

# 4.1 RAT

---

1. In classification, we can simplify Bayes Rule by assuming that which of the following is constant?
  - a) The total probability of a particular class.
  - b) The total probability of a particular document.
  - c) The likelihood of a particular document given a certain class.
  - d) The likelihood of a particular class given a certain document.
2. Without smoothing, the biggest problem with Maximum Likelihood Estimation would be that for a given class...
  - a) Finding a word that appears frequently in the training set for a class has too great an effect on the outcome.
  - b) Finding a word that appears infrequently in the training set for a class has too small an effect on the outcome.
  - c) Finding a word that didn't appear in the training set for a class has no effect on the outcome.
  - d) Finding a word that didn't appear in the training set for a class eliminates that class as a possible outcome.
3. Given this model for positive and negative reviews, which word creates the greatest relative bias towards **negative**?

|                       | pos  | neg  |
|-----------------------|------|------|
| <b><i>I</i></b>       | 0.09 | 0.16 |
| <b><i>always</i></b>  | 0.09 | 0.05 |
| <b><i>like</i></b>    | 0.19 | 0.04 |
| <b><i>foreign</i></b> | 0.03 | 0.11 |
| <b><i>films</i></b>   | 0.09 | 0.12 |

4. What is the  $F_1$  if  $P = 0.9$  and  $R = 0.6$ ?