Bayes Theorem

LIKELIHOOD
The probability of "B" being True, given "A" is True

PRIOR
The probability "A" being True. This is the knowledge.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

POSTERIOR
The probability of "A" being True, given "B" is True

MARGINALIZATION
The probability "B" being True.

10

$P_{fire} = 0.01$ or 1%

$P_{Smoke} = 0.10$ or 10%

$P_{Smoke/fire} = 0.9$ or 90%

$P_{fire/Smoke} = ?$

$$\frac{0.9 \times 0.01}{0.1} \rightarrow \frac{0.9 \times \frac{1 \times 1}{100}}{1} \rightarrow 0.09$$

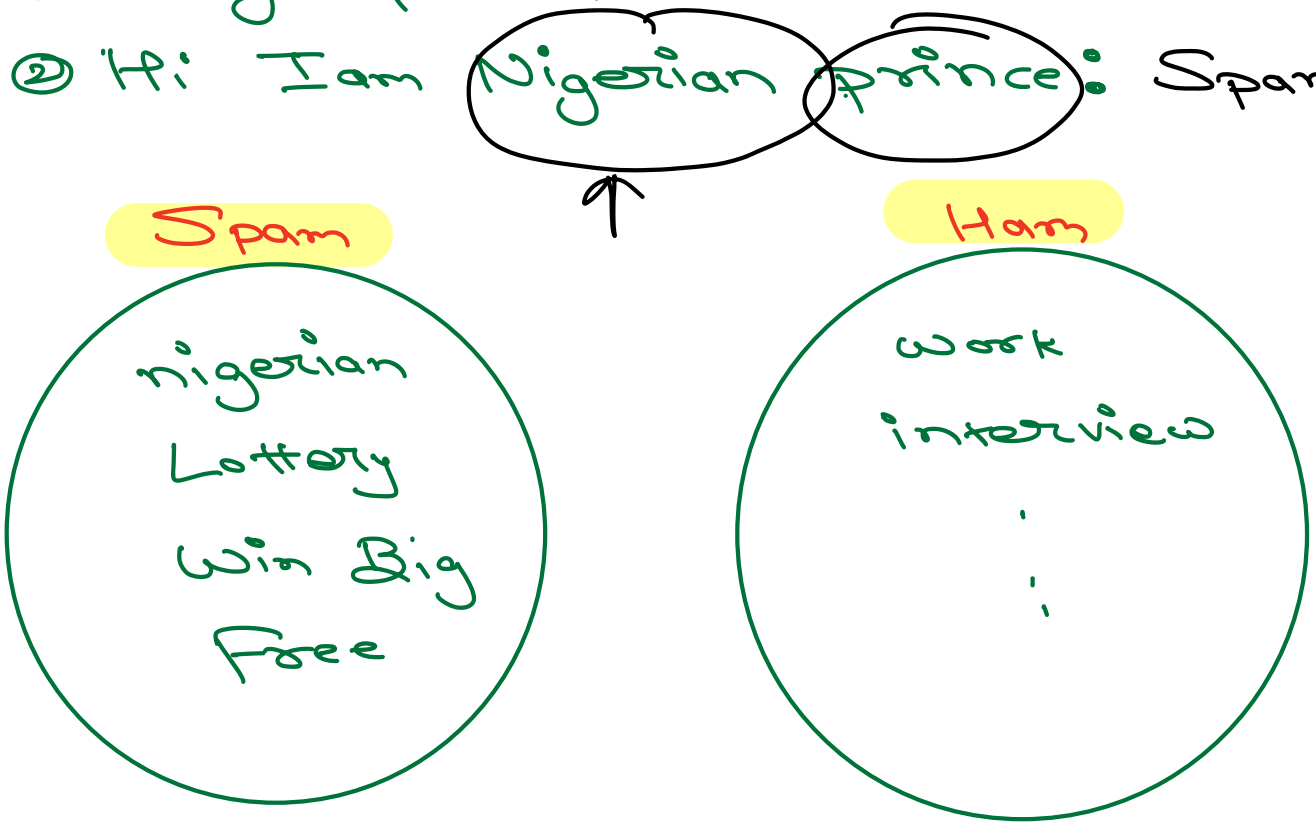
Spam Classifier

Objective : Build a Binary Text Classifier

Sample-row

① Can you please look at the Task . . . : Ham

② Hi Iam Nigerian prince : Spam



[Can, you, please]

