Practice 2

When investigating and analyzing recent two-week tweets of Mr. O, Ms. H, and Mr. T, and then focusing on 10 words, the following appeared.

Mr. O(10 tweets): USA great Democrats care mother

Ms. H(15 tweets): mother love Democrats care mother USA

Mr. T(25 tweets): Russia fake USA Mexico great Mexico haters

New tweets including "USA Democrats mother" were sent out. Based on the above data and the way of thinking of naïve Bayes classifier, predict the new tweet which belongs to whom among these three persons. Note that consider the number of tweets of each person as prior probability. And, use Laplace smoothing with 1 as the initial value of frequency.

Practice 2 (modified)

When investigating and analyzing recent two-week tweets of Mr. O, Ms. H, and Mr. T, and then focusing on 10 words, the following appeared.

Mr. O(10 tweets): USA great Democrats care mother

Ms. H(15 tweets): mother love Democrats care mother USA

Mr. T(25 tweets): Russia fake USA Mexico great Mexico haters

How to calculate if the new tweets including "USA Democrats mother mother"?

Ans. of Practice 2 (modified)

dm-04-practice2-ans-suppl.ipynb

1. Prior probability
   - P(O) = 10 / (10+15+25) = 1 / 5 = 0.2
   - P(H) = 15 / (10+15+25) = 3 / 10 = 0.3
   - P(T) = 25 / (10+15+25) = 1 / 2 = 0.5
2. Likelihood
   - P(USA|O) x P(Democrats|O) x P(mother|O) x P(mother|O) = 2/15 x 2/15 x 2/15 x 2/15
   - P(USA|H) x P(Democrats|H) x P(mother|H) x P(mother|H) = 2/16 x 2/16 x 3/16 x 3/16
   - P(USA|T) x P(Democrats|T) x P(mother|T) x P(mother|T) = 2/17 x 1/17 x 1/17 x 1/17

Ans. of Practice 2 (modified)

3. posterior prob. = Likelihood x prior prob.
   - P(O|Words) = 2/15 x 2/15 x 2/15 x 2/15 x 0.2 = 0.000063
     P(H|Words) = 2/16 x 2/16 x 3/16 x 3/16 x 0.3 = 0.000165
     P(T|Words) = 2/17 x 1/17 x 1/17 x 1/17 x 0.5 = 0.000012
4. Proportion
   - P(O|Words) / (P(O|Words) + P(H|Words) + P(T|Words)) = 0.2633…
   - P(H|Words) / (P(O|Words) + P(H|Words) + P(T|Words)) = 0.6867…
   - P(T|Words) / (P(O|Words) + P(H|Words) + P(T|Words)) = 0.0498…
5. The largest posterior probability points to Ms. H.

This calculation can also be confirmed by using naïve_bayes_small.ipynb in DM-05.