Assignment on:
Adaboost & Logistic Regression

Bangladesh University of Engineering and Technology

Course: CSE 472 (Machine Learning Sessional)
Date of Submission: 9.12.2023

Submitted by:
Abdur Rafi
1805008

How to Run Script

At line 670, file paths of the datasets can be specified

```
670     pathTelco = "WA_Fn-UseC_-Telco-Customer-Churn.csv"
671     pathAdultTrain = "./adult/adult.data"
672     pathAdultTest = "./adult/adult.test"
673     pathCreditCard = "./archive/creditcard.csv"
```

To run on a dataset, 2 things needs to be done

1. Choose the corresponding config

```
666     config = telcoConfig
667     # config = adultConfig
668     # config = creditCardConfig
```

2. Choose the corresponding dataset

```
676     dataset = Telco(pathTelco, -1)
677     # dataset = AdultDataset(pathAdultTrain, pathAdultTest, 50)
678     # dataset = CreditCard(pathCreditCard, -1)
679
```

To run logistic regression or adaboost

```
680     logistic(dataset, config)
681
682     # adaBoost(dataset, config)
```

Performance Scores of Logistic Regression

Telco-Customer-Churn Dataset (All features)

Metric	Train	Test
Accuracy	0.804757	0.810504
Recall	0.552387	0.581717
Specificity	0.896995	0.889313
Precision	0.662162	0.644172
False Discovery Rate	0.337838	0.355828
F1 Score	0.602314	0.611354

Adult Dataset (All features)

Metric	Train	Test
Accuracy	0.843617	0.833671
Recall	0.644943	0.723869
Specificity	0.906634	0.867632
Precision	0.686626	0.628442
False Discovery Rate	0.313374	0.371558
F1 Score	0.665132	0.672789

Credit Card Fraud Detection Dataset (All features)

Metric	Train	Test
Accuracy	0.995486	0.995853
Recall	0.824427	0.838384
Specificity	0.999687	0.99975
Precision	0.984802	0.988095
False Discovery Rate	0.0151976	0.0119048
F1 Score	0.897507	0.907104

Performance of Adaboost

Telco-Customer-Churn Dataset (All features)

k	Train	Test
5	0.599042	0.596167
10	0.617146	0.616749
15	0.637203	0.61462
20	0.663117	0.649397

Adult Dataset (Top 50 features)

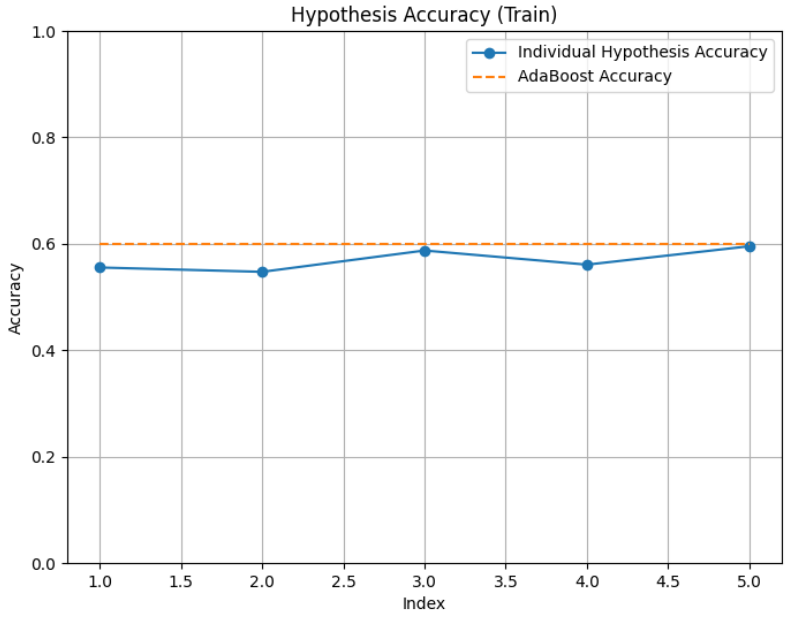
k	Train	Test
5	0.528792	0.53805
10	0.542919	0.545974
15	0.598016	0.607149
20	0.629188	0.635035

