# Predicting Rating of Yelp Review Text

Yaonan Zhong

June 12, 2014

# Love or hate?


PEOPLE
LOVE US
ON
yelp
www.yelp.com


PEOPLE
HATE US
ON
yelp
www.yelp.com

# How to learn?

I. Model
    1. Naive Bayes classifier
    2. Support Vector Machine

II. Feature
    1. Bag-of-words
    2. Part-of-speech

# Yelp open dataset

1. 15,585 businesses
2. 335,022 reviews
3. 11,434 check-in sets
4. 70,817 users

… …

We extract 11,355 reviews for all the Chinese restaurant.

# Generate feature

If:

- vocabulary = [A, B, C, D]
- Review text = AHBDJBB

Then:

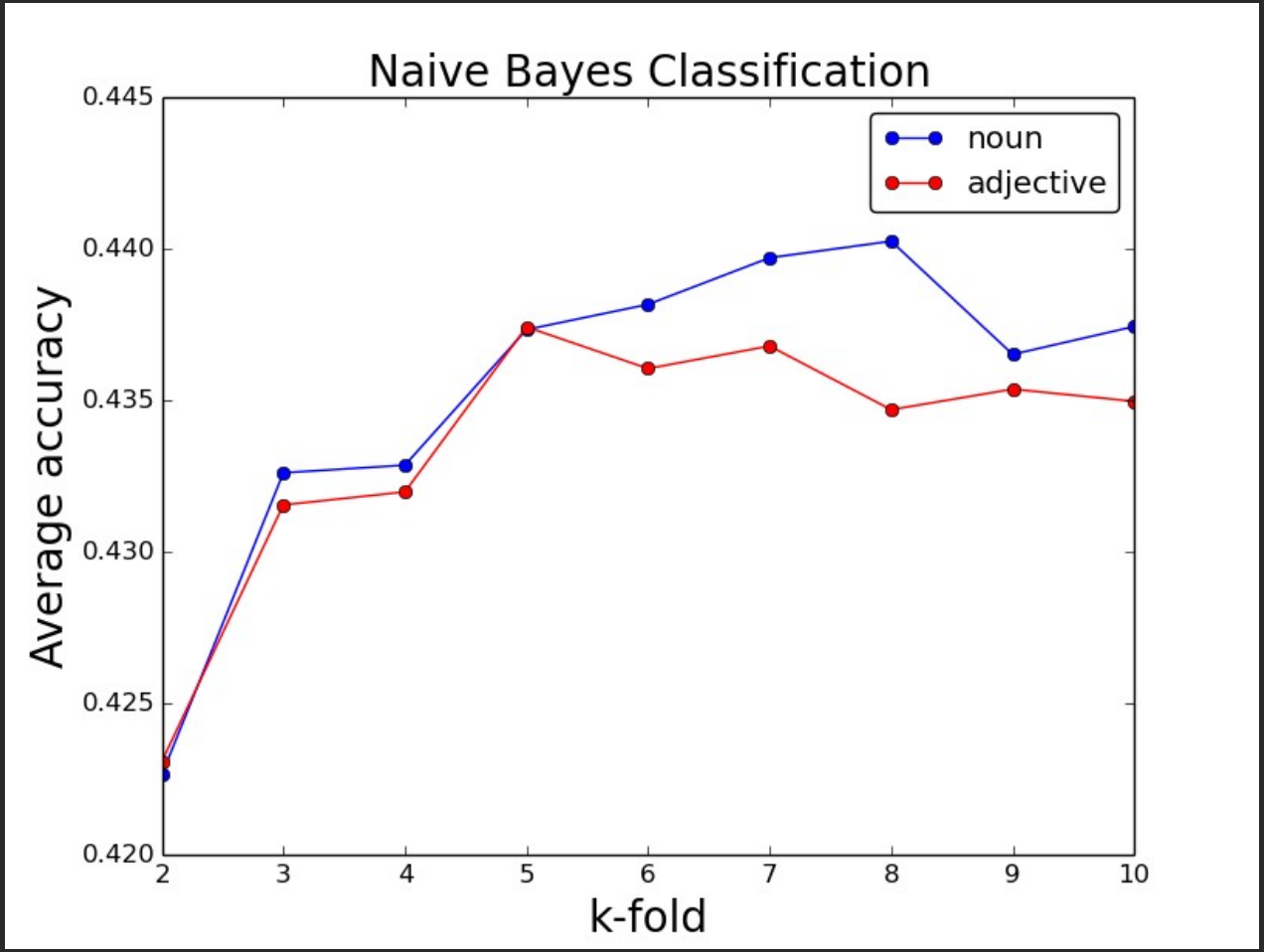
- Feature vector = [1, 3, 0, 1]

