Andrew Dyer

Dr. Dligach

Machine Learning

11/10/2016

Naive Bayes

I found great success in my Naive Bayes. Around 75% accuracy. I was really delighted with that, and to map that was painstaking. I had to compute the accuracy by hand, because I was having difficulty with data structures. I really had a successful moment when I realized I was over complicating, my original idea was to precalculate all NB values for each word as + or - then just reference them as needed. That turned out to be way to complicated for what was desired for the project. I think, given a bit more time, or if I was to start over from scratch I would have been able to refine my model and refactor it to be a lot simpler to understand.

Cleaning the data was also a chore, removing stopwords and punctuation ended up taking the majority of the time. The smartest thing I did was writing out the data after cleaning, that way testing time took a fraction of the time. Rather than load up the data and clean it with each iteration, instead it would just read from a text file each word precleaned.

My model was very successful (after the strain) and I feel like it will help me alot with the ARDS bag of words project. I really like summarizing my findings because it forces me to rethink everything I did, and allows me to develop better strategies in the future. My K Means writeup gave me much more clarity for this homework.