Assignment #4

CSE 447: Natural Language Processing
Eric Boris: 1976637
Collaborators: Allyson Ely

Implementation Problem

Problem 1
Explain how you generate a score for each possible choice and how you recover the best sequence.

My approach for generating a score for each possible choice is as follows: I create a table that holds columns. Each column contains a positive, $n > 1$ number of entries. Each entry is contains a character, a score, and a backpointer and is of the form (character, score, backpointer). I iterate over each character in a line. If that character is not the mask character I create a single entry column where the character is the one stored in the entry. Otherwise, if that character is the mask character, I create a column containing one entry per character in the vocabulary. In either case, the score in the entry is that score which maximizes the score between the character in the entry and some character from the previous column's entries. Similarly, the backpointer in the entry is that backpointer which points to the entry index in the previous column which contains the character that has the maximum score with the current character.

My approach for recovering the best sequenece is as follows. By design, the last column in the table is a single entry column. I get that column's backpointer. By design, that backpointer points to the entry index in the previous column with the maximum score. I store the last column's character. I follow that last column's backpointer. I store the character in that entry. I follow that entry's backpointer. And so on. When all the characters are added, I reverse the list that holds the recovered characters.