Assignment 3
Decision Trees

COMP 462 Introduction to Machine Learning   **Due: March 25ᵗʰ, 2020, 6am**

In this assignment, you will construct a decision tree to classify butterflies and birds. As shown in Figure 1a, you are given a 2D dataset. Construct a decision tree (max depth is three, excluding the leaf nodes) using information gain. Draw the decision boundaries as given in Figure 1b. For each decision boundary, provide 1) depth/level of the decision node, 2) which axis is selected and the corresponding boundary value such as x=2.33, and 3) the information gain obtained at that point.

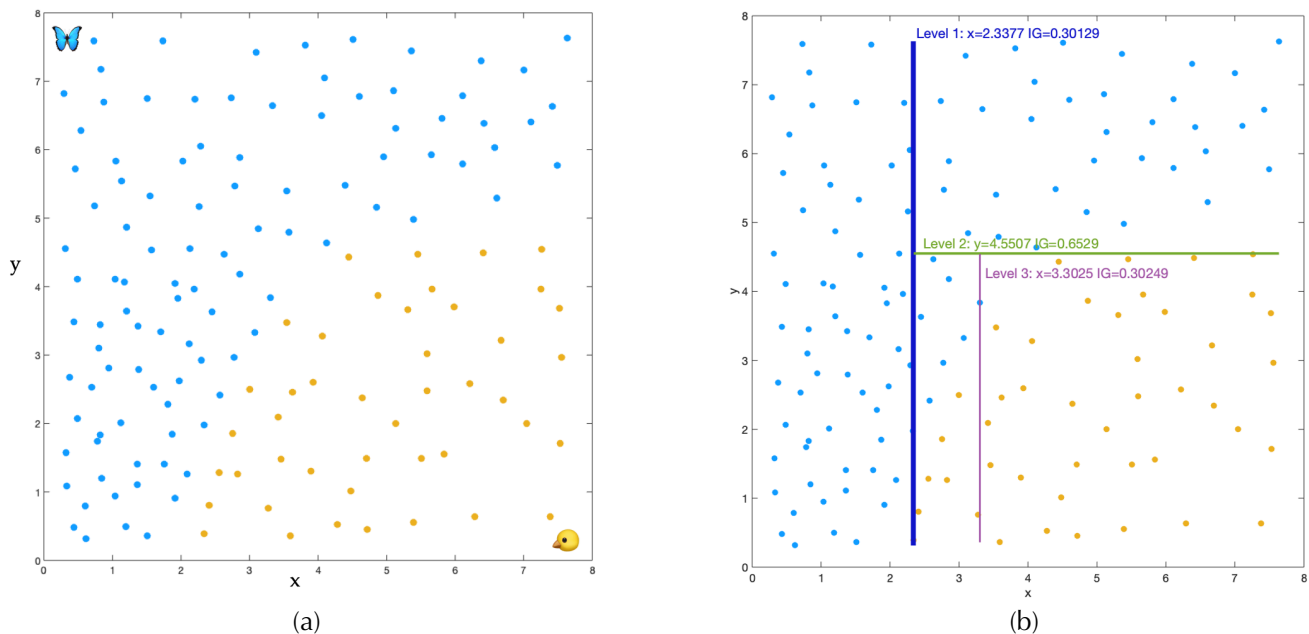(a)   (b)

Figure 1. (a) Dataset for butterflies (blue dots) and birds (orange dots) and (b) decision boundaries.

