COM6513, Lab4

Registration Number 170224545

08/03/2018

1. **Description**

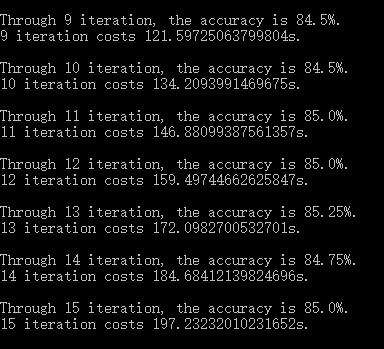
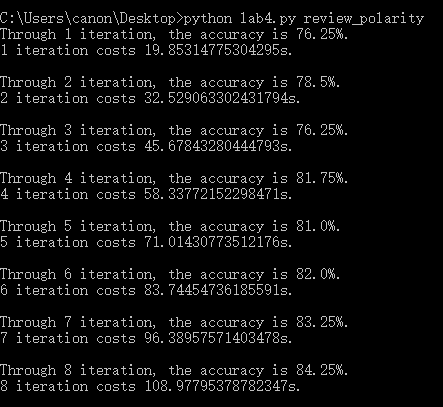
In this lab, the program successfully implements standard binary perceptron with bag-of-words representation.

This program takes the average of all the weight vectors for each class.

The program can be executable by running:

python3 lab4.py review\_polarity

And the result is as follows:



It shows that the accuracies through 15 iterations is : 76.25%, 78.5%, 76.25%, 81.75%, 81.0%, 82.0%, 83.25%, 84.25%, 84.5%, 84.5%, 85.0%,

85.0%, 85.25%, 84.75%, 85.0%.

The all 15 iteration costs 197.23s.

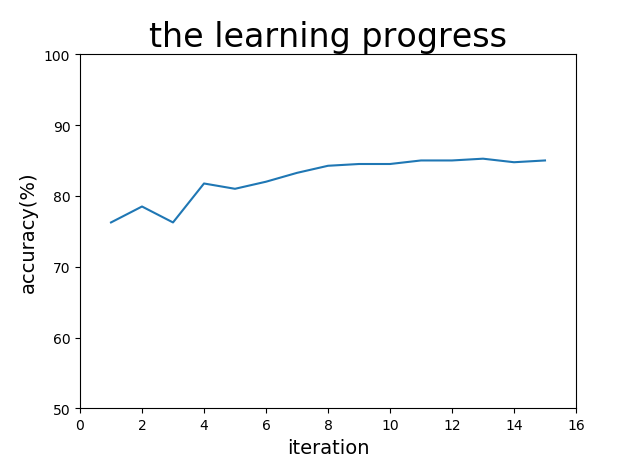
1. **Evaluation**
2. the top 10 features:

The top 10 positive features:

['seen', 'great', 'most', 'life', 'best', 'many', 'very', 'well', 'see', 'will']

The top 10 negative features:

['bad', 'script', 'only', 'plot', 'worst', 'boring', 'unfortunately', 'could', 'then', 'any']

1. the learning progress

As we can know, the accuracy increases with the iterations. When the iteration is sufficient, the accuracy converges to 85%.

1. Discussion of result

I think the features would generalize well for different domain because these words have obvious emotional tendencies. I propose words like “love”, “hate” and “favorite” which can show emotion can be good too.