Performance Report

Ruhul Azgor

December 2023

1 How to Run in local machine

1.1 Data

Please download all the dataset and unzip those and put the script at the same folder with the extracted data.

1.2 Requirements

Ensure that you have the following prerequisites installed on your machine:

- **Python version:** 3.10.12
- **NumPy version:** 1.23.5
- **Pandas version:** 1.5.3
- **Scikit-learn version:** 1.2.2

I think, my script won't run in any python version < 3.10.

1.3 Installation Instructions

```
pip install numpy==1.23.5
pip install pandas==1.5.3
pip install scikit-learn==1.2.2
```

Now, you can run the program, it will automatically interact with you through console. You will be able to choose dataset and parameters. It will train the model and output it's performance meassure.

1.4 How to Run in Google Colab

In google colab, there won't be any version related data.

1.4.1 Setting up Kaggle API in Colab

To access Kaggle datasets directly in Google Colab, you need to set up your Kaggle API key. Follow these steps:

1.4.2 Upload Kaggle API Key

1. Visit the Kaggle website and log in or create an account: `https://www.kaggle.com/`
2. Go to your Kaggle account settings page.
3. Scroll down to the "API" section and click on "Create New API Token."
4. This will download a file named `kaggle.json` to your local machine.
5. Upload the File in the Colab File system's folder.

Now, Uncomment the script from the top of the .py file(Before import). and run it in the Colab with the rest of the script.

2 Adaboost Implementation with Logistic Regression

2.1 Dataset: Telco Customer Churn

Feature=20

Number of Boosting Rounds	Training Accuracy	Test Accuracy
5	0.7593	0.7512
10	0.8005	0.7818
15	0.8005	0.7818
20	0.8009	0.7818

Table 1: Adaboost with Logistic Regression on Telco Customer Churn

2.2 Dataset: Adult

Feature 20

Number of Boosting Rounds	Training Accuracy	Test Accuracy
5	0.7658	0.7710
10	0.7656	0.7708
15	0.8135	0.8168
20	0.8195	0.8218

Table 2: Adaboost with Logistic Regression on Adult

2.3 Dataset: Credit Card Fraud

Feature=15

Number of Boosting Rounds	Training Accuracy	Test Accuracy
5	0.9919	0.9932
10	0.9919	0.9932
15	0.9919	0.9932
20	0.9919	0.9932

Table 3: Adaboost with Logistic Regression on Credit Card Fraud

2.4 Logistic Regression Performance on Telco Customer Churn

Feature=20

Performance Measure	Training Set				Test Set			
	K=5	K=10	K=15	K=20	K=5	K=10	K=15	K=20
Accuracy	0.7593	0.8005	0.8005	0.8005	0.7512	0.7818	0.7818	0.7818
True Positive Rate	0.1217	0.4221	0.4221	0.4221	0.1070	0.3930	0.3930	0.3930
True Negative Rate	0.9901	0.9375	0.9375	0.9380	0.9845	0.9226	0.9226	0.9226
Positive Predictive Value	0.8161	0.7098	0.7098	0.7114	0.7143	0.6476	0.6476	0.6476
False Discovery Rate	0.1839	0.2902	0.2902	0.2886	0.2857	0.3524	0.3524	0.3524
F1 Score	0.2119	0.5294	0.5294	0.5298	0.1860	0.4892	0.4892	0.4892

Table 4: Logistic Regression Performance on Telco Customer Churn

2.5 Logistic Regression Performance on Adult

Feature=20

Performance Measure	Training Set				Test Set			
	K=5	K=10	K=15	K=20	K=5	K=10	K=15	K=20
Accuracy	0.7658	0.7656	0.8135	0.8195	0.7710	0.7708	0.8168	0.8218
True Positive Rate	0.0295	0.0284	0.3239	0.3581	0.0328	0.0322	0.3242	0.3541
True Negative Rate	0.9994	0.9995	0.9688	0.9658	0.9993	0.9992	0.9691	0.9664
Positive Predictive Value	0.9390	0.9449	0.7671	0.7687	0.9333	0.9254	0.7646	0.7652
False Discovery Rate	0.0610	0.0551	0.2329	0.2313	0.0667	0.0746	0.2354	0.2348
F1 Score	0.0571	0.0552	0.4555	0.4886	0.0633	0.0623	0.4554	0.4842

Table 5: Logistic Regression Performance on Adult

2.6 Logistic Regression Performance on Credit Card Fraud

Feature=15

Performance Measure	Training Set				Test Set			
	K=5	**K=10**	**K=15**	**K=20**	**K=5**	**K=10**	**K=15**	**K=20**
Accuracy	0.9919	0.9919	0.9919	0.9919	0.9932	0.9932	0.9932	0.9932
True Positive Rate	0.8173	0.8173	0.8173	0.8173	0.8571	0.8571	0.8571	0.8571
True Negative Rate	0.9962	0.9962	0.9962	0.9962	0.9965	0.9965	0.9965	0.9965
Positive Predictive Value	0.8407	0.8407	0.8407	0.8407	0.8571	0.8571	0.8571	0.8571
False Discovery Rate	0.1593	0.1593	0.1593	0.1593	0.1429	0.1429	0.1429	0.1429
F1 Score	0.8288	0.8288	0.8288	0.8288	0.8571	0.8571	0.8571	0.8571

Table 6: Logistic Regression Performance on Credit Card Fraud