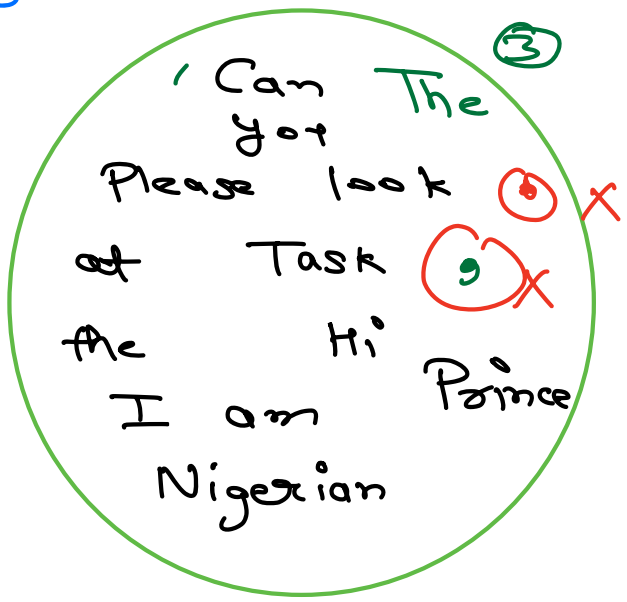
Sample-row

① Can you please look at the Task . . .

② Hi, Iam Nigerian prince.

Bag of Words

1) Set of all unique keywords in dataset



Embedding or Vectors ⇒ Text converted into Numerical features

	Can	you	please	the	The	Prince	,
①	1	1	1	1	0	0	
②	0	0	0	0	0	1	
...					1		

13

1000 rows ⇒ 1,00,000 (features)

Text Cleaning

⓪ Convert Sentences into words ⇒ Tokenization

① Convert all text to lowercase

② Remove Non-alphabetical features

③ Remove Stopwords ⇒ The, How, where

Try: Keep alphanumeric
Keep stopwords

Mathematical intuition of Naive Bayes

Sent 1 $\rightarrow$ [$w_1, w_2 \dots w_n$] $\rightarrow$ Class 0 Ham / Class 1 Spam

