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.

Q

$P_{fire} = 0.01$ or 1%

$P_{smoke} = 0.10$ or 10%

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

$P_{fire/smoke} = ?$

$$\frac{0.9 \times 0.01}{0.1} \Rightarrow \frac{0.9}{1} \times \frac{1 \times 1\cancel{0}}{10\cancel{0}}$$

$$\Rightarrow 0.09$$

Spam Classifier

Objective : Build a Binary Text Classifier

Sample-row

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

② Hi I am Nigerian prince : Spam

Spam

nigerian

Lottery

Win Big

Free

Ham

work

interview

.

.

[Can, you, please]

Sample-row

① Can you please look at the Task . . .

② Hi, I am Nigerian prince.

Bag of Words

⇒ Set of all unique keywords in dataset

' Can you The
Please look .
at Task ?
the Hi
Prince
I am
Nigerian

Embeddings
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 1s $[w_1, w_2 --- w_n]$

Class 0 Ham

Class 1 Spam

* Ham

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

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

* Spam

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

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

$$P\left(y=1/(w_1, w_2, --- w_n)\right)$$

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.

⇒ $P(A)$ ⇒ prior ⇒ $\dfrac{\#\, y==1}{\#\, y}$

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

* $P(B/A)$ ⇒ likelihood

⇒ $P(w_1, w_2, --- w_n / y==1)$

⇒ All hams where $(w_1, w_2 -- w_n)$ occur together / Total Hams

Joint probability

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

$\boxed{\text{* Naive Assumption: All words are independent of each other}}$

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

Happy New

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

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

$$\boxed{\prod_{i=1}^{d} P(w_i/y=1)} \leftarrow$$

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

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

?

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

?

Sent → Ham / Spam $\longrightarrow P(y=1/\text{sen}) \text{ vs } P(y=0/\text{sen})$

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

$$\frac{0.8}{\cancel{0.3}} \qquad \frac{0.6}{\cancel{0.3}}$$

good ∈ 3

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

$\boxed{\begin{array}{l}\text{* Semantic} \\ \text{* Contextual}\end{array}}$

Text Word is not present in Vocab

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

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

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

* Smoothing ⇒ Laplace Smoothing

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

distinct Possible Values of w_j

$w_y \Rightarrow \{0, 1\}$

$$P(w_y / y=1) \Rightarrow \frac{\#n_{y,1} + \alpha}{\#n_1 + 2\alpha}$$

α is hyperparameter that Controls Smoothening

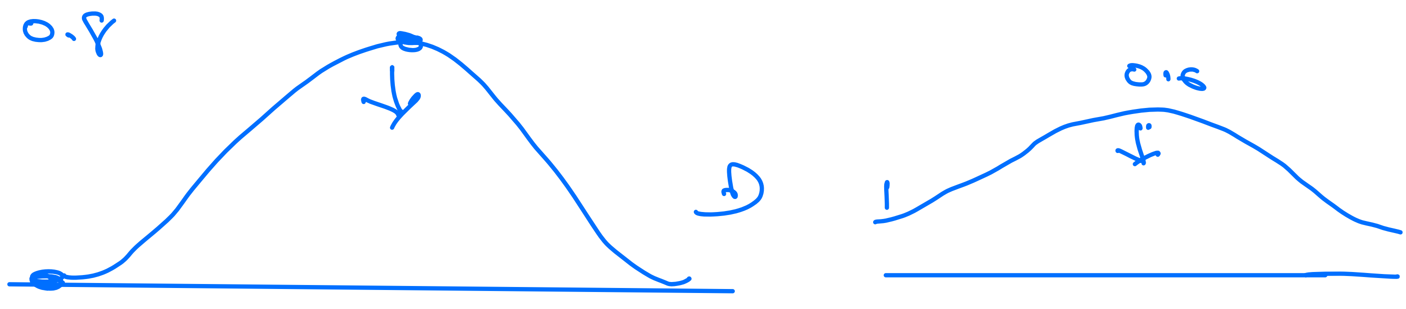
① w_y not present

$\alpha = 1 \Rightarrow \dfrac{0+1}{\underbrace{100}_{\text{Total Spams}} + 2\times 1} \Rightarrow \dfrac{1}{102} \Rightarrow \dfrac{0.01+1}{100+2}$

② Spam Class is Not present

n_i

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



Bernoulli vs Multinomial

Our features have only two possible distinct values (0, 1)

Bernoulli NB

	good	the	yes	no	- - -	y_i
x_i	0	1	0	1	- -	0

$\omega_{good}/_{y==1}$

$\times$

$\omega_{the}/_{y==1}$

$\omega_{yes}/_{y==1}$

⋮

Features can have k discrete values where k is frequency

→ Multinomial NB

	good	the	yes	no	$good_2$	$good_3$
x_1	1	0	0	0	1	0
x_2	1	0	0	0	0	0

x_1 → Good → 2 times

x_2 → Good → 1 times

$\omega_{good_1}/_{y==1}$

$\omega_{good_2}/_{y==1}$

$\omega_{good_3}/_{y==1}$

$\omega_{the}/_{y==1}$

$\omega_{yes}/_{y==1}$

Classes

2) OVR ~~X~~

$C \leftarrow$ Promotional, Social, Main Inbox

1) Directly calculate Likelihood

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

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

$$\frac{\prod_{i=1}^{d} P(w_i / y=2) \times P(y=2)}{P(w_1, w_2, w_3, \ldots)}$$