Naive Bayes classifier

Gian Marco Balia
Robotic Engineering - University of Genoa
s4398275@studenti.unige.it

***Abstract*—The Naive Bayes Classification is a family of algorithms for probabilistic classification based on Bayes' theorem. Relied on the probabilistic problem, naive Bayesian classifiers can be effectively trained in a supervised learning context to classify a set of observations. A well-known approach to smooth the Naive Bayes Classifier is the Laplace Smoothing, that consist in adds one observation to each input class. In this toy-study was considered a weather dataset to decide whether to go play outdoor.**

***Index Terms*—Naive Bayes Classifier, Laplace smoothing, Weather dataset**

I. Introduction

The Naive Bayes Classifier model is widely used [7] in machine learning application from medical diagnosis [1] to students' education [3]. Based on the specific characteristics of each probabilistic model, naive Bayesian classifiers can be trained in a supervised learning context to classify a set of observations [2]. In this toy-study we used it to predict a easy probabilistic problem. The aim of the model was to predict if would be a good day to play tennis outside given four classes of data, i.e., *outlook*, *temperature*, *humidity*, and *windy*. To avoid null probabilities and enhanced the accuracy, during the model's training were been introduced a Laplace Smoothing, like other studies [4], [5].

II. Material and methods

A. Data processing

Before working with the data it was needed to be processed. The first thing was to shuffle raw's dataset, preventing paths in itself avoiding biases in the trained model. At this point, the data reported in Table I, was splitted in four parts:

1) *training input data*, the 75 % of the input data (i.e., *Outlook*, *Temperature*, *Humidity*, and *Windy*);
2) *test input data*, the remaining 25 % of the input data;
3) *training output data*, the 75 % of the data in the column *Play*;
4) *test input data*, the remaining 25 % of the output data.

The input and output training data are used to fit the Naive Bayes Classifier. After, the testing data are used to test the models. In the end were evaluated the trained model with the *error rate*.

B. Naive Bayes Classifier

To simplify the problem we assumed that each feature of each class is independent from the others. This method called *Idiot's Bayes*, although being almost always wrong is extremely convenient [6]. In order to train the Naive Bayes Classifier were computed the *priors probability* and *likelihood probability*.

1) Priors probability: The first part of train the model is to compute the *priors probability*

$$P(C_j) = \frac{N_{C_j}}{\sum_{j=1}^{m} N_{C_j}}$$

where N_{C_j} is the total number of instances that belong to class C_j.

Outlook	Temperature	Humidity	Windy	Play
overcast	hot	high	False	yes
overcast	cool	normal	True	yes
overcast	mild	high	True	yes
overcast	hot	normal	False	yes
rainy	mild	high	False	yes
rainy	cool	normal	False	yes
rainy	cool	normal	True	no
rainy	mild	normal	False	yes
rainy	mild	high	True	no
sunny	hot	high	False	no
sunny	hot	high	True	no
sunny	mild	high	False	no
sunny	cool	normal	False	yes
sunny	mild	normal	True	yes

TABLE I
TABLE OF WEATHER'S DATASET

Each classe had different features: the *outlook* was describable as *overcast*, *rainy*, and *sunny*, the *temperature* with *hot*, *cool*, *mild*, the *humidity* level as *high*, *normal*, the *wind* could be present (*True*) or absent (*False*).

2) Conditional probability: The *conditional probability* was unused to fit the model as

$$P(x_i \mid C_j) = \frac{N_{x_i,C_j}}{N_{C_j}} \tag{1}$$

where N_{x_i,C_j} is the number of times the feature x_i appear in the instance of class C_j. In order to avoid $P(x_i \mid C_j) = 0$ was implemented the *Laplace smoothing* and the Equation 1 turns into

$$P(x_i \mid C_j) = \frac{N_{x_i,C_j} + \alpha}{N_{C_j} + \alpha v_i}$$

3) Likelihood probability: The prediction in this model is given by the *likelihood probability*

$$P(C_j \mid X) = \frac{P(C_j)\prod_{i=1}^{n} P(x_i \mid C_j)}{P(X)} \tag{2}$$

where α is the *Laplace smoothing parameter* and v_i is the number of possibles distinct values that the feature x_i can assume.
Although the probability $P(X)$ is often unknown, it is possible to choose which class C_i has more probability comparing numerator of the fraction in the Equation 2.

C. Model evaluation

The model accuracy was evaluated with the *error rate* r_e

III. Results

IV. Conclusion

References

[1] K. Al-Aidaroos, A. A. Bakar, and Z. Othman, "Medical data classification with naive bayes approach," *Information Technology Journal*, vol. 11, no. 9, pp. 1166–1174, 2012.

[2] A. Derbel and Y. Boujelbene, "Automatic classification and analysis of multiple-criteria decision making," in *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Vol. 1.* Springer, 2020, pp. 83–93.

[3] I. T. Nafea, "Machine learning in educational technology," *Machine learning-advanced techniques and emerging applications*, pp. 175–183, 2018.

[4] S. Narayan and E. Sathiyamoorthy, "Early Prediction of Heart Diseases using Naive Bayes Classification Algorithm and Laplace Smoothing Technique," *International Journal of Grid and High Performance Computing*, vol. 14, no. 1, pp. 1–14, 2023.

[5] F. F. Sabiq, A. Rahmatulloh, I. Darmawan, R. Rizal, R. Gunawan, and E. Haerani, “Performance Comparison of Multinomial and Bernoulli Naïve Bayes Algorithms with Laplace Smoothing Optimization in Fake News Classification,” in *2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*, Aug. 2024, pp. 19–24.

[6] M. Schonlau, *Applied Statistical Learning: With Case Studies in Stata*, ser. Statistics and Computing. Cham: Springer International Publishing, 2023.

[7] I. Wickramasinghe and H. Kalutarage, “Naive Bayes: Applications, variations and vulnerabilities: A review of literature with code snippets for implementation,” *Soft Computing*, vol. 25, no. 3, pp. 2277–2293, Feb. 2021.