Assignment 2 Report

Devansh Agarwal 200316
Utkarsh Aditya 201066
Harshit Singhai 2000434
Antariksh Choudhary 200159
Abhinav Zade 200019

Answers to Task-1

To develop our decision tree algorithm we first aim to choose the best possible query word which will give us the optimal decision tree. We achieve this goal by using the entropy minimization formula for each node of the tree, choosing the best split at the individual node level.

We also implement a look-ahead here in which we are minimizing the total entropy at the +2 level of each individual node. For this, we have defined a new function named look ahead which takes up the split dictionary starting from the 3rd level and then performs the look ahead entropy minimization for all the +1 levels after this. For all the other in between levels, normal entropy minimization is applied. We run a for loop for all the possible query words i.e all the words in that particular node and then use the entropy formula:

$$\sum -p * log(p)$$

where p is the fraction of words in child and total words at that node

We calculate the entropy of the next level for each word. We then choose the query word which gives us the minimum entropy for that level. We split the tree based on this query word.

For calculating the entropy we use the above formula. Also for making the node the query words giving the same response come under the same node. And p gives the fraction of words giving the same response on one particular level.

For pruning our code we have removed the history parameter from the nodes which was storing the link to the parent node. We did not use the history parameter. This helps in reducing the total time taken for running the code as well as memory allocated after using pickle.

On running the code on our system we achieved the following outcomes;

- The total runtime was about 6.8 sec
- Accuracy achieved was 100%
- And the average query was 3.99
- The total memory usage was 768 bytes