* Ham

$$P(y=1 / (w_1, w_2, \dots w_n))$$

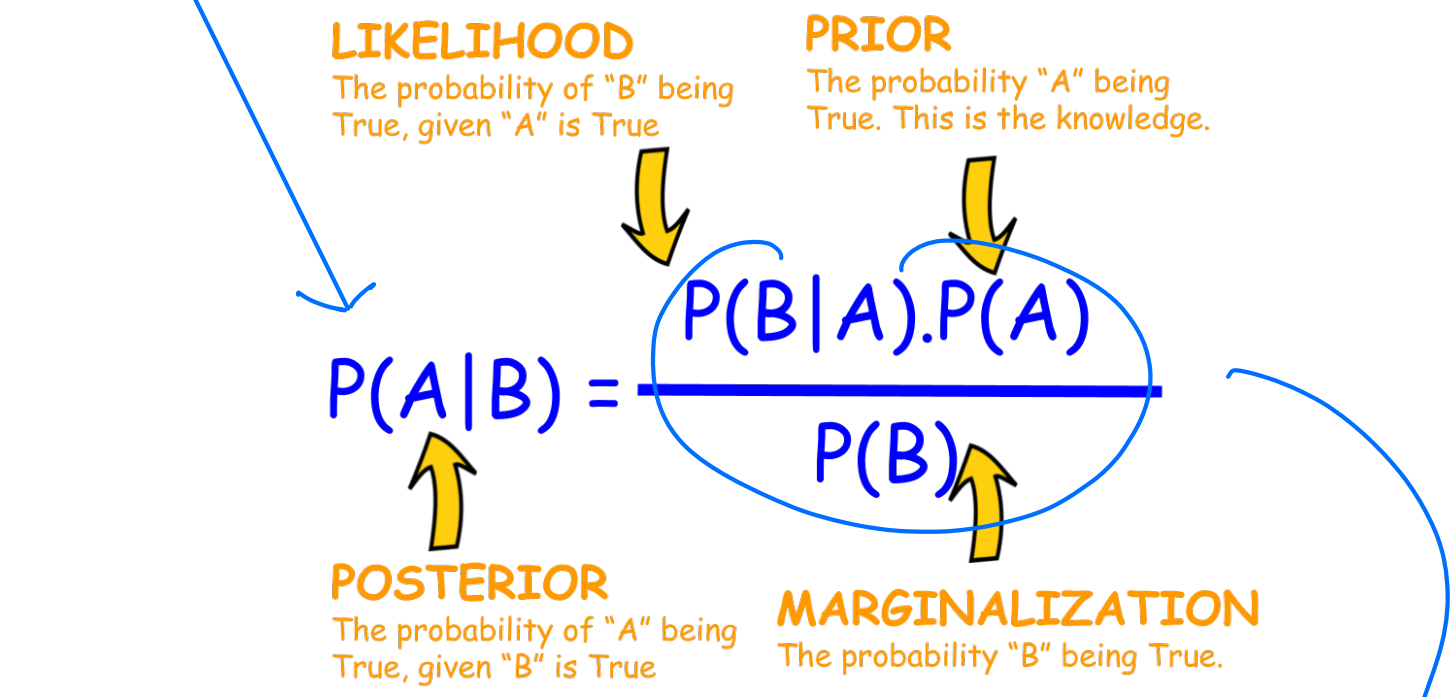
Conditional probability of $y \rightarrow 0$ given words present in sent

* Spam

$$P(y=0 / (w_1, w_2 \dots w_n))$$

Conditional probability of $y \rightarrow 1$ given words present in sent

$P(y = 1 / (w_1, w_2, \dots w_n))$



→ P(A) → prior → $\frac{\#\, y==1}{\#\, y}$

→ P(B) → marginalization → $\frac{(w_1, w_2, w_3)}{\#\text{Total Sents}}$

* $P(B/A)$ → likelihood

→ $P(w_1, w_2, \dots w_n / y==1)$

→ All hams where $(w_1, w_2 \dots w_n)$ occur together / Total Hams

Joint probability

$$P(w_1, w_2, \ldots w_n / y==1) \Rightarrow P(w_1 / y==1) \times P(w_2 / y==1, w_1) \times P(w_3 / y==1, w_1, w_2) \times \ldots \times P_n / y==1, w_1, w_2, w_3 \ldots w_{n-1}$$

* Naive Assumption: All words are independent of each other

$$P(w_2 / y==1, w_1) \Rightarrow P(w_2 / y==1)$$

Happy New

$$P(New / y==1, Happy) \Rightarrow P(New / y=1)$$

$P(w_1, w_2, \ldots w_n / y==1) \Rightarrow P(w_1/y==1) \times P(w_2/y==1) \times P(w_3/y==1) \times \ldots \times P(w_n/y==1)$

$\prod_{i=1}^{d} P(w_i/y=1)$ ←

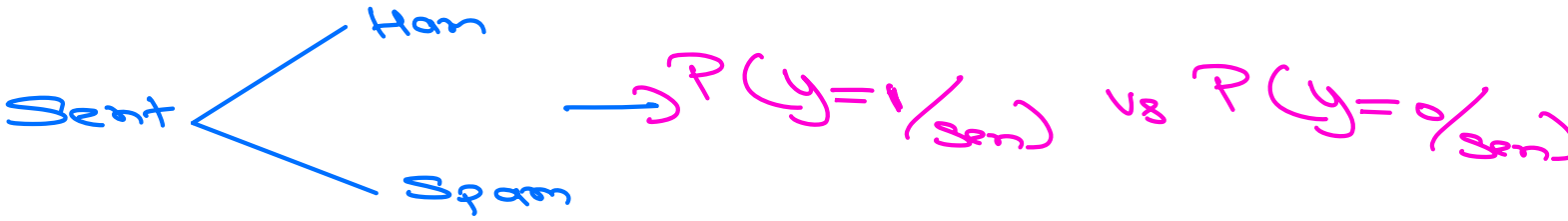
$P(w_i/y=1) \Rightarrow \frac{n_i}{n_1} \Rightarrow \frac{\text{all Spams containing } w_i}{\text{Total Spams}}$

$P(y=1/\text{Sent}) \Rightarrow \prod_{i=1}^{d} \frac{P(w_i/y=1) \times P(y=1)}{P(w_1, w_2, w_3, \ldots)}$

?

