

1.

“360 ”
16 50 10MB
500 100MB 500 10 5000 5000 5000 bits 7MB

2.

- - +60389585
 - QQ
 -
 -
- 2

1000 “ ” 1000 Stop words n

3.

A “ ” B “ ” 10

$$P(B|A)=\frac{2}{3} \times \frac{10}{4} = \frac{5}{2}$$

$$P(B)=\frac{4}{10} \times \frac{10}{3} = \frac{4}{3}$$

$$P(A)=\frac{3}{10} \times \frac{4}{2} = \frac{6}{5}$$

$$P(A|B)=\frac{2}{4} \times \frac{3}{2} = \frac{3}{4}$$

n

n



\$W_{1}\$ \$W_{2}\$ \$W_{3}\$... \$W_{n}\$ \$ n

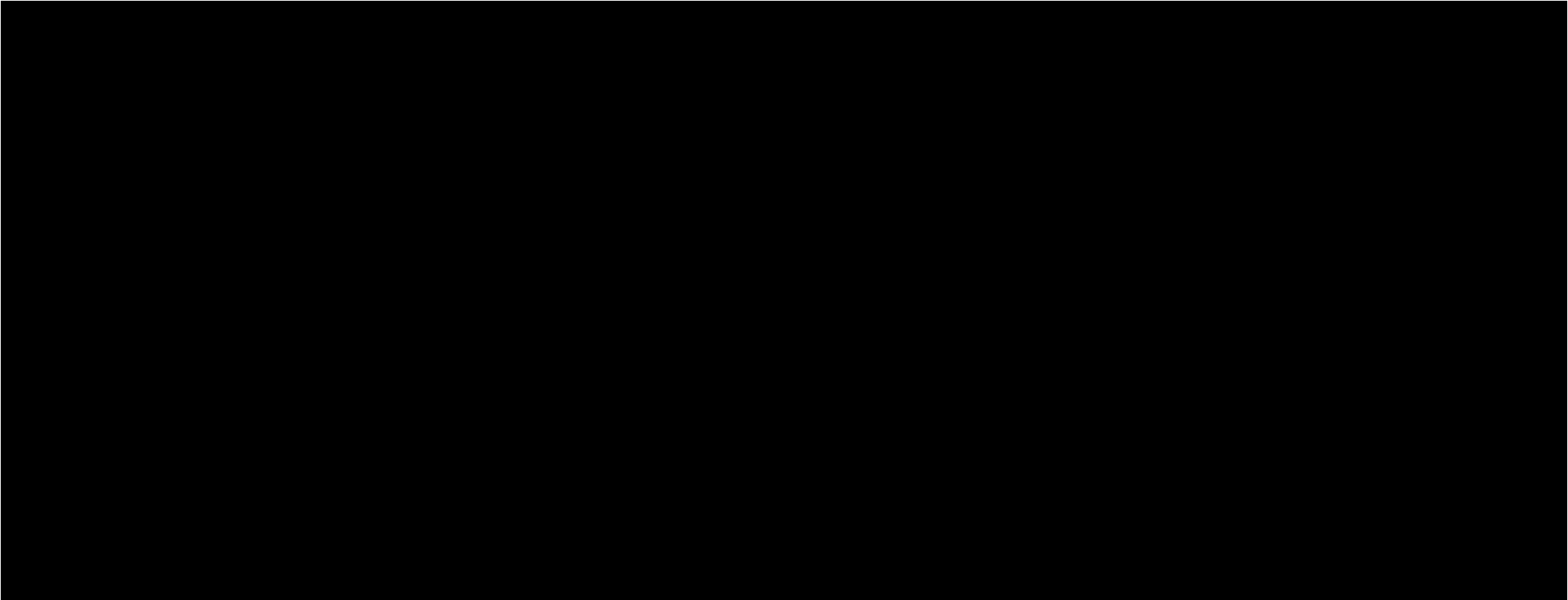
y/x

\$W_{1}\$ \$W_{2}\$ \$W_{3}\$... \$W_{n}\$ \$ n

x

y

\$W_{1}\$ \$W_{2}\$ \$W_{3}\$... \$W_{n}\$



P \$W_{1}\$ \$W_{2}\$ \$W_{3}\$... \$W_{n}\$ |

$P(A*B) = P(A)*P(B)$

A	B	A	P(A)	p1	B	P(B)	p2	P(A*B)	P(A)*P(B)
P W1 W2 W3 ... Wn									

$$P \quad SW_{\{i\}}\$ \quad | \quad SW_{\{i\}}\$$$

$$P \quad SW_{\{1\}}\$ \quad SW_{\{2\}}\$ \quad SW_{\{3\}}\$ \quad \dots \quad SW_{\{n\}}\$ \quad |$$

$$P$$

$$P \quad SW_{\{1\}}\$ \quad SW_{\{2\}}\$ \quad SW_{\{3\}}\$ \quad \dots \quad SW_{\{n\}}\$$$

$$SW_{\{1\}}\$ \quad SW_{\{2\}}\$ \quad SW_{\{3\}}\$ \quad \dots \quad SW_{\{n\}}\$ \quad n$$

$$p1 \quad p2 \quad 10$$

$$y$$

$$SW_{\{i\}}\$ \quad x$$

$$x/y$$

$$p1 \quad p2$$

$$p1$$

$$p1 \quad p2$$

$$P \quad \$W_{\{1\}}\$ \quad \$W_{\{2\}}\$ \quad \$W_{\{3\}}\$ \quad \dots \quad \$W_{\{n\}}\$$$

$$p1 \quad p2$$

$$p1 \quad p2$$

$$p1/p2$$

“ ”

- slvher 2019-01-11 09:08:58

NLP

anti-spam

logistic regression

naive-bayes

ml

[4]

- 2019-01-11 09:03:01

[4]

- C_love 2019-01-14 11:00:45

$P(W1W2...Wn|$)

$P(W1W2...Wn)$

[1]

- ban 2019-01-13 19:23:45

<https://www.jianshu.com/p/5cf3a155b2f0>

[1]

- 2019-01-11 12:57:10

[1]

- 2019-01-11 07:44:30

[1]

- Clement 2019-01-21 20:15:58

- William 2019-01-21 12:59:36

1.

naive

feature

feature

naive

“ ”

online

(*p

2.

0

laplacian smoothing

0

sample

7:1

• ken 2019-01-15 10:01:37

W

• spark 2019-01-14 09:49:58

2/3

.....

• Alexis 2019-01-13 13:40:55

hash

hashmap

• Geek_477c02 2019-01-13 13:38:35

P Wi | Wi

wi 0 0

• 2019-01-11 20:48:09

NLP

NLP

• Kudo 2019-01-11 14:03:30

w1, w2, ..., wn

• 2019-01-11 13:45:29

W_I

()

P1+P2=1 P1