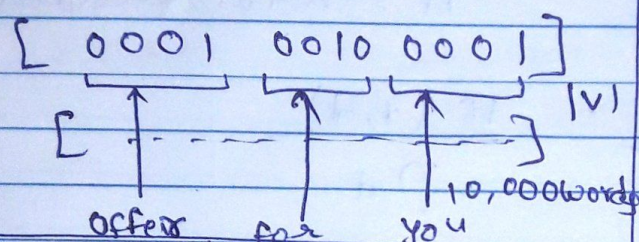


# ⑮ Multivariate Bernoulli vs Multinomial Naïve Bayes

## Multivariate Bernoulli Event model

① "offer for you"



$$p(x_i = 1 | y = \text{spam}) = p(x_i | y = \text{spam})$$

$y = \text{spam}$

$$p(x_i = 0 | y = \text{spam}) = 1 - p(x_i | y = \text{spam})$$

Likelihood

$$\prod_{i=1}^{|V|} p(x_i | y = \text{spam})^{x_i} (1 - p(x_i | y = \text{spam}))^{1-x_i}$$

Formula:

$$p(x_i | y = c)$$

=  $\frac{\text{count of docs having } C \text{ and containing } x_i + 1}{\text{count of docs in class } C + 2}$

+1 = Laplace smoothing  
+2 = Bernoulli's model

## Multinomial Event Model

① formula:

$$\hat{p}(x_i | w) = \frac{\sum_{dec} x_i \cdot \text{dec} + |\alpha|}{\sum N_{dec} + |\alpha| \cdot v}$$

$\alpha$  = Laplace smoothing

$|\alpha|$  = Hyper parameter

$v$  = vocab size

$N$  = No. of docs

"offer for you"

$$\prod p(x_i | y) \cdot p(y)$$

comes from multinomial distribution



$$\prod_{i=1}^{|V|} P(x_i | y) P(y)$$

included word  $i$   $j$  not included

[0 0 1 0 0 0 0 1]

$$[ \dots P(x_i | y) \dots (1 - P(x_j | y)) \dots ]$$

← |V| times  
vocab size times

- Can't

Here

$$P(\text{offering} | y = \text{"spam"})$$

$$= \prod_{i=1}^n$$

$p \ni x = \langle x_1, x_2, x_3 \rangle$   
query

$$= \prod_{j=1}^n P(x_j | y = S)$$

$$P(x_2 | y = S)$$

$$P(x_3 | y = S)$$

← (x size) →

- Can differ for every document.

- vocab large then this model performs better.

- But also depends upon choice of features:  
like %

- Bigrams

- Unigrams

- TFIDF weight

(Term frequency  
inverse document  
frequency)

- TF-IDF weight  
in place of term frequency