
Naïve Bayes

— Olivier Rochaix, Erick Sanmartin,
Logan Sell, Nolan Thomas —

How Does Naïve Bayes Work?



Brief Intro

- Set of supervised learning algorithms based on applying Bayes' theorem with a “naïve” assumption: each feature makes an independent and equal contribution to the outcome.
- A classifier that uses probabilities calculated from a training set
- Known as a decent classifier, but also known to be a bad estimator.
- Naïve Bayes algorithms are often used in **Sentiment Analysis, Spam Filtering, Recommendation Systems, Text Categorization**, etc.

Naïve Bayes Classifiers

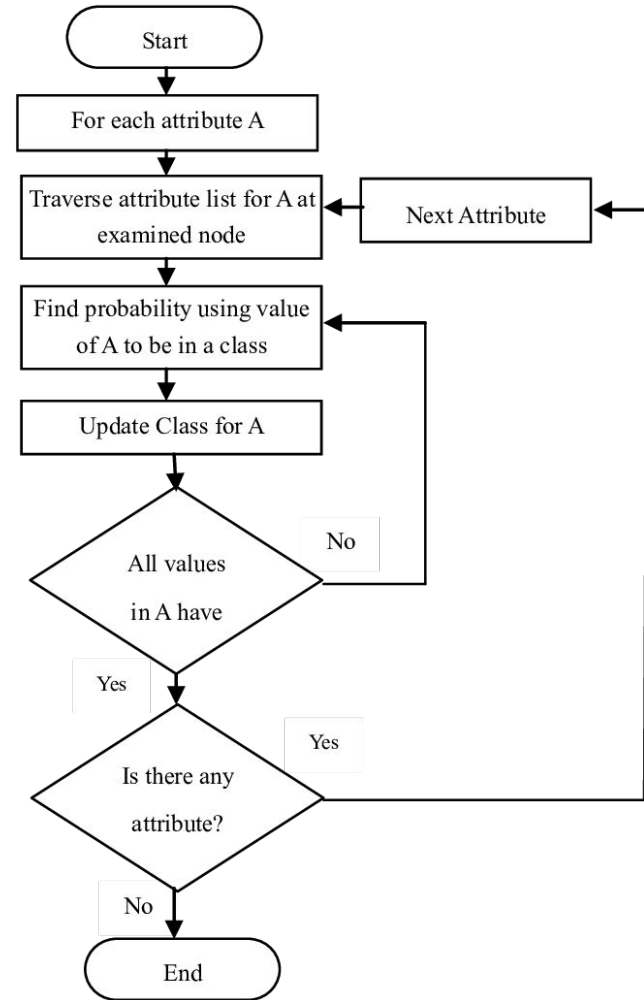
- **GAUSSIAN NAÏVE BAYES**
 - Continuous values associated with each feature are assumed to be distributed according to a Normal distribution.
- **MULTINOMIAL NAÏVE BAYES**
 - Suitable for classification with discrete features (e.g., word counts for text classification).
- **COMPLEMENT NAÏVE BAYES**
 - Designed to correct the “severe assumptions” made by the standard Multinomial Naive Bayes classifier.
- **BERNOULLI NAÏVE BAYES**
 - This classifier is suitable for binary/boolean features.
- **CATEGORICAL NAÏVE BAYES**
 - Suitable for classification with discrete features that are categorically distributed.

The Bayes Theorem



$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

$$P(dV \text{ given } iV) = \frac{P(iV \text{ given } dV) * P(dV)}{P(dV)}$$



Advantages / Disadvantages



Advantages

- Quick
- Simple to use
- Trainable with small test sets
- Scalable
- Resource efficient
- Noise tolerant

Disadvantages

- Assumption of feature independence
- Assumption of feature equality
- Relatively few hyperparameters
- Zero-frequency problem
- Less effective for applications that are highly complex or require very precise predictions

Required Data Processing Steps



Dataset Requirements

The data set is divided into two parts:

The feature matrix

- contains all the vectors(rows) of dataset
- each vector consists of the value of dependent features

The response vector

- contains the value of class variable(prediction or output) for each row of feature matrix

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No

Next Steps

- Separate training data by class
- `separate_by_class()` function assumes final column per row is the class value
- Create dictionary:
 - Keys: Class value
 - Values: List of all records

Hyperparameters



Hyperparameters - Gaussian

- **priors**: sets the probability for the output class.
 - Takes an array that adds up to 1 (array of probabilities).
 - Array is length of amount of classes
- **var_smoothing**: artificially adds a user-defined value to the distribution's variance to account for more samples further away from the distribution mean. When there is missing data or a class is not represented var_smoothing keeps model from breaking down.
 - Takes a small float
 - Default value is 1e-9

Code Snippet for Multinomial Naïve Bayes



```
1 def predict_category(s, train=train, model=model):  
2     pred = model.predict([s])  
3     return train.target_names[pred[0]]
```

[22]

```
1 predict_category('Jesus Christ')
```

[23]

```
... 'soc.religion.christian'
```

Appendix

- An introduction to the concept of Naïve Bayes
 - [1.9. Naive Bayes — scikit-learn 1.1.2 documentation](#) - This resource is the comprehensive documentation of the naïve Bayes model and its different classifiers. [Article]
 - [Naive Bayes, Clearly Explained!!!](#) - A very simple overview of the Bayes Theorem and how it is used for sorting classes. [Video]
 - [Naïve Bayes Algorithm: Everything You Need to Know - KDnuggets](#) - This resource explains the formula underlying the Naïve Bayes classifier and common problems that can be encountered when using it. [Article]
 - [What are Naive Bayes classifiers?](#) - Provides a quick glance at the 3 most popular Naïve Bayes classifiers and another look at the Bayes Theorem Formula [Article]
 - [\(PDF\) Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing](#) - This resource is a research paper from which the group pulled a handy flowchart to conceptualize the flow of the naïve Bayes model. [PDF]
 - [Naïve Bayes Classifier | Naive Bayes Algorithm](#) - This resource is a 45-minute comprehensive overview covering the theory of Naïve Bayes, different applications of the ML model, and a code-along example. [Video]

Appendix

- Code examples
 - [Naive Bayes Classifier From Scratch in Python](#) - This resource provides detailed code examples and a step-by-step explanation of implementing a Naïve Bayes classifier including data processing steps. It was a key source for our Data Processing section. [Article]
 - [Naive Bayes Classifier in Machine Learning - lavatpoint](#) - This resource provides a more in-depth view of the behind the scenes processes of the naïve Bayes model. [Article]
 - [Naive Bayes Classifiers - GeeksforGeeks](#) - This resource runs through an example of the data processing required for naïve bayes and provides some high level information about the model. [Article]
 - [Spam Filter in Python: Naive Bayes from Scratch - KDnuggets](#) - An example of one of the archetypal applications of Naïve Bayes. It includes details on data cleaning as well. [Article]
- Advantages/disadvantages
 - [Naive Bayes Pros & Cons - HolyPython.com](#) - A succinct and easy-to-understand breakdown of the benefits and drawbacks of Naïve Bayes. A useful starting point for learning about the algorithm. [Article]
 - [Gaussian Naive Bayes with Hyperparameter Tuning](#) - A code demonstration with examples of hyperparameter in context. It includes discussion of strengths and weaknesses (including the zero-frequency problem). [Code tutorial]
- Hyperparameters
 - [Naive Bayes Tuning - AlFinesse.com](#) - A summary of the hyperparameters and their use cases. This includes examples for how the var_smoothing parameter affects a model's performance. [Article]