The Overview of Bayes Classification Methods

Harshad M. Kubade

Assistant Professor,
Department of Information Technology, PIET, Nagpur, India

ABSTRACT

The Bayes methods are popular in classification because they are optimal. To apply Bayes methods, it is required that prior probabilities and distribution of patterns for class should be known. The pattern is assigned to highest posterior probability class. The three main methods under Bayes classifier are Byes theorem, the Naive Bayes classifier and Bayesian belief networks.

Keywords: Bayes theorem, Bayesian belief networks, Naive Bayes classifier, prior probability, posterior probability

I. INTRODUCTION

Bayes classifier is popular as it reduces the average probability of error. Bayes classification methods require prior probability and distribution of classes. The posterior probability is calculated and the test pattern is assigned to the class having highest posterior probability. The following are the various Bayes classification methods.

II. BAYES THEOREM

Assume t is the test pattern whose class is not known. Let prior probability of class C is given by P(C). To classify t, we have to determine posterior probability which is denoted by P (C|t).

$$P(C \mid t) = \frac{P(t \mid C)P(C)}{P(t)}$$

If the prior probability of the patient has malaria is P(M) = 0.2 . Then the probability that the patient has not malaria is 0.8. The probability of temperature P (t) is 0.4. The conditional probability P (t|M) is 0.6.Then by Bayes theorem, the posterior probability that the patient has malaria is given by

$$P(M \mid t) = \frac{P(t \mid M)P(M)}{P(t)}$$

Therefore, the probability that patient has malaria is given below

$$P(M \mid t) = \frac{0.6 \times 0.2}{0.4} = 0.3$$

III. Naive Bayes classifier

The Naive Bayes classifier is a probabilistic classifier based on Bayes theorem. The Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated of a presence(or absence) of any other feature, given the class variable. The advantage of Naive Bayes classifier is that it requires a small amount of training data to estimate the parameters for the classification. The probability model for a classifier is a conditional model
$P (C \mid F_1, ..., Fn)$
Using Bayes theorem, the above statement can be written as

$$p(C \mid F1, ..., Fn) = \frac{p(C)\, p(F1, ..., Fn \mid C)}{p(F1, ..., Fn)}$$

The denominator does not depend on C and the values of the features Fi are given, so the denominator is effectively constant. The numerator is equivalent to the joint probability model
$P (C, F1 , . . . , F n)$
which can be rewritten as follows, using repeated applications of the definition of conditional probability:
$p (C, F_1 , . . . , F n)$
$= p(C)\, p(F_1 , . . . , F n \mid C)$
$= p(C)\, p(F_1 \mid C)\, p(F_2 , . . . , Fn \mid C, F_1)$
$= p(C)\, p(F_1 \mid C)\, p(F_2 \mid C, F_1)\, p(F_3 , . . . , Fn \mid C, F_1 , F_2)$

$= p(C)\, p(F1\,|C)\, p(F_2\,|C, F_1)\, p(F3\,|C, F_1, F_2)\, p(F_4, \ldots, Fn\,|C, F_1, F_2, F_3)$

and so forth. Using the naive conditional independence assumptions, each feature Fi is conditionally independent of every other feature Fj for $j \neq i$. This means that

$p\,(F_i\,|C, F_j) = p(F_i\,|C)$

and so the joint model can be expressed as

$p(C, F1, \ldots, Fn) = p(C)\, p(F_1\,|C)\, p(F_2\,|C)\, p(F_3\,|C) \cdots p(Fn\,|C)$

$$= p(C) \prod_{i=1}^{n} p(Fi|C)$$

This means that under the above assumptions of independence, the conditional distribution over the class variable C can be expressed as

$$P(C|1,..Fn) = \frac{1}{z}\, P(C) \prod_{i=1}^{n} p(Fi|C)$$

where Z is a scaling factor dependent only on $F_1, \ldots, Fn$, i.e., a constant if the values of the feature variables are known.

The naive Bayes classifier combines the Bayes probability model with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the maximum a posterior or MAP decision rule. The corresponding classifier is the function ''classify'' defined as follows:

$$\text{classify}(f1, \ldots, fn) = \text{argmax}\; P(C = c) \prod_{i=1}^{n} p(Fi = fi\,|C = c)$$

IV. Bayesian Belief Network

A Bayesian network (or a belief network) is a probabilistic graphical model. This model represents a set of variables and their probabilistic dependencies. Bayesian Belief Networks provide the modeling and reasoning about the uncertainty. Bayesian Belief networks cover both subjective probabilities and probabilities based on objective data. For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. A belief network, also called a Bayesian network, is an acyclic directed graph (DAG), where the nodes are random variables. If there is an arc from node A to another node B, A is called a parent of B, and B is a child of A. The set of parent nodes of a node Xi is denoted by its parents (Xi). Thus, a belief network consists of

1. a DAG, where each node is labeled by a random variable;
2. a domain for each random variable; and
3. a set of conditional probability distributions giving P(Xi|parents(Xi)).

Full joint distribution is defined in terms of local conditional distributions as

$$P(X1, X2, \ldots\ldots, Xn) = \prod_{i=1}^{n} P(Xi|\text{ Parents }(Xi))$$

If node Xi has no parents, its local probability distribution is said to be unconditional, otherwise it is conditional. If the value of a node is observed, then the node is said to be an evidence node. Each variable is associated with a conditional probability table which gives the probability of this variable for different values of its parent nodes.

The value of P (X1, ..., Xn) for a particular set of values for the variables can be easily computed from the conditional probability table.

Consider the five events M, N, X, Y, and Z. Each event has a conditional probability table. The events M and N are independent events i. e. they are not influenced by any other events. The conditional probability table for X shows the probability of X,
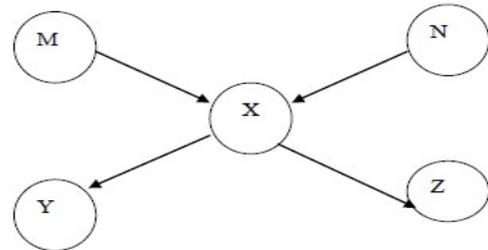

Figure: Bayesian Belief Network

given the conditions of the events M and N i. e. true or false. The event X occurs if event M is true. The event X also occurs if event N does not occur. The events Y and Z occur if the event X does not occur. The following figure shows the belief network for the given situation.

REFERENCES
1. V Susheela Devi and M Narasimha Murty, "Pattern Recognition: An Introduction", Universities Press (India) Private Limited.

2. Robert Schalkoff, "Pattern Recognition: Statistical, Structural And Neural Approaches", Wiley India (P.) Ltd.

3. Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", Wiley India (P.) Ltd.

4. Bayesian Networks Wikipedia.