INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 02, February 2024, pp : 439-442

e-ISSN : 2583-1062

Impact Factor : 5.725

AN ANALYSIS OF THE NAIVE BAYES CLASSIFIER EMPIRICALLY

B. Swetha[1]

[1]M. Sc. Department of computer science, Fatima college, Madurai, India.

ABSTRACT

The Naive Bayes classifier assumes that characteristics are independent of class, which considerably simplifies learning. While independence is often a bad assumption, naïve Bayes actually frequently competes effectively with more advanced classifiers. Our main objective is to comprehend the features of the data that influence naive Bayes' performance. Our methodology makes use of Monte Carlo simulations, which enable a methodical investigation of categorization accuracy over a number of classes of randomly produced problems. We examine how the distribution entropy affects the classification error and demonstrate that low-entropy feature distributions result in good naive Bayes performance. Additionally, we show that naïve Bayes performs optimally in two contradictory scenarios: fully independent features (as predicted) and functionally dependent features . Another unexpected finding is that there is no clear correlation between the degree of feature dependencies—which is defined as the class-conditional mutual information between the features—and the accuracy of naive Bayes. The amount of class information lost as a result of the independence assumption is a more accurate indicator of naive Bayes correctness.

1. INTRODUCTION

Bayesian classifiers assign the most likely class to a given example described by its feature vector. Learning such classifiers can be greatly simplified by assuming that features are independent given class, that is $P(X|C) = \pi \quad {}^{n} \; i = 1$ where is a feature vector and is a class. Despite this unrealistic assumption, the resulting classifier known as naive bayes is remarkably successful in practice, often competing with much more sophisticated techniques. Naive Bayes has proven effective in many practical applications, including text classification, medical diagnosis, and systems performance management.

The following explains why naive bayes performs well when feature dependencies are present: There is no guarantee that optimality and zero-one loss (classification error) are related to the appropriateness of the independence assumption, or the adequacy of the fit to a probability distribution. Instead, if the actual and predicted distributions concur on the most likely class, an optimal classifier is produced. For instance, demonstrated the prove Naive Bayes optimality for a number of problem classes, including disjunctive and conjunctive concepts, that have a high degree of feature dependencies.

But this explanation isn't really informative because it's too broad. In the end, our goal is to comprehend the features of the data that influence how well Naive Bayes. In contrast to most Naive Bayes research, which evaluates the algorithm's performance against other classifiers on specific benchmark problems (like UCI benchmarks), our method makes use of Monte Carlo simulations to enable a more methodical investigation of classification accuracy on parametric families of randomly generated problems. Furthermore, we are only examining the bias of the naive Bayes classifier in this analysis—not its variance.

In particular, we presume an unlimited quantity of data (i.e., perfect knowledge of data distribution), which enables us to distinguish between the error caused by the training sample set and the approximation error (bias) of naive Bayes.

We analyze the impact of the distribution entropy on the classification error, showing that certain almost deterministic, or low-entropy, dependencies yield good performance of naive Bayes. We show that the error of naive Bayes vanishes as the entropy **H(P(X|C)** approaches zero. Another class of almost-deterministic de pendencies generalizes functional dependencies between the features.

Particularly, we show that naive Bayes works best in two cases: completely independent features (as expected) and functionally dependent features. We also show that, surprisingly, the accuracy of naive Bayes is not directly correlated with the degree of feature de pendencies measured as the class-conditional mutual information between the feature **I(Xi;Xj|C).**

Instead, our experiments re veal that a better predictor of naive Bayes accuracy can be the loss of information that features contain about the class when assuming naïve Bayes model , namely

$I(C;(X_i,X_j)) - I_{NB}(C;(X_i,X_j))$, where **INB**

is the mutual information between features and class under naive Bayes assumption. This paper is structured as follows. In the next section we provide necessary background and definitions .

