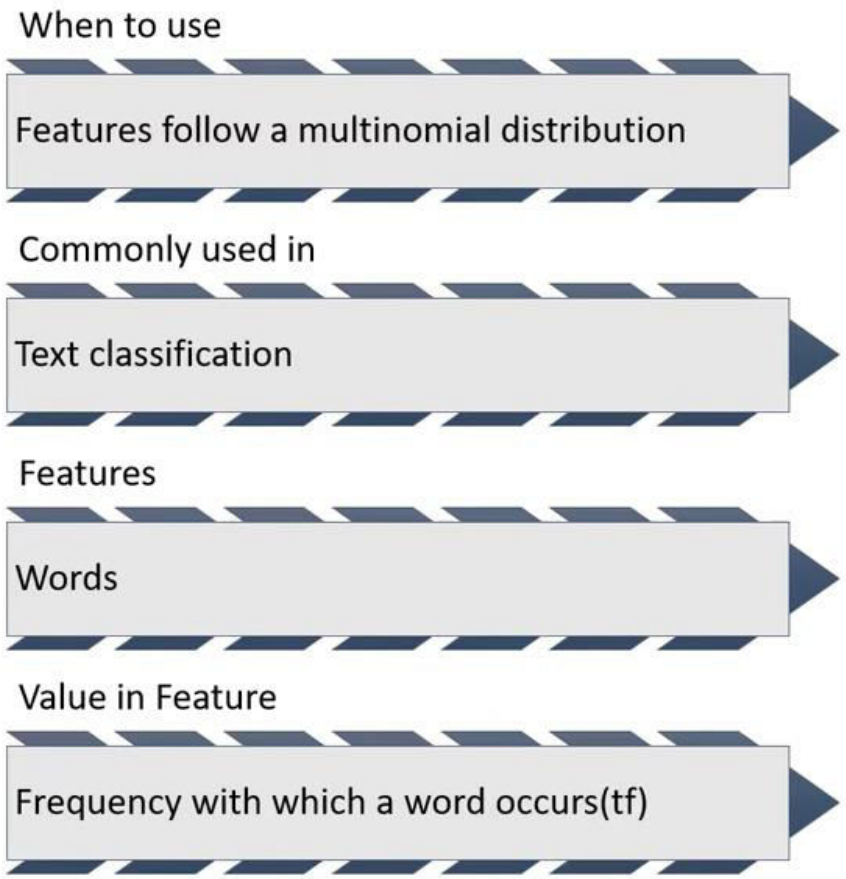
Sentiment Analysis using Multinomial and Bernoulli Naïve Bayes

