CSCI 544 – HW 3

In this assignment, we implemented the Hidden Markov Model to predict POS tags using the Penn Treebank dataset. To decode, we have used two algorithms – Greedy Decoding and Viterbi Algorithm.

Task 1 : Vocabulary Creation

- Threshold used for unknown word replacement = 2
- Total size of vocabulary = 23183
- Total occurrences of unknown words = 20011

Task 2 : Model Learning

- Number of transition parameters (“prev tag, current tag” pairs) – 1378
- Number of emission parameters (“current tag, current word” pairs) – 30304

Task 3 : Greedy Decoding

- Dev Accuracy – 0.9352877785198227
- Train accuracy- 0.9491895032863902

Task 4 : Viterbi Decoding

- Dev Accuracy – 0.9505187186473093
- Train - 0.9612896128961289