

# AI Assignment2: ID3 Decision Tree

Matthew Lee 1807789269

March 2019

## 1 ID3 Decision Tree

In this project an ID3 Decision Tree classifier was created based on the Psuedocode from.<sup>1</sup> Decision trees classify by splitting the data based on the value they have for a particular attribute. The algorithm considers each attribute that the data has not already been split on and selects the one that has the lowest total entropy, which maximises the information gain. Entropy is calculated using equation 1.

$$H(X) = - \sum_{i=0}^c p_i \log_2(p_i) \quad (1)$$

Equation 1 gives total entropy for a dataset, where X is the data, c is the number of classes,  $p_i$  is the probability of class i.  $P_i$  is simply the fraction of the data that has class i for this application of the ID3 algorithm.

The classifier uses training data to construct a tree, each node of the tree representing a different attribute to split the data. There is a branch for each value an attribute may have. During training, if the information gain from one split was equal to another, then the first one was selected. Some classifiers would choose randomly but that can introduce randomness that might fail the automated testing. This does bias the algorithm in favour of splitting on the earlier attributes. If the classifier splits the data on an attribute but finds no data that matches a possible value of that attribute, then it will create a leafnode that classifies data based on the most prevalent class in the pre-split data.

Once training is complete and a decision tree has been constructed, the algorithm can take test data into the "classify" function. This splits the test data into individual instances and classifies them one by one. The classification works by traversing down the tree, using the data's attribute values to decide which branches to follow. This continues until reaching a leafnode which will contain the instances predicted class.

New datasets were created to test the algorithm on non-binary classification and cases where training data could not be classified with 100% accuracy.

---

<sup>1</sup>Stuart J Russell and Peter Norvig. *p702, Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016.