CSCI 184 HW 1 -- Part 2: Programming

Due Date: Oct. 10, 11:59pm 2024

All assignments MUST have your name, student ID, course name/number at the beginning of your documents.
Your homework MUST be submitted via Gradescope with the file format and name convention as follows:
For programming questions, please upload your code and supporting files in "**HW#_Name.zip**".
You have 3 tasks to finish, please write all codes in a single jupyter-notebook file called "HW1.ipynb" instead of 3 files :)

If you have any questions, please don't hesitate to contact me :)

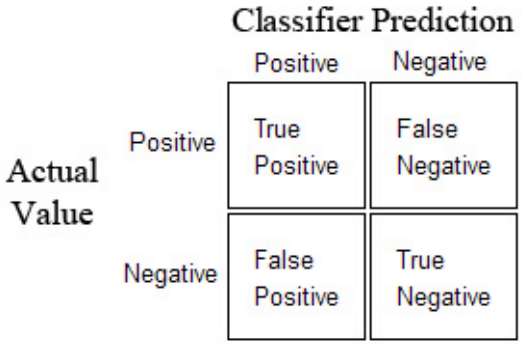
This assignment is totally 50 points. **You will need to write a report for your outputs of the following three questions.**

Problem: use **scikit-learn's DecisionTreeClassifier** [1] to learn a **fixed-depth decision tree**, that is, the input to the ID3 algorithm will include the training data and **maximum depth** of the tree to be learned.

Data Sets: The MONK's Problems were the basis of a first international comparison of learning algorithms. The training and test files for the three problems are named monks-X.train and monks-X.test. There are six attributes/features (columns 2–7), and binary class labels (column 1). See monks.names for more details.

a. (Learning Curve, 15 points) For max_depth = 1, .., 10, use DecisionTreeClassifier [1] to learn a decision tree using criterion='entropy', and compute the **average training and testing accuracy** on **each of the three MONK's problems**. Make **three plots**, one for each of the MONK's problem sets, plotting training and testing accuracy curves together for each problem, with tree depth on the x-axis and error on the y-axis.

b. (confusion matrix, 15 points) For **monks-1**, use **scikit-learn's DecisionTreeClassifier** [1] to learn a decision tree using **criterion='entropy'** for depth = 5. Report the visualized learned decision tree and the confusion matrix on the test set. You may use **scikit-learn's confusion matrix()** function [2].



c. (confusion matrix, 20 points) For **monks-1**, use **scikit-learn's DecisionTreeClassifier** [1] to learn a decision tree using **criterion='gini'** for depth = 5. Report the visualized learned decision tree and the confusion matrix on the test set. Compare the decision tree and confusion matrix obtained in this problem with the results obtained in problem b, and report your findings.

[1] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
[2] https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html