

**CS 124 / LING 180 From Languages to Information,  
Dan Jurafsky, Winter 2020  
Week 3: Group Exercises on Naive Bayes and Sentiment  
Jan 21, 2020**

**Part 1**

We want to build a naive bayes sentiment classifier using add-1 smoothing, as described in the lecture (not binary naive bayes, regular naive bayes). Here is our training corpus:

Training Set:

```
- just plain boring
- entirely predictable and lacks energy
- no surprises and very few laughs
+ very powerful
+ the most fun film of the summer
```

Test Set:

```
predictable with no originality
```

1. Compute the prior for the two classes + and -, and the likelihoods for each word given the class (leave in the form of fractions).
2. Then compute whether the sentence in the test set is of class positive or negative (you may need a computer for this final computation).
3. Would using binary multinomial Naive Bayes change anything?
4. Why do you add  $|V|$  to the denominator of add-1 smoothing, instead of just counting the words in one class?

5. What would the answer to question 2 be without add-1 smoothing?
6. Can you think of any other features (or preprocessing) that you could add that might be useful in predicting sentiment? (This will come in handy for PA3!).

## Part 2: Conceptual Problems

7. Data ethics is an important component of any supervised machine learning task, since the data has to come from somewhere. For example, all research involving human subjects in the United States must follow the 3 "Belmont Principles", which are:

- **justice** (the equitable distribution of benefits and burdens of research),
- **beneficence** (the obligation to not only "do no harm" but to actively maximize benefits and minimize harms to subjects)
- **respect for persons** (people should be able to make autonomous decisions),

(the full Belmont report is here: [https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c\\_FINAL.pdf](https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf))

Discuss the following in your groups

1. What is the age of the person who wrote this sentence in a movie review ?  
"There are many, many characters to enjoy in this fanciful tale; my personal favorite has always been Inigo Montoya."
2. Is it ok that we can figure this out? Is it ok that we are using this person's data to train our sentiment analyzers? Should researchers not use publically posted data like this? Is it ok that we are using IMDB data for a homework? Should IMDB.com own this review? How does this all relate to the Belmont Principles? It may be helpful to bring up cases you've seen in the news relating to privacy or fairness and machine learning.

8. Go to the Sentiment demo at <http://nlp.stanford.edu:8080/sentiment/rntnDemo.html>. Come up with 5 sentences that the classifier gets wrong. Can you figure out what is causing the errors?
9. It is sometimes the case that more complex features (like trigrams or bigrams) perform better than simple features (like unigrams) on the **training** set, but perform worse than simple features on the **test** set. This is a particular case of the phenomenon called 'overfitting' in machine learning. Discuss why this might be. Can you create a tiny training set with 2 3-word documents and a test set with one document for which this overfitting situation holds?
10. Binary multinomial NB seems to work better on some problems than full count NB, but full count works better on others. For what kinds of problems might binary NB be better, and why? (There is no known right answer to this question, but I'd like you to think about the possibilities.)