International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Use of Naive Bayes Algorithm for Classification in Supervised Machine Learning

SWATI MAHESH JOSHI

Lecturer, Department of Information Technology, V.P.M.'s Polytechnic, Thane, Maharashtra, India.

Abstract - *An Artificial Intelligence is a significant part of industrial revolution 4.0. It provides facilities to automate the processes by machines thinking like humans. Machine learning is a subset of Artificial Intelligence which actually uses data previously collected and by classifying, clustering, and associating like patterns comes to a decision. It uses different algorithms to build models and analyze data. The algorithm explained here is Naïve Byes algorithm which based on Bayes theorem. It uses probabilities from different classes of the previously collected dataset and creates classifiers. The classifier helps in predictions and taking decisions. The algorithm is explained with example at the end with a case study.*

Key Words: Artificial Intelligence, Machine learning, supervised learning, classification, Naïve Byes algorithm, prediction.

1. What is machine learning?

Machine learning is a technique which teaches the computes to learn from experiences just like humans. It is basically a subset of Artificial Intelligence. The idea here is to learn from data, identify various patterns and make decisions with minimum human intervention.

1.1 How machine learning works?

There are two techniques by which machine learning works. Supervised learning and unsupervised learning.

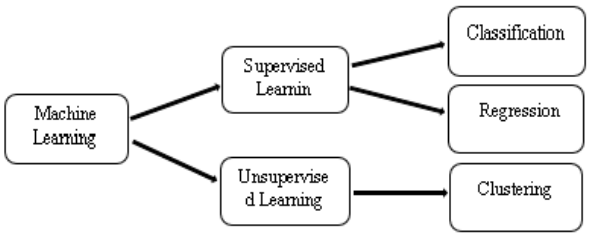

Fig 1.1 Types of machine learning

A supervised learning technique takes known input sets and designs a model to generate reasonable predictions for the response to new data. Supervised learning can be used when you have known data for the output you are trying to predict.

For example: In a basket of fruits, if we want to name the fruits, we can predict the name by known categories as color as red and shape as round is apple, or color as green yellow and shape as long curve we can name it as banana.

Supervised learning uses classification and regression techniques to find predictive models.

Unsupervised learning technique is that which uses no knows category or known data.it is a kind of self-learning technique in which system has to discover the features of the inputs by its own.

For example: in group of dogs and cats, if we don't have an idea about features of dogs and cats, we cannot categorize them, we can find some similarities, differences and patterns and form clusters to make two categories to separate them.

2. Different Algorithms used in supervised learning:

Linear Regression: It is used when linear relationship between input and output is to be predicted.

K-Nearest Neighbor: K-NN algorithm is mostly in classification which assumes the similarity between the new data and available data and accordingly puts the new data into the category that is most similar to the available categories.

Decision Trees: Decision trees are created with approach that identifies ways to split a data set on various conditions. It is one of the most widely used techniques for supervised learning.

Support Vector Machine (SVM): SVM technique includes multimedia information retrieval, bioinformatics, and pattern recognition.

Random Forest: Random forest algorithm consists of number of decision Trees. Each individual tree in the random forest spits out in a class prediction and the class with the most votes becomes the prediction

Naive Bayes: It is a probabilistic machine learning algorithm that can be used mainly in text classification and with the problems having multiple classes.

2. 1 Naive Bayes Algorithm:

As it is already discussed, Naïve Bayes algorithm is a supervised learning algorithm. It is mainly used in classification where text data is involved with high dimensional data set. It is the most commonly used classification algorithm for quick predictions.

Naïve Byes is a probabilistic algorithm which predicts on the basis of probabilities of individual entity from data set.

In the name Naïve Byes, the word Naïve is used to indicate that each feature in the classification is independent of other feature and Bayes is because it is based on the Bayes Theorem.

Bayes Theorem:

The formula for Bayes' theorem is given as:

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

where A and B are events and P(B) is not 0.
P(A) is called as the marginal probability of A.
P(B) is called as the marginal probability of B
P(A|B) is called as the conditional probability of A
P(B|A) is called as the conditional probability of B

3. Case study:

Working of Naïve Bayes Theorem can be well understood by following example. Let us consider a dataset containing car buyers. The dataset shows the values as age of the person, income and credit limit and can be used to predict by the probabilities of these features

Consider the data set:

Sr. No	age	income	credit limit	buys car
1	youth	high	fair	no
2	youth	high	excellent	no
3	middle age	high	fair	yes
4	senior	medium	fair	yes
5	senior	low	fair	yes
6	senior	low	excellent	no
7	middle age	low	excellent	Yes
8	youth	medium	fair	no
9	senior	medium	fair	Yes
10	youth	medium	excellent	Yes
11	middle age	medium	excellent	yes
12	middle age	high	fair	yes

Table 1: Data set

The data set shows features as age, income, credit limit and if the person has bought a car or not at the end. By considering the above dataset we can find a classifier for a new combination of feature. For eg : if we want to predict for a data set X={youth, low, fair} if the person can buy a car, we can apply Naïve Byes theorem and algorithm and find the classifier.

