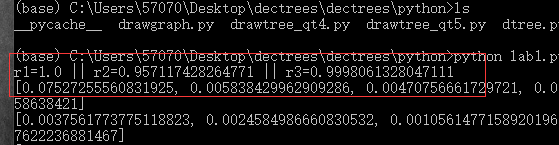
Assignment 0: Each one of the datasets has properties which makes them hard to learn. Motivate which of the three problems is most difficult for a decision tree algorithm to learn.

To obtain a specific answer, the Second dataset MONK-2 needs more questions to ask. It is more complicated to be classified according to the equation, which looks more uncertainty and means more branches in a decision tree. So the answer is MONK-2.

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.

The entropy of 3 datasets:



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

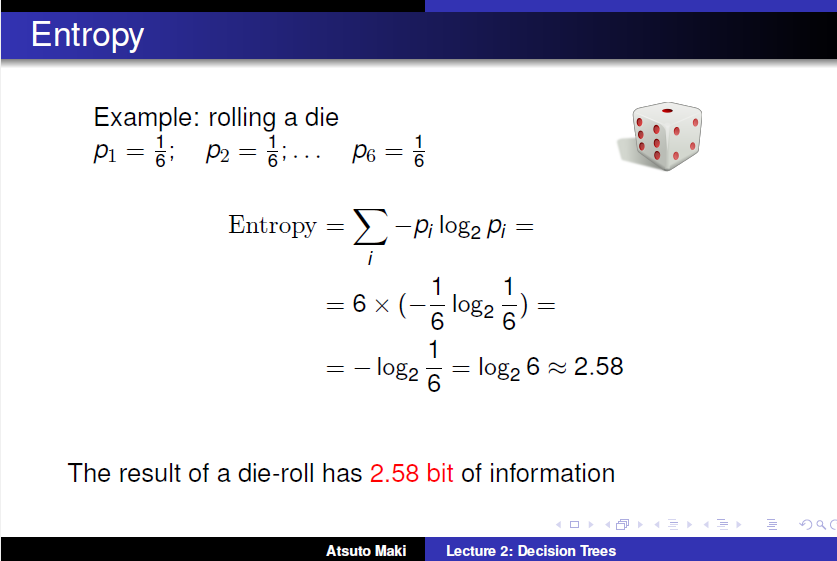
Uniform distribution got higher entropy

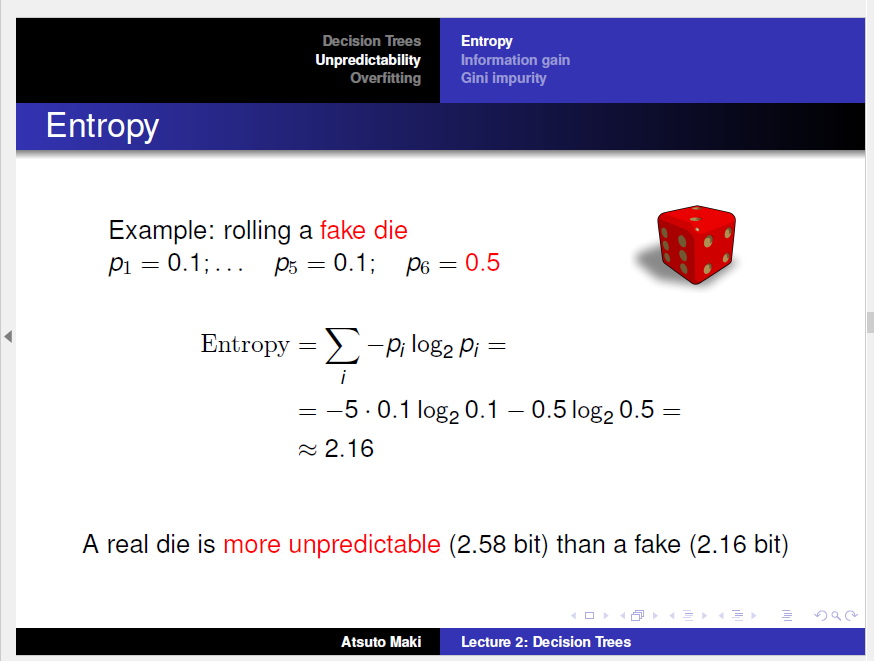
Non-uniform distribution got lower entropy

Uniform distribution receives more uncertainty because the possibility is same for every result and if it’s a non-uniform distribution, it receives less uncertainty because the possibility on certain result is bigger, means more predictable.

