Journal of Machine Learning Research 1 (2019) INSERT PAGE STUFF Submitted 9/19; Published DATE PUBLISHED

Learning with Naive Bayes

Andrew Kirby ANDY-KIRBY@LIVE.COM
Kevin Browder BROWDERKEVIN54@GMAIL.COM
Nathan Stouffer NATHANSTOUFFER1999@GMAIL.COM
Eric Kempf ERICKEMPF123@GMAIL.COM

Editor:

Abstract

Keywords:

1. Introduction

2. Problem Statement

The problem that we must solve is a classification problem. Given an input file that contains examples (each example consists of a list of attributes and an associated classification), our task is to implement a learning algorithm that is trained to classify examples. The algorithm we will implement is called Naive Bayes. The performance of our learning algorithm will be evaluated by two metrics of our choosing and 10-fold cross validation. When we have implemented the algorithm, we then perform our experiment. We are tasked with testing whether scrambling values in 10% of the features will affect the performance of Naive Bayes. This effectively eliminates the usefulness of 10% of features in a given data set.

2.1 Hypothesis

We predict that scrambling 10% of features will marginally affect performance.

3. Algorithm

Naive Bayes Naive Bayes is a low cost classifying algorithm that utilizes a probabilistic model based in Bayesian Decision Theory (Kaur and Oberai, 2014). Given an example $x \in X$ with attributes $a_1, a_2, ..., a_d$ and belonging to class $c \in C$, the algorithm will classify x by choosing the maximum probability $P(c|a_1, a_2, ..., a_d)$ from all $c \in C$.

Using the probability Calculating this probability becomes more apparent when rewritten using Bayes Theorem:

$$P(c|a_1, a_2, ..., a_d) = \frac{P(a_1, a_2, ..., a_d|c)P(c)}{P(a_1, a_2, ..., a_d)}$$

Because the algorithm chooses a classification $c = argmax_{c \in C} P(c|a_1, a_2, ..., a_d)$, the classification is still equivalent to

4. Experimental Design

4.1 Set Up

4.2 Tuning

Add stuff about bin size, what else?
Experimenting w/ using different attributes

4.3 Final Parameters

-Bin size
-Number of attributes used
-Which attributes used?

5. Results

6. Summary

References

Gurneet Kaur and Neelam Oberai. A review article on naive bayes classifier with various smoothing techniques. 2014.