Homework 2 Question 2

50/50 Points

21/10/2023

Attempt 1

Review Feedback
21/10/2023

Attempt 1 Score:
50/50

Add Comment

Anonymous Grading: **no**

Unlimited Attempts Allowed

Details

Introduction

This portion of Homework 2 will be done individually, not in a group. Same as in Question 1, your assignment should be submitted by uploading your code (in the form of a **Jupyter Notebook (.ipynb) AND pdf copy of the files** – so we can make comments directly on the file) to Canvas. **Be sure to run the file before committing so that we can directly see your results.** Please mention all the resources that were used to solve the problem (e.g., websites, books, research papers, other people, etc.). To complete the assignment, you can use any Python (or R) package that you want, but we recommend using Scikit-Learn.

Question

To gain a better understanding of the differences across datasets, perform the same tasks as in Question 1, but on a dataset of your choice (if you worked on a team for Question 1, please do not select the same dataset as your team members). The dataset should contain multiple features (attributes) and you can perform binary or multi-class classification. If your dataset already has train/validation/test or train/test split, you don't have to split it into 60/20/20 as in Question 1 - just make sure to have train, validation and test data (e.g., if you dataset only has train and test portion, you should take a subset of the train data and use it for validation).

View Rubric

Select Grader

Nakul Havaldar (TA)

(https://iu.instructure.com/courses/2165858/modules/items/30397202)

Attempt (https://iu.instructure.com/c

Assignment 2

Criteria	Ratings		Pts
Question -1	**2 to >0 pts** **Full Marks** Statistical descriptions and Visualizations :1.5 If any special treatment required :0.5	**0 pts** **No Marks**	2 / 2 pts
Question-2	**3 to >0 pts** **Full Marks** Computing the PCC:1.5 Scatter Plots :1.5	**0 pts** **No Marks**	3 / 3 pts
Question-3	**5 to >0 pts** **Full Marks** Splitting the data in testing , validation and training sets correctly 2.5 Verification of splitting 2.5	**0 pts** **No Marks**	5 / 5 pts
Question 4 a Multinomial Logistic Regression	**10 to >0 pts** **Full Marks** Model is implemented correctly :2 Different hyperparameters (C, solver,max number of iterations) have been tried:3 Training, Validation and Testing Performance have been reported :3 Discussion on the impact of different hyper parameters has been done :2	**0 pts** **No Marks**	10 / 10 pts
Question 4 b view longer description	**10 to >0 pts** **Full Marks** Model is implemented	**0 pts** **No Marks**	10 / 10 pts

(https://iu.instructure.com/courses/2165858/modules/items/30397202) (https://iu.instructure.com/c

Assignment 2

Criteria	Ratings		Pts
	tried:3 Training, Validation and Testing Performance have been reported :3 Discussion on the impact of different hyper parameters has been done :2		
Question 4 c view longer description	**10 to >0 pts** **Full Marks** Model is implemented correctly :2 Different hyperparameters(no. of trees, max depth ,the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node) have been tried:3 Training, Validation and Testing Performance have been reported :3 Discussion on the impact of different hyper parameters has been done :2	**0 pts** **No Marks**	10 / 10 pts
Question 5 view longer description	**10 to >0 pts** **Full Marks** Ensemble classifier has been implemented via all the models with the best hyperparameters :4 Accuracy of the ensemble is greater than all the individual classifiers :2 Test set Accuracy :1 Discussion on Findings	**0 pts** **No Marks**	10 / 10 pts
			Total Points: 50

File Name	Size

(https://iu.instructure.com/courses/2165858/modules/items/30397202) (https://iu.instructure.com/c

	File Name	Size	
	WebsitePhishing.pdf	720 KB	
	WebsitePhishing.ipynb	996 KB	

(https://iu.instructure.com/courses/2165858/modules/items/30397202) (https://iu.instructure.com/c

(https://iu.instructure.com/courses/2165858/modules/items/30397202)

(https://iu.instructure.com/c