Name; Claire Danaher

Assignment: HW1 Part 2

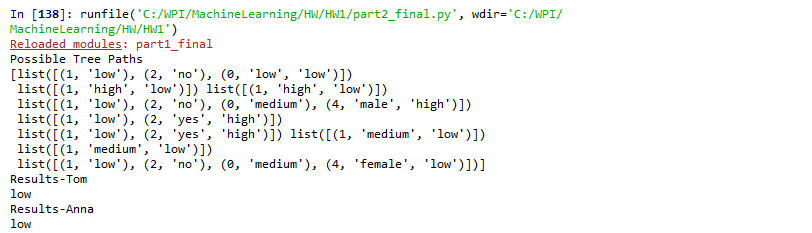
Part 1: Notes

I collaborated with Alex Shoop on the approach for the recursive binary tree creation.

Part 2: Credit Risk Prediction Let’s assume that you work for a credit card company. Given the sample credit dataset (credit.txt) as a training set, your job is to build a decision tree and make risk prediction of individuals. The target/class variable is credit risk described as high or low. Features are debt, income, marital status, property ownership, and gender.

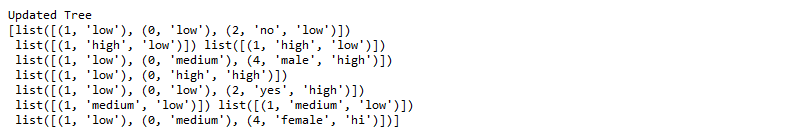
Task 2-1: Draw your decision tree and report it. You may use visualization tools (e.g., Graphviz) or use text. You might find it easier if you turn the decision tree on its side, and use indentation to show levels of the tree as it grows from the left. For example:

Results



Interpretation: each list is a branch to a leaf node. The tuple with three values is the lead node and it’s primary Y.

Task 2-2: How does your decision tree change if Sofia's credit risk is high instead of low as recorded in the training data? Given the decision tree constructed from the original dataset, if existing, name any feature not playing a role in the decision tree.



The same features played a role but the order of the splits and the depth changed. The original tree was depth 4 and the split two occured on attr2. The revised tree was depth 3 and the second split occurred on attr2.