Logistic Regression and AdaBoost for Classification

Instructions

You have to follow the below steps in order to run the script for both training and testing purposes.

1. Make sure that you have these Python modules installed beforehand.
   - numpy can be installed with `pip install numpy`
   - pandas can be installed with `pip install pandas`
   - scikit-learn can be installed with `pip install scikit-learn`
2. In your current working directory, download these files.
   - Telco customer churn dataset from https://www.kaggle.com/blastchar/telco-customer-churn and rename it `WA_Fn-UseC_-Telco-Customer-Churn.csv`
   - Adult income dataset from https://archive.ics.uci.edu/ml/datasets/adult and rename it `adult.data` and `adult.test`
   - Credit card fraud dataset from https://www.kaggle.com/mlg-ulb/creditcardfraud and rename it `creditcard.csv`
3. Run the `1805061.py` file

Report

Logistic Regression on Telco Customer Churn Dataset

Performance measure	Training (%)	Test (%)
Accuracy	80.28	82.11
True positive rate (sensitivity, recall, hit rate)	54.08	60.05
True negative rate (specificity)	89.75	90.06
Positive predictive value (precision)	65.61	68.5
False discovery rate	34.39	31.5
F1 score	59.29	64.0

Logistic Regression on Adult Income Dataset

Performance measure	Training (%)	Test (%)
Accuracy	85.17	85.14
True positive rate (sensitivity, recall, hit rate)	59.61	58.79
True negative rate (specificity)	93.27	93.29
Positive predictive value (precision)	73.76	73.03
False discovery rate	26.24	26.97
F1 score	65.93	65.14

Logistic Regression on Credit Card Fraud Dataset (Sampled)

Performance measure	Training (%)	Test (%)
Accuracy	99.53	99.63
True positive rate (sensitivity, recall, hit rate)	81.41	87.27
True negative rate (specificity)	99.96	99.97
Positive predictive value (precision)	98.11	98.97
False discovery rate	1.89	1.03
F1 score	88.98	92.75

Accuracy of AdaBoost on Telco Customer Churn Dataset

Number of Boosting Rounds	Training (%)	Test (%)
5	79.13	80.7
10	78.88	81.05
15	79.04	80.48
20	78.88	80.41

Accuracy of AdaBoost on Adult Income Dataset

Number of Boosting Rounds	Training (%)	Test (%)
5	84.0	83.9
10	84.17	84.06
15	84.12	84.01
20	84.19	84.1

Accuracy of AdaBoost on Credit Card Fraud Dataset (Sampled)

Number of Boosting Rounds	Training (%)	Test (%)
5	99.46	99.51
10	99.43	99.46
15	99.46	99.46
20	99.54	99.54