Name Surname : **Göktuğ YILDIRIM**
Student ID : **041502012**

COMP462- Introduction to Machine Learning

Assignment 3

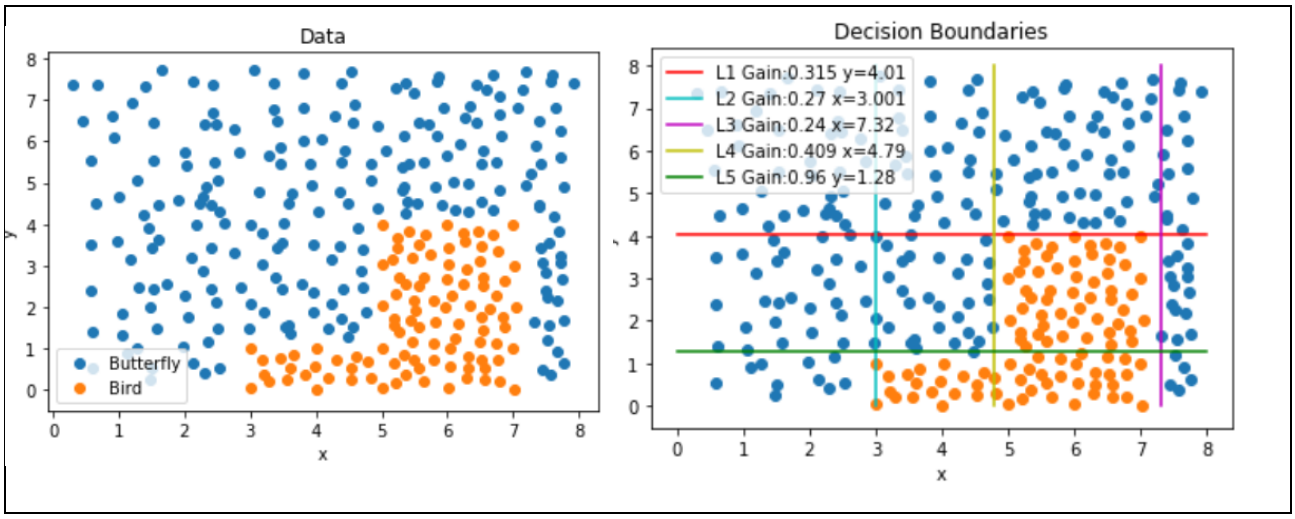

Figure 1. (a) Dataset for butterflies (blue dots) and birds (orange dots) and (b) decision boundaries.

