CSE 472 Machine Learning

Offline 2 - Logistic Regression and AdaBoost

Name: Siam Al Mujadded
Department: CSE
Student ID: 1805051
Section: A2
Date: 09/12/2023

Data tables:

Logistic Regression

For adult dataset (using all the features)

Performance measure	Training	Test
Accuracy	81.33%	81.31%
Recall	31.75%	31.68%
Specificity	96.88%	97.07%
Precision	76.17%	77.47%
False Discovery Rate	23.83%	22.53%
F1 Score	0.448	0.4496

For credit card dataset (using top 5 features)

Performance measure	Training	Test
Accuracy	99.87%	99.86%
Recall	26.65%	22.45%
Specificity	99.99%	99.99%
Precision	85.37%	84.62%
False Discovery Rate	14.63%	15.38%
F1 Score	0.406	0.3548

For telco-customer dataset (using all the features)

Performance measure	Training	Test
Accuracy	79.18%	80.27%
Recall	39.04%	42.63%
Specificity	93.69%	93.82%
Precision	69.11%	71.30%
False Discovery Rate	30.89%	28.70%
F1 Score	0.4989	0.533

AdaBoost:

For Telco-customer-churn dataset (using all the features)

Number of boosting rounds	Training	Test
5	79.20%	80.98%
10	78.86%	80.41%
15	79.14%	80.27%
20	78.93%	80.70%

For adult dataset (using all the features)

Number of boosting rounds	Training	Test
5	82.84%	82.71%
10	83.29%	83.09%
15	83.70%	83.86%
20	83.31%	83.17%

For credit-card-fraud dataset (using only 5 top features)

Number of boosting rounds	Training	Test
5	99.86%	99.86%
10	99.86%	99.86%
15	99.87%	99.86%
20	99.86%	99.86%

Instructions for running the model:

Before executing the 1805051.py file install necessary packages. Especially for the adult dataset to be loaded, install the '**ucimlrepo**' package using the command '**pip install ucimlrepo'.** After that run the file and see the results. At the bottom of the file you can configure different parameters. I have specified which is which in the code.