# Feature reduction

- Dimension depend on the size of vocabulary
- Use all unique words as vocabulary
- Or, we can filter the keywords by PoS

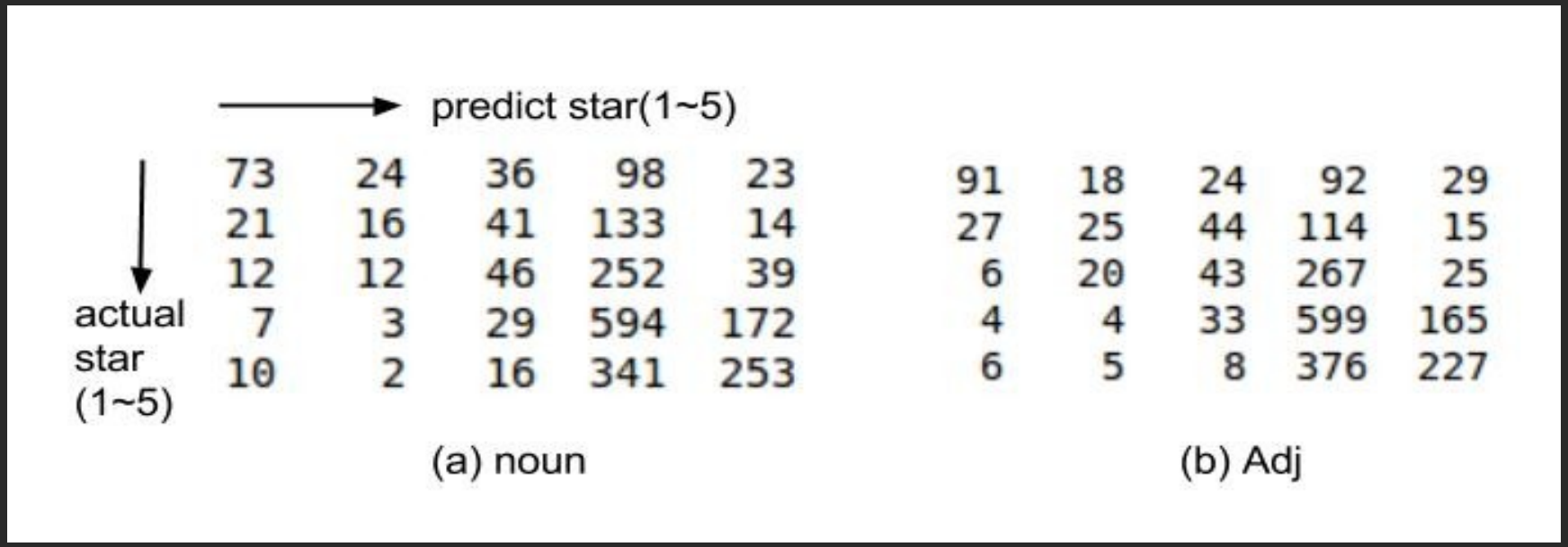
- “The best sweet and sour soup ever!”
- [(u'The', 'DT'), (u'best', 'JJS'), (u'sweet', 'NN'), (u'and', 'CC'), (u'sour', 'PRP$'), (u'soup', 'NN'), (u'ever', 'RB'), (u'!', '.')]

