

INFO284 - First Obligatory Assignment

Naives Bayes for classification

For better documentation, see the python code. Several detailed comments are describing each function in the code. You can also use the tweets below as input, and get a better explanation for the prediction.

Wrong predicted tweets:

Sentiment	Tweet
Original: Positive Predicted: Negative	@USAirways thanks for getting me rescheduled on a direct flight to NOLA tonight in 10 minutes--and hopefully an hour earlier getting there!
Original: Neutral Predicted: Negative	@USAirways I just hope that pilot had a good day off

Explanation:

The first tweet has several words that are weighted negatively. Only the word *thanks* are predicted positively by the algorithm. This will result in a false prediction of the tweet because all the other words are weighted negatively. The probability for each word occurring multiplied with the probability for each class.

The second tweet have no words that are predicted neutral. All the words are predicted negatively. This results in a negative prediction, when the original tweet is neutral.

Correct predicted tweets:

Sentiment	Tweet
Original: Neutral Predicted: Neutral	@USAirways http://t.co/AXRYeIWzh0 , vuelo24.es, vuelos24.es sell too cheap MAD-NYC with USAirways. Are they trusted webs or is it a scam?
Original: Negative Predicted: Negative	@USAirways the voucher you give us for a hotel is useless you call the number they say no rooms available what do you expect me to do?

Explanation:

The first tweet has four words that have highest probability with neutral, and nine words that have highest probability with negative. But because the neutral words have a higher probability and is weighted most compared to the other words, it is natural for the algorithm to choose neutral.

The second tweet is predicted negative, because every word are weighted most with the negative sentiment.