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 02, February 2024, pp : 439-442

e-ISSN : 2583-1062

Impact Factor : 5.725

2. WHEN DOES NAIVE BAYES WORK WELL? EFFECTS OF SOME NEARLY-DETERMINISTIC DEPENDENCIES

In this section, we discuss known limitations of naive Bayes and then some conditions of its optimality and near optimality, that include low-entropy feature distributions and nearly-functional feature dependencies. We focus first on concepts with **P(C=i|x)=0 or 1** or for anyand (i.e. no noise), which therefore have zero Bayes risk.), which therefore have zero Bayes risk. The features are assumed to have finite domains (i -th feature has values), and are often called nominal. (A nominal feature can be transformed into a numeric one by imposing an order on its domain.) Our attention will be restricted to binary classification problems where the class is either 0 or 1. When K>I for some features, naive Bayes is able to learn (some) polynomial discriminant functions ; thus, polynomial separability is a necessary, although not sufficient , condition of naive Bayes optimality for concepts with finite-domain features. Despite its limitations, naive Bayes was shown to be optimal for some important classes of concepts that have a high degree of feature dependencies, such as disjunctive and conjunctive concepts. These results can be generalized to concepts with any nominal features.

3. THEOREM 1

The naive Bayes classifier is optimal for any two-class concept with nominal features that assigns class 0 to exactly one example, and class 1 to the other examples, with probability 1.

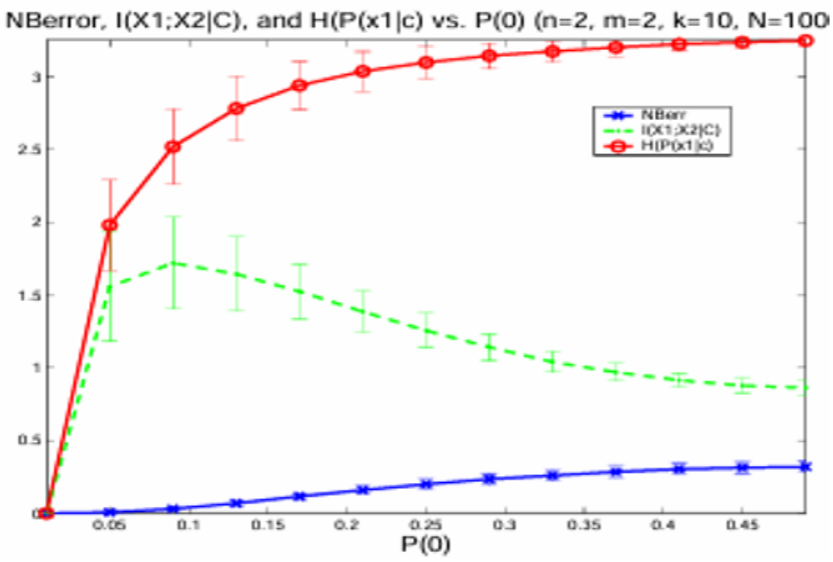
The performance of naive Bayes degrades with increasing number of class-0 examples (i.e., with increasing prior **P(C=0)**, also denoted P(0)), as demonstrated in Figure 1a. This figure plots average naive Bayes error computed over 1000 problem instances generated randomly for each value of **P(C=0).**

The problem generator, called zerobayesrisk, assumes m features (here we only consider two features), each having k values.

As expected, larger P(C=0) yield a wider range of problems with various dependencies among features. Which result into increased errors of bayes a closer look at the data shows no other cases of optimality besides P(C=0)=1/N). Surprisingly, the strength of inter feature dependencies, measured as the class conditional mutual information I(X1;X2|C), is not a good predictor of naïve bayes performance: while average naïve bayes error increases monotonically with P(0), the mutual information is non-monotone, reaching its maximum around P(0)=0.1.

