Name : Ammaar Ahmad
Roll No: 1801CS08
CS564 Assignment 3

In this assignment I have designed a HMM model for Part of Speech tagging on the Brown dataset. We use the part of speech tags as the hidden states and the tokens as the observed variable of the HMM. We code the Viterbi algorithm for the decoding problem and obtain the POS for the test data. We also report the f1 score, the precision and recall for all the POS tags.

Tags: - PRON, VERB, DET, NOUN, ., ADJ, ADP, PRT, ADV, NUM, CONJ, X

The recall, precision and F1 score for the tags are as follows: -

Precision: - [0.63531353, 0.97007293, 0.90665803, 0.94966915, 0.98982751,

0.8676573 , 0.90895197, 0.92457421, 0.9146437 , 0.99219969,

0.98421808, 0.00366331]

Recall: -[0.7815095 , 0.64989329, 0.6582576 , 0.55137322, 0.62790208,

0.51233891, 0.6255917 , 0.63039151, 0.62728249, 0.47462687,

0.57857745, 0.91608392]

F1 Score: -[0.70086885, 0.77834202, 0.76274341, 0.69767803, 0.76837904,

0.64425481, 0.74110997, 0.74965476, 0.74418605, 0.64209995,

0.72875354, 0.00729744]

The overall accuracy for the POS tagging task using the bigram model is around 60%.