CS771: Assignment 2

Kajal Deep - 200483 **Kuldeep Singh Chouhan - 200530** **Kumar Arpit - 200532**

Sandeep Kumar Bijarnia - 200856 **Yash Raj Mittal - 2001148**

1 Overview

The heart of the problem lies within how we guess the new word at each node. In the code, the Instructor gave, it was simply selecting the word randomly from the remaining list of words at each level. This simply was giving a much less win rate. To improve the model word to be guessed should be selected according to some metric, and at each node, the decision tree will split the dictionary according to the given response.

2 Towards the Algorithm

The most basic thing we know is to reduce Entropy. Out of all the remaining words, we tried to choose a word that provides splitting of Minimum Entropy at each step, a greedy method. The difference between this method and the following method is that the difference in sizes of the child nodes would be minimum in entropy reduction; we noticed it was taken more queries and large training time. Due to this, we instead used the method based on the Positional Frequency of Characters in the remaining list.
This method focuses on the most frequent letter occurring at a position and tries to choose a word that reduces a Score, capturing the similarities between words.

3 Algorithm for guessing word at each level

- We started by calculating the cumulative frequency of each letter at a given position of the word from the remaining list of words.
- Stores it in a dictionary name **frequencies** where keys are characters in the alphabet and value is a list storing the frequency of the character at each position of the word.
- Now, at a given position, we calculate the maximum frequency of any letter at that position and store in **maxFreq**
- For each word, we find a **Score** that captures the **difference in frequency** of the maximum frequent letter at the position and the letter of the word.
- Finally, we choose the word with a minimum above score.
- This word is further used to split the wordlist according to the given response.
- If the guessed word is correct, we stop and win the game otherwise, this word is further used to split the wordlist according to the given response.

.