \frac{P(C_3) * P(x_1/C_3) * P(x_2|C_3) * P(x_3|C_3)}{\cancel{P(x_1)\, P(x_2)\, P(x_3)}}$$

max Prob will be predicted class

$$P(yes\ x_1x_2x_3) = \frac{P(yes) * P(x_1/yes) * P(x_2|yes) * P(x_3/yes)}{\cancel{P(x_1)P(x_2)P(x_3)}} \text{ - constant}$$

$$P(No/(x_1x_2x_3)) = \frac{P(No) * P(x_1/No) * P(x_2/No) * P(x_3/No)}{\cancel{P(x_1)\cdot P(x_2)\cdot P(x_3)}} \text{ - constant}$$

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No
sunny	Hot			?

Outlook	yes	No	P(E/y)	P(E/N)
Sunny	2	3	2/9	3/5
overcast	4	0	4/9	0
rain	3	2	3/9	2/5
tota	9	5	1	1

Temperature	yes	No	P(E/y)	P(E/N)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Tota	9	5	1	1

Play (Y/No)

Y - 9 $P(Yes) = \frac{9}{14}$

N - 5 $P(No) = \frac{5}{14}$

Test day (Sunny, Hot)

(overcast, cool)

$$P(Yes/S,H) = P(Y)\cdot P(S/Y)\cdot P(H/Y)$$
$$= \cancel{9/14}_{7} \cdot \cancel{2/9} \cdot \frac{2}{9} = \frac{2}{63} = 0.031$$

$$P(N/S,H) = P(N)\cdot P(S/N)\cdot P(Hot/No)$$
$$= 5/14 \cdot 3/4 \times 2/5 = 3/35 = 0.085$$

* Normalise the prob to get in percentage for

$$P(Y / \text{Sunny}, \text{Hot}) = \frac{0.031}{0.031 + 0.085} \Rightarrow 0.27 = 27\%$$

$$P(No / \text{sunny} / \text{hot}) = \frac{0.085}{0.031 + 0.085} = 0.73 \Rightarrow 73\%$$

Prediction → The person will not go to play football.

* Why Naive Bayes ⇒
  - → Bayes theorem
  - → Why Naive → It assume that features are independent of each other
    - → No multicollinearity.

* Advantage
  - → based on probability.
  - → fast Algo.
  - → Work well with less training data
  - → Performance is good

* disadvantage
  - → assumes features are independent.
  - → It doesn't work for Regression Problem.

* Variants of Naive Bayes

① Bernoulli Naive Bayes → I.V. follow a Bernoulli, Bernoulli Naive Bayes.

f_1	f_2	f_3
M	Yes	Good
F	NO	Better
M	–	–

$$\Rightarrow \begin{bmatrix} f_1 & f_2 & f_3 & y \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & & & 0 \\ & & & 1 \end{bmatrix} \Rightarrow \text{Sparse matrix} \Rightarrow \text{leads to overfitting}$$

② Multinomial Naive Bayes

I am going to play football

inputs are ⇒ text data

Email classification ⇒ Spam | ham

input → Email body	y (output)
→ You got a discount	spam
→ Millions of Rs	Spam
→ offer letter	Ham

⇓

Numerical values ⟹ NLP technique
(vectors)

① Tf - Idf
② BoW
③ Word2Vec.

* Multinomial and Bernoulli is use for textual data in NLP.

③ Gaussian Naive Bayes

→ if features are continuous & follow ND ⇒ GNB

Age	Salary	fraud
		Yes
		No

⇓

GNB

* f_1 → BD | f_2 → ND - - - f_{10}

→ Use that variant whichever feature is higher