As the theorem states, P(A|B) can be found out for both yes and no cases of buying for all the three features.

Consider X={youth,low.fair} ,Then,

Probability P(youth|yes) can be found as

$$\frac{P(yes|youth)P(youth)}{P(yes)}$$

So substituting the values from above data sheet we get

$$P(\text{youth}|\text{yes}) = \frac{\frac{1}{4}*\frac{4}{12}}{\frac{8}{12}} = \frac{1}{8}$$

Same way we can find for

$$P(\text{low}|\text{yes}) = \frac{\frac{2}{3}*\frac{3}{12}}{\frac{8}{12}} = \frac{2}{8}$$

$$\text{And } P(\text{fair}|\text{yes}) = \frac{\frac{5}{7}*\frac{7}{12}}{\frac{8}{12}} = \frac{5}{8}$$

Find out probabilities of posterior, evidence, proposition and likelihood as shown in tables:

Total rows = 12		P(X)
yes	8	$\frac{8}{12}$
no	4	$\frac{4}{12}$

Table 2: evidence P(yes) and P(no)

age	yes	no	P(X)	P(X/yes)	P(X/no)
youth	1	3	$\frac{4}{12}$	$\frac{1}{8}$	$\frac{3}{4}$
middle age	4	0	$\frac{4}{12}$	$\frac{4}{8}$	0
senior	3	1	$\frac{4}{12}$	$\frac{3}{8}$	$\frac{1}{4}$
total	8	4	1	1	1

Table 3.1: proposition and likelihood (age)

income	yes	no	P(X)	P(X/yes)	P(X/no)
high	2	2	$\frac{4}{12}$	$\frac{2}{8}$	$\frac{2}{4}$
medium	4	1	$\frac{5}{12}$	$\frac{4}{8}$	$\frac{1}{4}$
low	2	1	$\frac{3}{12}$	$\frac{2}{8}$	$\frac{1}{4}$
total	8	4	1	1	1

Table 3.2 proposition and likelihood (income)

Credit limit	yes	no	P(X)	P(X/yes)	P(X/no)
Excellent	3	2	$\frac{5}{12}$	$\frac{3}{8}$	$\frac{2}{4}$
fair	5	2	$\frac{7}{12}$	$\frac{5}{8}$	$\frac{2}{4}$
total	8	4	1	1	1

Table 3.3 proposition and likelihood (credit limit)

For given X=(youth,low,fair),
The probability P(X|yes) =
P(youth|yes) *P(low|yes)*P(fair|no)*P(yes)

$$= \frac{1}{8}*\frac{2}{8}*\frac{5}{8}*\frac{8}{12}$$

= 0.125*0.25*0.625*0.666 = $\boxed{0.0130}$ →(1)

The probability P(X|no) =
P(youth|no) *P(low|no)*P(fair|no)*P(no)

$$= \frac{3}{4}*\frac{1}{4}*\frac{2}{4}*\frac{4}{12}$$

= 0.75*0.25*0.5*0.33= $\boxed{0.03093}$ →(2)

Comparing (1) and (2) we see that 0.0393 >0.0130, so prediction is "no" and we can predict for X= {youth, low, fair} the person is predicted as not buying a car.

4. Conclusion:

Naive Bayes algorithms are used for classification in supervised machine learning. Mostly they are used in sentiment analysis, spam detection, recommendation systems etc. It's a fast and easy to implement algorithm but the requirement that the features required to be independent of other becomes a disadvantage sometimes. Because in most of the real-life cases, the predictors are dependent. But because of less complexity and easy to implement Naïve Bayes algorithm is mainly used in supervised machine learning.

References:

[1] Ethem Alpaydin "Introduction to machine learning ".
[2] P. Langley. "Elements of Machine Learning", Morgan Kaufmann Publishers, Inc. 1996.
[3] https://www.mathworks.com
[4] https://www.geeksforgeeks.org