CSE 494: Artificial Intelligence for Cyber Security: Spring 2019

Homework 2: (20 points)

These question tests your understanding on using supervised learning techniques for predicting exploited CVEs. We will be using Decision Trees and Logistic Regression as the model for prediction.

Download the list of CVEs from the file *CVE_list_hw2.csv* – the file contains a list of CVEs and the *isExploited* column indicates whether a CVE has been exploited (a 1 indicates exploited).

1. Extract the relevant attributes from the JSON file of the CVE data corresponding to the CVE list in the file and test the Decision Trees classifier to report the Precision, Recall and F1 on a 10 fold Cross Validation
    a. Report the classification metrics using the best attributes (you may choose to compare all the attributes against one another or group them) for predicting whether a CVE would be exploited (4 points)
    b. Report the results by setting the maximum depth of the Decision Tree to various values and plot a chart to compare the F1 scores. Additionally, split the data vendor wise (you may only consider cves for which the vendors are available)and perform a classification for each vendor separately and report the results. (4 points)
    c. Report the decision tree rules for the following CVEs: (2 points)
        i. CVE-2016-9079
        ii. CVE-2015-2483

2. Use the same list of CVEs and test the Logistic Regression classifier this time.
    a. Report the Precision, Recall and F1 metrics on 10 fold Cross Validation (you can leave the hyper parameters as default). 4 points
    b. Vary the hyperparameter "C" that controls the regularization strength and plot the train-test error curve. 2 points.
    c. Compare your results with the Decision Tree classifier and explain why and which features work better with Logistic Regression compared to Decision Trees and vice versa. 4 points

BONUS QUESTION:

Repeat the same experiments as above but with the Random Forest classifier. Report the results by varying different parameters and compare your results to the Decision Tree classifier.