Naive Bayes

- * Classification

  ↳ Multi-class

- * 18th Century → Thomas Bayes ⇒ Bayes Theorm

  Probability concepts

- * "Naive" ⇒ multicollinearity X

  Assumption

Mathematical

Email	(X_1) Discount	(X_2) Win	(Y) Spam/Not spam
1	Y	Y	Spam
2	Y	N	Not spam
3	N	Y	Spam
4	Y	Y	Spam
5	N	N	Not spam
6	Y	Y	spam
7	N	Y	Not spam

Two phases → Training , Prediction

Training Phase

Step ① : Calculate probability of categories of o/p colum

$$P = \frac{\text{No of obs of that category}}{\text{Total no of observation}}$$

(a) Spam

$P(\text{spam}) = \frac{4}{7}$

(b) Not spam

$P(\text{not spam}) = \frac{3}{7}$

Step ②: Calculate Conditional Probability / Likelihood

for spam

(i) $P(X_1 = \text{Yes} / \text{spam}) = \dfrac{X_1 = \text{Yes but belong to spam}}{\text{Total no of spam}}$

$$= \frac{3}{4}$$

(ii) $P(X_1 = \text{No} / \text{spam}) = \dfrac{1}{4}$

(iii) $P(X_2 = \text{Yes} \mid \text{Spam}) = \frac{4}{4}$

(iv) $P(X_2 = \text{No} \mid \text{Spam}) = 0$

for Not Spam

(i) $P(X_1 = \text{Yes} \mid \text{Not spam}) = \frac{1}{3}$

(ii) $P(X_1 = \text{No} \mid \text{Not spam}) = \frac{2}{3}$

(iii) $P(X_2 = \text{Yes} / \text{Not spam}) = \frac{1}{3}$

(iv) $P(X_2 = \text{No} / \text{Not spam}) = \frac{2}{3}$

Training Phase Completed

Prediction Phase

$$\left.\begin{cases} X_1 = \text{Yes} \\ X_2 = \text{Yes} \end{cases}\right\}$$ new data (Bayes Theorm)

Spam

$$P(\text{Spam} \mid X_1 = \text{Yes},\ X_2 = \text{Yes}) = P(\text{spam}) \times P(X_1 = \text{Yes} \mid \text{spam}) \times P(X_2 = \text{Yes} \mid \text{Spam})$$

$$= \frac{4}{7} \times \frac{3}{4} \times \frac{4}{4} = \frac{3}{7} = \underline{\underline{0.42}}$$

Not Spam

$$P(\text{Not spam} / x_1 = \text{Yes}, x_2 = \text{Yes}) = P(\text{Not spam}) \times P(x_1 = \text{Yes} / NS) \times P(x_2 = \text{Yes} / NS)$$

$$= \frac{3}{7} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{21}$$

$$= \underline{\underline{0.04}}$$

Compare

$$\boxed{0{\cdot}42 > 0{\cdot}04}$$

(spam) (Not spam)

User new input email is a "spam"

Types

Gaussian NB

I/P feature are continuous and fallow a normal distribution

eg. Iris

Multinomial NB

I/P features are discrete.

eg. Sentiment Analysis

Bernoulli NB

IP features are binary.

eg. Spam detection

When to Use

* Multi-class classification

* Simple, fast & easy

* Best for small & medium size datasets

* Text Classification

⇒ Spam detection
Sentiment Analysis

When Not to Use

* Multicollinearity in data

* If your data has many features also most of features are irrelevant

* Imbalanced dataset