Oblig2, Instructions and comments about naive Bayes usefulness as classifier

15 March 2019 10:38

Running the script

- **Access the terminal**
- **To run script, type in terminal:** python "insertScriptName".py "fileToBeRead".txt

Conclusion

Naive Bayes is clearly a useful measure of the reliability of a text.
Given a well developed vocabulary, any text can be assigned a reliable/unreliable label with a fair degree of accuracy.
However, this leads to assumptions that may not be generally true.
When featured words occur in a text, some measure of classification accuracy can always be achieved. How accurate this classification is will largely depend on how well the vocabularies correspond to the unreliable/reliable labelling of the training set.
Whether this is useful for solving the "fake news"-problem for all texts is not something that can be proven using the tools we have developed at this time.

As stated in [source 1], fake news often have grammatical errors, emotional wording, manipulative intent, and untrue content.
These are not things that can be directly derived from a vocabulary of separate words.
While this is not an issue with naive Bayes itself, a much deeper model could utilize sentiment analysis, source tracing and so forth.
This is currently far beyond the scope of this assignment, but it appears that a simple naive Bayes classifer would not be the most useful way to reliably classify fake news considering the great diversity of text types and writing styles.

Sources:

1. https://ieeexplore.ieee.org/document/8100379