Multinomial Naïve Bayes

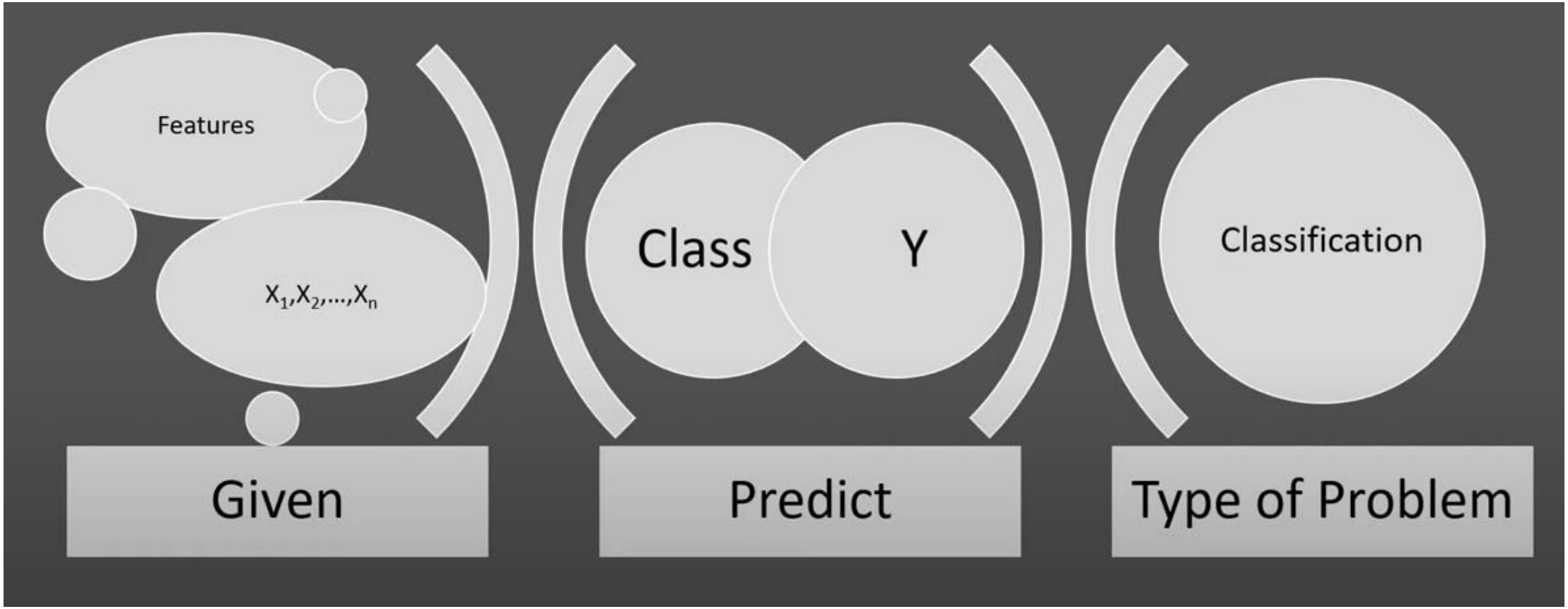


Sentiment Analysis using Multinomial NB

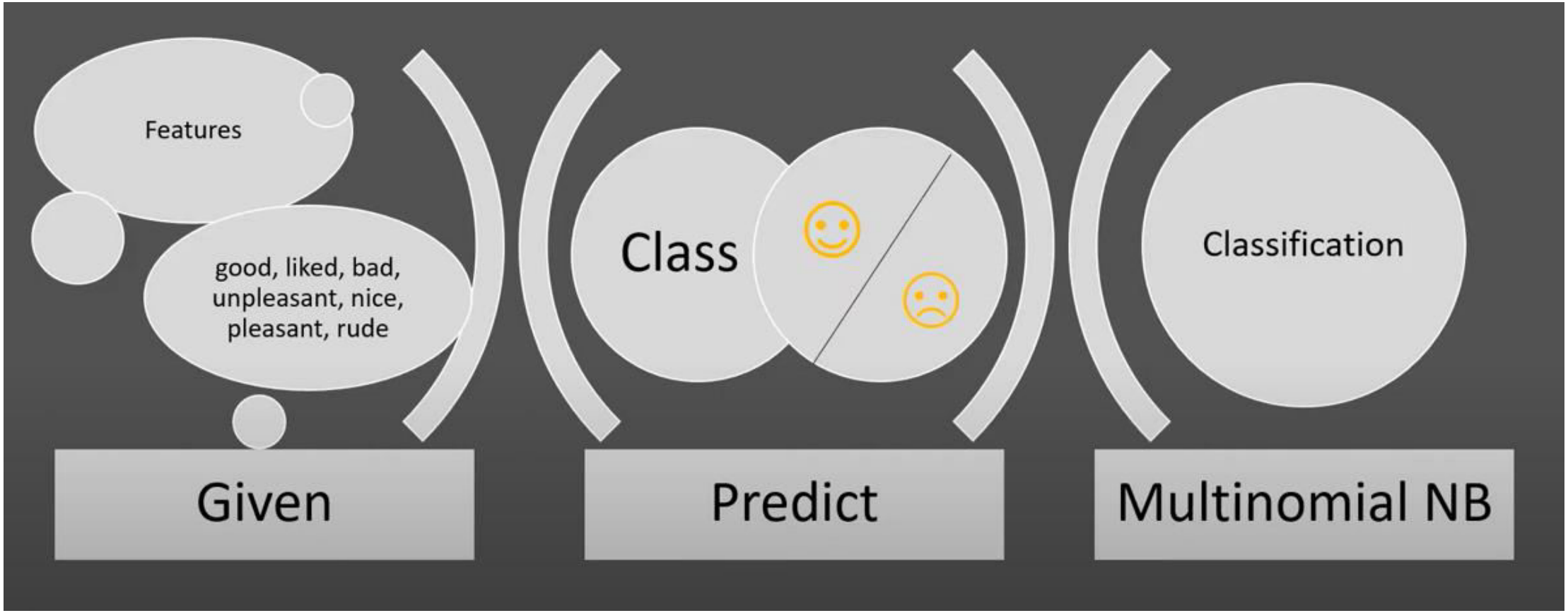
Dataset

	REVIEW	SENTIMENT
Training Set	The rooms were good and I liked the location since it was good	
	The hotel was very bad and the stay was unpleasant	
	Liked the huge play area and the food was nice	
	The stay was good and pleasant	
	The location was good but was bad overall because The staff were rude	
Test Set	The rooms were good and the staff were nice	?

Sentiment Analysis using Multinomial NB



Sentiment Analysis using Multinomial NB



Sentiment Analysis using Multinomial NB

Features

DOCUMENT	REVIEW	CLASS
1	The rooms were good and I liked the location since it was good	+
2	The hotel was very bad and the stay was unpleasant	-
3	Liked the huge play area and the food was nice	+
4	The stay was good and pleasant	+
5	The location was good but was bad overall because The staff were rude	-

Term Document Matrix

Features \ Documents	D1	D2	D3	D4	D5
good	2	0	0	1	1
liked	1	0	1	0	0
bad	0	1	0	0	1
unpleasant	0	1	0	0	0
nice	0	0	1	0	0
pleasant	0	0	0	1	0
rude	0	0	0	0	1

Number of Features in '+' = 7
good : 3, liked : 2, nice : 1, pleasant : 1

Number of Features in '-' = 5
good: 1, bad :2, unpleasant :1, rude :1

Sentiment Analysis using Multinomial NB

Probability of getting class c

- $P(\text{Class=c}|\text{Features}) = \frac{P(\text{Features}|\text{Class=c}) \cdot P(Class=c)}{P(Features)} \alpha\ P(\text{Features}|\text{Class} = \text{c}) \cdot P(Class = c)$

