*Lorenzo Mazza*

*DEC TREE LAB - ML Course*

## *Assignment 0: Each one of the datasets has properties which makes them hard to learn. Motivate which of the three problems is most difficult for a decision tree algorithm to learn.*

The most difficult to learn is the dataset Monk2 because it requires exactly two attributes to assume the same value, so if we were to translate the formula in a logic circuit it would be the most complex one.

## *Assignment 1: The file dtree.py defines a function entropy which calculates the entropy of a dataset. Import this file along with the monks datasets and use it to calculate the entropy of the training datasets.*

MONK 1: 1.0

MONK 2: 0.957117428264771

MONK 3: 0.9998061328047111

## *Assignment 2: Explain entropy for a uniform distribution and a non-uniform distribution, present some example distributions with high and low entropy.*

Entropy is maximized for a uniform distribution because uncertainty is maximal when all possible events are equiprobable. On the other hand entropy is minimized when there is no uncertainty, for example when the probability of an event is 1.

## *Assignment 3: Use the function averageGain (defined in dtree.py) to calculate the expected information gain corresponding to each of the six attributes. Note that the attributes are represented as in- stances of the class Attribute (defined in monkdata.py) which you can access via m.attributes[0], ..., m.attributes[5]. Based on the results, which attribute should be used for splitting the examples at the root node?*

MONK1: A5

Information gain of a5 is 0.28703074971578435

MONK2: A5

Information gain of a5 is 0.01727717693791797

MONK3: A2

Information gain of a2 is 0.29373617350838865

*Assignment 4: For splitting we choose the attribute that maximizes the information gain, Eq.3. Looking at Eq.3 how does the entropy of the subsets, $S_k$, look like when the information gain is maximized? How can we motivate using the information gain as a heuristic for picking an attribute for splitting? Think about reduction in entropy after the split and what the entropy implies.*

The entropy of the subset is minimized when the information gain is maximized. We motivate it because reducing the entropy of the subset we also reduce the uncertainty of the samples and the number of bits that we need to represent the results.

*Assignment 5: Compute the train and test set errors for the three Monk datasets for the full trees. Were your assumptions about the datasets correct? Explain the results you get for the training and test datasets.*

MONK1: Training Error: 0.0 Test Error: 0.17129629629629628

MONK2: Training Error: 0.0 Test Error: 0.30787037037037035

MONK3: Training Error: 0.0. Test Error: 0.05555555555555558

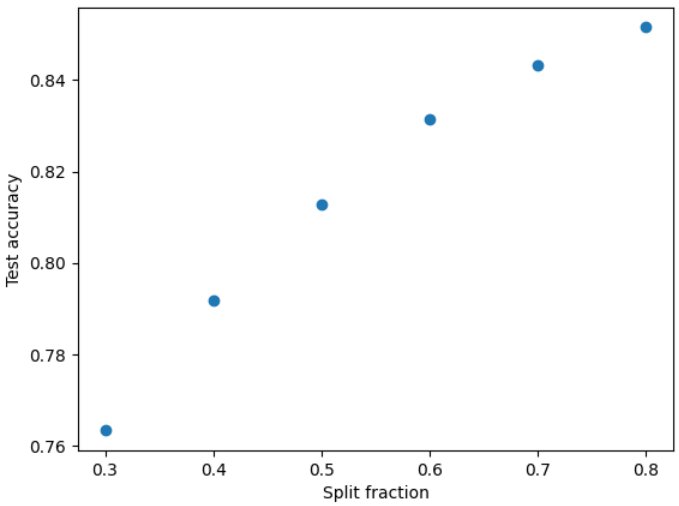
As we can see the assumptions were correct because the second dataset is the one that has the highest test error, being the most complex one to classify.

*Assignment 6: Explain pruning from a bias variance trade-off perspective.*

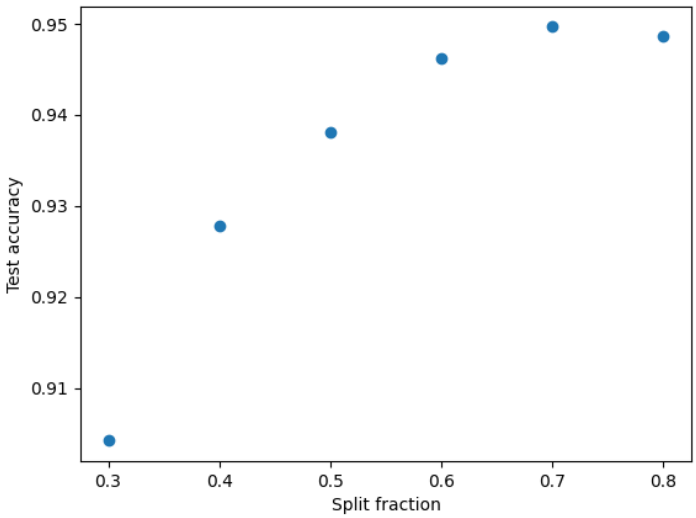
From a Bias-Variance trade-off perspective pruning decreases the variance and increases the bias of the decision tree. This because pruning decreases the complexity of the decision tree and the branching factor as well: so the model becomes smoother and simpler and, doing so, it avoids overfitting. However if the model gets too pruned it risks easily to become too simple and to underfit data increasing the bias.

*Assignment 7: Evaluate the effect pruning has on the test error for the monk1 and monk3 datasets, in particular determine the optimal partition into training and pruning by optimizing the parameter fraction. Plot the classification error on the test sets as a function of the parameter fraction* $\in \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$.

***MONK 1***




***MONK 3***




For each fraction I did 1000 iterations and got the average test accuracy of them. As we can see there is a slight difference in the two datasets regarding the best fraction to split training data with: The best fraction for Monk1 is 0.8 while the best one for Monk3 is 0.7. Anyway in general we can see from the plots that it seems reasonable to consider a good fraction for pruning any number greater than 0.6