

Naive Bayes for Sentiment Analysis: Takeaways



by Dataquest Labs, Inc. - All rights reserved © 2019

Concepts

- The Naive Bayes classifier figures out how likely data attributes are associated with a certain class.
- The classifier is based on Bayes' theorem, which is

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}, P(B) \neq 0$$

where:

- A and B are events.
- $P(A|B)$ is a conditional probability. Specifically, the likelihood of event A occurring given the B is true.
- $P(B|A)$ is also a conditional probability. Specifically, the likelihood of event B occurring given the A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other.
- Bayes' Theorem describes the probability of an event based on prior knowledge of conditions that might be related to the event.
- Naive Bayes extends Bayes' theorem to handle the case of multiple data points by assuming each data point is independent.
- The formula for the classifier is the following

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)}$$

- To find the "right classification", we find out which classification ($P(y | x_1, \dots, x_n)$) has the highest probability.

Resources

- [Bayes' theorem](#)
- [Probability theory](#)



Takeaways by Dataquest Labs, Inc. - All rights reserved © 2019