# Naive Bayes Classifier



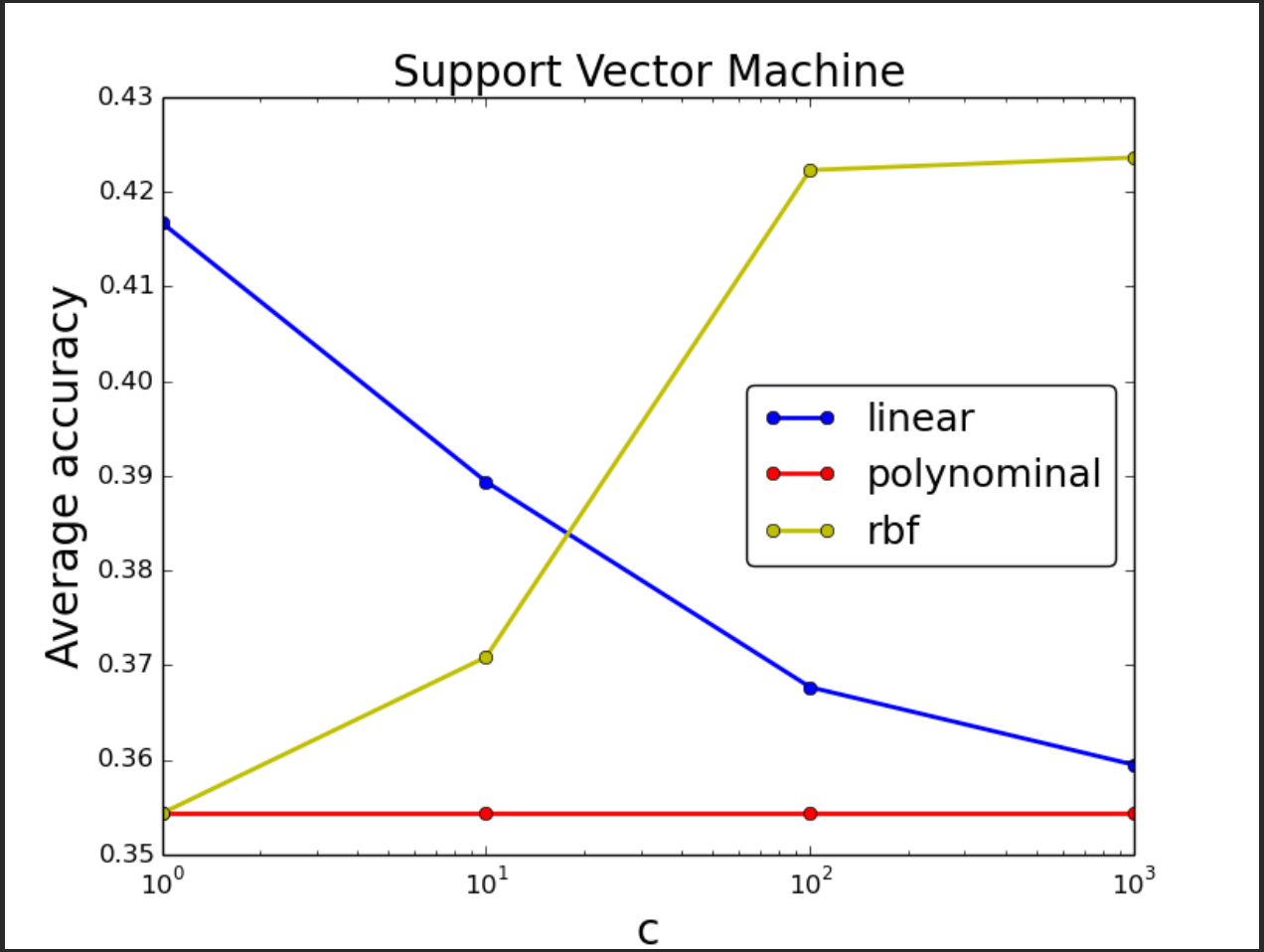
# Naive Bayes Classifier

Confusion matrix




9000 training data, 2000 testing data

# Support Vector Machine



# More to do

- Keep reduce feature dimension
- Adaboost
- Different smoothing in NB

……

# Thank you