Pembelajaran Mesin

Naive Bayes

22 Oktober 2018

Dr. rer. nat. Akmal Junaidi, S.Si., M.Sc.

Jurusan Ilmu Komputer
FMIPA – Universitas Lampung

Introduction

Naive Bayes is a classification algorithms based on Bayes Theorem.

The assumption of Naive Bayes is that among features must be:

- independent
- equal

It’s relatively a simple idea indeed. However Naive Bayes can often outperform other more sophisticated algorithms.

Pros and Cons of Naive Bayes

Pros and cons of Naive Bayes:

Pros

🎬It’s relatively simple to understand and build

🎬It’s easily trained, even with a small dataset

🎬It’s fast!

🎬It’s not sensitive to irrelevant features

Cons

🎬 It assumes the every feature is independent, which isn’t always the case in the reality.

Example of Observation

Consider a fictional dataset that describes the weather conditions for playing a game of golf. Given the weather conditions, each tuple classifies the conditions as fit(“Yes”) or unfit(“No”) for playing golf

The dataset is divided into two parts, namely, feature matrix and the response vector.

🎬 Feature matrix contains all the vectors(rows) of dataset in which each vector consists of the value of dependent features. In above dataset, features are ‘Outlook’, ‘Temperature’, ‘Humidity’ and ‘Windy’.

🎬 Response vector contains the value of class variable(prediction or output) for each row of feature matrix. In above dataset, the class variable name is ‘Play golf’.

A Fictional Dataset

	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
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Interpretation of Assumptions

Based on this dataset, the concept of both assumptions is understood as follows:

🎬 We assume that no pair of features are dependent. For example, the temperature being ‘Hot’ has nothing to do with the humidity or the outlook being ‘Rainy’ has no effect on the winds. Hence, the features are assumed to be **independent**.

🎬 Secondly, each feature is given the same weight (or importance). For example, knowing only temperature and humidity alone can’t predict the outcome accuratey. None of the attributes is irrelevant and assumed to be contributing **equally** to the outcome.

Note

The assumptions required by Naive Bayes are not generally correct in reality. In fact, the independence assumption does not always meet but often works well in practice.

In a nutshell, the algorithm allows us to predict a class, given a set of features using probability.

Theorem of Bayes

Bayes’ Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes’ theorem is stated mathematically as the following equation:

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

where A and B are events and $P(B) \neq 0$

Note

🎬 Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as **evidence**.

🎬 P(A) is the **priori** of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance (here, it is event B).

🎬 P(A|B) is a **posteriori** probability of B, i.e. probability of event after evidence is seen.

Applying the Theorem of Bayes

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

where, y is class variable and X is a dependent feature vector (of size n) where:

$$X = (x_1, x_2, x_3 . \ldots, x_n)$$

Note

Just to clear, an example of a feature vector and corresponding class variable can be: (refer 1st row of dataset)

X = (Rainy, Hot, High, False)

y = No

So basically, P(X|y) here means, the probability of “Not playing golf” given that the weather conditions are “Rainy outlook”, “Temperature is hot”, “high humidity” and “no wind”.

Naive Assumption

Now, its time to put a naive assumption to the Bayes’ theorem, which is, independence among the features. So now, we split evidence into the independent parts.

Now, if any two events A and B are independent, then, P(A,B) = P(A)P(B)

Hence, we reach to the result:

$$P\left(y \mid x_1 x_2 x_3 \ldots, x_n\right) = ?$$

Naive Assumption

Which can be expressed as:

$$P\left(y \mid x_1, x_2, x_3, \ldots, x_n\right) = \frac{P\left(y\right) \prod_{i=1}^{n} P\left(x_i \mid y\right)}{P\left(x_1\right) P\left(x_2\right) \ldots P\left(x_n\right)}$$

Due to the denominator remains constant for a given input, it can be removed from such term to be:

$$P\left(y \mid x_1, x_2, x_3, \ldots, x_n\right) \propto P\left(y\right) \prod_{i=1}^{n} P\left(x_i \mid y\right)$$

Classifier Model

Based on such term, how the model of classifier can be created?

The model can be used to classify the a set of given input by selecting class variable y with maximum probability. This expression can be denoted as the following mathematical term:

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i | y)$$

The only calculation for this formula is to compute the probability P(y) and P(x_i | y).

P(y) is also called **class probability** and P(x_i | y) is called **conditional probability**.

Weather Dataset

What should be done?

- Performing some precomputation on weather dataset.
- The intention of computation is to find the probability $P(x_i \mid y_j)$ for each x_i in X and y_j in y
- The calculation is done for probability of outlook, temperature, humidity, wind and probability of class.

Outlook

	Yes	No	P (yes)	P (no)
Sunny				
Overcast				
Rainy				
Total				

P(outlook = overcast | play golf = Yes) = 4/9.

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Class probability

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Example

Let have a condition called **today** = (Sunny, Hot, Normal, False)

Let y: playing golf

X1 : outlook

X2 : temperature

X3 : humidity

X4 : wind

How is the probability of playing golf today? Is the decision to play or not to play?

Example

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

$$P(y|X)=\frac{P(x_1=sunny|y)\,P(x_2=hot|y)\,P(x_3=normal|y)\,P(x_4=false|y)\,P(y)}{P(X=today)}$$

Since the question is whether the decision to play or not to play, the given information (prior information) is y = yes or y = no if the condition is **today** (evidence). Therefore, the computation must be done for P(y = yes | X) and P(y = no | X).

Example

$$P(y=yes|X)=\frac{P(x_1=sunny|y=yes)P(x_2=hot|y=yes)P(x_3=normal|y=yes)P(x_4=false|y=yes)P(y=yes)}{P(X=today)}$$

$$P(y=no|X)=\frac{P(x_1=sunny|y=no)P(x_2=hot|y=no)P(x_3=normal|y=no)P(x_4=false|y=no)P(y=no)}{P(X=today)}$$

Note that in both probabilities, the denominator P(X = today) is common for both so that the probability can rely on computation of nominator only.

Example

$$P(y=yes|X) \propto P(x_1=sunny|y=yes)P(x_2=hot|y=yes)P(x_3=normal|y=yes)P(x_4=false|y=yes)P(y=yes)$$

P(y=yes|X = today) = ?

$$P(y=no|X) \propto P(x_1=sunny|y=no)P(x_2=hot|y=no)P(x_3=normal|y=no)P(x_4=false|y=no)P(y=no)$$

P(y=no|X = today) = ?

Play or not to play ....

Question ??

Thank you for your attention