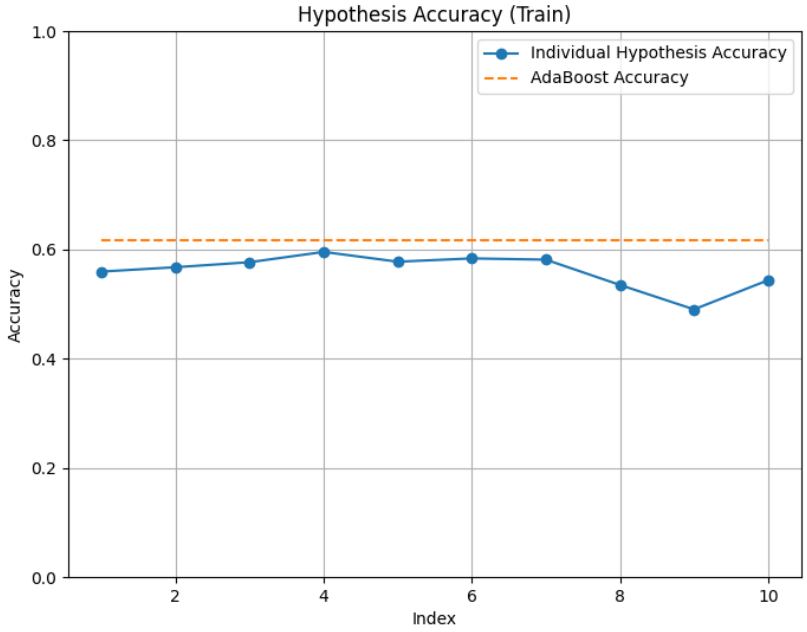
Credit Card Fraud Detection Dataset (All features)

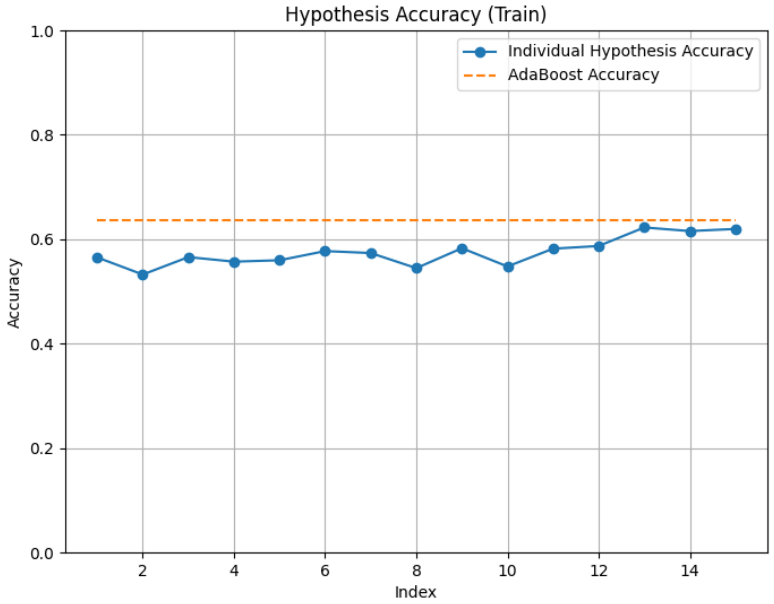
k	Train	Test
5	0.5542	0.56404
10	0.589947	0.588436
15	0.61386	0.61332
20	0.668334	0.675775

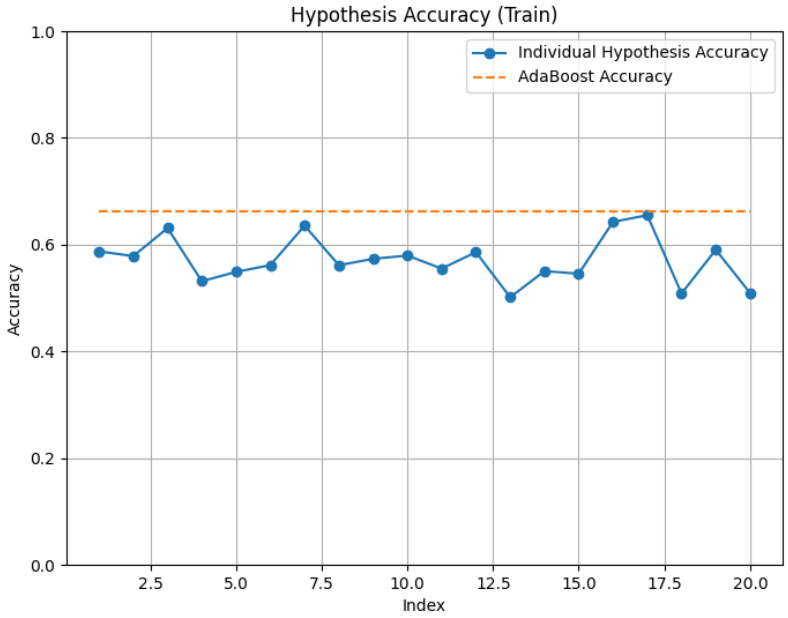
Observations

Performance of Adaboost is quite worse than logistic regression. However, this is *expected* since we are using weak learners in Adaboost, in order to demonstrate its ability to learn decision boundaries using multiple weak learners. We can see that Adaboost indeed works from the following graphs. We see that the accuracy of the Adaboost is higher than all of its constituent hypotheses.









Also it is seen that the accuracy of the Adaboost algorithm increases with increased k. This is somewhat maintained in all of the datasets.