Probability of getting class '+'

$P(\text{Y=+}|\text{X=good, liked, bad, unpleasant, nice, pleasant, rude}) \ \alpha$
$P(X = \text{good, liked, bad, unpleasant, nice, pleasant, rude}|Y) \cdot P(Y = +)$

Probability of getting class '-'

$P(\text{Y=-}|\text{X=good, liked, bad, unpleasant, nice, pleasant, rude}) \ \alpha$
$P(X = \text{good, liked, bad, unpleasant, nice, pleasant, rude}|\text{Y}) \cdot P(Y = -)$

Sentiment Analysis using Multinomial NB

$$P(Y = +)$$

$$\frac{Number\ of\ samples\ of\ the\ class\ '+'}{total\ number\ of\ samples} = \frac{3}{5}$$

$$P(Y = -)$$

$$\frac{Number\ of\ samples\ of\ the\ class\ '-'}{total\ number\ of\ samples} = \frac{2}{5}$$

Sentiment Analysis using Multinomial NB

$P(X=x_1, x_2, \ldots, x_n|Y) = \prod_{i=1}^{n} P(x_i|Y)$

P(good, ..., pleasant, rude|Y='+'|) = P(good|Y='+') *... * P(pleasure|Y='+')* P(rude|Y='+')

P(good, ..., pleasant, rude|Y='-') = P(good|Y='-') *... * P(pleasure|Y='-')* P(rude|Y='-')

$P(X=x_i|Y='c')$

$$\frac{Frequency\ with\ which\ feature\ x_i\ occurs\ in\ class\ 'c'}{Total\ number\ of\ features\ in\ class\ 'c'}$$

Sentiment Analysis using Multinomial NB

Term Document Matrix

Documents / Features	D1 +	D2 -	D3 +	D4 +	D5 -
good	2	0	0	1	1
liked	1	0	1	0	0
bad	0	1	0	0	1
unpleasant	0	1	0	0	0
nice	0	0	1	0	0
pleasant	0	0	0	1	0
rude	0	0	0	0	1

Number of Features in '+' = 7
good : 3, liked : 2, nice : 1, pleasant : 1

Number of Features in '-' = 5
good: 1, bad :2, unpleasant :1, rude :1

P(Y=+) = 3/5

P(Y=-) = ?

