

The Naive Bayes Algorithm: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2021

Concepts

- When a new message " w_1, w_2, \dots, w_n " comes in, the Naive Bayes algorithm classifies it as spam or non-spam based on the results of these two equations:

$$\begin{aligned} P(\text{Spam} | w_1, w_2, \dots, w_n) &\propto P(\text{Spam}) \cdot \prod_{i=1}^n P(w_i | \text{Spam}) \\ P(\text{Spam}^C | w_1, w_2, \dots, w_n) &\propto P(\text{Spam}^C) \cdot \prod_{i=1}^n P(w_i | \text{Spam}^C) \end{aligned}$$

- To calculate $P(w_i | \text{Spam})$ and $P(w_i | \text{Spam}^C)$, we need to use the additive smoothing technique:

$$\begin{aligned} P(w_i | \text{Spam}) &= \frac{N_{w_i | \text{Spam}} + \alpha}{N_{\text{Spam}} + \alpha \cdot N_{\text{Vocabulary}}} \\ P(w_i | \text{Spam}^C) &= \frac{N_{w_i | \text{Spam}^C} + \alpha}{N_{\text{Spam}^C} + \alpha \cdot N_{\text{Vocabulary}}} \end{aligned}$$

- Below, we see what some of the terms in equations above mean:

```
\begin{aligned} N_{w_i | \text{Spam}} &= \text{the number of times the word } w_i \\ &\text{occurs in spam messages} \\ N_{w_i | \text{Spam}^C} &= \text{the number} \\ &\text{of times the word } w_i \text{ occurs in non-spam messages} \\ N_{\text{Spam}} &= \text{total number of words in spam messages} \\ N_{\text{Spam}^C} &= \text{total number of words in non-spam messages} \\ N_{\text{Vocabulary}} &= \text{total number of words in the vocabulary} \\ \alpha &= 1 \end{aligned}
```

(α is a smoothing parameter)

Resources

- [A technical intro to a few version of the Naive Bayes algorithm](#)
- [An intro to conditional independence](#)