



Benoulli Naïve Bayes for Text Categorization

School of Information Studies
Syracuse University

Benoulli Model

Multinomial model estimates the probability of the event that one of the N unique words occurs in a position.

Benoulli model estimates the probability of the event that a word is present or absent.

Comparison 1: Priors Are the Same

Table 13.1: Data for parameter estimation examples.

	docID	words in document	in $c = \text{China?}$ —
training set	1	Chinese Beijing Chinese	yes
	2	Chinese Chinese Shanghai	yes
	3	Chinese Macao	yes
	4	Tokyo Japan Chinese	no
test set	5	Chinese Chinese Chinese Tokyo Japan	?

Priors are the same: $\frac{3}{4}$ and $\frac{1}{4}$

Comparison 2: Conditional Probabilities Are Different

In multinomial model, conditional probabilities are based on word frequency, smoothed over the vocabulary.

$$\hat{P}(\text{Chinese}|c) = (5 + 1) / (8 + \textcircled{6}) = 6/14 = 3/7$$

In Benoulli model, conditional probabilities are based on document frequency, smoothed over two events, either presence or absence.

$$\hat{P}(\text{Chinese}|c) = (3 + 1) / (3 + \textcircled{2}) = 4/5$$

Comparison 3: Posteriors Are Different

test set	5	Chinese Chinese Chinese Tokyo Japan	?
----------	---	-------------------------------------	---

Posterior in Benoulli model

$$\begin{aligned}\hat{P}(c|d_5) &\propto \hat{P}(c) \cdot \hat{P}(\text{Chinese}|c) \cdot \hat{P}(\text{Japan}|c) \cdot \hat{P}(\text{Tokyo}|c) \\ &\quad \cdot (1 - \hat{P}(\text{Beijing}|c)) \cdot (1 - \hat{P}(\text{Shanghai}|c)) \cdot (1 - \hat{P}(\text{Macao}|c)) \\ &= 3/4 \cdot 4/5 \cdot 1/5 \cdot 1/5 \cdot (1-2/5) \cdot (1-2/5) \cdot (1-2/5) \\ &\approx 0.005\end{aligned}$$

Posterior in multinomial model

$$\begin{aligned}\hat{P}(c|d_5) &\propto 3/4 \cdot (3/7)^3 \cdot 1/14 \cdot 1/14 \approx 0.0003. \\ \hat{P}(\bar{c}|d_5) &\propto 1/4 \cdot (2/9)^3 \cdot 2/9 \cdot 2/9 \approx 0.0001.\end{aligned}$$

Which NB to Choose?

McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naïve Bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41–48).

- Bernoulli for shorter texts (can take Boolean representation only)
- Multinomial for longer texts (can take word count, tfidf)