Feature	P(Feature|+)	P(Feature|-)
good	$\frac{3}{7}$	$\frac{1}{5}$
liked	$\frac{2}{7}$	$\frac{0}{5}$
bad	$\frac{0}{7}$	?
unpleasant	$\frac{0}{7}$	$\frac{1}{5}$
nice	$\frac{1}{7}$	$\frac{0}{5}$
pleasant	?	$\frac{0}{5}$
rude	$\frac{0}{7}$	$\frac{1}{5}$

Sentiment Analysis using Multinomial NB

Classification Model

P(Y=+) = 3/5		
P(Y=-) = 2/5		
Feature	**P(Feature|+)**	**P(Feature|-)**
good	$\frac{3}{7}$	$\frac{1}{5}$
liked	$\frac{2}{7}$	$\frac{0}{5}$
bad	$\frac{0}{7}$	$\frac{2}{5}$
unpleasant	$\frac{0}{7}$	$\frac{1}{5}$
nice	$\frac{1}{7}$	$\frac{0}{5}$
pleasant	$\frac{1}{7}$	$\frac{0}{5}$
rude	$\frac{0}{7}$	$\frac{1}{5}$

Sentiment Analysis using Multinomial NB

$$P(X=x_1, x_2, \ldots, x_n|Y) = \prod_{i=1}^{n} P(x_i|Y)$$

Laplace Smoothing

P(Feature|class=c) is calculated as $$\frac{Number\ of\ times\ the\ feature\ occurs\ in\ c+1}{Total\ number\ of\ features\ in\ class\ 'c'\ +number\ of\ features}$$

Sentiment Analysis using Multinomial NB

Classification Model with Laplace Smoothing

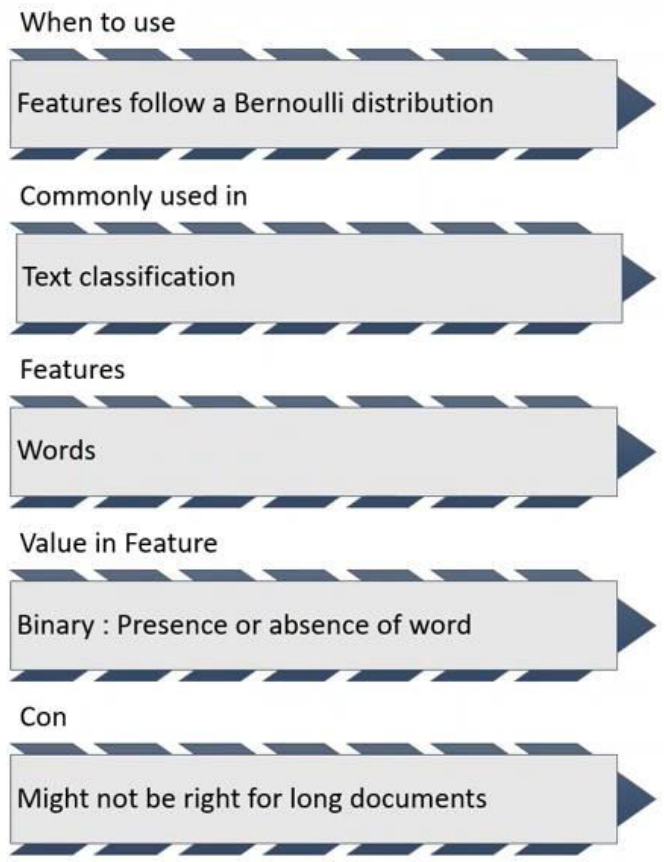
P(Y=+) = 3/5		
P(Y=-) = 2/5		
Feature	**P(Feature|+)**	**P(Feature|-)**
good	$\frac{3+1}{7+7} = \frac{4}{14}$	$\frac{1+1}{5+7} = \frac{2}{12}$
liked	$\frac{2+1}{7+7} = \frac{3}{14}$	$\frac{0+1}{5+7} = \frac{1}{12}$
bad	$\frac{0+1}{7+7} = \frac{1}{14}$	$\frac{2+1}{5+7} = \frac{3}{12}$
unpleasant	$\frac{0+1}{7+7} = \frac{1}{14}$	$\frac{1+1}{5+7} = \frac{2}{12}$
nice	$\frac{1+1}{7+7} = \frac{2}{14}$	$\frac{0+1}{5+7} = \frac{1}{12}$
pleasant	$\frac{1+1}{7+7} = \frac{2}{14}$	$\frac{0+1}{5+7} = \frac{1}{12}$
rude	$\frac{0+1}{7+7} = \frac{2}{14}$	$\frac{1+1}{5+7} = \frac{2}{12}$

