Naïve Bayes Classification

Ruchika Rana[1] Jyoti pruthi[2]

Abstract—The health care industry collects huge amounts of health care data which unfortunately are not mined to discover hidden information for effective decision making..Discovery of hidden patterns and relationships often goes unexploited.Naive bayes assumes that the presence or absence of particular feature of a class is unrelated to the presence or absence of any other feature.This algorithm is based on conditional probabilities .It uses bayes theorem concepts that tells us the occurrence of an event.We assumes that this algorithm is best among all other classifiers like decision tree, data miming etc.

Keywords:- Naïve classifier, supervised training, text mining, document ,independence relations

I. INTRODUCTION

Naïve bayes uses bayes theorem that calculates a probability by counting frequency of values and combination of values in the historic data.It assumes that the presence or absence of particular feature of a class is unrelated to the presence or absence of any other feature. Naïve Bayes text classifier has been widely used because of its simplicity in both the training and classifying stage . Although it is less accurate than other discriminative methods (such as SVM), numerous researchers proved that it is effective enough to classify the text in many domains .Naïve Bayes models allow each attribute to contribute towards the final decision equally and independently from other attributes, in which it is more computational efficient when compared with other text classifiers. Thus, the present study focuses on employing Naïve Bayes approach as the text classifier for document classification and thus evaluates its classification performance

Naïve bayes is a statistical classifier which assigns no dependency between attributes. To determine the class the posterior probability should be maximized.It requires training data or training set for its feature space.naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods

An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

II. BASICS OF NAÏVE CLASSIFIERS

Naïve claasifiers are based on Bayesian networks. It makes the independence assumption that the input features are conditionally independent of each other in a given classification.The independence of the naïve Bayesian claasifier is embedded in aparticular belief network where the features are the nodes and the target variable has no parent , and the claasification is only parent of each input variable. The belief network requires the probability distribution.

Let us consider the figure we take a node A which is itself a node we start with attribute and traverse attribute list of Aand then update value of A if a is numeric then we split the values of node and then evaluate each and every attribute and then we put the values of nods in database and we are independent of other relations in function A is our training data and we determine the matrix only for a rest are independent for A.

IN This way we consider naïve classification now we take another example of fruit which explain in detail how it works.

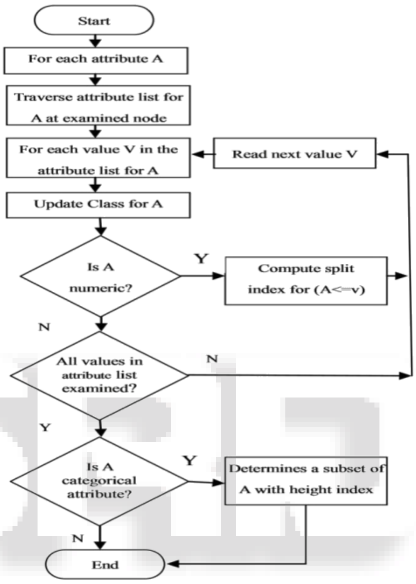

Fig. 1:

A. *Fruit Example*

Let's try it out on an example to increase our understanding: The OP asked for a 'fruit' identification example.

Let's say that we have data on 1000 pieces of fruit. They happen to be Banana, Orange or some Other Fruit. We know 3 characteristics about each fruit:

- Whether it is Long
- Whether it is Sweet and
- If its color is Yellow.

This is our 'training set.' We will use this to predict the type of any new fruit we encounter.

Type Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	Total
Banana 500	400	100	350	150	450	50
Orange 300	0	300	150	150	300	0
Other Fruit 200	100	100	150	50	50	150
Total 1000	500	500	650	350	800	200

We can pre-compute a lot of things about our fruit collection.

The so-called "Prior" probabilities. (If we didn't know any of the fruit attributes, this would be our guess.)

