(15) Multivariate Bernoulli vs Multinomial Naive Bayes:

Multivariate Bernoulli Event model	Multinomial Event Model

Multivariate Bernoulli Event model

① "Offer for you"

[0001 0010 0001] $|V|$

[. . . . . .] 10,000 words

offer for you

$$\boxed{p(x_i = 1 \mid y = s) = p(x_i \mid y = spam)}$$

s = spam

$$\boxed{p(x_i = 0 \mid y = spam) = 1 - p(x_i \mid y = spam)}$$

$$L(\theta) = \prod_{i=1}^{|V|} p(x_i \mid y = spam)^{b} \cdot (1 - p(x_i \mid y = spam))^{1-b} \cdot p(spam)$$

Likelihood

formula:

$$p(x_i \mid y = c) = \frac{\text{count of docs having } C \text{ and containing } x_i + 1}{\text{count of docs in class } c + 2}$$

+1 = laplace smoothing

+2 = bernoulli's model

Multinomial Event Model

① formula:

$$\hat{p}(x_i \mid \omega) = \frac{\sum tf(x_i, d \in c) + |\alpha| 1}{\sum N_{d \in c} + |\alpha| V}$$

1 = laplace smoothing

$|\alpha|$ = Hyper parameter

V = vocab size

N = No. of docs

"Offer for you"

$$\underbrace{\prod p(x_i \mid y)}_{\text{comes from multinomial distribution}} \; p(y)$$

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

included word — i; not included — j

$$[\,0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\,]$$

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

$\longleftrightarrow$ $|V|$ times

vocab size times

- Can't

Here

$$P(\text{offers} \mid y = \text{"spam"})$$

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

$P \ni x = \langle x_1, x_2, x_3 \rangle$

query

$$= \prod_{j=1}^{n} P(x_1|y=S) \cdot P(x_2|y=S)\, P(x_3|y=S)$$

$\longleftrightarrow$ (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