

CS221 Fall 2018 Homework [Reconstruct]

SUNet ID: prabhjot

Name: Prabhjot Singh Rai

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1

- (a) Assuming that the language model is a uni-gram model. Total cost assigned to words $[w_1, w_2 \dots w_n]$ is defined as $\sum_i^n u(w_i)$, where

$$u(w) = c, \text{ if } w \text{ in given corpus, } c > 0$$

Greedy algorithm, as described in the problem, focuses on traversing in the tree, and stopping when it finds a solution (when all the words created by adding spaces are in the corpus). But it doesn't consider all the possible actions which may lead to an overall lowest cost for the segmentation. For example, considering an input "antioxidantswerepresent" would lead to a total cost of $4c$ ("anti oxidants we represent"), whereas, the solution for lowest cost for segmentation is $3c$ ("antioxidants we represent" or "antioxidants were present").

- (b) (your solution)

Problem 2

- (a) (your solution)
- (b) (your solution)