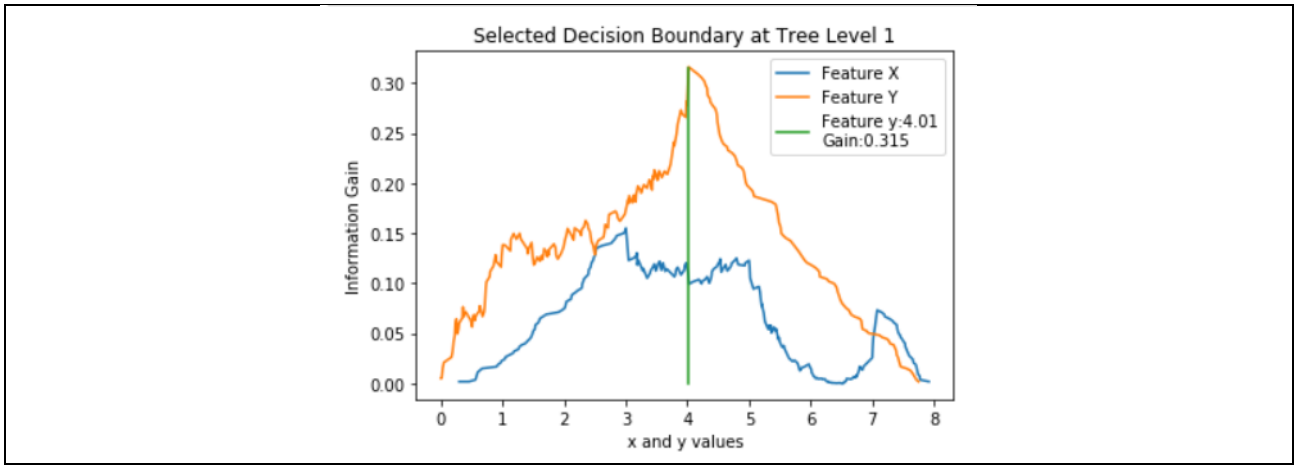

Figure 2. Information gain values for all x and y values. Information gain values for the x and y values are denoted by blue and orange colors, respectively. Maximum gain (0.315) is obtained at y = 4.01.