Sentiment Analysis using Multinomial NB

Classify : “The rooms were good and the staff were nice”

$P(+\mid good,nice)$		$P(-\mid good,nice)$
$P(good,nice\mid +)*P(+)$		$P(good,nice\mid -)*P(-)$
$= P(good\mid +)*P(nice\mid +)*P(+)$		$= P(good\mid -)*P(nice\mid -)*P(-)$
$= \frac{4}{14}*\frac{2}{14}*\frac{3}{5}$		$= \frac{2}{12}*\frac{1}{12}*\frac{2}{5}$
$= 0.2449$	$>$	$= 0.0055$

Bernoulli Naïve Bayes



Sentiment Analysis using Bernoulli NB

Features

DOCUMENT	REVIEW	CLASS
1	The rooms were good and I liked the location since it was good	+
2	The hotel was very bad and the stay was unpleasant	-
3	Liked the huge play area and the food was nice	+
4	The stay was good and pleasant	+
5	The location was good but was bad overall because The staff were rude	-

Term Document Matrix

Features \ Documents	D1	D2	D3	D4	D5
good	1	0	0	1	1
liked	1	0	1	0	0
bad	0	1	0	0	1
unpleasant	0	1	0	0	0
nice	0	0	1	0	0
pleasant	0	0	0	1	0
rude	0	0	0	0	1

Sentiment Analysis using Bernoulli NB

$P(X=x_1, x_2, \ldots, x_n|Y) = \prod_{i=1}^{n} P(x_i|Y)$

P(good, ..., pleasant, rude|Y='+'|) = P(good|Y='+') *... * P(pleasure|Y='+')* P(rude|Y='+')

P(good, ..., pleasant, rude|Y='-') = P(good|Y='-') *... * P(pleasure|Y='-')* P(rude|Y='-')

$P(X=x_i|Y='c')$

$$\frac{Number\ of\ documents\ of\ class\ 'c'\ with\ feature\ x_i}{Total\ number\ of\ documents\ of\ class\ 'c'}$$

Sentiment Analysis using Bernoulli NB

Term Document Matrix

Documents / Features	D1 +	D2 -	D3 +	D4 +	D5 -
good	1	0	0	1	1
liked	1	0	1	0	0
bad	0	1	0	0	1
unpleasant	0	1	0	0	0
nice	0	0	1	0	0
pleasant	0	0	0	1	0
rude	0	0	0	0	1

