

## **CSCI 544 HW2 Questions and Answers**

- 1) What threshold value did you choose for identifying unknown words for re-Placement?

Ans - **2**

- 2) What is the overall size of your vocabulary, and how many times does the special token "< unk >" occur following the replacement process?

Ans - The total length of vocabulary is : **21158**

Total number of words replaced by <unk> are **17401**

- 3) How many transition and emission parameters in your HMM?

Transmission Parameters - **1351**

Emission Parameters - **29679**

- 4) What is the accuracy on the dev data?

Accuracy of Greedy Algorithm is : **0.9201399429299982**

- 5) What is the accuracy on the dev data?

Accuracy of Viterbi Algorithm is : **0.9360390990225245**