Sentiment Analysis and Media Monitoring Exercise 1

Heath Gordon

March 15, 2016

1 Features

Bigrams were way more informative, even if they gave me a paltry 60 percent accuracy. And 'bad movie' was at the top of the list like every time. Also clearly these were all about similar movies, because 'toy story' and 'dark city' came up a lot. And then would disappear for one loop. You can check the output for the last time I ran it on github, if you're at all interested: Exercise 1 it's called 'output.txt'.

2 Notes

Not much to say here. This was pretty interesting, implementing the cross-fold validation. Turns out Klenner taught it to me incorrectly, go figure. That was good practice.
It needs a bit of a cleanup. There are a couple spots where I used nested 'for' loops where I should have used list constructors if I want to be a *pythonista fantastica* . So I think one big difference between this and like PCLII, is I'm pretty sure back in PCLII I would use functions to manipulate global variables, which I'm feeling a little nauseous about right now.