

Naive Bayes

Prof. Eric A. Suess

February 10, 2020

Introduction

Today we will begin discussing **Naive Bayes**, a **Classification** algorithm that uses **probability**.

According to the author, these estimates are based on probabilistic methods, or methods concerned with describing uncertainty. They use the data from past events to extrapolate to future events.

Naive Bayes

Naive Bayes has been used successful for working with email.

Spam Filtering

- ▶ spam filtering
- ▶ SpamBayes
- ▶ Spam Filtering with Naive Bayes - Which Naive Bayes?
- ▶ Why Bayesian filtering is the most effective anti-spam technology
- ▶ A Bayesian Approach to Filtering Junk E-Mail

Naive Bayes

Naive Bayes has been used successful for working with email.

Prioritizing

Folderizing

Naive Bayes

Naive Bayes has been used successful for Text Classification.

- ▶ text classification
- ▶ text classification
- ▶ text classification
- ▶ text classification

Naive Bayes

Naive Bayes has been used for Sentiment Analysis.

- ▶ Citius

Naive Bayes

Naive Bayes has been used for Intrusion Detection.

- ▶ Network Intrusion Detection

Naive Bayes

Naive Bayes has been used Medical Diagnosis

- ▶ Diagnosis
- ▶ Diagnosis

Naive Bayes

- ▶ Coursera Stanford Videos on Graphical Models

Probability

- ▶ event
- ▶ trial
- ▶ mutually exclusive
- ▶ Venn Diagrams
- ▶ joint probability
- ▶ independent events
- ▶ dependent events
- ▶ conditional probability

Probability

- ▶ Bayes' Theorem
- ▶ prior probability
- ▶ likelihood
- ▶ posterior probability

Bayes Theorem

$$P(\textit{spam}|\textit{Viagra}) = \frac{P(\textit{Viagra}|\textit{spam})P(\textit{spam})}{P(\textit{Viagra})}$$

prior: $P(\textit{Viagra})$

likelihood: $P(\textit{Viagra}|\textit{spam}) = L(\textit{spam})$

posterior: $P(\textit{spam}|\textit{Viagra})$

The classification is done using the posterior probability.

The class with the *highest probability* is the *classification* for that observation/example.

The naive Bayes algorithm

The **naive Bayes (NB)** algorithm describes a simple application using Bayes' theorem for classification.

NB is the de facto standard for much *text classification*.

See page 97/95 for the Strengths and Weaknesses.

Why naive?

The naive Bayes algorithm is named naive because it makes a couple of “naive” assumptions about the data.

1. It assumes that all of the features in the dataset are **equally important**.
2. It assumes that all of the features in the dataset are **independent**.

The naive Bayes classification

Naive Bayes assumes *class-conditional independence*, which means the events are independent so long as they are conditioned on the same class value.

Using numeric features with naive Bayes

Features need to be in **categories**

- ▶ discretize
- ▶ bin
- ▶ cut points

(Reminder: All variables/features need to be **numeric** for kNN.)

Or use a different algorithm from a different package.

Example

Next time we will work with the example in the book - filtering mobile phone (SMS) spam with the naive Bayes algorithm.