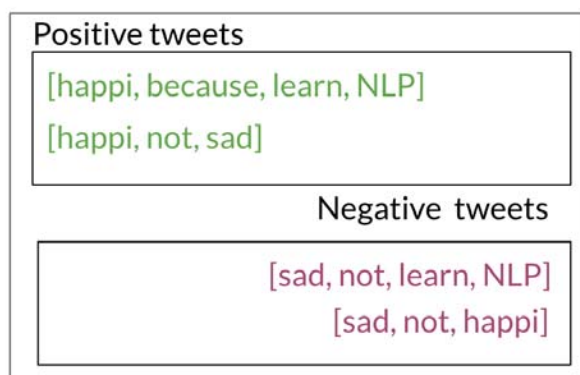coursera

☰ Menu

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.