

Text Classification and Naïve Bayes

Naïve Bayes (I)



Naïve Bayes Intuition

- Simple (“naïve”) classification method based on Bayes rule
- Relies on very simple representation of document
 - Bag of words



The bag of words representation

Y (

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet.

) = C





The bag of words representation

Y (

I **love** this movie! It's **sweet**, but with **satirical** humor. The dialogue is **great** and the adventure scenes are **fun**... It manages to be **whimsical** and **romantic** while **laughing** at the conventions of the fairy tale genre. I would **recommend** it to just about anyone. I've seen it **several** times, and I'm always **happy** to see it **again** whenever I have a friend who hasn't seen it yet.

) = C





The bag of words representation: using a subset of words

Y (

x love xxxxxxxxxxxxxxxxxxxx sweet
 xxxxxxxx satirical xxxxxxxxxxxx
 xxxxxxxxxxxx great xxxxxxxx
 xxxxxxxxxxxxxxxxxxxxxxxxxxxx fun xxxx
 xxxxxxxxxxxxxxxxxxxx whimsical xxxx
 romantic xxxx laughing
 xx
 xxxxxxxxxxxxxxxxxxxx recommend xxxxxx
 xx
 xx several xxxxxxxxxxxxxxxxxxxxxxxxxxxx
 xxxxxx happy xxxxxxxxxxxx again
 xx
 xx

)

= C







The bag of words representation

Y (

great	2
love	2
recommend	1
laugh	1
happy	1
...	...

) = C



Bag of words for document classification

Test
document

parser
language
label
translation
...

?

Machine
Learning

learning
training
algorithm
shrinkage
network...

NLP

parser
tag
training
translation
language...

Garbage
Collection

garbage
collection
memory
optimization
region...

Planning

planning
temporal
reasoning
plan
language...

GUI

...



Text Classification and Naïve Bayes

Naïve Bayes (I)