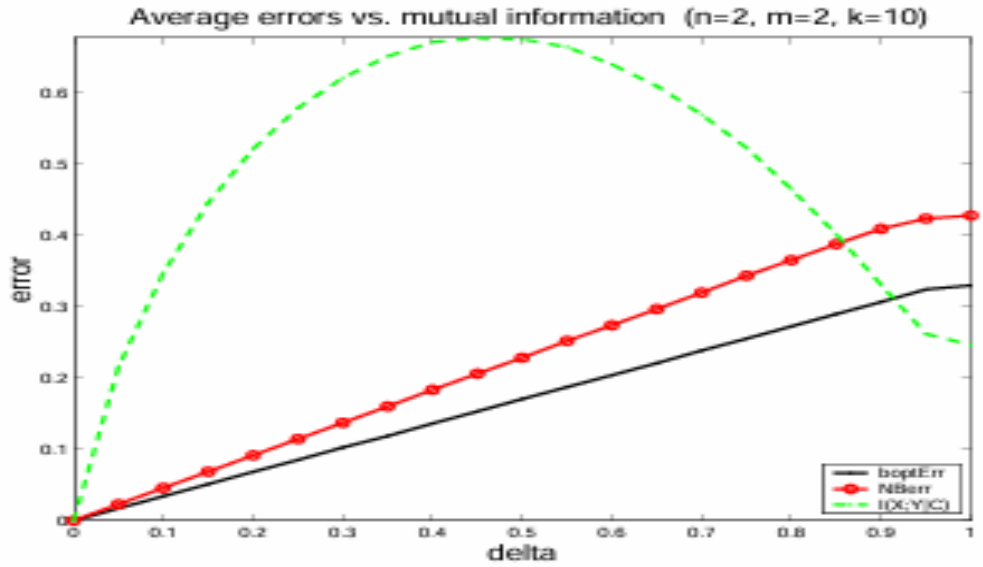
This observation is consistent with previous empirical results on UCI benchmarks that also revealed low correlation between the degree offeature dependence and relative performance of naïve Bayes with respect too other classifiers, such asC4.5,CN2,andPEBLS. It turns out that the entropy of class-conditional marginal distributions, P(Xi|C), is a better predictor of naïve bayes performance. Intuitively ,low entropy of P(Xi|0) means that most of 0s are "concentrated around" one state.

Indeed plot average H(P(X1|0) in figure 1a demonstrates that both average error and average entropy increase monotonically in P(0).



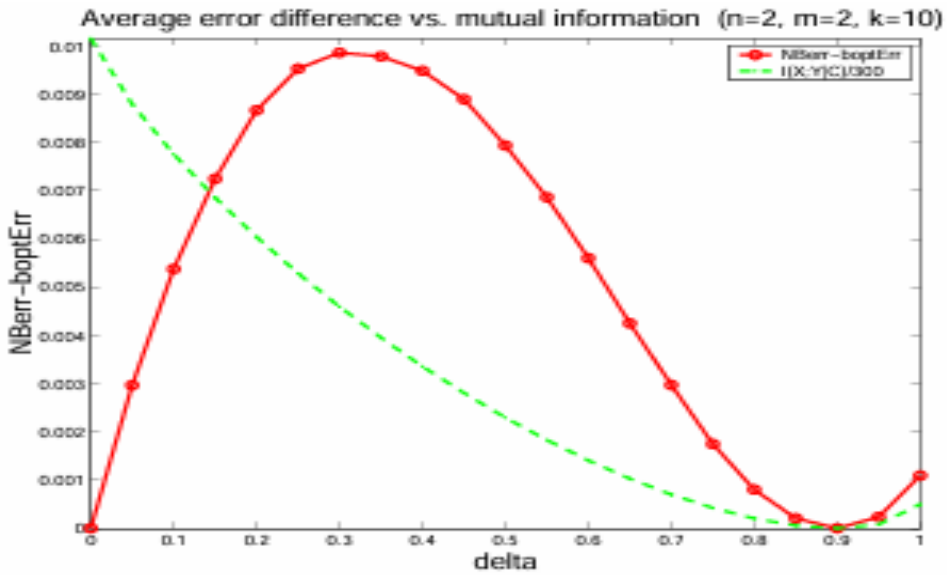
(a)

Fig. 1



(b)

Fig. 2



(c)

Fig. 3

(a) Results for the generator zerobayes risk (k=10, 1000 instances): average naïve bayes error, class conditional mutual information between features (I(X1|X2|C)), and entropy of marginal distribution ,H(P(X1|0)); the error bars correspond to the standard deviation of each measurement across 1000 problems instances;

(b) Results for the generator EXTREME: average bayes and naïve bayes errors and average I(X1;X2|C);

(c) Results for the generator FUNCI :average difference between naïve bayes error and bayes error(=0.336 – constant for all δ), and scaled I(X1;X2_C)(divided by 300)