Number of documents in '+' = 3
D1, D3, D4

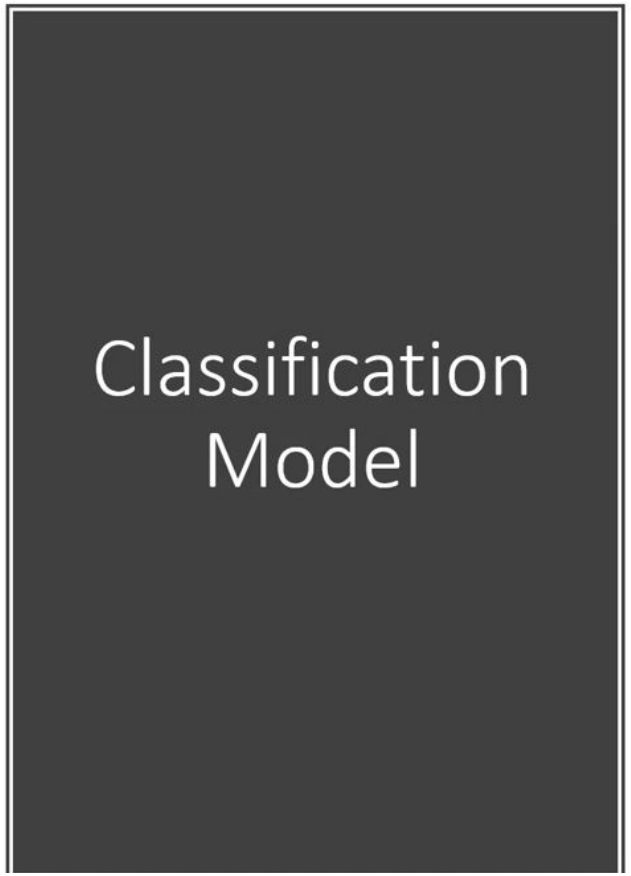
Number of documents in '-' = 2
D2, D5

P(Y=+) = 3/5

P(Y=-) = ?

Feature	P(Feature|+)	P(Feature|-)
good	$\frac{2}{3}$	$\frac{1}{2}$
liked	$\frac{2}{3}$	$\frac{0}{2}$
bad	$\frac{0}{3}$	?
unpleasant	$\frac{0}{3}$	$\frac{1}{2}$
nice	$\frac{1}{3}$	$\frac{0}{2}$
pleasant	?	$\frac{0}{2}$
rude	$\frac{0}{3}$	$\frac{1}{2}$

Sentiment Analysis using Bernoulli NB



P(Y=+) = 3/5		
P(Y=-) = 2/5		
Feature	**P(Feature|+)**	**P(Feature|-)**
good	$\frac{2}{3}$	$\frac{1}{2}$
liked	$\frac{2}{3}$	$\frac{0}{2}$
bad	$\frac{0}{3}$	$\frac{2}{2}$
unpleasant	$\frac{0}{3}$	$\frac{1}{2}$
nice	$\frac{1}{3}$	$\frac{0}{2}$
pleasant	$\frac{1}{3}$	$\frac{0}{2}$
rude	$\frac{0}{3}$	$\frac{1}{2}$

Sentiment Analysis using Bernoulli NB

$$P(X=x_1, x_2, \ldots, x_n|Y) = \prod_{i=1}^{n} P(x_i|Y)$$

Laplace Smoothing

P(Feature|class=c) is calculated as $\frac{Number\ of\ documents\ of\ class\ 'c'\ with\ feature\ x_i\ +1}{Total\ number\ of\ documents\ of\ class\ 'c'\ +2}$

Sentiment Analysis using Bernoulli NB

Classification Model with Laplace Smoothing

P(Y=+) = 3/5		
P(Y=-) = 2/5		
Feature	**P(Feature|+)**	**P(Feature|-)**
good	$\frac{2+1}{3+2}=\frac{3}{5}$	$\frac{1+1}{2+2}=\frac{2}{4}$
liked	$\frac{2+1}{3+2}=\frac{3}{5}$	$\frac{0+1}{2+2}=\frac{1}{4}$
bad	$\frac{0+1}{3+2}=\frac{1}{5}$	$\frac{2+1}{2+2}=\frac{3}{4}$
unpleasant	$\frac{0+1}{3+2}=\frac{1}{5}$	$\frac{1+1}{2+2}=\frac{2}{4}$
nice	$\frac{1+1}{3+2}=\frac{2}{5}$	$\frac{0+1}{2+2}=\frac{1}{4}$
pleasant	$\frac{1+1}{3+2}=\frac{2}{5}$	$\frac{0+1}{2+2}=\frac{1}{4}$
rude	$\frac{0+1}{3+2}=\frac{2}{5}$	$\frac{1+1}{2+2}=\frac{2}{4}$