Name Surname : **Göktuğ YILDIRIM**
Student ID : **041502012**

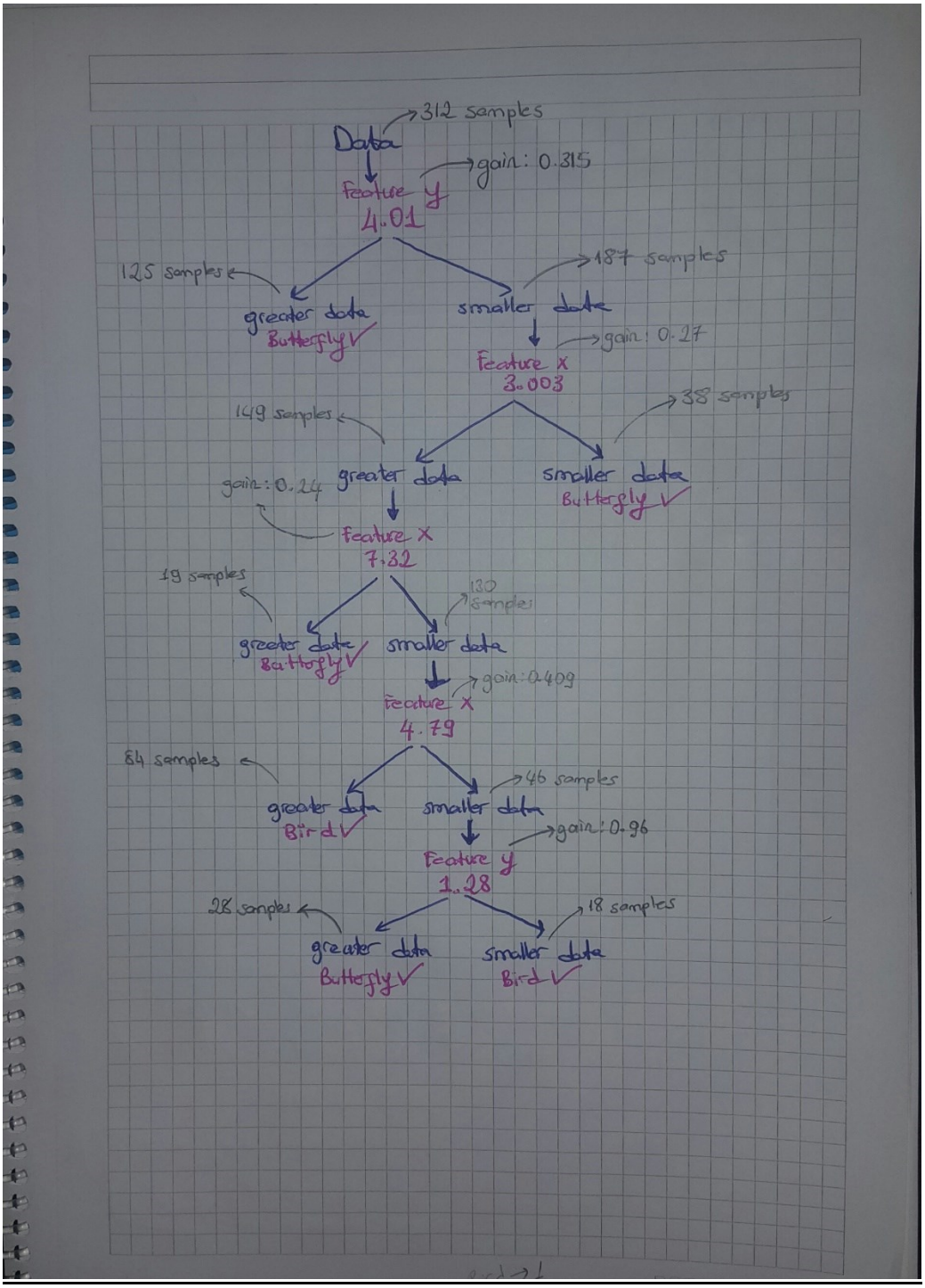

Figure 3. Decision tree constructed from the dataset.

Name Surname : **Göktuğ YILDIRIM**
Student ID : **041502012**

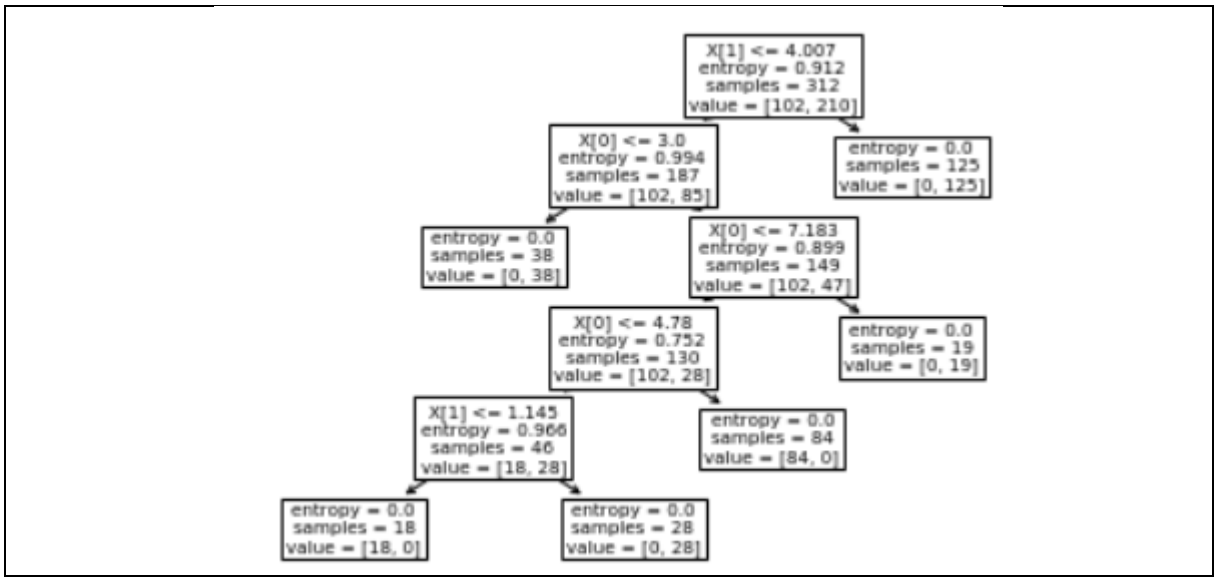

Figure 4. Decision tree constructed from the dataset by Scikit-learn.

Comment: My decision boundaries and Scikit-learn decision boundaries are very similiar but not the same. Because I did not search by making the axis grid to find best split. I made the feature values set then subtracting an epsilon value from each feature values to provide a faster algorithm. In relation to the criterion of the Scikit-learn, I could not set the criterion as information gain so that criterion was entropy. Since decision boundaries are very similiar, my information gain function works well. Also number of samples in each level is are equal to each other. Also, the number of samples at each level is are equal to each other.