$P(y=0/\text{Sent}) \Rightarrow \prod_{i=1}^{d} \frac{P(w_i/y=0) \times P(y=0)}{P(w_1, w_2, w_3, \ldots)}$

?



$$\prod_{i=1}^{d} \frac{P(w_i/y=1) \times P(y=1)}{P(w_1, w_2, w_3, \ldots)} \quad \text{vs} \quad \prod_{i=1}^{d} \frac{P(w_i/y=0) \times P(y=0)}{P(w_1, w_2, w_3, \ldots)}$$

0.8 ~~0.32~~ $\qquad$ 0.6 ~~0.3~~

* Limitations:
  1) It doesn't understand the meaning of text
  2) Order of words doesn't

* Semantic
* Contextual

Text ⊕ word is not present in Vocab

$P(y=1 / w_1, w_2, w_3) \Rightarrow$

$P(w_1 / y==1) * P(w_2 / y==1) * P(w_3 / y==1)$

if w_3 is not present $\rightarrow 0$

$P(y=1 / w_1, w_2, w_3) \Rightarrow 0$

Handle Outlier ⇒ word Not present in Vocab

$P(\text{word\_unknown} / y=1) \Rightarrow t$

* Smoothing ⇒ Laplace Smoothing

$$P(w_j / y=1) \Rightarrow \frac{\#n_{j,1} + \alpha}{\#n_1 + \alpha C}$$

Distinct Possible Values of w_j

$w_y \Rightarrow [0, 13$

$$P(w_y / y=1) \Rightarrow \frac{\#_{y,1} + \alpha}{\#_{y,1} + 2\alpha}$$

α is hyperparameter that Controls Smoothing

$$\frac{0.01 + 1}{100 + 2}$$

① w_y not present

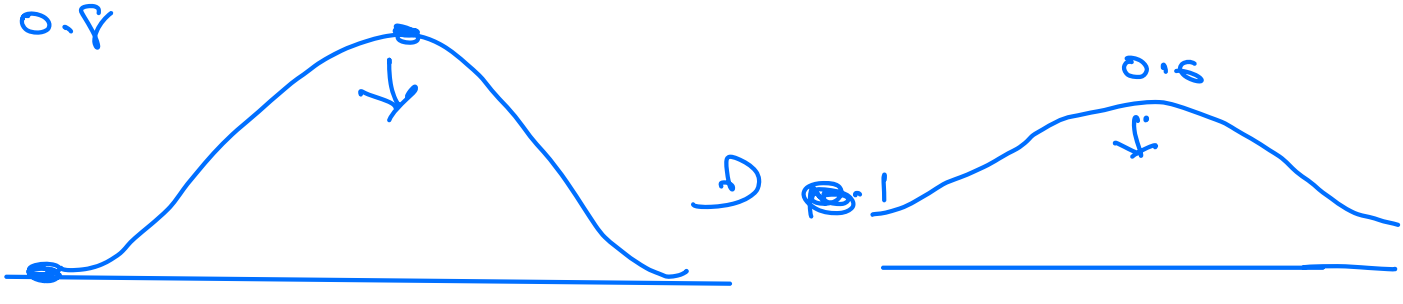
$\alpha = 1 \Rightarrow \frac{0+1}{100+2x1} \Rightarrow \frac{1}{102} \Rightarrow$

$\downarrow$

Total Spams

② Spam Class is Not present in:

$\alpha \Rightarrow \frac{0+1}{0+2} \Rightarrow 0.5$



Bernoulli vs. Multinomial