Project: Who is the real author of Hamlet?

Table of Contents

- Introduction
- Finding Prior Probabilities
- Finding Conditional Probabilities
- Test
- Conclusion
- References

Introduction

- In this presentation we will be using bayes theorem to a text classification.

Finding Prior Probabilities

- We find prior probabilities
- P(C) = 3/7
- P(W) = 2/7
- P(F) = 2/7

$$\hat{P}(c) = \frac{N_c}{N} \qquad \hat{P}(w|c) = \frac{count(w,c)+1}{count(c)+|V|}$$

- Finding Conditional Probabilities

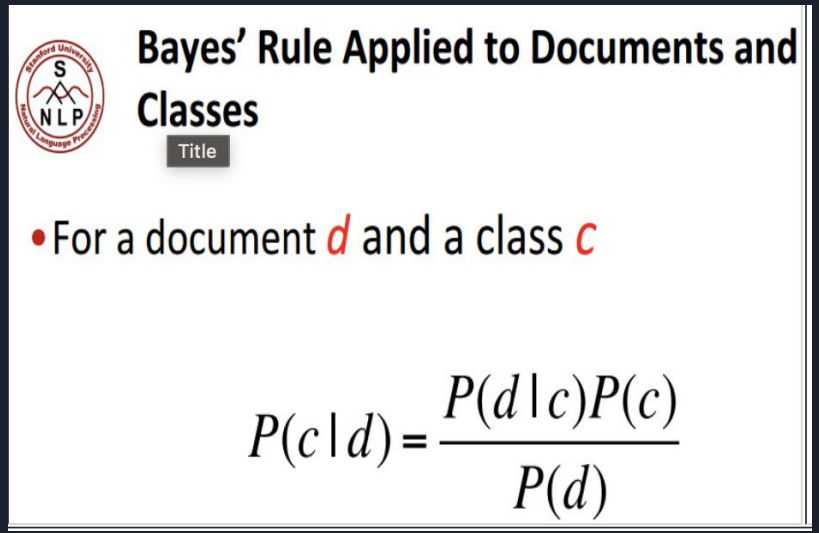
- P(W1|C) = 4+1/12+6 = 5/18
- P(W4|C) = 2+1/12+6 = 3/18
- P(W6|C) = 0+1/12+6 = 1/18
- P(W5|C) = 2+1/12+6 = 3/18
- P(W3|C) = 2+1/12+6 = 3/18
- P(W1|W) = 1+1/8+6 = 2/14
- P(W4|W) = 1+1/8+6 = 2/14
- P(W6|W) = 2+1/8+6 = 3/14
- P(W5|W) = 2+1/8+6 = 3/14
- P(W3|W) = 1+1/8+6 = 2/14

Continued

- P(W1|F) = 0+1/9+6 = 1/15
- P(W4|F) = 2+1/9+6 = 3/15
- P(W6|F) = 1+1/9+6 = 2/15
- P(W5|F) = 2+1/9+6 = 3/15
- P(W3|F) = 2+1/9+6 = 3/15

Test

- The probability of the author being C
- P(C|D8) = P(C)*P(D8|C) / P(D8)
- P(C|D8) = P(C)* P(W1|C)* P(W4|C)* P(W6|C)* P(W5|C)* P(W3|C)/P(D8)
- P(C|D8) = 3/7* 5/18* 3/18*1/18*3/18*3/18 = 0.000031



Continued

- The probability of the author being W
- P(W|D8) = P(W)* P(W1|W)* P(W4|W)* P(W6|W)* P(W5|W)* P(W3|W) P(W|D8) = 2/7* 2/14*2/14*3/14*3/14*2/14 = 0.000038

Continued

- The probability of the author being F
- P(F|D8) = P(F)* P(W1|F)* P(W4|F)* P(W6|F)* P(W5|F)* P(W3|F) P(F|D8) = 2/7* 1/15*3/15*2/15*3/15*3/15 = 0.000020

Conclusion

- P(C|D8) =0.000031
- P(W|D8) = 0.000038
- P(F|D8) = 0.000020
- In this case the author should W, William Stanley.

Bibliography

- https://hc.labnet.sfbu.edu/~henry/sfbu/course/mllib/naive_bayes/slide/exercise_naive_bayes.html
- https://github.com/Alami64/Machine-Learning/tree/main/End-To-End-Machine-Learning-Algorithm