

CLASSIFICATION ALGORITHMS(Problem 1)

Nitin K (2017csb1093)

30.10.2020

CSE FINAL YEAR, UG

IIT ROPAR

INTRODUCTION

This is a report for three classifier algorithms (Bayes classifier, Naive Bayes Classifier, K-NN classifier). We are going to analyse how these algorithms perform on the given dataset (Credit card fraud dataset).

Algorithms Details:

I have taken all the algorithms from the books. That is why I am providing the reference only as the algorithm has already been discussed in the class.

Bayes Classifier

Reference book for the algorithm is **Data Mining and Analysis: Fundamental Concepts and Algorithms**, Mohammed J. Zaki and Wagner Meira Jr.

Naive Bayes Classifier

Reference book for the algorithm is **Data Mining and Analysis: Fundamental Concepts and Algorithms**, Mohammed J. Zaki and Wagner Meira Jr.

K-NN Classifier

Reference book for the Algorithm is **Data Mining and Analysis: Fundamental Concepts and Algorithms**, Mohammed J. Zaki and Wagner Meira Jr.

DATA

DataSet Descriptions

Here We have Provided DataSet:

- **Credit card fraud dataset DataSet**
(<https://www.kaggle.com/mlg-ulb/creditcardfraud>)

	Time	V1	...	Amount	Class
count	284807.000000	2.848070e+05	...	284807.000000	284807.000000
mean	94813.859575	3.919560e-15	...	88.349619	0.001727
std	47488.145955	1.958696e+00	...	250.120109	0.041527
min	0.000000	-5.640751e+01	...	0.000000	0.000000
25%	54201.500000	-9.203734e-01	...	5.600000	0.000000
50%	84692.000000	1.810880e-02	...	22.000000	0.000000
75%	139320.500000	1.315642e+00	...	77.165000	0.000000
max	172792.000000	2.454930e+00	...	25691.160000	1.000000

RESULTS

1. Bayes Classifier

	precision	recall	f1-score	support
0	1.00	0.97	0.98	56864
1	0.03	0.60	0.06	98
accuracy			0.97	56962
macro avg	0.52	0.78	0.52	56962
weighted avg	1.00	0.97	0.98	56962

2. Naive Bayes Classifier

	precision	recall	f1-score	support
0	1.00	0.98	0.99	56864
1	0.06	0.79	0.11	98
accuracy			0.98	56962
macro avg	0.53	0.88	0.55	56962
weighted avg	1.00	0.98	0.99	56962

3. K-NN Classifier

I am not providing any results because when I ran it for a very small sample of 10% test data it consumed too much time.

Here I run my dataset on three different classifiers. After classifying through these algorithms I found that Naive-Bayes Classifier was performing very well on the provided dataset with 98% accuracy whereas Bayes classifier provides 97% accuracy on the given dataset. Its (**Naive-Bayes Classifier**) run time and accuracy is much better than Bayes-Classifier and K-NN Classifier. With respect to run time K-NN classifier is the worst among these.

The reason for Naive-Bayes to perform better was the dataset itself. Most of the features of the dataset were independent to each other which makes it a good fit dataset for the

Naive-Bayes classifier. Although Naive-Bayes classifiers consider dataset to be independent to each other hence it was producing very good results compared to Bayes classifier.

CONCLUSION

If features of the dataset are independent then Naive-Bayes Classifier is going to work really well. And If not then Bayes would work better. But Bayes Classifier consumes too much time than Naive-Bayes Classifier. Similarly K-NN consumes too much time to classify very small datasets. In K-NN the value of K should be chosen wisely to get the best output.