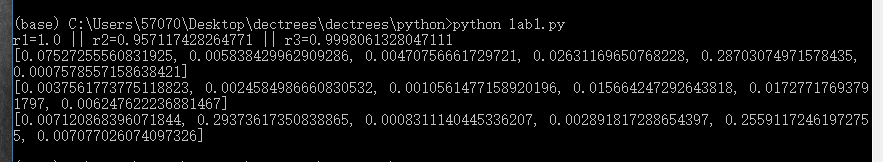
As a uniform distribution, Assume that are a group of discrete observation: X1, ..., Xn the possibility for each result is same, i.e., 1/n. so, the entropy of this dataset can be calculated as: log2 N.

Examples





answer Assignment 3:



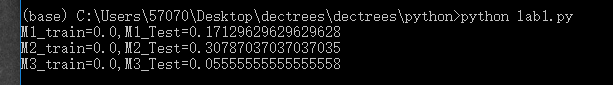
The attribute with the largest information gain should be selected as the root node. So: 'Tree1's root node: A5, Tree2's root node: A5, Tree3's root node: A2

Ass4: For splitting we choose the attribute that maximizes the information gain, Eq.3. Looking at Eq.3 how does the entropy of the subsets, Sk, 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.

When the entropy is minimized, the information gain is maximized. Each split increase the purity of the data in subsets t, showing as entropy reduction. The larger the entropy is, the more disordered the dataset is.

answer Assignment 5:

The test error is larger than train error. The value of zero on train error means overfitting. That's why we need to prune some branches, sacrifice a little bit bias for the improvement of variance.



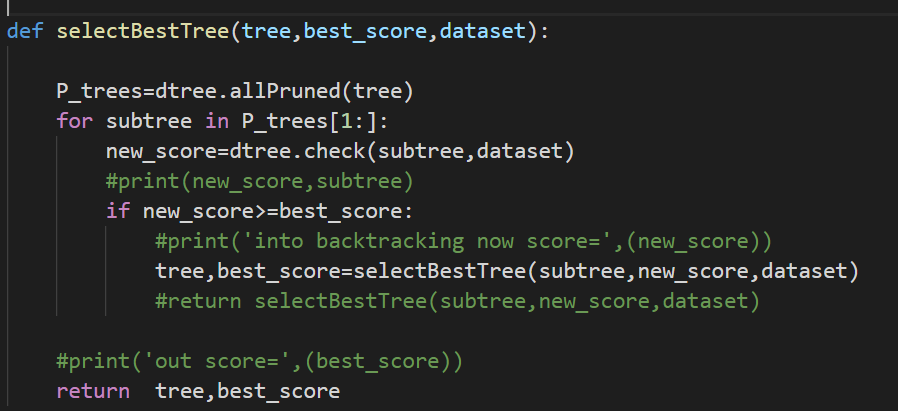
Assignment 6: Explain pruning from a bias variance trade-off perspective.

A smaller tree with fewer splits (that is, fewer regions ) might lead to lower variance and better interpretation at the cost of a little bias.

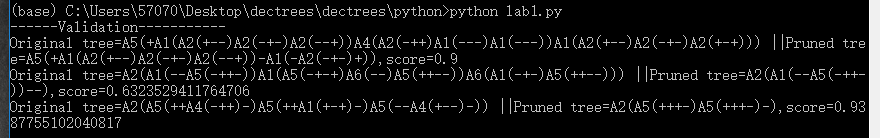
The pruning process will decrease the complexity of the tree and the possibility of overfitting. When there are too many splits in trees, the bias will be small, but it tends to overfit and gain bigger variance in test dataset, because the tree are built more complex with too much focus on the original training data. So pruning can lead to lower variance and better interpretation at the cost of a little bias.

**Assignment 7:**

Recursion



Validation for dataset when fraction is 0.6



When fraction=0.8 or 0.7. the model may have the better performance in test dataset.

