Assignment 2
Decision Trees

CMPE 481 Data Analysis and Visualization **Due: December 20th, 2021, 6am**

In this assignment, you will construct a simple decision tree to classify butterflies and birds. You are given a 2D dataset as shown in Figure 1. Construct a decision tree with two decision boundaries using entropy calculations.

Step 1. Decision Boundaries

Draw the final decision boundaries as shown in Figure 1. For each decision boundary in the figure, provide 1) level of the decision node, e.g., one or two, and 2) which axis is selected and the corresponding boundary value such as x=5.10. For each decision boundary, report the following entropy values: 1) E_{left}, 2) E_{right} (or E_{top}, E_{bottom}) and 3) their weighted average $E_{total} = w_{left}E_{left} + w_{right}E_{right}$.

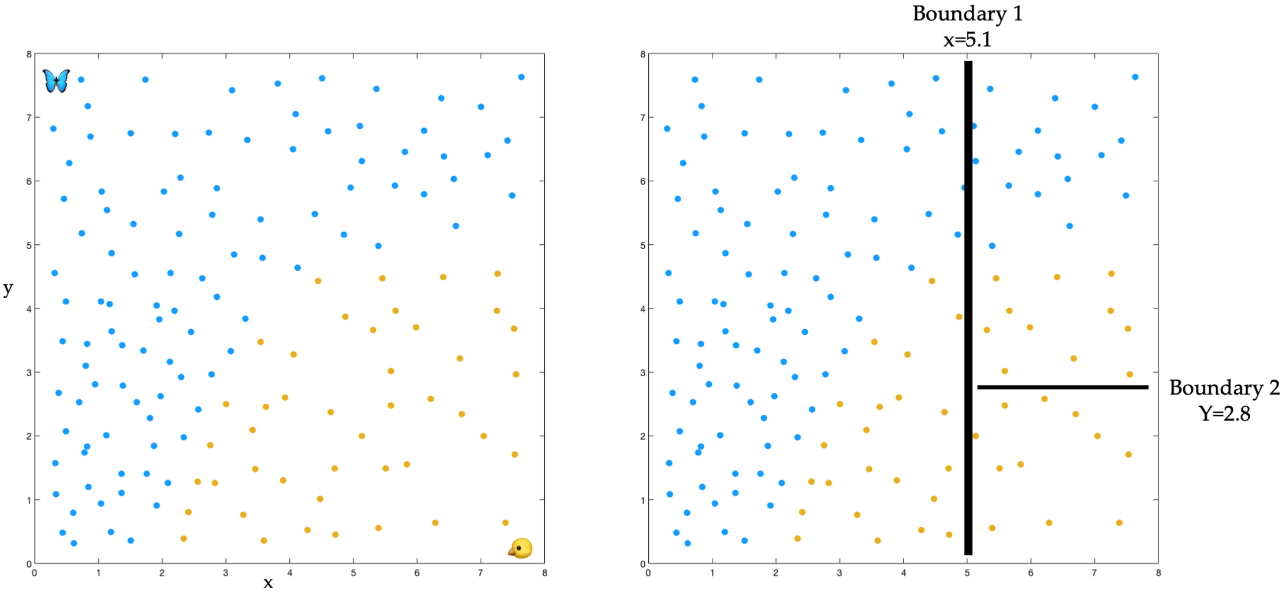

Figure 1. Dataset for butterflies (blue dots) and birds (orange dots) is shown on the left. Arbitrary decision boundaries are shown on the right. First boundary at x=5.1 is thicker than the second boundary.

Step 2. Plot of the Decision Tree

Plot the decision tree as shown in Figure 2. The tree is randomly created for illustrative purposes. Provide decision boundary conditions at the decision nodes such as x > 186. You can draw the decision tree manually, i.e., there is no need to draw it automatically.

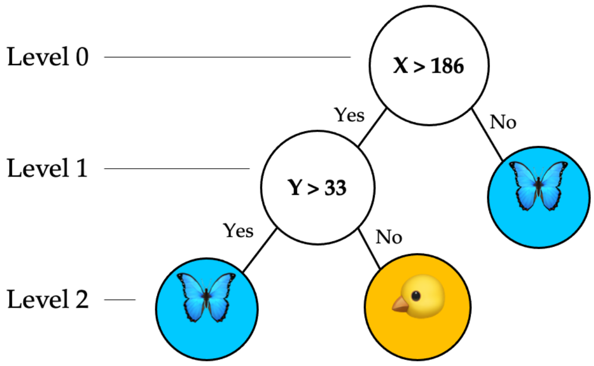

Figure 2. Decision tree constructed from the dataset shown in Figure 1.

Step 3. Comparison with scikit-learn

Draw the decision boundaries and the decision tree obtained by scikit-learn (use sklearn.tree.DecisionTreeClassifier) and compare it with your decision tree.

Step 4. Your Own Dataset

Repeat the steps 1 to 3 using your own 2D dataset. Be creative in generating your own dataset, e.g., with complicated decision boundaries. In this step, if you want, you can generate more than two levels (depth of the tree) for the decision tree.

Evaluation Criteria

		Points
Decision tree algorithm		50
Report (Contents, completeness, format, etc.)		40
Compliance to Submission Rules (Directory structure, file formats/naming, organization, etc.)		10
	TOTAL	**100**

Submission Guide

Submission Files

Submit a single compressed (.zip) file, named as name_surname.zip, to the Moodle. It should contain all source code files (under the \code directory), report (in PDF format, under the \report directory) and all other files if needed (under \misc directory)

File Naming

Name your report as name_surname.pdf. Name the main code which is used to run your assignment as assignmentX.py, where X is the assignment number and .py is the extension for Python, given as an example.

Late Submission Policy

Maximum delay is two days. Late submission will be graded on a scale of 50% of the original grade.

Mandatory Submission

Submission of assignments is mandatory. If you do not submit an assignment, you will fail the course.

Plagiarism

Leads to grade F and YÖK regulations will be applied