For the root node selection at tree level 1, plot all information gain values for both x and y axes and show the maximum information gain at the selected point, i.e., x=2.33 where the information gain is 0.30, as shown in Figure 2.
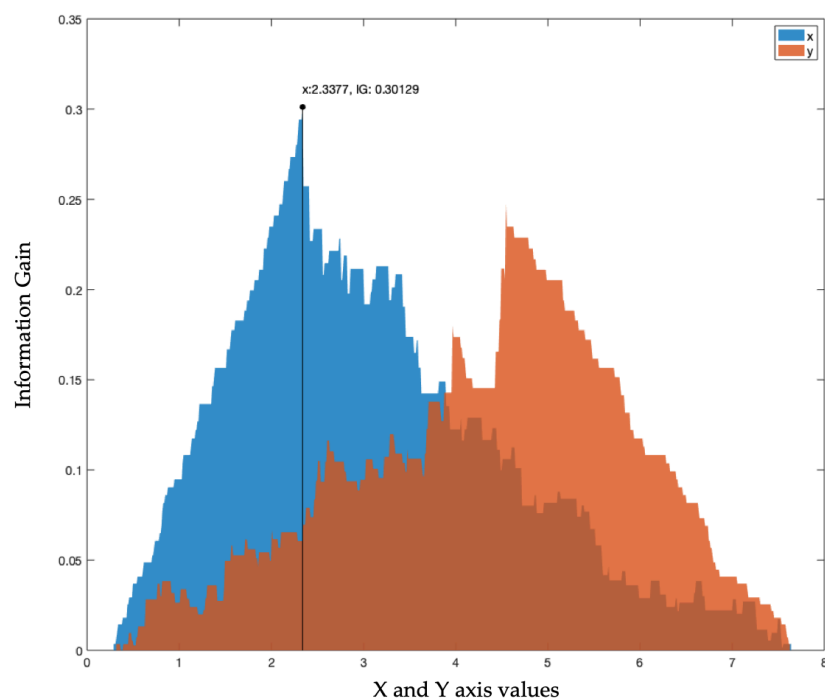

Figure 2. Information gain values for all x and y values. Information gain values for the x and y values are denoted by blue and orange colors, respectively. Maximum gain (0.30) is obtained at x=2.33.

Plot the decision tree as shown in Figure 3. Tree should have a maximum depth of three, excluding the leaf nodes. Provide decision boundary conditions at the decision nodes such as x > 2.33. You can draw the decision tree manually, i.e., there is no need to draw it automatically.
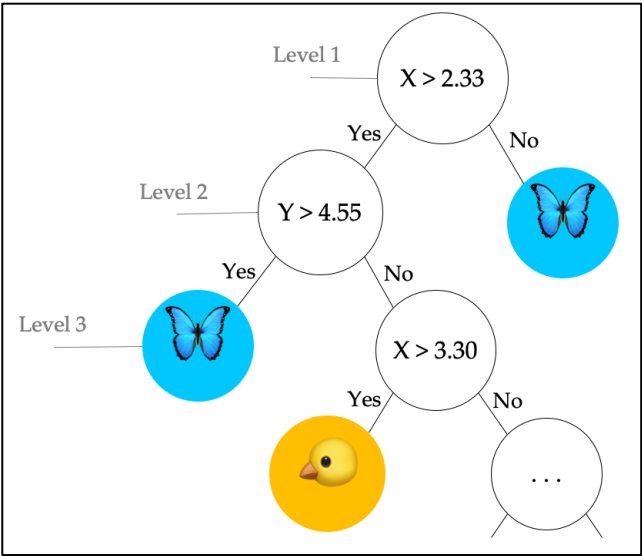

Figure 3. Decision tree constructed from the dataset.

Bonus
Use a machine learning library such as Scikit-learn to compare the results of your solution. Draw the decision boundaries and the decision tree obtained by the machine learning toolbox.

Evaluation Criteria

	Points
Decision tree algorithm	60
Report (Contents, completeness, format, etc.)	30
Compliance to Submission Rules (Directory structure, file formats/naming, organization, etc.)	10
Bonus: Comparison with a machine learning library	20
TOTAL	**120**

Submission Guide

Submission Files
Submit a single compressed (.zip) file, named as name_surname.zip, to Blackboard. It should contain all source codes/data 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 start/run your assignment as assignmentX.py, where X is the assignment number and .py is the extension of 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