These are our base rates.
P(Banana) = 0.5 (500/1000)
P(Orange) = 0.3
P(Other Fruit) = 0.2

Probability of "Evidence"
p(Long) = 0.5
P(Sweet) = 0.65
P(Yellow) = 0.8

Probability of "Likelihood"
P(Long/Banana) = 0.8
P(Long/Orange) = 0 [Oranges are never long in all the fruit we have seen.]
P(Yellow/Other Fruit) = 50/200 = 0.25
P(Not Yellow/Other Fruit) = 0.75

Given a Fruit, how to classify it?

Let's say that we are given the properties of an unknown fruit, and asked to classify it. We are told that the fruit is Long, Sweet and Yellow. Is it a Banana? Is it an Orange? Or Is it some Other Fruit?

We can simply run the numbers for each of the 3 outcomes, one by one. Then we choose the highest probability and 'classify' our unknown fruit as belonging to the class that had the highest probability based on our prior evidence (our 1000 fruit training set):

P(Banana/Long, Sweet and Yellow) = P(Long/Banana) p(Sweet/Banana).P(Yellow/Banana) x P(banana)
____
P(Long). P(Sweet). P(Yellow)

= 0.8 x 0.7 x 0.9 x 0.5
____
P(evidence) = 0.252/P(evidence)

P(Orange/Long, Sweet and Yellow) = 0
P(Other Fruit/Long, Sweet and Yellow) = P(Long/Other fruit) x P(Sweet/Other fruit) x P(Yellow/Other fruit) x P(Other Fruit) = (100/200 x 150/200 x 50/150 x 200/1000)

III. Steps For Building A Bayesian Classifier

A. Formula:

Bayes theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

Likelihood — Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability — Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

1P(c|x) is the posterior probability of class (target) given predictor (attribute).
P(c) is the prior probability of class.
P(x|c) is the likelihood which is the probability of predictor given class.
P(x) is the prior probability of predictor.

B. The Requirements For A Naive Bayes Model Are As Follows:

– A single key column Each model must contain one numeric or text column that uniquely identifies each record. Compound keys are not allowed.
– Input columns In a Naive Bayes model, all columns must be either discrete or discretized columns. For information about discretizing columns, see Discretization Methods (Data Mining).
– For a Naive Bayes model, it is also important to ensure that the input attributes are independent of each other. This is particularly important when you use the model for prediction.
– The reason is that, if you use two columns of data that are already closely related, the effect would be to multiply the influence of those columns, which can obscure other factors that influence the outcome.
– Conversely, the ability of the algorithm to identify correlations among variables is useful when you are exploring a model or dataset, to identify relationships among inputs.
– At least one predictable column The predictable attribute must contain discrete or discretized values.

The values of the predictable column can be treated as inputs. This practice can be useful when you are exploring a new dataset, to find relationships among the columns. Creating, Designing, and Linking It Up:QR codes can be customised according to your brand.

C. Prediction

Atlast we predict that is the final step is the prediction in which we predict all the variables and attributes independently which solve our purpose and we get the final classifier.

IV. Advantages

– It converge quicker than any other Models
– Provides practical learning algorithm'
– It is easy to implement.
– It shows good result optimization
– It works efficiently with large training sets
– It has goodness of classification rules.
– DISADVANTAGES
– .Loss of accuracy because of class conditional independence
– It takes much time to implement.
– Compression based learning
– It has poor performance in domain knowledge.

V. Conclusion

The system extract hidden knowledge from a historical data.We require training sets that corporate with the real world . Data preparation needs to construct the final dataset and form subset from hidden information.we can combine results of naïve bayes and other models and provides users all attribute values that relate to predictable state.

References

[1] S.L.TANG W.H. Ip,Albert H.L.T"Is naïve bayes a good classifier or document claasifier"
[2] Microsoft Developer Network (MSDN)
[3] Jiangtao Ren———————————————, Sau"Naïve based classification of uncertain data