Worksheet: Basic sentiment analysis

COSC 101, 2018-04-27

1) Warm-up
Write a function called get_words that takes a string as a parameter and returns a list of words in the string. Only digits and letters should be kept (all other characters should be ignored). All letters must be converted to lowercase.

2) Training a word sentiment model
Example training set
5 This product is good.
5 This product is made very well.
4 This product functions well.
3 This is a mediocre product.
2 A piece broke off. I am dissatisfied with this product.
1 This product is junk.
1 This product does not function as it is advertised.

Write a function called train that takes a list of tuples as input and returns a dictionary of words and their numeric rating. Each tuple contains a numeric rating (an integer) and a list of words (i.e., a list of strings).
3) Applying a word sentiment model
Write a function called `compute_score` that takes a list of words (i.e., a review) and a dictionary of word ratings (i.e., the output from `train`) and returns an estimated rating for the review.

4) Testing
Example testing set
5 This product works very well.
1 This product is a piece of junk.
1 Very poorly made product!

Write a function called `test` that takes a review score, a list of words (i.e., a review), and a dictionary of word ratings and returns True if the estimated review score is within the same “bucket” as the provided review score. The buckets are: < 2.5 Positive, > 3.5 Negative, 2.5-3.5 (inclusive) Neutral.