4. INFORMATION LOSS: A BETTER ERROR PREDICTOR THAN FEATURE DEPENDENCIES?

As we observed before, the strength of feature dependencies(i.e. the class-conditional mutual information between the features) 'ignored' by naive Bayes is not a good predictor of its classification error. This makes us look for a better parameter that estimates the impact of independence assumption onclassification. We start with a basic question: which dependencies between features can be ignored when solving a classification task? Clearly, the dependencies which do not help distinguishing between different classes, i.e. do not provide any information about the class. Formally, let I(C;(X1,X2)) be the mutual information between the features and the class given the "true" distribution P(X1,X2,C),PNB(X1,X2,C)=P(X1|C)P(X2|C)P(C), the naïve bayes approximation of P(X1,X2,C).

Then the parameter I diff=I(C;X1,X2))-INB(C;X1,X2)) measures the amount of information about the class which is "lost" due tonaive Bayes assumption.

5. CONCLUSIONS

Despite its unrealistic independence assumption, the naive Bayes classifier is surprisingly effective in practice since its classification decision may often be correct even if its probability estimates are inaccurate. Although some optimality conditions of naive Bayes have been already identified in thepast , a deeper understanding of data characteristics that affect the performance of naive Bayes is still required.

Our broad goal is to understand the data characteristics which affect the performance of naive Bayes. Our approach uses Monte Carlo simulations that allow a systematic study of classification accuracy for several classes of randomly generated problems. We analyze the impact of the distribution entropy on the classification error, showing that certain almost- deterministic, or low-entropy, dependencies yield good performance of naive Bayes. Particularly, we demonstrate that naive Bayes works best in two cases: completely independent features (as expected) and functionally dependent features (which is surprising). Naive Bayes has its worst performance between these extremes.

Surprisingly, the accuracy of naive Bayes is not directly correlated with the degree of feature dependencies measured as the class-conditional mutual information between the features. Instead, a better predictor of accuracy is the loss of information that features contain about the class when assuming naive Bayes model. However, further empirical and theoretical study is required to better understand the relation between those information-theoretic metrics and the behavior of naive Bayes.

Further directions also include analysis of naive Bayes on practical application that have almost-deterministic dependencies, characterizing other regions of naive Bayes optimality and studying the effect of various data parameters on the naive Bayes error. Finally, a better understanding of the impact of independence assumption on classification can be used to devise better approximation techniques for learning efficient Bayesian net classifiers, and for probabilistic inference **e.g**., for finding maximum-likelihood assignments.

6. REFERENCES

[1] T.M. Cover and J.A. Thomas. Elements of information theory.New York:John Wiley & Sons, 1991.

[2] P. Domingos and M. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29:103–130, 1997.

[3] R.O. Duda and P.E. Hart. Pattern classification and scene anal- ysis. New York: John Wiley and Sons, 1973.

[4] N. Friedman, D. Geiger, and Goldszmidt M. Bayesian networkclassifiers. Machine Learning, 29:131–163, 1997.

[5] J. Hellerstein, Jayram Thathachar, and I. Rish. Recognizing end-user transactions in performance management. In Pro- ceedings of AAAI-2000, pages 596–602, Austin, Texas, 2000.

[6] J. Hilden. Statistical diagnosis based on conditional indepen- dence does not require it. Comput. Biol. Med., 14(4):429–435,1984.

[7] R. Kohavi. Wrappers for performance enhancement and obliv- ious decision graphs. Technical report, PhD thesis, Department of Computer Science, Stanford, CA, 1995.

[8] P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. In Proceedings of the Tenth National Conference on Artificial Intelligence, pages 399–406, San Jose, CA, 1992. AAAI Press.

[9] Tom M. Mitchell. Machine Learning. McGraw-Hill, 1997.

[10] I. Rish, J. Hellerstein, and T. Jayram. An analysis of data char- acteristics that affect naive Bayes performance. Technical Re-port RC21993, IBM T.J. Watson Research Center, 2001.

[11] H. Schneiderman and T. Kanade. A statistical method for 3d detection applied to faces and cars. In Proceedings of CVPR- 2000, 2000.