Lecture: Naive Bayes

- **Video:** Week Introduction 27 sec
- **Video:** Probability and Bayes' Rule 3 min
- **Reading:** Probability and Bayes' Rule 10 min
- **Video:** Bayes' Rule 4 min
- **Reading:** Bayes' Rule 10 min
- **Video:** Naïve Bayes Introduction 5 min
- **Reading:** Naive Bayes Introduction 10 min
- **Video:** Laplacian Smoothing 2 min
- **Reading:** Laplacian Smoothing 10 min
- **Video:** Log Likelihood, Part 1 6 min
- **Reading:** Log Likelihood, Part 1 10 min
- **Video:** Log Likelihood, Part 2 2 min
- **Reading:** Log Likelihood Part 2 10 min
- **Video:** Training Naïve Bayes 3 min
- **Reading:** Training naïve Bayes 10 min
- **Lab:** Visualizing likelihoods and confidence ellipses 1h
- **Video:** Testing Naïve Bayes 4 min
- **Reading:** Testing naïve Bayes 10 min
- **Video:** Applications of Naïve Bayes 3 min
- **Reading:** Applications of Naive Bayes 10 min
- **Video:** Naïve Bayes Assumptions 3 min
- **Reading:** Naïve Bayes Assumptions 10 min
- **Video:** Error Analysis 3 min
- **Reading:** Error Analysis 10 min
- **Video:** Week Conclusion 44 sec

Lecture Notes (Optional)

Practice Quiz

Assignment: Naive Bayes

Training naïve Bayes

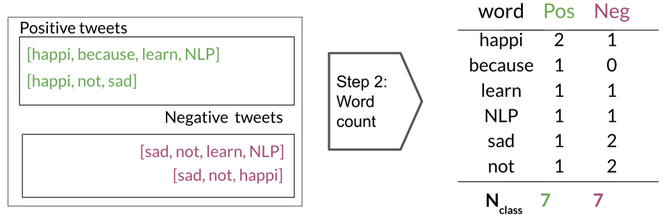
To train your naïve Bayes classifier, you have to perform the following steps:

1) Get or annotate a dataset with positive and negative tweets

2) Preprocess the tweets: process_tweet(tweet) → [w1, w2, w3, ...]:

- Lowercase
- Remove punctuation, urls, names
- Remove stop words
- Stemming
- Tokenize sentences

3) Compute freq(w, class):



word	Pos	Neg
happi	2	1
because	1	0
learn	1	1
NLP	1	1
sad	1	2
not	1	2
N_{class}	7	7

freq(w, class)

4) Get $P(w|pos), P(w|neg)$

You can use the table above to compute the probabilities.

5) Get $\lambda(w)$

$\lambda(w) = \log \frac{P(\mathrm{w}|\mathrm{pos})}{P(\mathrm{w}|\mathrm{neg})}$

6) Compute $logprior = \log(P(pos)/P(neg))$

$\mathrm{logprior} = \log \frac{Dpos}{Dneg}$, where $Dpos$ and $Dneg$ correspond to the number of positive and negative documents respectively.

Mark as completed

Like Dislike Report an issue