Sentiment Analysis using Bernoulli NB

Classify : “The rooms were good and the staff were nice”

P(+|good, liked, bad, unpleasant, nice, pleasant, rude)

P(good, liked, bad, unpleasant, nice, pleasant, rude|+)*P(+)

= P(good|+) * (1-P(liked|+)) * (1-P(bad|+)) * (1-P(unpleasant|+)) * P(nice|+) * (1-P(pleasant|+)) * (1-P(rude|+)) * P(+)

$= \frac{3}{5} * (1-\frac{3}{5}) * (1-\frac{1}{5}) * (1-\frac{1}{5}) * \frac{2}{5} * (1-\frac{2}{5}) * (1-\frac{2}{5}) * \frac{3}{5}$

= 0.013271

>

P(-|good, liked, bad, unpleasant, nice, pleasant, rude)

P(good, liked, bad, unpleasant, nice, pleasant, rude|-)*P(-)

= P(good|-) * (1-P(liked|-)) * (1-P(bad|-)) * (1-P(unpleasant|-)) * P(nice|-) * (1-P(pleasant|-)) * (1-P(rude|-)) * P(-)

$= \frac{2}{4} * (1-\frac{1}{4}) * (1-\frac{3}{4}) * (1-\frac{2}{4}) * \frac{1}{4} * (1-\frac{1}{4}) * (1-\frac{2}{4}) * \frac{2}{5}$

= 0.001758

Multinomial Naïve Bayes Classifier

Mail Content	Class
This is a promotion campaign. Promotion offers for selected customers.	Spam
Special offers for Holiday season	Spam
Have you completed the task? Can you provide the task status?	Non-Spam
Thank you for your mail. We have considered you for next steps.	Non-Spam

Distinct words in the training data become features.

All the training instances are transformed accordingly.

$$p(Wordi|Classk) = \frac{\sum_{TrainingSet} Count_{ik}}{\sum_{TrainingSet} Count_{k}}$$

Sum of frequencies for word$_i$ across Class$_k$

Sum of frequencies across all words in Class$_k$

Calculating the priors

P(Spam) = 2/4 *P(Non-Spam) = 2/4*

Calculating the Likelihood (Example)

P(promotion | Spam) = 2/15

P(task | Non-Spam) = 2/23

provide	steps	season	next	offers	have	we	is	campaign	special	your	status	the	promotion	can	you	customers	for	a	task	thank	completed	mail	this	selected	holiday	considered	Class
0	0	0	0	1	0	0	1	1	0	0	0	0	2	0	0	1	1	1	0	0	0	0	1	1	0	0	Spam
0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	Spam
1	0	0	0	0	1	0	0	0	0	0	1	2	0	1	2	0	0	0	2	0	1	0	0	0	0	0	Non-Spam
0	1	0	1	0	1	1	0	0	0	1	0	0	0	0	2	0	2	0	0	1	0	1	0	0	0	1	Non-Spam

Count of all frequencies of m-words for Spam Class = 15, Count of all frequencies of m-words for Non-Spam Class = 23

Multinomial Naïve Bayes Classifier

Zero Frequency Problem
Naïve Bayes faces the 'zero-frequency problem', where it assigns zero probability to a word in the test data set wasn't available in the training dataset.

Example:
Word 'promotions' never appeared for Non-Spam Class in training data.
So, the likelihood i.e. *P ('promotions' | Non-Spam) = 0*

Solution: Adaptive Smoothing or popularly known as Laplace Smoothing.

Usually, λ is a hyper-parameter and mostly taken as 1.

Adding every distinct word λ number of times to each class in the data.

K-Classes

$$p(Wordi|Classk) = \frac{\sum_{TrainingSet} Count_{ik} + \lambda}{\sum_{TrainingSet} Count_k + n\lambda}$$

Number of distinct words present in the data